IBM Content Analytics with Enterprise Search Versão 3.0

Manual de Administração



IBM Content Analytics with Enterprise Search Versão 3.0

Manual de Administração



Nota

Antes de utilizar estas informações e o produto que suportam, leia as informações incluídas na secção "Avisos" na página 583.

Esta edição aplica-se à versão 3, edição 0, modificação 0 do IBM Content Analytics with Enterprise Search (número de produto 5724-Z21) e a todas as edições e modificações seguintes até indicação em contrário em novas edições.

Índice

ibm.com e recursos relacionados	ix
Como enviar comentários	. x
Contactar a IBM	. x
Novidades na Versão 3.0	1
Análico do contoúdo o procuro	
Analise de conteudo e procura	-
empresarial	9
Conteúdo empresarial	10
Encadeamento de análise aberto e escalável	10
Interface do utilizador de prospecção de conteúdo	
interactiva	11
Análise de frequência e correlações	11
Análises periódicas e altamente escaláveis	12
Exportar facetas para melhorar análises e	
comunicações.	12
Interface de utilizador de procura empresarial	
interactiva	13
Descrição geral dos componentes	15
Pesquisadores	16
Analisadores, processadores de documentos e	
anotadores	17
Indexar	18
Servidores de procura	19
Consola de administração	20
Supervisionar o sistema	20
Ficheiros de registo.	21
Personalizar o sistema	21
Aplicações para procura empresarial e prospecção	
de conteúdo	22
Como os dados fluem ao longo do sistema	23
Origens de dados suportadas	24
Administração do sistema	25
Iniciar sessão na consola de administração	25
Alterar a palavra-passe de administrador na	
configuração de um servidor	26
Alterar a palavra-passe de administrador na	
configuração de múltiplos servidores	27
Adicionar servidores ao sistema	29
Configuração do número da porta	31
Alterar o número de portas do sistema	32
Alterar o número da porta da aplicação	34
Alterar o número da porta e a raiz de contexto	
do URL do Centro de Informações.	34
Alterar nomes de sistema central do servidor ou	
enderecos de IP	35
Configuração do suporte para endereços de IP	
duplos	37
Activar o suporte para o protocolo IPv6 em sistemas	
Windows	38
Desactivar suporte para o protocolo IPv6	40

Activar o suporte para o protocolo IPv6 em sistemas	
AIX ou Linux	41
Desactivar suporte para o protocolo IPv6	43
Administração da colecção	45
Diferenças entre tipos de colocçãos	45
Crier une colocie	4J
	51
Clonar uma colecção	52
Editar uma colecção	53
Eliminar uma colecção	54
Determinar o ID de colecção	55
·	
Administração do pesquisador	57
Crier um necquise der	50
	39
Eliminar um pesquisador.	60
Agendamentos de pesquisadores	60
Pesquisadores do Agente para sistemas de ficheiros	
Windows	61
Comandos de administração do servidor de	
agente	63
Pesquisadores do Case Manager	63
Configurar a consider de pasquice deres pare	05
Configurar o servidor de pesquisadores para	
suportar pesquisadores de Case Manager e	
$FileNet P8. \dots \dots$	65
Pesquisadores do Content Integrator	65
Configurar um conector para aceder a	
repositórios do Content Integrator	67
Configurar o servidor do pesquisador para	
pesquisadores do Content Integrator	67
Pesquisar origons de Decumentum no Linux no	07
Continue of generation Documentum no Linux no	(0
System z	68
Pesquisadores do Content Manager	68
Configurar o servidor do pesquisador no UNIX	
para pesquisadores Content Manager	70
Configurar o servidor do pesquisador no	
Windows para pesquisadores Content Manager .	71
Pesquisadores de DB2	74
Configurar o servidor do pesquisador para DB?	
posquisadoros	76
Configuration of Data From the Data in the second	70
Configurar o Data Event Publisher para	
pesquisadores do DB2	77
Configurar o WebSphere MQ para pesquisadores	
de DB2	79
Pesquisar bases de dados DB2 num servidor de	
origens de dados clássico	81
Configurar o nível de isolamento para pesquisar	
bases de dados DB2	81
Pasquisadares de Domine Decument Manager	87
Pesquisadores do Domino Document Manager	02
Pesquisadores do Exchange Server	85
Definir o Exchange Web Service para os	
pesquisadores do Exchange Server	87
Implementar o serviço da Web	
ESExchangeServices no servidor Exchange Server	88
Implementar o serviço Web ESCommonServices	
, , , , , , , , , , , , , , , , , , ,	90
no servidor Exchange Server	

Metadados do Exchange Server	2
Proteger a procura do conteúdo do Exchange	,
Described and Euclidean as Control 2000 - 2002) -
Verificar o acesso aos documentos Exchange	,
Server protegidos $\ldots \ldots \ldots$)
Pesquisadores do FileNet P8)
Manag de relações para bases de dados de IDPC 100	י ר
Pesquisar tabelas de bases de dados de JDBC	,
Opcões para pesquisar várias tabelas de base de	-
dados IDBC estruturadas	5
NNTP pesquisadores)
Pesquisadores do Notes)
Configurar o servidor do pesquisador no UNIX	
para pesquisar origens do Lotus Domino 114	ł
Configurar o servidor do pesquisador no	
Windows para pesquisar origens do Lotus	
Domino \ldots \ldots \ldots \ldots \ldots \ldots \ldots 116	5
Configurar servidores que utilizam o protocolo	_
DIIOP	7
Configurar a porta de conclusão E/S no AIX	
para pesquisar origens no Lotus Domino 118	3
Melhorar a forma como os documentos Notes	
são novamente pesquisados)
Sugestões para pesquisar campos nas bases de	
dados do Lotus Domino	L
Sugestoes para pesquisar anexos para	
documentos do Notes	1
Pesquisadores do Quickr for Domino	-
Configurar um utilizador de administrador de	,
procura para o IBM Connections	3
Configurar campos de índice para valores de	
campo que estão publicados numa lista de	_
valores gerados do IBM Connections 128	3
Pesquisadores do SharePoint)
Implementar serviços da Web para	
pesquisadores SharePoint	L
Pesquisadores do Sistema de ficheiros UNIX 133	3
Pesquisadores da Web	Ł
Configuração do agente do utilizador 135	5
Como o pesquisador da Web utiliza o protocolo	_
Robots Exclusion $\ldots \ldots \ldots$)
Suporte para JavaScript	5
Regras para limitar o espaço de pesquisa da	2
Web	~
lestar ligações de UKL com o pesquisador da)
W/ob 143	2
Web	3
Web	2 3 1
Web	5 3 4 5
Web	5 3 1 5
Web	5 3 1 5 5
Web 143 Nova pesquisa de definições de intervalo no pesquisador da Web 144 Opções para visitar URLs com o pesquisador da Web 144 Como o pesquisador da Web processa páginas de erros esporádicos 144 Suporte para pesquisa de sítios da Web protegidos 146 Sítios da Web que são servidos por servidores	5 3 1 5 5
Web 143 Nova pesquisa de definições de intervalo no pesquisador da Web 144 Opções para visitar URLs com o pesquisador da Web 144 Como o pesquisador da Web processa páginas de erros esporádicos 144 Suporte para pesquisa de sítios da Web protegidos 146 Sítios da Web que são servidos por servidores proxy 149	5 3 1 5 5

Configuração do espaço global da pesquisa Web	151
Directivas para não seguir ou não indexar	153
Substituir directivas para não seguir ou não	
indexar em páginas da Web	154
Pesquisadores do Web Content Management	154
Copiar o URL para pesquisar no WebSphere	155
Portal	156
Pesquisadores do Sistema de ficheiros Windows Pesquisadores do sistema de ficheiros: configurar	157
Varios modulos.	158
resquisadores do sistema de ficheiros: aumentar o	160
Configurar suporte para aplicações de Data	100
Listener	160
Suplementos do pesquisador personalizados	161
Suporte para pesquisa de ficheiros de arquivo	163
Formatos de URI no índice	164
Administração de análise e indexação	177
Configuração de serviços de análise e indexação	177
Configurar módulos para serviço de índice	178
Activar a analise avançada para termos	170
Activar suporto para procura do XMI pativa	170
Facetas para coleccões de procura empresarial	179
Configurar facetas hierárquicas e simples para	17 /
uma colecção de procura empresarial	182
Configurar facetas de intervalo decimal para	
uma colecção de procura empresarial	184
Configurar facetas de intervalo de data para	
uma colecção de procura empresarial Dicionários lexicais para colecções de procura	186
empresarial	189
Adicionar dicionários lexicais ao sistema	190
Associar um dicionario lexical personalizado a	101
Campos de índice	191
Criar campos de índice	194
Atributos do campo de índice.	195
Nomes de campos de índice reservados	
· · · · · · · · · · · · · · · · · · ·	199
Importar e exportar campos de índice	199 205
Importar e exportar campos de índice Remover informações sobre as origens de entrada e facetas do campo de índice	199 205 206
Importar e exportar campos de índice Remover informações sobre as origens de entrada e facetas do campo de índice Correlacionamento de campo de índice para	199 205 206
Importar e exportar campos de índice Remover informações sobre as origens de entrada e facetas do campo de índice Correlacionamento de campo de índice para elementos XML	199 205 206 206
Importar e exportar campos de índice Remover informações sobre as origens de entrada e facetas do campo de índice Correlacionamento de campo de índice para elementos XML Correlacionar elementos XML com campos de	199 205 206 206
Importar e exportar campos de índice Remover informações sobre as origens de entrada e facetas do campo de índice Correlacionamento de campo de índice para elementos XML	199 205 206 206 207
Importar e exportar campos de índice Remover informações sobre as origens de entrada e facetas do campo de índice Correlacionamento de campo de índice para elementos XML Correlacionar elementos XML com campos de índice	199 205 206 206 207 209
Importar e exportar campos de índice Remover informações sobre as origens de entrada e facetas do campo de índice Correlacionamento de campo de índice para elementos XML Correlacionar elementos XML com campos de índice	199 205 206 206 207 209
Importar e exportar campos de índice Remover informações sobre as origens de entrada e facetas do campo de índice Correlacionamento de campo de índice para elementos XML Correlacionar elementos XML com campos de índice	199 205 206 206 207 209 209
Importar e exportar campos de índice Remover informações sobre as origens de entrada e facetas do campo de índice Correlacionamento de campo de índice para elementos XML Correlacionar elementos XML com campos de índice	199 205 206 206 207 209 209
Importar e exportar campos de índice Remover informações sobre as origens de entrada e facetas do campo de índice Correlacionamento de campo de índice para elementos XML	199 205 206 206 207 209 209 210
Importar e exportar campos de índice Remover informações sobre as origens de entrada e facetas do campo de índice Correlacionamento de campo de índice para elementos XML Correlacionar elementos XML com campos de índice	199 205 206 206 207 209 209 210
Importar e exportar campos de índice Remover informações sobre as origens de entrada e facetas do campo de índice Correlacionamento de campo de índice para elementos XML. elementos XML.	199 205 206 206 207 209 209 210 211
Importar e exportar campos de índice Remover informações sobre as origens de entrada e facetas do campo de índice Correlacionamento de campo de índice para elementos XML. Correlacionar elementos XML com campos de índice	199 205 206 206 207 209 209 210 211
Importar e exportar campos de índice Remover informações sobre as origens de entrada e facetas do campo de índice Correlacionamento de campo de índice para elementos XML Correlacionar elementos XML com campos de índice	199 205 206 206 207 209 209 210 211 212
Importar e exportar campos de índice Remover informações sobre as origens de entrada e facetas do campo de índice Correlacionamento de campo de índice para elementos XML	199 205 206 206 207 209 209 210 211 212 212
Importar e exportar campos de índice Remover informações sobre as origens de entrada e facetas do campo de índice Correlacionamento de campo de índice para elementos XML	199 205 206 206 207 209 209 210 211 212 213 214

sistema	216
Associar um motor de análise de texto a uma	
colecção	217
Correlacionar elementos de XML com a	
estrutura de análise comum	218
Correlacionar a estrutura de análise comum com	
o índice	219
Correlacionar a estrutura de análise comum com	
uma base de dados relacional	220
Filtros de campos	221
Configurar filtros de campos	222
Campos de data e formatos de data personalizados	223
Remover URIs do índice.	225

Processos globais										227
-------------------	--	--	--	--	--	--	--	--	--	-----

Análise de ligações	227
Detecção de documentos duplicados	228
Criação de miniaturas	229
Gerar miniaturas em AIX	231
Alterar o tamanho das imagens de miniatura	232
Análise global personalizada	232
Categorias	234
Categorias baseadas em regras	235
Configurar e implementar conjuntos de	
documentos	239
Anotadores	243

Activar anotadores para uma colecção			. 244
Activar o anotador Voz do cliente .			. 245

Recursos de análise e análise de

conteúdo
Facetas para colecções de análise de conteúdo 249
Configurar a árvore da faceta para colecções de
análise de conteúdo
Configurar facetas de intervalo decimal para
uma colecção de análise de conteúdos 252
Configurar facetas de intervalo de data para
uma colecção de análise de conteúdos 255
Configurar dicionários do utilizador para colecções
de análise de conteúdo
Importar um ficheiro CSV para um dicionário
de utilizador
Configurar ficheiros de regras para colecções de
análise de conteúdo
Indice da faceta opcional para colecções de análise
de conteúdo
Termos de facetas de interesse para colecções de
análise de conteúdos
Configurar análise de postura para colecções de
análise de conteúdo
Configurar vistas contextuais para colecções de
análise de conteúdo
Deteccão do formato do documento 265

Tipos de documentos suportados por predefinição	266
Tipos de documentos associados com analisadores	
de colecção e extractor de texto	267

Mapeamento do tipo MIME e do tipo de	
analisador	268
Analisar tipos de documentos desconhecidos	269
Associar tipos de documentos ao extractor de texto	269
Regras de análise predefinidas para o extractor	
de texto	272
Configurar sessões múltiplas para o extractor de	
texto	275
Analisar texto bidireccional em ficheiros PDF	275
Indexação de documentos .docm com URLs que	
incluam consultas	276

Suporto do idiomas o páginas do

Suporte de Idiomas e paginas de	
códigos	279
Detecção automática de idiomas	280
Detecção automática de páginas de códigos	281
Análise linguística de documentos em chinês,	
japonês e coreano	282
Categorizar caracteres SBCS e numéricos como	
categorias n-gram	282
Expansão de sinónimos ao procurar colecções	
n-gram	283
0	
Administração do servidor de procura	285
Influenciar o rendimento da consulta e a correcção	
ortográfica	286
Caches de procura	287
Configurar uma cache de procura	287
Resumo dinâmico	288
Personalizar resumos de documentos	288
Dicionários de sinónimos personalizados	289
Adicionar dicionários de sinónimos ao sistema	291
Associar um dicionário de sinónimos a uma	-/1
colecção	291
Dicionários de palavras de paragem personalizados	292
Adicionar dicionários de palavras de paragem	2/2
ao sistema	293
Associar um dicionário de palavras de paragem	270
a uma colecção	294
Trabalhar com ligações rápidas	294
Ligações rápidas	294
Configurar ligações rápidas	295
Caracteres globais em consultas	296
Configurar opcões para caracteres globais em	
consultas.	297
Âmbitos	297
Configurar âmbitos	298
URIs contraídos	299
Contrair URIs nos resultados da procura	300
Suporte de introdução adiantada para consultas	301
Configurar suporte de introdução adiantada .	303
Configurar sinalizadores de documentos	303
Exportar documentos de um colecção de procura	-
empresarial para uma colecção de análise de	
conteúdos	304

Inspecção profunda	 307
Exportar resultados de inspecção profunda .	 308
Formato de ficheiros XML para relatórios de	
inspecção profunda	 309

Classificação de documentos 313
Classificação baseada em texto
Classificação estática e importância de documentos 314
Classificação de documentos através de dicionários
de palavras hierárquicas
Adicionar dicionários de palavras hierárquicas
ao sistema
Associar um dicionário de palavras hierárquicas
a uma colecção
Classificação de documento baseada em padrões
URI
Influenciar as pontuações dos documentos que
correspondem aos padrões de URI
Regras para optimizar consultas e classificar
resultados

Aplicações de procura empresarial e

de prospecção de conteúdo 32	23
Aceder a aplicações	324
Abrir a prospecção de análise de conteúdo com	
condições de arrangue	324
Personalizar aplicações	325
Ficheiros de configuração da aplicação 3	325
Personalizar aplicações através da utilização de	
uma interface de personalizador	347
Personalizar a vista Vista-resumo	351
Personalizar a imagem do cabeçalho de	
apresentação da aplicação	355
Personalizar o modo como os títulos e URLs são	
apresentados nos resultados	356
Personalizar URLs utilizando um filtro de URL	
de expressão global	357
Configurar opções de expansão de grupo para	
melhorar tempos de resposta da procura protegida. 3	360
Reter dados da sessão para deslocar mais depressa	
através de resultados protegidos	361

Segurança no IBM Content Analytics with Enterprise Search

with Enterprise Search	363
Activar segurança no servidor de aplicação da Web	,
incorporado	. 365
Configurar a segurança do início de sessão da	
aplicação no servidor de aplicação da Web	
incorporado	. 365
Configurar suporte de início de sessão único no	
servidor de aplicação da Web incorporado.	. 367
Activar a segurança no WebSphere Application	
Server.	. 369
Configurar o suporte SSO no WebSphere	
Application Server	. 370
Activar segurança para um sistema de servidor	
único no WebSphere Application Server	. 372
Activar a segurança para um sistema de	
servidor distribuído no WebSphere Application	
Server	. 373
Segurança ao nível do sistema.	. 375
Autenticação versus controlo de acesso	. 375
ID de administrador predefinido	. 376
Configurar funções administrativas	. 376
Configurar privilégios de utilizador da aplicação	378

Configurar credenciais de ligação.	380
Segurança ao nível da colecção	381
Associar aplicações a colecções	381
Identificadores de aplicações	382
Análise de documentos duplicados e segurança	
de colecção	382
Segurança ao nível do documento	383
Pré e pós-filtragem de resultados da procura	384
Facetas em colecções protegidas	385
Validação por tokens de segurança armazenados	386
Validação das credenciais actuais durante o	
processamento de consultas	387
Análise de texto âncora	395
Requisitos de configuração do pesquisador para	
suporte de segurança	397
Suportar procura protegida quando o servidor	
do pesquisador não está disponível	402
Verificar o acesso aos documentos Exchange	
Server protegidos	404
Cumprimento da segurança ao nível do	
documento para documentos do Lotus Domino .	405
Cumprimento da segurança ao nível do	
documento para documentos do sistema de	
ficheiros do Windows	411
Desactivar a segurança	414
Desactivar a segurança de nível de documento	414
Desactivar a segurança para resultados da	
procura contraídos	416
1	
Iniciar o parar o sistema	
	417
Iniciar o sistema	417 417
Iniciar o sistema Parar os servidores de colecção ou todo o sistema	417 417 419
Iniciar o sistema Parar os servidores de colecção ou todo o sistema Parar todo o sistema	417 417 419 420
Iniciar o sistema Parar os servidores de colecção ou todo o sistema Parar todo o sistema Controlar quais os componentes iniciados ou	417 417 419 420
Iniciar o sistema Parar os servidores de colecção ou todo o sistema Parar todo o sistema Controlar quais os componentes iniciados ou parados	417 417 419 420 420
Iniciar o sistema Parar os servidores de colecção ou todo o sistema Parar todo o sistema Controlar quais os componentes iniciados ou parados	417 417 419 420 420
Iniciar o sistema Parar os servidores de colecção ou todo o sistema Parar todo o sistema Controlar quais os componentes iniciados ou parados Administrar os servidores de procura no modo autónomo	417 417 419 420 420 421
Iniciar o sistema Iniciar o sistema Parar os servidores de colecção ou todo o sistema Parar todo o sistema Parar todo o sistema Controlar quais os componentes iniciados ou parados Administrar os servidores de procura no modo autónomo Iniciar e parar serviços locais	417 417 419 420 420 421 422
Iniciar o sistema Iniciar o sistema Parar os servidores de colecção ou todo o sistema Parar todo o sistema Parar todo o sistema Controlar quais os componentes iniciados ou parados Administrar os servidores de procura no modo autónomo Iniciar e parar serviços locais	417 417 419 420 420 421 422
Iniciar o sistema Iniciar o sistema Parar os servidores de colecção ou todo o sistema Parar todo o sistema Parar todo o sistema Controlar quais os componentes iniciados ou parados Administrar os servidores de procura no modo autónomo Iniciar e parar serviços locais Supervisionar o sistema	417 417 419 420 420 421 422 423
Iniciar o sistema Iniciar o sistema Parar os servidores de colecção ou todo o sistema Parar todo o sistema Parar todo o sistema Controlar quais os componentes iniciados ou parados Administrar os servidores de procura no modo autónomo Iniciar e parar serviços locais Supervisionar o sistema	417 417 419 420 420 421 422 423 423
Iniciar o sistema Iniciar o sistema Parar os servidores de colecção ou todo o sistema Parar todo o sistema Parar todo o sistema Controlar quais os componentes iniciados ou parados Administrar os servidores de procura no modo autónomo Iniciar e parar serviços locais Supervisionar o sistema Supervisionar pesquisadores Como o pesquisador do Notes determina o	417 417 419 420 420 421 422 423 423
Iniciar o sistema	417 417 419 420 420 420 421 422 423 423 425
Iniciar o sistema	417 417 419 420 420 420 421 422 423 423 423
Iniciar o sistema	417 417 419 420 420 421 422 423 423 425 426
Iniciar o sistema Iniciar o sistema Parar os servidores de colecção ou todo o sistema Parar todo o sistema Parar todo o sistema Controlar quais os componentes iniciados ou parados Administrar os servidores de procura no modo autónomo Iniciar e parar serviços locais Supervisionar o sistema Como o pesquisadores Como o pesquisador do Notes determina o número de documentos pesquisados Visualizar detalhes sobre a actividade de pesquisadores da Web Detalhes do módulo do pesquisador da Web	417 417 419 420 420 421 422 423 423 423 425 426 427
Iniciar o sistema Iniciar o sistema Parar os servidores de colecção ou todo o sistema Parar todo o sistema Parar todo o sistema Controlar quais os componentes iniciados ou parados Administrar os servidores de procura no modo autónomo Iniciar e parar serviços locais Supervisionar o sistema Supervisionar pesquisadores Como o pesquisador do Notes determina o número de documentos pesquisados Visualizar detalhes sobre a actividade de pesquisadores da Web Detalhes do módulo do pesquisador da Web Sítios activos do pesquisador da Web	417 417 419 420 420 421 422 423 423 423 425 426 427 427
Iniciar o sistema Iniciar o sistema Parar os servidores de colecção ou todo o sistema Parar todo o sistema Parar todo o sistema Controlar quais os componentes iniciados ou parados Administrar os servidores de procura no modo autónomo Iniciar e parar serviços locais Supervisionar o sistema Supervisionar pesquisadores Como o pesquisador do Notes determina o número de documentos pesquisados Visualizar detalhes sobre a actividade de pesquisadores da Web Detalhes do módulo do pesquisador da Web Sítios activos do pesquisador da Web	417 417 419 420 420 421 422 423 423 425 426 427 427 428
Iniciar o sistema Iniciar o sistema Parar os servidores de colecção ou todo o sistema Parar todo o sistema Parar todo o sistema Controlar quais os componentes iniciados ou parados Administrar os servidores de procura no modo autónomo Iniciar e parar serviços locais Supervisionar o sistema Supervisionar pesquisadores Como o pesquisador do Notes determina o número de documentos pesquisados Visualizar detalhes sobre a actividade de pesquisadores da Web Detalhes do módulo do pesquisador da Web Sítios activos do pesquisador da Web Criar relatórios de pesquisadores da Web	417 417 419 420 420 421 422 423 423 423 425 426 427 427 428 429
Iniciar o sistema Iniciar o sistema Parar os servidores de colecção ou todo o sistema Parar todo o sistema Parar todo o sistema Controlar quais os componentes iniciados ou parados Administrar os servidores de procura no modo autónomo Iniciar e parar serviços locais Supervisionar o sistema Supervisionar pesquisadores Como o pesquisador do Notes determina o número de documentos pesquisados Visualizar detalhes sobre a actividade de pesquisadores da Web Detalhes do módulo do pesquisador da Web Sítios activos do pesquisador da Web Criar relatórios de pesquisadores da Web Criar relatórios de pesquisadores da Web Criar relatórios de pesquisadores da Web	417 417 419 420 420 421 422 423 423 423 425 426 427 428 429
Iniciar o sistema Iniciar o sistema Parar os servidores de colecção ou todo o sistema Parar todo o sistema Parar todo o sistema Controlar quais os componentes iniciados ou parados Administrar os servidores de procura no modo autónomo Iniciar e parar serviços locais Supervisionar o sistema Supervisionar pesquisadores Como o pesquisador do Notes determina o número de documentos pesquisados Visualizar detalhes sobre a actividade de pesquisadores da Web Detalhes do módulo do pesquisador da Web Sítios activos do pesquisador da Web Criar relatórios de pesquisadores da Web Criar relatórios de pesquisadores da Web Códigos de estado de HTTP devolvidos ao pesquisador da Web	417 417 419 420 420 421 422 423 423 423 423 425 426 427 427 428 429 430
Iniciar o sistema Iniciar o sistema Parar os servidores de colecção ou todo o sistema Parar todo o sistema Parar todo o sistema Parar todo o sistema Parar todo o sistema Controlar quais os componentes iniciados ou parados Administrar os servidores de procura no modo autónomo Iniciar e parar serviços locais Supervisionar o sistema. Supervisionar pesquisadores Como o pesquisador do Notes determina o número de documentos pesquisados Visualizar detalhes sobre a actividade de pesquisadores da Web Velocidade da pesquisador da Web Sítios activos do pesquisador da Web Velocidade da pesquisa do pesquisador da Web Criar relatórios de pesquisadores da Web Códigos de estado de HTTP devolvidos ao pesquisador da Web	417 417 419 420 420 421 422 423 423 423 423 425 426 427 427 428 429 430 434
Iniciar o sistema Iniciar o sistema Parar os servidores de colecção ou todo o sistema Parar todo o sistema Iniciar os servidores de colecção ou todo o sistema Parar todo o sistema Iniciar os servidores iniciados ou parados Iniciar os servidores de procura no modo Administrar os servidores de procura no modo Iniciar e parar serviços locais Iniciar e parar serviços locais Iniciar Supervisionar o sistema. Iniciar o pesquisadores Como o pesquisador do Notes determina o Inúmero de documentos pesquisados Visualizar detalhes sobre a actividade de Iniciar detalhes sobre a actividade de Detalhes do módulo do pesquisador da Web Iniciar relatórios de pesquisadores da Web Criar relatórios de pesquisadores da Web Iniciar relatórios de pesquisadores da Web Supervisionar os serviços de análise e de indexação	417 417 419 420 420 421 422 423 423 423 423 425 426 427 427 428 429 430 434 436
Iniciar o sistema Iniciar o sistema Parar os servidores de colecção ou todo o sistema Parar todo o sistema Parar todo o sistema Controlar quais os componentes iniciados ou parados Administrar os servidores de procura no modo autónomo Iniciar e parar serviços locais Supervisionar o sistema. Supervisionar pesquisadores Como o pesquisador do Notes determina o número de documentos pesquisados Visualizar detalhes sobre a actividade de pesquisadores da Web Detalhes do módulo do pesquisador da Web Sítios activos do pesquisadores da Web Velocidade da pesquisa do pesquisador da Web Criar relatórios de pesquisadores da Web Códigos de estado de HTTP devolvidos ao pesquisador da Web Supervisionar os serviços de análise e de indexação Visualizar detalhes sobre um URI Supervisionar documentos abandonados	417 417 419 420 420 421 422 423 423 423 423 425 426 427 428 429 430 434 436 437
Iniciar o sistema Iniciar o sistema Parar os servidores de colecção ou todo o sistema Parar todo o sistema Parar todo o sistema Parar todo o sistema Controlar quais os componentes iniciados ou parados Administrar os servidores de procura no modo autónomo Iniciar e parar serviços locais Supervisionar o sistema. Supervisionar pesquisadores Como o pesquisador do Notes determina o número de documentos pesquisados Visualizar detalhes sobre a actividade de pesquisadores da Web Detalhes do módulo do pesquisador da Web Sítios activos do pesquisador da Web Criar relatórios de pesquisadores da Web Códigos de estado de HTTP devolvidos ao pesquisador da Web Supervisionar os serviços de análise e de indexação Visualizar detalhes sobre um URI Supervisionar documentos abandonados	417 417 419 420 420 421 422 423 423 423 423 425 426 427 427 428 429 430 434 436 437 438
Iniciar o sistema Iniciar o sistema Parar os servidores de colecção ou todo o sistema Parar todo o sistema Parar todo o sistema Parar todo o sistema Controlar quais os componentes iniciados ou parados Administrar os servidores de procura no modo autónomo Iniciar e parar serviços locais Supervisionar o sistema. Supervisionar pesquisadores Como o pesquisador do Notes determina o número de documentos pesquisados Visualizar detalhes sobre a actividade de pesquisadores da Web Sítios activos do pesquisador da Web Sítios activos do pesquisador da Web Criar relatórios de pesquisadores da Web Códigos de estado de HTTP devolvidos ao pesquisador da Web Supervisionar os serviços de análise e de indexação Visualizar detalhes sobre um URI Supervisionar documentos abandonados Supervisionar documentos abandonados	417 417 419 420 420 421 422 423 423 423 423 425 426 427 427 427 427 427 428 429 430 434 436 437 438 439
Iniciar o sistema Iniciar o sistema Parar os servidores de colecção ou todo o sistema Parar todo o sistema Parar todo o sistema Controlar quais os componentes iniciados ou parados Administrar os servidores de procura no modo autónomo Iniciar e parar serviços locais Supervisionar o sistema. Supervisionar pesquisadores Como o pesquisador do Notes determina o número de documentos pesquisados Visualizar detalhes sobre a actividade de pesquisadores da Web Velocidade da pesquisador da Web Stitos activos do pesquisador da Web Criar relatórios de pesquisadores da Web Códigos de estado de HTTP devolvidos ao pesquisador os serviços de análise e de indexação Visualizar detalhes sobre um URI Supervisionar os serviços de análise e de indexação Visualizar detalhes sobre um URI Supervisionar os serviços de análise e de indexação Visualizar detalhes sobre um URI Supervisionar os serviços de análise e de indexação Visualizar detalhes sobre um URI Supervisionar os serviços de análise e de indexação <td>417 417 419 420 420 421 422 423 423 423 423 423 425 426 427 427 427 428 429 430 434 436 437 438 439 441</td>	417 417 419 420 420 421 422 423 423 423 423 423 425 426 427 427 427 428 429 430 434 436 437 438 439 441
Iniciar o sistema Iniciar o sistema Parar os servidores de colecção ou todo o sistema Parar todo o sistema Parar todo o sistema Controlar quais os componentes iniciados ou parados Administrar os servidores de procura no modo autónomo Iniciar e parar serviços locais Supervisionar o sistema. Supervisionar pesquisadores Como o pesquisador do Notes determina o número de documentos pesquisados Visualizar detalhes sobre a actividade de pesquisadores da Web Sítios activos do pesquisador da Web Sítios activos do pesquisadores da Web Criar relatórios de pesquisadores da Web Códigos de estado de HTTP devolvidos ao pesquisador da Web Supervisionar os serviços de análise e de indexação Visualizar detalhes sobre um URI Supervisionar os serviços de análise e de indexação Visualizar detalhes sobre um URI Supervisionar os serviços de análise e de indexação Visualizar detalhes sobre um URI Supervisionar os serviços de análise e de indexação Visualizar detalhes sobre um URI Supervisionar os serviços de análise. Super	417 417 419 420 420 421 422 423 423 423 423 423 425 427 427 427 427 427 427 429 430 434 436 437 438 439 441

Ficheiros de registo e alertas			. 445
--------------------------------	--	--	-------

Alentee		445
Alertas	•	. 445
Configurar alertas ao nível da colecção		. 446
Configurar alertas ao nível do sistema		. 447
Configurar ficheiros de registo		. 448
Configurar informações de servidor SMTP		. 449
Receber correio electrónico sobre mensagens		
registadas		. 450
Visualizar ficheiros de registo		. 451
Visualizar ficheiros de registo utilizando a conso	ola	
de administração		. 453

Efectuar cópia de segurança e

restaurar o sistema		455
Efectuar a cópia de segurança do sistema		. 456
Restaurar o sistema		. 457
Exportar e importar configurações da colecção		. 458

Integração com o IBM Cognos BI. . . 463

Gerar relatórios IBM Cognos BI		. 464
Exportar documentos para IBM Cognos BI		. 466

Exportar documentos para utilização

noutras aplicações	469
Exportar documentos pesquisados ou analisados	472
Exportar documentos a partir de resultados de	
consulta	. 475
Exportar para um directório Windows partilhado	479
Caminhos de saída para documentos exportados	480
Formato de ficheiro XML para documentos	
exportados	. 481
Correlações de bases de dados relacionais para	
documentos exportados de colecções de procura	405
Empresarial	. 485
Exportar documentos como incheiros CSV	. 492
exportados	195
	. 495
Importar ficheiros CSV	497
	T J1
Importantichoirea CSV para uma calaccão	407
Importar ficheiros CSV para uma colecção	. 497
Importar ficheiros CSV para uma colecção Integração no Lotus Notes Versão 8	. 497 499
Importar ficheiros CSV para uma colecção Integração no Lotus Notes Versão 8 Criar o sítio de actualização do suplemento	. 497 499 499
Importar ficheiros CSV para uma colecção Integração no Lotus Notes Versão 8 Criar o sítio de actualização do suplemento Instalar o suplemento de pesquisa no cliente do	. 497 499 . 499
Importar ficheiros CSV para uma colecção Integração no Lotus Notes Versão 8 Criar o sítio de actualização do suplemento Instalar o suplemento de pesquisa no cliente do Lotus Notes versão 8	. 497 499 . 499 . 499
Importar ficheiros CSV para uma colecção. . Integração no Lotus Notes Versão 8 Criar o sítio de actualização do suplemento . Instalar o suplemento de pesquisa no cliente do Lotus Notes versão 8. .	. 497 499 . 499 . 499
Importar ficheiros CSV para uma colecção Integração no Lotus Notes Versão 8 Criar o sítio de actualização do suplemento Instalar o suplemento de pesquisa no cliente do Lotus Notes versão 8	. 497 499 . 499 . 499 501
Importar ficheiros CSV para uma colecção.Integração no Lotus Notes Versão 8Criar o sítio de actualização do suplementoInstalar o suplemento de pesquisa no cliente doLotus Notes versão 8.Integração com o WebSphere Portal	. 497 499 . 499 . 499 501
Importar ficheiros CSV para uma colecção. . Integração no Lotus Notes Versão 8 Criar o sítio de actualização do suplemento . Instalar o suplemento de pesquisa no cliente do Lotus Notes versão 8. . Integração com o WebSphere Portal Integração com o IBM Content	. 497 499 . 499 . 499 501
Importar ficheiros CSV para uma colecção Integração no Lotus Notes Versão 8 Criar o sítio de actualização do suplemento Instalar o suplemento de pesquisa no cliente do Lotus Notes versão 8	. 497 499 . 499 . 499 501 503
Importar ficheiros CSV para uma colecção Integração no Lotus Notes Versão 8 Criar o sítio de actualização do suplemento Instalar o suplemento de pesquisa no cliente do Lotus Notes versão 8	. 497 499 . 499 . 499 501 503
Importar ficheiros CSV para uma colecção. Integração no Lotus Notes Versão 8 Criar o sítio de actualização do suplemento Instalar o suplemento de pesquisa no cliente do Lotus Notes versão 8 Integração com o WebSphere Portal Integração com o IBM Content Classification Anotador do Content Classification Planos de decisão e pontuações por categoria do	. 497 499 . 499 . 499 501 503 . 503
Importar ficheiros CSV para uma colecção. Integração no Lotus Notes Versão 8 Criar o sítio de actualização do suplemento Instalar o suplemento de pesquisa no cliente do Lotus Notes versão 8 Integração com o WebSphere Portal Integração com o IBM Content Classification Anotador do Content Classification Planos de decisão e pontuações por categoria do Content Classification	. 497 499 . 499 501 503 . 503 . 504
Importar ficheiros CSV para uma colecção. Integração no Lotus Notes Versão 8 Criar o sítio de actualização do suplemento Instalar o suplemento de pesquisa no cliente do Lotus Notes versão 8. Integração com o WebSphere Portal Integração com o IBM Content Classification Anotador do Content Classification Planos de decisão e pontuações por categoria do Content Classification Content Classification Content Classification	. 497 499 . 499 501 503 . 503 . 504
Importar ficheiros CSV para uma colecção. Integração no Lotus Notes Versão 8 Criar o sítio de actualização do suplemento Instalar o suplemento de pesquisa no cliente do Lotus Notes versão 8. Integração com o WebSphere Portal Integração com o IBM Content Classification Planos de decisão e pontuações por categoria do Content Classification	. 497 499 . 499 501 503 . 503 . 504 . 506
Importar ficheiros CSV para uma colecção. Integração no Lotus Notes Versão 8 Criar o sítio de actualização do suplemento Instalar o suplemento de pesquisa no cliente do Lotus Notes versão 8. Integração com o WebSphere Portal Integração com o IBM Content Classification Anotador do Content Classification Planos de decisão e pontuações por categoria do Content Classification Configurar os campos de índice e pontuações do Content Classification Treinar bases de conhecimento do Content	. 497 499 . 499 501 503 . 503 . 504 . 506
Importar ficheiros CSV para uma colecção. Integração no Lotus Notes Versão 8 Criar o sítio de actualização do suplemento Instalar o suplemento de pesquisa no cliente do Lotus Notes versão 8 Integração com o WebSphere Portal Integração com o IBM Content Classification Anotador do Content Classification Planos de decisão e pontuações por categoria do Content Classification Configurar os campos de índice e pontuações do Content Classification Treinar bases de conhecimento do Content Classification	. 497 499 . 499 501 503 . 503 . 504 . 506 . 507
Importar ficheiros CSV para uma colecção. Integração no Lotus Notes Versão 8 Criar o sítio de actualização do suplemento Instalar o suplemento de pesquisa no cliente do Lotus Notes versão 8. Integração com o WebSphere Portal Integração com o IBM Content Classification Anotador do Content Classification Planos de decisão e pontuações por categoria do Content Classification Configurar os campos de índice e pontuações do Content Classification Treinar bases de conhecimento do Content Classification Content Classification Melhorar pontuação com base em categorias e	. 497 499 . 499 501 503 . 503 . 504 . 506 . 507

classificação de pesquisa conceptual. 509

Comandos, códigos de retorno e IDs	511
	011
Dependência de maiúsculas e	
minúsculas	541
Detecção e resolução de problemas	
da administração	543
Ferramentas para solucionar problemas	543
Recolher informações para a análise de	
problemas	543
Obter informação sobre erros de CCL	545
Problemas ao iniciar e parar o sistema	546
Um servidor do sistema pára	546
O comando esadmin stop tem de preceder o	
comando stopcel	546
Problemas com pesquisa de documentos	547
Pesquisador da Web pára inesperadamente	547
Visualizar valor de início de sessão do	- 10
pesquisador Web no navegador Firetox	548
Controlar redirecção de HITP do pesquisador	= 10
	549
O botao Seguinte nao esta disponível durante a	E 40
configuração do pesquisador	549
De sum en fune	FEO
Documentum	550
documentos	551
Não é possível iniciar o applicador	551
Analisador permanece po estado de espera	551
Analisador falha ao iniciar depois de associar	551
Circumping and falle and the international de la colecção	552
Sincronização faina quando ficheiros maiores de	550
O extractor de texte falha guando detecta	552
documentos corrompidos	553
Os dados de saída de rastreio de pilha	555
aparecem depois de o extractor de texto parar o	
processamento	554
Problemas com aplicações e servidores de procura	554
Nenhum processo de procura configurado (erro	555
Iniciar sessão numa anlicação demora muito	555
tempo	556
Tempos de resposta de procura lentos no Linux	557
Não é possível registar o ID da aplicação	558
Clonar a portlet Pesquisa para WebSphere	000
Portal	559
Problemas operacionais	560
Conflitos de portas causam erros de socket	560
Problemas de atribuição de sockets causam	
falha de operações com erro FFQO0143E	560
Recuperar de uma falha de disco cheio	562
Não é possível efectuar cópia de segurança ou	
restaurar a informação da base de dados depois de executar a cópia de segurança e restaurar	
scripts.	562
Reiniciar registo centralizado	563

Definir a palavra-passe de administrador codificada para ser a mesma em todos os					
servidores	•	•	•		. 564
Documentação do produto					565
Funções de acessibilidade					567

Glossário											•		569
Avisos . Marcas come	• rcia	• ais	•	•	•	•	•	•	•	•	•	•	583 . 585
Índice Rer	ni	ssi	ivo	э.									587

ibm.com e recursos relacionados

Está disponível suporte e documentação de produtos a partir de ibm.com.

Suporte e assistência

Está disponível suporte de produtos na Web.

IBM[®] Content Analytics with Enterprise Search

http://www.ibm.com/support/entry/portal/Overview/Software/ Enterprise_Content_Management/Content_Analytics

IBM Cognos Business Intelligence

http://www.ibm.com/support/entry/portal/Overview/Software/ Cognos/Cognos_Business_Intelligence

IBM Content Classification

http://www.ibm.com/support/entry/portal/Overview/Software/ Enterprise_Content_Management/Classification_Module

IBM Content Integrator

http://www.ibm.com/support/entry/portal/Overview/Software/ Enterprise_Content_Management/Content_Integrator

IBM InfoSphere BigInsights

http://www.ibm.com/support/entry/portal/Overview/Software/ Information_Management/InfoSphere_BigInsights

Centro de Informações

Pode ver a documentação do produto num Centro de Informações baseado no Eclipse com um navegador da Web. O Centro de Informações permite-lhe procurar em toda a documentação e poderá estar mais actualizado do que as publicações PDF. Para obter informações sobre as actualizações de documentos, caso aplicável, consulte a página inicial do Centro de Informações em http://pic.dhe.ibm.com/ infocenter/analytic/v3r0m0/.

Publicações em PDF

Pode ver os ficheiros PDF em linha utilizando o Adobe Acrobat Reader para o seu sistema operativo. Se não tem o Adobe Reader instalado, pode transferi-lo a partir do sítio da Web da Adobe em http://www.adobe.com.

Consulte os seguintes sítios de Web com publicações em PDF:

Produto	Endereço do sítio da Web
IBM Content Analytics with	http://www.ibm.com/support/
Enterprise Search Versão 3.0	docview.wss?uid=swg27023678
IBM Cognos Business Intelligence	http://www.ibm.com/support/
Versão 10.1	docview.wss?uid=swg27018884
IBM Content Classification Versão	http://www.ibm.com/support/
8.8	docview.wss?uid=swg27020843
IBM Content Integrator Versão 8.6	https://www.ibm.com/support/ docview.wss?uid=swg27016951

Produto	Endereço do sítio da Web
IBM InfoSphere BigInsights Versão	Os PDFs não estão disponíveis. Para obter mais
1.5	publib.boulder.ibm.com/infocenter/bigins/v1r3/

Como enviar comentários

Os seus comentários são importantes para nos ajudar a fornecer as informações mais exactas da mais alta qualidade.

Envie os seus comentários utilizando o formulário para os comentários dos leitores em linha em https://www14.software.ibm.com/webapp/iwm/web/signup.do?lang=en_US&source=swg-rcf.

Não hesite em comunicar à IBM como melhorar utilização de produtos de software. Se quiser ajudar a IBM a tornar mais fácil a utilização do IBM Content Analytics with Enterprise Search, responda ao Consumability Survey em http://www.ibm.com/software/data/info/consumability-survey/.

Contactar a IBM

Para contactar a assistência a clientes IBM nos E.U.A. ou Canadá, ligue 1-800-IBM-SERV (1-800-426-7378).

Para obter informações sobre as opções de assistência disponíveis, telefone para um dos seguintes números:

- Nos E.U.A.: 1-888-426-4343
- No Canadá: 1-800-465-9600

Para obter mais informações sobre como contactar a IBM, consulte o sítio da Web de Contacto da IBM em http://www.ibm.com/contact/us/.

Novidades na Versão 3.0

O IBM Content Analytics with Enterprise Search Versão 3.0 introduz muitas funcionalidades novas para planeadores, administradores, utilizadores e programadores de aplicações.

Alterações de terminologia

O nome do produto e os nomes de diversos componentes foram alterados para reflectir melhor a convergência das duas tecnologias chave: análise de conteúdos e procura empresarial.

Termos em versões	
anteriores	Termos novos na Versão 3.0
IBM Content Analytics	IBM Content Analytics with Enterprise Search
LanguageWare Resource Workbench	Content Analytics Studio
aplicação de prospecção de texto	Content Analytics Miner
colecção de análise de texto	colecção de análise de conteúdo
aplicação de procura	aplicação de procura empresarial
colecção de procura	colecção de procura empresarial
Analisador Stellent	extractor de texto
campos de procura	campos de índice
palavras-chave da faceta	valores da faceta
Pesquisador do QuickPlace	Pesquisador do Quickr for Domino
Nomes de produtos da IBM	Nomes de produtos novos:
relacionados:	IBM Content Classification
• IBM Classification Module	IBM Connections
Lotus Connections	IBM Web Content Manager
Lotus Web Content Management	Ŭ

Tabela 1. Alterações de terminologia

Melhoramentos para planeadores e administradores

Estão disponíveis novas funcionalidades para administrar o sistema e configurar colecções.

Tabela 2. Melhoramentos para planeadores e administradores

Melhoramento	Descrição
Suporte do Linux expandido no System z	Num sistema Linux em System z, pode pesquisar tipos de origens de dados adicionais. A assistência inclui o conteúdo pesquisado por estes pesquisadores: Agente para sistemas de ficheiros Windows, Content Integrator (EMC/Documentum 6.5 origens apenas), Base de dados JDBC (DB2 e DB2 para bases de dados zOS apenas), SharePoint, Web e Sistema de ficheiros UNIX.

Melhoramento	Descrição
IBM InfoSphere BigInsights	Num sistema Linux de 64-bits, pode adicionar um servidor IBM InfoSphere BigInsights à topologia do sistema para assistir na análise de grande escala e na indexação de conteúdo, tais como uma colecção que contém terabytes de dados.
Vistas-resumo de administração	No passado, os administradores alternavam entre o modo de edição e o modo de supervisão quando administravam colecções ou o sistema. Agora, a interface do utilizador é simplificada em três vistas-resumo para colecções, o sistema e a segurança. As vistas-resumo permitem aos administradores editar facilmente definições e supervisionar a actividade a partir da mesma vista.
Suporte expandido do servidor de aplicação da Web	Quando instala o IBM Content Analytics with Enterprise Search, pode especificar que pretende utilizar o WebSphere Application Server para suporte da aplicação da Web em vez do servidor da aplicação incorporado. A consola de administração e o servidor REST API podem ser implementados no WebSphere Application Server. No passado, estes componentes necessitavam de um servidor da aplicação incorporado.
Suporte de idioma expandido	As colecções de análise de conteúdos agora suportam checo, hebraico, Polaco e russo, para além dos idiomas que eram suportados anteriormente.
Procura segura	Diversas melhoras expandem a respectiva capacidade de procurar colecções seguras:
	• A procura de facetas é suportada quando os utilizadores consultam colecções seguras. A pós-filtragem dos resultados da procura não se aplica às facetas devolvidas. Se a pós-filtragem estiver activada para uma colecção, os administradores podem configurar quais são as facetas que devem ser devolvidas quando os utilizadores consultam a colecção.
	• A procura segura é suportada se utilizar múltiplas instâncias do WebSphere Portal que requerem formatos de credenciais diferentes. O nível de suporte depende da respectiva versão do WebSphere Portal entre outros factores.
	• Suporte de início de sessão único (SSO) expandido.
Configuração de servidor LDAP	Para controlar o acesso a aplicações e para fornecer início de sessão único, os novos assistentes e botões de teste ajudam-no a configurar o servidor de LDAP e o ficheiro de chave LTPA. Pode testar a capacidade do servidor para ligar ao servidor LDAP. Também pode verificar que o formato das entradas de utilizadores e grupos corresponde aos requisitos do registo de utilizadores.
Funções de utilizador de aplicação e administrador expandidas	Para controlar o acesso às funções administrativas, pode limitar alguns administradores a partes específicas de uma colecção, tais como a configuração da árvore da faceta ou dos dicionários. Também pode controlar quais as funções que os utilizadores podem desempenhar numa aplicação através da atribuição de privilégios a utilizadores ou grupos. Por exemplo, os novos privilégios permitem-lhe limitar quais os utilizadores que podem alterar o esquema dos painéis na aplicação de procura empresarial.

Tabela 2. Melhoramentos para planeadores e administradores (continuação)

Melhoramento	Descrição
Clonar colecções	Pode criar a colecção ao clonar uma colecção existente. Clonar uma colecção copia todos os dados de configuração de uma colecção para uma colecção nova, incluindo todos os pesquisadores, opções para pesquisa, análise e indexação de conteúdos, anotadores e opções para procurar a colecção. Pode, em seguida, modificar apenas as definições que pretende alterar para a nova colecção, tais como a pesquisa de origens de dados diferentes ou a activação de um anotador personalizado diferente. Para procurar a colecção clonada, tem de pesquisar ou importar dados para que o índice desta nova colecção possa ser construído.
Melhorias do campo de índice	A interface para definir campos de índice foi melhorada para apresentar mais informações sobre a configuração de cada campo de índice, incluindo quais são as origens de entrada que estão correlacionadas para cada campo. As ligações permitem-lhe aceder rapidamente a uma origem de entrada para alterar como um campo de origem ou uma faceta se correlaciona com o campo de índice.
Correlacionamento directo da faceta	Pode criar facetas ao correlacionar estruturas das funcionalidades numa estrutura de análise comum (CAS) com o índice em vez de correlacionar campos de índice com uma faceta.
Correlacionamento do campo de índice personalizado para conteúdo binário	O componente do extractor de texto pode extrair texto e metadados de documentos binários, tais como ficheiros .doc e .pdf e correlacionar automaticamente os metadados extraídos para campos de índice. Pode controlar este comportamento de correlação automático. Por exemplo, pode configurar o extractor de texto para correlacionar o nome do ficheiro para um campo de índice em vez do título do documento.
Campos de data personalizados	A data do documento é crítica para a análise de texto, especialmente para explorar como os dados são alterados ao longo do tempo e para observar desvios e tendências. Quando configura o analisador para uma colecção, pode especificar formatos de data personalizados para garantir que os dados da data que adiciona à colecção são correlacionados com o campo de índice de data fornecido e correctamente indexados.
Suporte de conteúdo multimédia personalizado	Pode adicionar algoritmos de análise personalizada para conteúdo multimédia, como um anotador da UIMA. Por exemplo, poderá necessitar de um algoritmo personalizado para detectar objectos de imagem ou para classificar imagens. Pode correlacionar os resultados de análise personalizada com campos de índice ou facetas.
Regras e dicionários personalizados para expansão da consulta	Quando configura uma colecção de procura empresarial, pode adicionar regras e dicionários personalizados para expandir consultas quando os utilizadores procuram a colecção. Se uma consulta incluir palavras que correspondem a um padrão que especificou numa regra, a consulta é expandida automaticamente para aplicar a regra antes de devolver os resultados. Por exemplo, os termos da consulta pode ser substituídos ou aumentado por palavras de um dicionário. A regra também pode controlar a classificação, tais como classificar documentos que correspondem à consulta expandida com uma classificação mais elevada do que documentos que correspondem aos termos da consulta original.

Tabela 2. Melhoramentos para planeadores e administradores (continuação)

Melhoramento	Descrição
Reconhecimento de entidade nomeada personalizada	O anotador de reconhecimento de entidade nomeada pode ser activado para colecções de procura empresarial. No passado, este anotador estava disponível apenas para colecções de análise de conteúdos. Por predefinição, o sistema correlaciona entidades com facetas que são nomeadas Pessoa, Localização e Organização. Para ambos os tipos de colecções, pode agora especificar nomes personalizados para novas facetas ou correlacionar as entidades com facetas simples existentes. Os idiomas suportados são inglês, francês, alemão, japonês e espanhol. O reconhecimento de entidades nomeadas é suportado em chinês, mas a definição de entidade personalizada não é.
Análise global personalizada para uma colecção no BigInsights	Quando configura uma colecção de análise de conteúdos num servidor IBM InfoSphere BigInsights, pode integrar a respectiva lógica personalizada para análise global. Por exemplo, pode aplicar acções a todos os documentos na colecção em oposição a documentos que foram adicionados por um pesquisador. Pode aplicar a sua análise global personalizada, para além das funções de análise global fornecidas, tais como a análise de ligação, a criação de imagens em miniatura e a detecção de documentos duplicados.
Melhorias do pesquisador do Exchange Server	O pesquisador suporta tanto as estruturas de Pasta pública como de Pasta pessoal do Microsoft Exchange Server 2007 SP1 e do Exchange Server 2010. No passado, eram suportadas apenas pastas públicas.
Melhorias do pesquisador do Notes	Se configurar o pesquisador para pesquisar vistas e pastas específicas numa base de dados, pode agora especificar se o pesquisador deve pesquisar os valores dos campos em vistas (os valores das colunas) para além de pesquisar campos de documentos. Pesquisar os valores da coluna em vistas da base de dados fornece uma forma eficaz de recolher o conteúdo a partir da base de dados.
Melhorias do pesquisador da Lista de valores geradores	Em vez de utilizar os registos do histórico do pesquisador para calcular os documentos novos, actualizados e eliminados, o pesquisador pode ser configurado para utilizar a funcionalidade da lista de valores geradores incremental no IBM Web Content Manager. Quando activada, o servidor da lista de valores geradores fornece as diferenças da pesquisa anterior. Também foram efectuadas melhorias para garantir que as correlações de campo de título predefinidas para os documentos do IBM Connections, tais como os títulos em wikis, fóruns e blogues estejam correctas.
Melhorias do pesquisador da SharePoint	O suporte é expandido ao Microsoft SharePoint 2010. Para além disso, para o ajudar a incluir múltiplos sítios de nível superior no espaço de pesquisa, todos os sítios de nível superior que pertencem a uma farm num servidor SharePoint podem ser detectados automaticamente. Com o suporte com noção de farm, pode configurar um pesquisador para recolher conteúdo a partir de múltiplos sítios de nível superior em vez de configurar um pesquisador para cada servidor SharePoint.
Melhoria do pesquisador da Web	Para limitar o espaço de pesquisa, pode controlar quantas ligações o pesquisador da Web deve seguir à medida que segue ligações de um documento para outro.

Tabela 2. Melhoramentos para planeadores e administradores (continuação)

Melhorias para a prospecção de conteúdo

Várias novas funcionalidades estão disponíveis para o Content Analytics Miner.

Melhoramento	Descrição
Flexibilidade de escala de tempo	Estão disponíveis mais opções para o ajudar a explorar o conteúdo nas vistas temporais (Série Cronológica, Desvios e Tendências). Por exemplo, pode seleccionar a partir de uma grande variedade de escalas temporais e de escalas temporais cíclicas (tais como por anos, meses, dias, meses do ano, dias do mês e dias da semana). Também pode controlar a quantidade de dados que são apresentados numa altura ao especificar o número de unidades de tempo por gráfico e uma data de início para a escala temporal.
Análise de postura	Uma nova vista de análise permite-lhe explorar expressões positivas e negativas à medida que explora uma colecção a partir de diferentes facetas. Também pode ver uma cronologia que demonstra como a postura se altera ao longo do tempo. Na vista Documentos, são realçadas expressões positivas, negativas e ambíguas no texto para uma fácil identificação e revisão. Os idiomas suportados são inglês, holandês, francês, alemão, japonês e espanhol.
Vistas contextuais	Um administrador pode configurar o sistema para analisar campos específicos numa colecção de análise de conteúdo. Quando explora a colecção, pode optar por explorar apenas as secções dos documentos que correlacionam para estes campos, que lhe permite focar no conteúdo que ocorre num determinado contexto. Por exemplo, a expressão componente de substituição pode ter diferentes conotações dependendo se ocorre num campo que contém reclamações de clientes ou um campo que contém resoluções.
Visualizações de análise personalizada	Os programadores podem criar suplementos personalizados para o Content Analytics Miner. Após a configuração do sistema para utilizar o suplemento, os analistas podem utilizar a vista personalizada para além das vistas de análise que são fornecidas com o produto.

Tabela 3. Melhorias para a prospecção de conteúdo

Melhorias para procura empresarial

Estão disponíveis novas funcionalidades para as aplicações de procura empresarial. Para estarem disponíveis, algumas dessas funcionalidades têm de ser activadas por um administrador. Por exemplo, os administradores podem personalizar a aplicação para activar ou desactivas as funções de análise. Os administradores também podem restringir funcionalidades ao activar e desactivas privilégios de utilizador.

Tabela 4. Melhorias para procura empresarial

Melhoramento	Descrição
Procura de facetas federada	Quando procura mais do que uma colecção de procura empresarial de cada vez, os utilizadores podem aperfeiçoar os resultados ao seleccionar facetas. A árvore da faceta apresenta a contagem da frequência total e a contagem da frequência para cada colecção. No passado, as facetas estavam disponíveis apenas quando procurava uma única colecção.

Melhoramento	Descrição
Esquema personalizado	Os utilizadores autorizados podem arrastar painéis da interface da aplicação de procura empresarial para novas localizações e redimensionar os mesmos. Por exemplo, um utilizador pode querer deslocar o painel de navegação da faceta para a extremidade direita ou deslocar a consola que apresenta a sintaxe da consulta actual para o topo.
Análise de correlação	Para ajudar os utilizadores a identificar quais os valores da faceta são provavelmente os mais relevantes, o painel de navegação da faceta pode indicar quais os valores que têm classificações de correlação mais elevadas ao apresentar as ligações da faceta em tons de cores gradativos, partido do amarelo, que indica uma classificação de correlação inferior, até ao vermelho, que indica elevada correlação. Esta função está disponível apenas se um administrador activar o modo de análise para a aplicação.
Análise de série cronológica	Um novo gráfico cronológico pode ajudar os utilizadores a visualizar como os documentos correspondem às condições de consulta actuais ao longo do tempo. Os utilizadores podem seleccionar diferentes intervalos de tempo para consultar, tais como por ano, por mês, por meses do ano ou por dias do mês. Esta função está disponível apenas se um administrador activar o modo de análise para a aplicação.
Árvore de consultas melhorada	Os melhoramentos na árvore de consulta facilita ver e modificar a sintaxe da consulta. Por exemplo, os utilizadores podem inverter um operador (tal como alterar de AND para AND NOT), mudar o operador (tal como alterar de AND para OR) e adicionar ou remover condições de procura. À medida que um utilizador edita a sintaxe da consulta, é apresentado o número de documentos que correspondem a cada expressão da consulta.
Sinalização de documentos	Os utilizadores autorizados podem filtrar os resultados ao seleccionar sinalizadores da mesma forma que estes seleccionam as facetas. Por exemplo, os utilizadores podem limitar os resultados a documentos que estejam sinalizados como Confidenciais. Os utilizadores podem adicionar, alterar ou remover sinalizadores e aplicar a alteração aos documentos seleccionados ou a todos os documentos nos resultados.
Exportar documentos sinalizadores para uma colecção de análise de conteúdos	Os utilizadores autorizados podem sinalizar documentos e exportá-los para colecções de análise de conteúdos. Um administrador tem de especificar quais são as colecções às quais os documentos devem ser adicionados e configurar correlacionamentos entre campos nos campos de documentos exportados e de índice.
Construtor de consultas interactivo	Um novo diálogo permite aos utilizadores autorizados construir consultas complexas destacando os valores de texto de documento e da faceta; especificando se a consulta será executada como uma procura de palavra-chave, procura de proximidade ou outro tipo de procura e seleccionando uma pesquisa guardada para adicionar à consulta. À medida que a consulta é construída, o número de documentos que correspondem é apresentada.
Categorias baseadas em análise de agrupamentos de documentos	Os administradores podem analisar agrupamentos de documentos no índice para classificar documentos e atribuí-los a categorias. Os utilizadores podem seleccionar as categorias para aperfeiçoar a respectiva procura.

Tabela 4. Melhorias para procura empresarial (continuação)

Melhoramentos para programadores de aplicações

As melhorias estão disponíveis para o ajudar a personalizar a análise de conteúdos e a desenvolver aplicações personalizadas.

Melhoramento	Descrição
Anotadores personalizados	Para desenvolver anotadores personalizados, pode utilizar o Content Analytics Studio como uma alternativa para desenvolver anotadores manualmente com o UIMA Software Development Kit (SDK). O Content Analytics Studio é um componente instalado separadamente do IBM Content Analytics with Enterprise Search.
REST APIs	Os APIs de administração do SIAPI e os APIs de serviços da Web já não são suportados. A depreciação destes APIs foi anunciada anteriormente. As APIs SIAPI que não são de administração estão a ser depreciados e não serão suportados em publicações futuras. Para desenvolver aplicações personalizadas, utilize os APIs REST de Administração e Procura:
	 Estão disponíveis informações sobre a utilização das APIs REST no directório ES_INSTALL_ROOT/docs/api/rest.
	 Estão disponíveis situações exemplo para executar tarefas de administração no directório ES_INSTALL_ROOT/samples/rest.
	A versão 3.0 inclui mais exemplos de API REST e documentação melhorada sobre como utilizar os programas exemplo.Foram alterados diversos números de portas predefinidas. Para se certificar que as aplicações que migra funcionam com as portas da Versão 3.0 predefinidas, poderá necessitar de alterar as respectivas portas da aplicação para a nova configuração do número da porta.
Análise em tempo real	As colecções de procura empresarial agora suportam a API de processamento de idioma natural (NLP) em tempo real. No passado, esta API estava disponível apenas para colecções de análise de conteúdos. Com esta API, pode executar análises de texto ad-hoc em documentos para analisar conteúdos sem adicionar os documentos ao índice.

Tabela 5. Melhoramentos para programadores de aplicações

Conceitos relacionados:

- Diferenças entre tipos de colecções
- E Configuração do número da porta

Análise de conteúdo e procura empresarial

A análise de conteúdos ajuda as organizações a obter uma nova compreensão e visibilidade empresarial dos conteúdos e contextos de informações não estruturadas. A procura empresarial ajuda as organizações a produzir conteúdo a partir de múltiplas origens estruturadas e não estruturadas passíveis de serem procuradas por utilizadores empresariais.

O IBM Content Analytics with Enterprise Search combina os pontos positivos destas duas tecnologias. Através de uma interface de prospecção de conteúdo fornecida, os analistas empresariais podem explorar dados de forma interactiva a partir de diferentes facetas e descobrir relações entre vários valores da faceta. Através de uma interface concebida para a procura, os utilizadores empresariais podem consultar o índice para encontrar e obter documentos relevantes a partir de uma lista de resultados classificada.

Análise de conteúdos

Numa vasta gama de indústrias, a análise de conteúdos pode ajudar a enfrentar vários desafios de informação, tais como a necessidade de:

- Melhorar a satisfação do cliente através de análises de grandes volumes de comentários e retornos da satisfação cliente
- Ganhar melhor visibilidade no mercado através de notícias, inquéritos e análises de marca automatizados
- Melhorar previsão das necessidades de clientes através da identificação das tendências em comunicações de clientes desestruturadas
- Optimizar os processos que requerem muita documentação através da classificação inteligente e do encaminhamento de itens dos conteúdos
- Antecipar problemas de qualidade do produto através da análise de queixas, garantias, reparações e senhas de apoio
- Reduzir a fraude através da análise inteligente de formulários, documentos e comunicações
- Melhorar pesquisas e investigações através de dados combinados e análise de conteúdos

As ferramentas de análise de conteúdos da IBM permitem-lhe formular perguntas que saem do âmbito dos serviços de informação empresarial baseados em dados normais e ajudam-o a identificar novos tópicos, áreas e perguntas que carecem de maior investigação.

Procura empresarial

Através da análise semântica, o IBM Content Analytics with Enterprise Search pode identificar conceitos, significados latentes, relações, factos e outros dados relevantes que frequentemente estão ocultos em texto não estruturado. A informação que é extraída durante a análise pode ser utilizada para melhorar a qualidade dos resultados da procura. Também pode ser utilizado para melhorar a qualidade de outras aplicações, tais como informação empresarial e prospecção de dados.

"Descrição geral dos componentes" na página 15

"Como os dados fluem ao longo do sistema" na página 23

"Conteúdo empresarial"

"Encadeamento de análise aberto e escalável"

"Interface do utilizador de prospecção de conteúdo interactiva" na página 11

"Interface de utilizador de procura empresarial interactiva" na página 13

Referências relacionadas:

"Origens de dados suportadas" na página 24

Conteúdo empresarial

A análise e procura de conteúdos é iniciada com a recolha de conteúdo empresarial relevante.

O IBM Content Analytics with Enterprise Search faculta pesquisadores para uma vasta variedade de origens de conteúdos empresariais que permitem que os conteúdos, metadados e dados estruturados e desestruturados sejam recolhidos e alimentados num encadeamento da análise. A configuração do pesquisador é flexível e escalável, permitindo-lhe recolher as informações de origem que precisa para responder de forma inteligente às perguntas baseadas nos conteúdos.

Para além dos pesquisadores fornecidos, pode adicionar conteúdo ao importar ficheiros de valor separado por vírgulas (CSV). Os assistentes ajudam-o a seleccionar os ficheiros a adicionar e especificam quais são os valores a serem indexados e a disponibilizar nas respectivas aplicações.

Referências relacionadas:

🕩 Origens de dados suportadas no IBM Content Analytics with Enterprise Search

Encadeamento de análise aberto e escalável

À medida que o conteúdo é pesquisado, os documentos e registos são alimentados no encadeamento de análise do IBM Content Analytics with Enterprise Search.

Cada item é convertido em texto e enviado por uma série de passos de processamento e análise. Cada passo *anota* o item de conteúdo com mais informações e limpa, clarifica e extrai significado do item.

Tal como fornecido, o IBM Content Analytics with Enterprise Search inclui um poderoso conjunto de anotadores para anotar e extrair significado do conteúdo. Os anotadores:

- Detectam a linguagem de origem e a codificação de caracteres do conteúdo
- Segmentam o texto em palavras
- Identificam as categorias gramaticais de diversas palavras
- · Localizam segmentos de palavras, como expressões ou pares nome-adjectivo
- Extraem nomes de entidades do texto automaticamente, como pessoas, localizações e organizações
- Categorizar e classificar automaticamente os itens de conteúdo através de IBM Content Classification
- Localizam padrões personalizados no textos mediante expressões globais definidas pelo cliente

 Procuram termos relevantes do dicionário definido pelo cliente, tais como nomes de produtos e marcas

Como o IBM Content Analytics with Enterprise Search implementa o contexto Unstructured Information Management Architecture aberto (UIMA), pode personalizar completamente o processamento de conteúdo e tirar partido dos anotadores UIMA existentes, tanto de código aberto como comercial, para melhorar as funções de análise de conteúdo automatizadas.

Para desenvolver anotadores personalizados, pode utilizar o Content Analytics Studio como uma alternativa para desenvolver anotadores manualmente com o UIMA Software Development Kit (SDK). O Content Analytics Studio é um componente instalado separadamente do IBM Content Analytics with Enterprise Search.

A análise aprofundada e automatizada dos conteúdos pode ser uma operação que requer muitos recursos informáticos. Para apoiar a análise rápida e eficiente de grandes quantidades de conteúdos, o IBM Content Analytics with Enterprise Search fornece uma implementação altamente escalável da especificação UIMA, que lhe permite distribuir a análise de conteúdos por várias máquinas.

Conceitos relacionados:

Anotadores

Content Analytics Studio

Interface do utilizador de prospecção de conteúdo interactiva

O encadeamento de análise do IBM Content Analytics with Enterprise Search anota ou realça conteúdos desestruturados, extraindo eficazmente dados com sentido de textos desestruturados.

Para além dos dados extraídos, podem existir metadados ou outros itens estruturados anteriormente associados ao conteúdo, tais como datas, palavras-chave e atributos ou elementos XML. Tanto os elementos de dados extraídos como os elementos de metadados previamente existentes são todos combinados num único conjunto de *facetas* para cada item de conteúdo.

A interface do utilizador do Content Analytics Miner permite-lhe explorar as facetas do seu conjunto de conteúdo. Em tempo real, pode procurar por todas as facetas e o texto original do seu conteúdo, efectuar uma pesquisa detalhada numa determinada faceta, ver as relações entre facetas e observar como as facetas do conteúdo mudam com o tempo.

Conceitos relacionados:

- "Análise de frequência e correlações"
- "Análises periódicas e altamente escaláveis" na página 12
- "Exportar facetas para melhorar análises e comunicações" na página 12

Análise de frequência e correlações

Através de vistas diferentes, a interface do utilizador do Content Analytics Miner pode demonstrar-lhe tendências e anomalias no conjunto de conteúdos.

Muitas destas estatísticas baseiam-se na frequência em que ocorre determinado valor de uma faceta. Por exemplo, pode ir vendo ao longo do tempo um número crescente de referências a um determinado produto competitivo nos registos do centro de atendimento. Ou, outro exemplo, pode observar a utilização frequente de uma determinada abreviatura ou palavra por determinadas pessoas durante uma auditoria ou investigação. Este tipo de análise de frequência bruta pode ser útil para rapidamente identificar os termos e tópicos em voga.

O Content Analytics Miner, contudo, vai mais longe para o ajudar a localizar os dados de elevada frequência que de facto interessam. No exemplo de auditoria precedente, o simples facto de um termo ser usado frequentemente por alguém não significa que seja importante. A verdadeira questão é saber se esse termo é usado com maior frequência por esse indivíduo do que qualquer outra pessoa. Com as avaliações de correlação que são processadas pelo Content Analytics Miner, pode detectar rapidamente estes tipos de anomalias, ainda que sejam pouco frequentes.

Conceitos relacionados:

"Interface do utilizador de prospecção de conteúdo interactiva" na página 11

Análises periódicas e altamente escaláveis

Há determinadas análises que são mais bem eficazes quando feitas continuamente. O IBM Content Analytics with Enterprise Search permite análises agendadas e contínuas de um conjunto de dados para identificar novas perspectivas à medida que vão sendo adicionados itens de conteúdos.

Podem ser enviados relatórios XML sobre os resultados da análise de conteúdos a outras aplicações, incluindo as bases de dados relacionais. Para além disso, as consultas de análise altamente escalável, que contêm um elevado número de critérios de faceta e de valor da faceta, podem ser emitidas através da utilização do mesmo mecanismo de agendamento. O IBM Content Analytics with Enterprise Search permite não só que estas consultas altamente escaláveis sejam agendadas, como também podem ser automaticamente decompostas e distribuídas para vários servidores durante o tempo de execução para um processamento mais rápido.

Conceitos relacionados:

"Interface do utilizador de prospecção de conteúdo interactiva" na página 11

Exportar facetas para melhorar análises e comunicações

O IBM Content Analytics with Enterprise Search pode ser integrado com as aplicações de armazém de dados e de informação empresarial tais como o InfoSphere Warehouse e o Cognos Business Intelligence através de funções de exportação da faceta e comunicação.

Quando o IBM Content Analytics with Enterprise Search extrai dados estruturados de textos não estruturados, as facetas extraídas podem ser exportadas directamente para um armazém num modelo de esquema em estrela. Estas exportações podem ser agendadas e incrementadas, o que facilita a criação de relatórios operacionais regularmente actualizadas e com sentido.

Depois das facetas terem sido exportadas, pode utilizar as funções do InfoSphere Warehouse e do Cognos para obter mais capacidades de extracção-transformaçãocarregamento, criação de modelos, granulação e elaboração de relatórios.

Na interface do utilizador do Content Analytics Miner, pode criar relatório e abrir os relatórios no IBM Cognos Business Intelligence, que lhe permite explorar correlações, tendências, desvios e relações entre diferentes facetas com ferramentas de informação empresarial. Também pode guardar relatórios num formato de valor separado por vírgulas (CSV) e, em seguida, importar os ficheiros CSV para a folha de cálculo e outras aplicações.

Conceitos relacionados:

"Interface do utilizador de prospecção de conteúdo interactiva" na página 11

Interface de utilizador de procura empresarial interactiva

O IBM Content Analytics with Enterprise Search fornece uma aplicação de procura de exemplos que pode personalizar ou utilizar como um modelo para as próprias aplicações.

A interface do utilizador para a procura empresarial ajuda-o a aperfeiçoar rapidamente os resultados para documentos que são os mais relevantes para a consulta. As funcionalidades de introdução adiantada e de criação de consultas ajudam-o a especificar consultas exactas. Pode facilmente aperfeiçoar resultados ao seleccionar facetas ou categorias, pré-visualizar conteúdo do documento e visualizar imagens em miniatura de um determinado tipo de documentos. Pode escolher activar as funcionalidades de análise, tais como ver os valores de correlação para valores da faceta devolvidos e ver como os documentos expandem um gráfico cronológico.

Descrição geral dos componentes

Os componentes do sistema recolhem dados de toda a empresa, analisam-nos, examinam-nos e extraem significado a partir da informação e criam um índice de texto que os utilizadores podem consultar.

Uma *colecção* representa o conjunto de origens que os utilizadores podem procurar e prospectar com uma única consulta. Quando cria uma colecção, especifica quais as origens que pretende incluir e configura as opções sobre como os utilizadores podem consultar os dados indexados.

Pode criar múltiplas colecções e cada colecção pode conter dados a partir de várias origens de dados. Por exemplo, pode criar uma colecção que inclua documentos de bases de dados de IBM DB2 e IBM Content Manager Enterprise Edition ou uma colecção que inclua documentos de arquivos de objectos IBM FileNet P8 ou repositórios Microsoft SharePoint. Quando os utilizadores procuram uma colecção, os resultados, provavelmente, incluem documentos de cada uma das origens de dados.

O tipo de colecção criada vai determinar quais as funções disponíveis para configurar a colecção:

Colecção de procura empresarial

Estas colecções suportam funções de procura e obtenção, incluindo a capacidade de procurar e limitar os resultados através da selecção de facetas, ordenar documentos por relevância ou data, prever documentos nos resultados de procura e ver imagens de miniatura de determinados tipos de documentos. Pode optar por activar algumas funcionalidades de análise para procurar estas colecções, tais como a capacidade de visualizar classificações de correlação e a forma como os resultados fluem ao longo de um gráfico cronológico.

Colecção de análise de conteúdo

Para além da procura, estas colecções suportam funções de prospecção de conteúdo, tais como a capacidade de explorar correlações, desvios e tendências nos dados. Também pode exportar resultados de análise para armazém de dados ou aplicações de informação empresarial e criar relatórios que podem ser guardados no formato de valor separado por vírgulas (CSV) ou abertos com as ferramentas do IBM Cognos Business Intelligence.

Criar e administrar uma colecção envolve as seguintes actividades:

Recolher dados

Os componentes do *pesquisador* recolhem documentos de origens de dados, tanto continuamente como de acordo com um programa especificado pelo utilizador. A pesquisa frequente assegura que os utilizadores têm sempre acesso às informações mais recentes.Para além de pesquisar origens de dados, pode adicionar conteúdo a uma colecção ao importar ficheiros CSV.

Analisar dados

O *encadeamento de análise* extrai texto de documentos, efectua análise linguística, encontra palavras e expressões significativas, extrai entidades e

executa análises personalizadas em cada documento. A análise de conteúdo detalhada fornece facetas de dados que podem ser utilizadas para explorar o conteúdo.

Indexar dados

Os componentes do *índice* adicionam dados dos documentos novos e modificados ao índice. Os componentes de índice também efectuam análises globais dos documentos numa colecção para determinar a classificação de correlação ou para detectar documentos duplicados e quase duplicados. Numa colecção de análise de conteúdos, pode ser criado um índice separado para as facetas.

Conteúdo de procura e prospecção

Uma aplicação de *procura empresarial* fornece uma interface gráfica interactiva para encontrar e obter documentos específicos. O Content Analytics Miner fornece uma interface gráfica interactiva para explorar conteúdo analisado para descobrir relações e anomalias.

Conceitos relacionados:

"Como os dados fluem ao longo do sistema" na página 23

"Descrição geral dos componentes" na página 15

"Pesquisadores"

"Analisadores, processadores de documentos e anotadores" na página 17

"Indexar" na página 18

"Servidores de procura" na página 19

"Consola de administração" na página 20

"Supervisionar o sistema" na página 20

"Ficheiros de registo" na página 21

"Personalizar o sistema" na página 21

"Aplicações para procura empresarial e prospecção de conteúdo" na página 22

Referências relacionadas:

"Origens de dados suportadas" na página 24

Pesquisadores

Os pesquisadores recolhem documentos de origens de dados de forma a que os documentos possam ser analisados, indexados, procurados e pesquisados.

Os pesquisadores têm as funções seguintes:

- Ao configurar um pesquisador, os processos de *identificação* encontram informações sobre as origens disponíveis para serem pesquisadas, tais como os nomes de todas as vistas e pastas de uma base de dados do Lotus Notes ou os nomes de todos os sistemas de ficheiros num servidor UNIX.
- Depois de seleccionar as origens que pretende pesquisar e iniciar o pesquisador, os componentes do pesquisador recolhem dados das origens de forma a que os dados possam ser analisados e indexados.

Uma única colecção pode ter vários pesquisadores, e cada pesquisador está concebido para recolher dados de um determinado tipo de origem de dados. Por exemplo, pode criar três pesquisadores para combinar dados de sistemas de ficheiros, de bases de dados do Notes, e de bases de dados relacionais na mesma colecção. Ou, pode criar vários pesquisadores do mesmo tipo e definir diferentes programas de pesquisa para as ferramentas de acordo com a frequência de alterações aos dados que estão a ser pesquisados por cada pesquisador.

Pode iniciar ou parar manualmente pesquisadores, ou pode definir programas de pesquisa. Se programar um pesquisador, especifique o momento inicial em que é executado e a frequência com que deve voltar a visitar as origens de dados para pesquisar documentos novos e alterados.

Pode configurar o pesquisador da Web para ser continuamente executado. Especifique quais os Uniform Resource Locators (URLs) que pretende pesquisar, e o pesquisador regressa periodicamente para procurar dados novos ou alterados.

As *Propriedades do pesquisador* são um conjunto de regras que determinam o comportamento de um determinado pesquisador quando esta pesquisa. Por exemplo, especifica regras para controlar a forma como o pesquisador utiliza os recursos do sistema. O conjunto de origens que cumpre os requisitos para ser pesquisado constitui o *espaço de pesquisa* de um pesquisador. Depois de criar um pesquisador, pode editar as propriedades do pesquisador em qualquer momento para alterar a forma como esse pesquisador recolhe dados. Também pode editar o espaço de pesquisa do pesquisador, adicionar origens novas ou remover origens que não pretende mais que sejam procuradas ou pesquisadas.

Conceitos relacionados:

Administração do pesquisador

Analisadores, processadores de documentos e anotadores

Os analisadores, processadores de documentos e anotadores analisam documentos que foram recolhidos por um pesquisador e preparamos para a indexação.

Os resultados da análise são armazenados num arquivo de dados para serem acedidos pelo componente de indexação. O processamento de documento inclui as seguintes tarefas:

- Detecção da codificação do conjunto de caracteres de cada documento. Antes executar qualquer análise linguística, o analisador utiliza estas informações para converter todo o texto para Unicode.
- Detecção da linguagem de origem de cada documento.
- Extracção de texto de qualquer formato de documento. Por exemplo, o analisador extrai texto de controlos em documentos XML e HTML. O extractor de texto extrai texto a partir de formatos binários tais como documentos Microsoft Word e Adobe portable document format (PDF).
- Extracção de texto e tokens para melhorar a obtenção e prospecção de dados. Durante esta fase, o analisador executa as seguintes tarefas:
 - Normalização de caracteres, tal como normalização de maiúsculas e marcas diacríticas como o trema alemão.
 - Analisar a estrutura de parágrafos, frases, palavras e espaço em branco. Através da análise linguística, o analisador decompõe palavras compostas e atribui tokens que permitem procura de dicionários e sinónimos.
- Aplica regras de processamento de documentos especificadas para a colecção. Por exemplo, pode configurar facetas para explorar conteúdo e aplicar análise de texto empresarial específica através de anotadores e dicionários personalizados.

- 🕞 Administração de análise e indexação
- Anotadores
- Content Analytics Studio
- Integração de análise de texto personalizada
- Suporte linguístico no IBM Content Analytics with Enterprise Search

Indexar

Depois de analisar e examinar os documentos, os processos de indexação vão adicionando de forma incremental informações sobre os documentos novos, alterados e eliminados ao índice para assegurar que os utilizadores têm sempre acesso às informações mais recentes.

Ao criar um índice, os processos de indexação executam análises globais a documentos. Durante esta fase são aplicados algoritmos para identificar documentos duplicados, para analisar a estrutura da ligação de documentos, e para executar um processamento especial sobre o texto âncora (o texto que descreve a página de destino numa ligação de hipertexto) em documentos da Web.

Quando cria uma colecção de procura empresarial, pode optar por criar múltiplas partições de índice. As colecções de análise de conteúdos são automaticamente particionadas. As partições permitem ao sistema equilibrar o processamento de milhões de documentos em várias partições, em que cada uma processa sensivelmente o mesmo número de documentos. As partições são acedidas como se estivessem num índice.

Ao configurar uma colecção, pode especificar opções para as seguintes actividade de indexação:

- Pode permitir que os utilizadores especifiquem consultas que incluam caracteres globais. Um carácter global pode ser ilimitado ou permitido apenas como caracteres de seguimento.
- Pode configurar âmbitos. Um *âmbito* permite-lhe limitar o que os utilizadores podem ver na colecção. Por exemplo, pode criar um âmbito que inclua os URIs para documentos no departamento de Assistência Técnica e outro âmbito para os URIs de documentos no departamento de Recursos Humanos. Se a respectiva aplicação suportar âmbitos, os utilizadores podem procurar e obter documentos a partir de apenas estes subconjuntos da colecção.
- Pode especificar opções para contrair documentos de resultados da procura que tenham o mesmo prefixo de URI. Também pode especificar um nome de grupo de forma a que os documentos com prefixos de URI diferentes possam ser contraídos juntos nos resultados.
- Depois de um índice remissivo ser criado, pode remover URIs que pretende evitar que os utilizadores procurem e prospectem.

- 🕞 Caracteres globais em consultas
- 🕩 Âmbitos
- URIs contraídos
- 🕩 Classificação de documento baseada em padrões URI

Tarefas relacionadas:

Remover URIs do índice

Servidores de procura

Os servidores de procura funcionam com as aplicações para processar consultas, procurar o índice e devolver resultados.

Ao configurar os servidores de procura para uma colecção, pode especificar opções para a forma como a colecção deve ser procurada ou pesquisada:

- Pode configurar uma cache de procura para manter os resultados de pedidos efectuados com frequência. Uma cache de procura pode melhorar o desempenho de obtenção e procura.
- Se os programadores de aplicações criarem dicionários personalizados, pode associar os dicionários a colecções:
 - Quando os utilizadores consultam uma colecção que utilize um *dicionário de* sinónimos, os documentos que contêm sinónimos dos termos da consulta são incluídos nos resultados.
 - Quando os utilizadores consultam uma colecção que utilize um *dicionário de* palavras de paragem, as palavras de paragem são removidas da consulta antes de esta ser processada.
 - Quando os utilizadores consultam uma colecção que utilize um *dicionário de* palavras hierárquicas, a importância de documentos que contenham as palavras é diminuída ou aumentada, dependendo do factor hierárquico associado à palavra no dicionário.
- Se definir previamente que determinados documentos são relevantes para determinadas consultas, pode configurar ligações rápidas. Uma *ligação rápida* associa um URI específico a palavras-chave e frases específicas. Se uma consulta contiver qualquer uma das palavras-chave ou frases especificadas numa definição de ligação rápida, o URI associado é devolvido automaticamente nos resultados.
- Pode configurar o suporte de *introdução adiantada*. Ao escrever uma consulta, os utilizadores podem ver sugestões para potenciais correspondências de termos da consulta e seleccionar uma das correspondências sugeridas para executarem a consulta. Também são facultadas correspondências sugeridas para valores da faceta. Os utilizadores podem seleccionar os valores da faceta que pretende adicionar a uma procura seleccionando uma ou mais sugestões.

Numa configuração de múltiplos servidores, a capacidade de mudança de recurso está disponível ao nível da colecção, não apenas ao nível do servidor. Se, por qualquer razão, uma colecção em um dos servidores da procura se tornar indisponível, então as consultas para essa colecção são encaminhadas automaticamente para outro servidor da procura.

- Administração do servidor de procura
- Aplicações de procura empresarial e de prospecção de conteúdo

Consola de administração

A consola de administração é executada num navegador, o que significa que os utilizadores administrativos podem aceder-lhe a partir de qualquer localização em qualquer altura. Os mecanismos de segurança asseguram que apenas os utilizadores autorizados a acederem a funções administrativas o fazem.

A consola de administração inclui assistentes que o podem ajudar a executar várias das principais tarefas administrativas. Por exemplo, os assistentes do pesquisador são específicos de um tipo de origem de dados e ajudam-no a seleccionar as origens que pretende permitir que os utilizadores procurem.

Para outras tarefas administrativas, as vistas-resumo permitem-lhe visualizar e aceder rapidamente a uma parte específica do sistema que necessita de administrar ou supervisionar. Por exemplo:

- Na vista Colecções, pode configurar opções para pesquisar, analisar e procurar conteúdo e, em seguida, supervisionar o progresso de cada um destes processos.
- Na vista Sistema, pode visualizar a topologia do sistema, adicionar e remover servidores e configurar controlos em todo o sistema, tais como ficheiros de registo para mensagens em todo o sistema.
- Na vista Segurança, pode configurar controlos, por exemplo, para limitar o acesso a aplicações, à consola de administração e a colecções.

Conceitos relacionados:

Administração do sistema

Tarefas relacionadas:

🕩 Iniciar sessão na consola de administração

Supervisionar o sistema

Pode utilizar a consola de administração para supervisionar actividades do sistema e ajustar operações conforme necessário.

Após ter instalado o IBM Content Analytics with Enterprise Search e ter criado, pelo menos, uma colecção, pode visualizar as estatísticas detalhadas para cada actividade, tal como a pesquisa, análise, indexação e o processamento das consultas. A informação inclui tempos de resposta médios e informações sobre o progresso, tais como o número de documentos pesquisados ou indexados durante determinada sessão.

Pode parar e iniciar grande maioria das actividades a partir da vista Colecções na vista-resumo de administração. Por exemplo, pode parar uma actividade, modificar a configuração ou resolver problemas, e, reiniciar o processamento quando estiver pronto para permitir que a actividade continue.

Também pode configurar alertas que lhe permitem receber mensagens de correio electrónico sobre determinadas actividades sempre que ocorre um evento supervisionado. Por exemplo, pode receber um alerta se o tempo de resposta da consulta exceder um limiar especificado.

Se for largado um documento do sistema, pode rastreá-lo e determinar quando, onde e porquê o documento foi largado. Por exemplo, um documento pode ser largado se o analisador não conseguir reconhecer o idioma do documento.

Pode visualizar estatísticas detalhadas sobre o processamento da consulta actual e antiga. Por exemplo, pode seleccionar uma cronologia para a estatística da consulta, ver as consultas mais populares e ver consultas que não devolveram resultados. Também pode exportar o histórico da consulta num ficheiro de valor separado por vírgulas (CSV) que, em seguida, pode importar numa folha de cálculo ou noutra aplicação para uma maior análise estatística.

Conceitos relacionados:

- 🖙 Supervisionar o sistema
- Iniciar e parar o sistema

Ficheiros de registo

Os ficheiros de registo são criados para colecções individuais e para sessões de nível de sistema.

Ao configurar opções de registo para uma colecção ou para o sistema, especifica os tipos de mensagens que pretende registar, tais como mensagens de erro e mensagens de aviso. Também especifica a frequência com que pretende que o sistema analise ficheiros de registo mais antigos para arranjar espaço para mensagens recentes. Pode seleccionar opções para receber uma mensagem de correio electrónico sobre mensagens específicas (alertas incluídos) ou sobre todas as mensagens de erro, sempre que estas ocorrerem.

Ao visualizar ficheiros de registo, selecciona o ficheiro de registo que pretende visualizar. O nome do ficheiro inclui informações sobre quando o ficheiro foi criado e qual o componente que emitiu as mensagens. Também pode especificar filtros de visualização. Por exemplo, pode escolher ver apenas mensagens de erro ou apenas mensagens de uma determinada sessão do sistema.

Conceitos relacionados:

- Ficheiros de registo e alertas
- Alertas

Im Mensagens para o IBM Content Analytics with Enterprise Search

Tarefas relacionadas:

- Configurar ficheiros de registo
- Configurar informações de servidor SMTP
- 🖙 Receber correio electrónico sobre mensagens registadas
- literative de registo

Personalizar o sistema

As interfaces de programação de aplicações permitem-lhe criar aplicações personalizadas para administrar o sistema e para procurar ou prospecção de colecções. Também pode criar aplicações personalizadas para actualizar o conteúdo das colecções, dos programas personalizados para análise de texto e dicionários personalizados e dos ficheiros de regras para melhorar a análise de texto e semântica. Depois de instalar IBM Content Analytics with Enterprise Search, as seguintes famílias de APIs estão disponíveis para extensão de colecções:

REST API (SIAPI)

Utilize esta API para criar aplicações de procura empresarial personalizada, aplicações de prospecção de conteúdo e aplicações de administração.

API Procurar e Indexar (SIAPI)

Utilize esta API para criar aplicações de procura empresarial personalizada e aplicações de prospecção de conteúdo.

Suplementos do pesquisador

Crie suplementos personalizados para adicionar metadados a documentos quando os documentos são pesquisados ou para associar tokens de segurança que fazem cumprir as regras empresariais e de segurança.

Exportar suplementos

Crie suplementos personalizados para documentos de exportação a partir do IBM Content Analytics with Enterprise Search. Em seguida, pode importar documentos e resultados de análises para outras aplicações, tais como aplicações de armazém de dados, informação empresarial, classificação e eDiscovery.

API componente de gestão de identidade

Utilize esta API para processar a autenticação de utilizador. Pode permitir que os utilizadores procurem em vários repositórios com uma única consulta e vejam apenas os documentos que estão autorizados a ver. Pode construir este componente nas respectivas aplicações personalizadas para que os utilizadores possam iniciar sessão com apenas um ID de utilizador e palavra-passe ao acederem a uma colecção segura.

Pode melhorar a qualidade da procura e análise ao integrar programas personalizados e ficheiros para análise linguística com as colecções. Após ter adicionado motores de análise de texto personalizado, anotadores, dicionários e ficheiro de regras de análise de texto ao sistema, pode associá-los às colecções. Quando os utilizadores consultam uma colecção, estes beneficiam das associações de palavras que os programas personalizados criam no índice. Por exemplo, os utilizadores podem descobrir conceitos e relações entre termos para além de localizarem documentos que contêm os termos da consulta especificados.

Também pode utilizar o Content Analytics Studio para facilmente criar e implementar a análise de texto personalizada para as aplicações do IBM Content Analytics with Enterprise Search. O Content Analytics Studio é um componente instalado separadamente do IBM Content Analytics with Enterprise Search.

Conceitos relacionados:

- L⁺⁺ Content Analytics Studio
- 🕩 Aplicações de procura empresarial e de prospecção de conteúdo

L⁺⁺ Suplementos do pesquisador

Tarefas relacionadas:

Personalizar aplicações através da utilização de uma interface de personalizador

Aplicações para procura empresarial e prospecção de conteúdo

O IBM Content Analytics with Enterprise Search fornece aplicações para procura e prospecção de conteúdo.

Aplicação de procura empresarial

Uma aplicação de procura empresarial de amostra é instalada quando instala o IBM Content Analytics with Enterprise Search. Esta aplicação é um exemplo funcional que pode utilizar com colecções de procura empresarial e como um modelo para desenvolver as próprias aplicações de procura empresarial. Não é possível utilizar esta aplicação para consultar colecções de análise de conteúdos.

Pode utilizar esta aplicação de amostra para testar colecções novas antes de as disponibilizar aos utilizadores. A aplicação apresenta um suporte para procura federada através da capacidade para consultar uma ou mais colecções de cada vez.

Para personalizar a aplicação de procura empresarial fornecida, pode utilizar o Search Customizer que é uma interface gráfica do utilizador que lhe permite visualizar os efeitos das alterações à medida que as executa. Também pode personalizar a aplicação ao editar um ficheiro de configuração.

Prospecção de análise de conteúdos

O Content Analytics Miner é instalado quando instala o IBM Content Analytics with Enterprise Search. Esta aplicação é uma aplicação funcional que lhe permite pesquisar conteúdo numa colecção de análise de conteúdos. Não pode utilizar esta aplicação com colecções de procura empresarial.

Pode utilizar esta aplicação para testar uma nova colecção antes de disponibilizar a colecção aos utilizadores. Quando explora uma colecção, pode explorar uma colecção de cada vez.

A aplicação facultada apresenta as várias formas para explorar e procurar os resultados de análise de conteúdos, como avaliar desvios e tendências nos dados ao longo do tempo ou comparar facetas dos dados num a vista bidimensional.

Para personalizar o Content Analytics Miner, pode utilizar o Personalizador de análise, que é uma interface gráfica do utilizador que lhe permite ver os efeitos das alterações à medida que as efectua. Também pode personalizar a aplicação ao editar um ficheiro de configuração.

Conceitos relacionados:

Aplicações de procura empresarial e de prospecção de conteúdo **Tarefas relacionadas**:

Aceder a aplicações

Personalizar aplicações através da utilização de uma interface de personalizador

Como os dados fluem ao longo do sistema

Os componentes do IBM Content Analytics with Enterprise Search interagem para recolher, analisar, verificar e indexar conteúdos para assegurar que as informações mais recentes estão disponíveis para procura e exploração.

O seguinte diagrama demonstra o fluxo de informação através de um sistema do IBM Content Analytics with Enterprise Search.



Figura 1. Como os dados fluem ao longo do sistema

Origens de dados suportadas

O IBM Content Analytics with Enterprise Search pode recolher e extrair conteúdo a partir de várias origens.

Está disponível suporte predefinido para vários sistemas de gestão de conteúdos, sistemas de base de dados, sistemas de ficheiros, sistemas de correio electrónico e produtos baseados na Web. Para obter mais informações sobre os produtos e versões de produtos suportados, consulte *Origens de dados suportados* no sítio da Web do IBM Content Analytics with Enterprise Search.

Conceitos relacionados:

"Como os dados fluem ao longo do sistema" na página 23

Referências relacionadas:

Corigens de dados suportadas no IBM Content Analytics with Enterprise Search
Administração do sistema

Após instalar o IBM Content Analytics with Enterprise Search, pode ter que alterar a configuração do sistema de formas que têm um impacto que abrange todo o sistema.

Para administrar a maioria dos aspectos do sistema, inicie sessão na consola de administração. Outras alterações de configuração podem requerer que edite os ficheiros de configuração e reinicie o sistema. As tarefas que abrangem todo o sistema incluem:

- · Adicionar e remover servidores da topologia do sistema
- Alterar a palavra-passe para o administrador predefinido do IBM Content Analytics with Enterprise Search
- Alterar o número de portas atribuídas aos componentes do IBM Content Analytics with Enterprise Search
- Configurar suporte para diferentes protocolos do sistema, tal como o suporte para endereços de IP e IPv6 duplos

Conceitos relacionados:

"Administração do pesquisador" na página 57

"Supervisionar o sistema" na página 423

Tarefas relacionadas:

"Iniciar o sistema" na página 417

"Parar os servidores de colecção ou todo o sistema" na página 419

"Criar uma colecção" na página 51

"Administrar os servidores de procura no modo autónomo" na página 421

Iniciar sessão na consola de administração

Para administrar o sistema, inicie sessão na consola de administração.

O uso simultâneo da consola de administração por vários utilizadores não é suportado. O sistema não impede vários utilizadores de iniciar sessão ao mesmo tempo, mas não existe qualquer mecanismo para impedir a inconsistência se operações simultâneas colidirem. Se vários utilizadores acederem à consola de administração ao mesmo tempo, uma operação poderá ser bem-sucedida, mas é provável que as outras operações falhem.

Restrição: Não pode especificar caracteres suplementares Unicode na consola de administração.

Tem de iniciar sessão com um ID de utilizador que esteja autorizado a aceder à consola de administração:

 Se permitir a aplicação de segurança no início de sessão na consola de administração, pode configurar as funções administrativas. Ao configurar as funções administrativas, permite que mais utilizadores iniciem sessão na consola de administração, mas consegue controlar as funções e as colecções às que cada utilizador administrativo pode aceder. Se não acrescentar utilizadores administrativos, apenas o ID de administrador predefinido especificado quando o IBM Content Analytics with Enterprise Search foi instalado, pode aceder à consola de administração.

Para iniciar sessão na consola de administração:

1. Escreva o URL da consola de administração no navegador da Web. Por exemplo:

http://index_server.com:port/ESAdmin/

em que *index_server.com* é o nome do sistema central ou endereço de IP do servidor do controlador. O número predefinido da porta é 8390, mas pode ser diferente para a sua instalação.

 Insira o seu ID do utilizador e palavra-passe e, de seguida, faça clique em Iniciar sessão. Na primeira vez que iniciar sessão, deve especificar o ID de administrador predefinido do IBM Content Analytics with Enterprise Search. É apresentada a vista Colecções que é o ponto de entrada para administrar o

sistema e as colecções. As acções que pode executar, bem como as colecções que pode visualizar, dependem da sua função administrativa.

Tarefas relacionadas:

"Configurar funções administrativas" na página 376

Alterar a palavra-passe de administrador na configuração de um servidor

A palavra-passe para o administrador está armazenada num formato codificado. Para alterar a palavra-passe, utilize o script **eschangepw**.

O IBM Content Analytics with Enterprise Search utiliza dois níveis de IDs de utilizador para o utilizador administrativo. Um ID de utilizador está ao nível do sistema operativo. O outro ID de utilizador, que está ao nível do IBM Content Analytics with Enterprise Search, está armazenado no ficheiro ES_INSTALL_ROOT/nodeinfo/es.cfg. As palavras-passe para ambos os IDs do utilizador devem corresponder. Se o ID de utilizador ao nível do sistema operativo expirar, por exemplo, também tem de alterar a palavra-passe ao nível do IBM Content Analytics with Enterprise Search.

É especificada a palavra passe para o ID do administrador padrão quando é instalado o IBM Content Analytics with Enterprise Search. Para alterar a palavra-passe, tem de executar o script **eschangepw** para disseminar a alteração através do sistema.

Caracteres especiais

As palavras-passe podem incluir caracteres alfabéticos, caracteres numéricos e os seguintes caracteres especiais:

! @ # \$ % ^ & * () - _ = + , . / < > ?

Em sistemas AIX e Linux, se especificar uma palavra-passe que inclua caracteres especiais, tem de colocar toda a palavra-passe entre plicas.Por exemplo: 'mypwd@\$%'

Num sistema Windows, se especificar uma palavra-passe que inclua caracteres especiais, tem de colocar toda a palavra-passe entre aspas.Por exemplo: "minha?+!palavrapasse"

Para alterar a palavra-passe de administrador numa configuração do servidor único:

1. Inicie sessão como administrador predefinido do IBM Content Analytics with Enterprise Search e pare o sistema:

esadmin system stopall

Importante: Quando o sistema é interrompido, os utilizadores não podem consultar as colecções.

- 2. Altere a palavra-passe do sistema para o ID de utilizador de administrador utilizando comandos do sistema operativo (em AIX ou Linux) ou ao utilizar a função de alteração de palavra-passe (em Windows).
- 3. Execute o seguinte script, em que *newValue* é a palavra-passe especificada no passo 2:

AIX ou Linux

eschangepw.sh newValue

Windows

eschangepw newValue

- 4. Em Windows, inicie o serviço da camada de comunicação comum (CCL) no fundo:
 - a. Inicie os Serviços do Windows: Iniciar > Programas > Ferramentas Administrativas > Serviços.
 - b. Faça clique com o botão direito do rato em **IBM Content Analytics with Enterprise Search** e faça clique em **Propriedades**.
 - c. Faça clique no separador Iniciar sessão.
 - d. Altere a palavra-passe especificando o novo valor de palavra-passe e, em seguida, faça clique em **OK**.
 - e. Faça clique com o botão direito do rato em **IBM Content Analytics with Enterprise Search** e faça clique em **Iniciar**.
- 5. Reinicie o sistema IBM Content Analytics with Enterprise Search:

esadmin system startall

Referências relacionadas:

"Comandos, códigos de retorno e IDs da sessão" na página 511

Alterar a palavra-passe de administrador na configuração de múltiplos servidores

A palavra-passe para o administrador está armazenada num formato codificado. Para alterar a palavra-passe, utilize o script **eschangepw** para a alterar em todos os servidores no sistema.

O IBM Content Analytics with Enterprise Search utiliza dois níveis de IDs de utilizador para o utilizador administrativo. Um ID de utilizador está ao nível do sistema operativo. O outro ID de utilizador, que está ao nível do IBM Content Analytics with Enterprise Search, está armazenado no ficheiro ES_INSTALL_ROOT/nodeinfo/es.cfg. As palavras-passe para ambos os IDs do utilizador devem corresponder. Se o ID de utilizador ao nível do sistema operativo expirar, por exemplo, também tem de alterar a palavra-passe ao nível do IBM Content Analytics with Enterprise Search. É especificada a palavra passe para o ID do administrador padrão quando é instalado o IBM Content Analytics with Enterprise Search. Para alterar a palavra-passe, tem de executar o script **eschangepw** para disseminar a alteração através do sistema.

Este procedimento sugere uma ordem para alterar a palavra-passe em todos os servidores. Esta ordem não é obrigatória, mas é necessário completar os passos requeridos para cada tipo de servidor.

Caracteres especiais

As palavras-passe podem incluir caracteres alfabéticos, caracteres numéricos e os seguintes caracteres especiais:

! @ # \$ % ^ & * () - _ = + , . / < > ?

Em sistemas AIX e Linux, se especificar uma palavra-passe que inclua caracteres especiais, tem de colocar toda a palavra-passe entre plicas.Por exemplo: 'mypwd@\$%'

Num sistema Windows, se especificar uma palavra-passe que inclua caracteres especiais, tem de colocar toda a palavra-passe entre aspas.Por exemplo: "minha?+!palavrapasse"

Para alterar a palavra-passe do administrador numa configuração de servidores múltiplos:

 No servidor principal, inicie sessão como o administrador do IBM Content Analytics with Enterprise Search e pare o sistema: esadmin system stopall

Importante: Quando o sistema é interrompido, os utilizadores não podem consultar as colecções.

- a. Altere a palavra-passe do sistema para o ID de utilizador de administrador utilizando comandos do sistema operativo (em AIX ou Linux) ou ao utilizar a função de alteração de palavra-passe (em Windows).
- b. Execute o seguinte script, em que *newValue* é a palavra-passe especificada no passo 1a:

AIX ou Linux

eschangepw.sh newValue

Windows

eschangepw *newValue*

- 2. Execute os seguintes passos em todos os outros servidores do seu sistema:
 - a. Inicie sessão como administrador do IBM Content Analytics with Enterprise Search.
 - b. Pare o common communication layer (CCL):

AIX ou Linux

stopccl.sh

Linha de comandos do Windows

stopccl

Ferramenta administrativa de Serviços Windows

- Inicie os Serviços do Windows: Iniciar > Programas > Ferramentas Administrativas > Serviços.
- 2) Faça clique com o botão direito do rato em **IBM Content Analytics with Enterprise Search** e faça clique em **Parar**.

- c. Altere a palavra-passe do sistema para o ID de utilizador de administrador utilizando comandos do sistema operativo (em AIX ou Linux) ou ao utilizar a função de alteração de palavra-passe (em Windows). Esta palavra-passe tem de corresponder à palavra-passe especificada no passo 1a na página 28.
- d. Execute o seguinte script, em que *newValue* é a palavra-passe especificada no passo 1a na página 28:

AIX ou Linux

eschangepw.sh newValue

Windows

eschangepw *newValue*

e. Reinicie o CCL:

AIX ou Linux

startccl.sh -bg

Linha de comandos do Windows startccl

Ferramenta administrativa de Serviços Windows

Para iniciar o CCL em segundo plano:

- Inicie os Serviços do Windows: Iniciar > Programas > Ferramentas Administrativas > Serviços.
- Faça clique com o botão direito do rato em IBM Content Analytics with Enterprise Search e faça clique em Propriedades.
- 3) Faça clique no separador Iniciar sessão.
- 4) Altere a palavra-passe especificando o novo valor de palavra-passe e, em seguida, faça clique em OK.
- 5) Faça clique com o botão direito do rato em **IBM Content Analytics with Enterprise Search** e faça clique em **Iniciar**.
- **3**. No servidor principal, inicie sessão como administrador predefinido do IBM Content Analytics with Enterprise Search e reinicie o sistema do IBM Content Analytics with Enterprise Search:

esadmin system startall

Referências relacionadas:

"Comandos, códigos de retorno e IDs da sessão" na página 511

Definir a palavra-passe de administrador codificada para ser a mesma em todos os servidores

Adicionar servidores ao sistema

Pode melhorar a cópia de segurança de alta disponibilidade e o suporte de mudança de recurso adicionando servidores para suportar funções específicas. Por exemplo, poderá configurar servidores de procura para distribuir o volume de processamento da consulta por processadores.

Deve instalar o IBM Content Analytics with Enterprise Search no servidor que pretende adicionar à topologia do sistema antes de especificar o objectivo do servidor na consola de administração.

É necessário assegurar que todo o software que é necessário para suportar a função de servidor pretendida está instalado. Por exemplo, se acrescentar um servidor de

procura e pretender utilizar o WebSphere Application Server para suportar as aplicações de procura, é necessário que uma versão suportada do WebSphere Application Server está instalada.

Se adicionar um servidor de alta disponibilidade, o servidor de alta disponibilidade e o servidor principal devem partilhar o mesmo directório de dados (ES_NODE_ROOT). Os servidores de alta disponibilidade são suportados através do IBM PowerHA para AIX e Microsoft Cluster Service (MSCS) em Windows.

Os tipos de servidores que pode adicionar dependem da forma como instalar os componentes essenciais:

Servidor único

Numa instalação de um servidor único, tem um servidor de controlador principal. Posteriormente pode adicionar mais servidores de procura e servidores de processamento de documentos. Em plataformas AIX ou Windows, também pode adicionar um servidor do controlador de alta disponibilidade para suportar pesquisa, indexação e procura.

Servidores distribuídos

Numa instalação de um servidor distribuído, tem um servidor de controlador principal um servidor do pesquisador principal e, pelo menos, um servidor de procura. Posteriormente pode adicionar mais servidores de procura e servidores de processamento de documentos. Em plataformas AIX ou Windows, também pode adicionar um servidor de cópia de segurança de alta disponibilidade para suportar a pesquisa e um servidor de controlador para suportar indexação.

Para adicionar um servidor ao sistema e atribuir uma função para a forma como o servidor será utilizado:

- Verifique se o IBM Content Analytics with Enterprise Search e todo o software necessário estão instalados no servidor que pretende adicionar, e verifique se o novos servidor está em execução. Para adicionar o servidor, o sistema tem de ser capaz de estabelecer ligação com o servidor.
- 2. Faça clique em Sistema para abrir a vista Sistema.
- 3. Faça clique em Adicionar Servidor.
- 4. Na página Adicionar um Servidor, insira o nome totalmente qualificado do servidor que pretende adicionar. Além disso, especifique a porta do Common Communication Layer (CCL) que especificou para esse servidor. O valor predefinido é 6002.
- Seleccione a função do servidor e faça clique em OK. O servidor que adicionou aparece na topologia do sistema e fica disponível para processar pedidos neste sistema.

Pode supervisionar o estado do servidor, iniciar e parar o servidor ou remover o servidor da topologia do sistema. Para um servidor de procura, também pode ver estatísticas detalhadas sobre o processamento da consulta.

Restrição: Após de acrescentar um servidor, não pode alterar o nome do sistema central, a porta ou a função. Se pretender alterar a configuração do servidor, é necessário remover o servidor da topologia do sistema e, em seguida, adicioná-lo novamente.

Conceitos relacionados:

E Configurações de sistema suportadas

Tarefas relacionadas:

- 🕩 Instalar servidores adicionais
- Configurar elevada disponibilidade em AIX
- Configurar elevada disponibilidade em Windows

Configuração do número da porta

Reveja os números de porta predefinidos, para poder evitar conflitos de portas ao configurar recursos ou atribuir números de portas a outras aplicações.

Se configurar uma firewall, tem de activar explicitamente o acesso a números de porta em particular. Também é necessário assegurar que todos os servidores do IBM Content Analytics with Enterprise Search se encontram dentro da firewall.

Números de Nome da porta portas Onde estão configurados ES INSTALL ROOT/nodeinfo/es.cfg e Nível de comunicação 6002 comum (Common ES_NODE_ROOT/master_config/nodes.ini em communication layer, CCL) todos os servidores Porta de escuta de dados Para alterar esta porta, utilize a vista Sistema 6668 na IBM Content Analytics with Enterprise Search consola de administração para configurar a aplicação escuta de dados. Servidor de aplicação da 8390 ES INSTALL ROOT/nodeinfo/es.cfg Web incorporado: consola administrativa e REST API para administração WebSphere Application 9081 / 80 ES INSTALL ROOT/nodeinfo/es.cfg Server: consola de administração e REST API para administração 8394 ES INSTALL ROOT/nodeinfo/es.cfg Servidor de procura Servidor de aplicação da 8393 ES INSTALL ROOT/nodeinfo/es.cfg Web incorporado: aplicação de procura empresarial, prospecção de análise de conteúdo, e REST API para procura WebSphere Application ES INSTALL ROOT/nodeinfo/es.cfg 9081 / 80 Server: aplicação de procura empresarial, prospecção de análise de conteúdo, e REST API para procura HTTP SERVER ROOT/conf/http.conf nos Servidor de aplicação da 80 servidores de procura Web incorporado: HTTP HTTP SERVER ROOT/conf/http.conf nos WebSphere Application 9081 Server: HTTP servidores de procura

Tabela 6. Configuração predefinida do número de porta

Nome da porta	Números de portas	Onde estão configurados
WebSphere Application Server consola de administração	9060	Nos servidores de procura, se utilizar o WebSphere Application Server
Pesquisador do DB2	6000, 6001, 6002, 60003, 50000	No servidor do pesquisador
Centro de Informações	8889	Nos servidores de procura
Portas anónimas ou dinâmicas para CCL, transferências de ficheiros (ESFTP) e cópia de índice	49152 a 65535	Em todos os servidores
Apache Derby Network Server	1527	No servidor do pesquisador
Comunicações personalizadas	8890	No servidor de pesquisador. A porta de escuta predefinida não está definida, o que significa que a sessão obtém uma porta disponível. Se pretender definir um número de porta fixo, defina a propriedade de porta no ficheiro ES_INSTALL_ROOT/configurations/ customcommunication /customcommunication.properties.
Base de dados IBM Content Integrator FastObjects	6001 (6002 como alternativa)	Na consola de administração do IBM Content Integrator
Ligação proxy de Remote Method Invocation (RMI) do IBM Content Integrator	1251 (porta de RMI)	Na consola de administração do IBM Content Integrator

Tabela 6. Configuração predefinida do número de porta (continuação)

Tarefas relacionadas:

"Alterar o número de portas do sistema"

"Alterar o número da porta da aplicação" na página 34

"Alterar o número da porta e a raiz de contexto do URL do Centro de Informações" na página 34

Alterar o número de portas do sistema

Se o número de portas que o sistema IBM Content Analytics with Enterprise Search utiliza para comunicação entrar em conflito com um número de porta utilizado por outro produto, tem de alterar o número da porta.

É especificado um número da porta para o sistema quando for instalado o IBM Content Analytics with Enterprise Search. Numa configuração de servidores múltiplos, o mesmo número de porta é especificado em todos os servidores.

Se o número de porta não for utilizável (por exemplo, se tiver sido atribuído a outro produto no mesmo servidor), o conflito resulta na seguinte mensagem de erro no ficheiro CCLServer_data.log, em que *data* especifica a data de criação do ficheiro de registo:

FFQ00273W Ocorreu um aviso interno - Mensagem de Excepção: {0} at java.net.PlainSocketImpl.socketBind(Native Method) at java.net.PlainSocketImpl.bind(PlainSocketImpl.java:357)

```
at java.net.ServerSocket.bind(ServerSocket.java:341)
```

at java.net.ServerSocket.<init>(ServerSocket.java:208)

at java.net.ServerSocket.<init>(ServerSocket.java:120)

Para alterar o número da porta que é usada pelo IBM Content Analytics with Enterprise Search:

 Aceda ao computador cujo número da porta tem de ser alterado, inicie sessão como administrador predefinido do IBM Content Analytics with Enterprise Search e pare o sistema:

esadmin system stopall

Importante: Quando o sistema é interrompido, os utilizadores não podem consultar as colecções.

 Edite o ficheiro ES_INSTALL_ROOT/nodeinfo/es.cfg, localize a propriedade seguinte, especifique um novo valor do número da porta e depois guarde e feche o ficheiro:

CCLPort=novo_número_porta

3. Reinicie o common communication layer (CCL):

AIX ou Linux

startccl.sh

Linha de comandos do Windows

Ferramenta administrativa de Serviços Windows

Para iniciar o CCL em segundo plano:

- a. Inicie os Serviços do Windows: Iniciar > Programas > Ferramentas Administrativas > Serviços.
- b. Faça clique com o botão direito do rato em **IBM Content Analytics** with Enterprise Search e faça clique em Iniciar.
- 4. Aceda ao servidor do controlador, inicie sessão como o administrador predefinido do IBM Content Analytics with Enterprise Search e pare o CCL:

AIX ou Linux

stopccl.sh

Linha de comandos do Windows

stopccl

Ferramenta administrativa de Serviços Windows

- a. Inicie os Serviços do Windows: Iniciar > Programas > Ferramentas Administrativas > Serviços.
- b. Faça clique com o botão direito do rato em **IBM Content Analytics** with Enterprise Search e faça clique em Parar.
- 5. Edite o ficheiro ES_NODE_ROOT/master_config/nodes.ini:
 - a. Localize a seguinte propriedade, em que nome_computador é o nome do computador onde modificou o número de porta no passo 2. O N na propriedade nodeN é um número que identifica o servidor. nodeN.destination=nome computador
 - b. Localize a seguinte subpropriedade, especifique o mesmo número de porta que especificou para o servidor no passo 2, e depois guarde e feche o ficheiro:

nodeN.port=novo_número_porta

6. Reinicie o sistema IBM Content Analytics with Enterprise Search:

esadmin system startall

Referências relacionadas:

"Comandos, códigos de retorno e IDs da sessão" na página 511

"Configuração do número da porta" na página 31

Alterar o número da porta da aplicação

Se pretender alterar o número da porta que é utilizada pelas aplicações após IBM Content Analytics with Enterprise Search estar instalado, tem de editar o ficheiro de configuração es.cfg para especificar o novo número da porta.

Actualização: Alguns números de porta predefinidos alterados no IBM Content Analytics with Enterprise Search Versão 3.0. Se efectuou a actualização para esta versão, altere o número da porta que está configurado para as suas aplicações para 8393 de modo a assegurar que as suas aplicações funcionam com as portas predefinidas do IBM Content Analytics with Enterprise Search.

Sistemas Linux: Se alterar o número da porta para um número inferior a 1024 (denominada porta privilegiada), a aplicação pode não iniciar com o comando esadmin searchapp.*node_number* ou com o comando esadmin system startall. Para executar a aplicação com o número da porta privilegiada e com o privilégio de administrador predefinido de IBM Content Analytics with Enterprise Search, tem de configurar o sistema operativo para suportar acesso a esse número da porta.

Para alterar o número da porta que é utilizado pela aplicação de procura empresarial e prospecção de análise de conteúdo:

- Inicie sessão como o administrador predefinido de IBM Content Analytics with Enterprise Search no servidor do controlador e edite o ficheiro ES_INSTALL_ROOT/nodeinfo/es.cfg.
- Localize a seguinte propriedade, especifique um novo valor de número da porta e, de seguida, guarde e feche o ficheiro: SearchApplicationPort=new_port_number
- 3. Opcional: No Windows, edite o ficheiro ES_INSTALL_ROOT/bin/shortcuts/ searchapplication.url para especificar o novo número da porta. Execute este passo apenas se utilizar o atalho fornecido para aceder à aplicação.
- 4. Reinicie o sistema IBM Content Analytics with Enterprise Search:

```
esadmin system stopall
esadmin
system startall
```

5. Aceda à aplicação com o novo número da porta e confirme que abre. Por exemplo, se alterar o número da porta de 8393 para 80, o URL para aceder à aplicação de procura empresarial muda de http://example.server.com:8393/search/para http://server.example.com/search/.

Referências relacionadas:

"Configuração do número da porta" na página 31

Alterar o número da porta e a raiz de contexto do URL do Centro de Informações

É possível alterar o número da porta e raiz de contexto para ligações para o Centro de Informações do IBM Content Analytics with Enterprise Search. As alterações

afectam o modo como acede directamente ao Centro de Informações e o modo como os tópicos no Centro de Informações são acedidos a partir da ajuda online.

Para alterar a raiz de contexto e o número da porta utilizados pelo Centro de Informações do IBM Content Analytics with Enterprise Search:

- Inicie sessão como o administrador predefinido do IBM Content Analytics with Enterprise Search no servidor do controlador e pare o sistema: esadmin system stopall
- 2. Edite os ficheiros de arranque e de encerramento do Centro de Informações:

AIX ou Linux

ES_INSTALL_ROOT/ibm_help/IC_start.sh e IC_stop.sh

Windows

ES_INSTALL_ROOT/ibm_help/IC_start.bat e IC_stop.bat

3. Altere o caminho predefinido para a raiz de contexto, -Dorg.eclipse.equinox.http.jetty.context.path=/help, para o seu caminho preferido. Por exemplo:

-Dorg.eclipse.equinox.http.jetty.context.path=/myHelp

4. Altere a porta predefinida, -porta 8889, para o seu valor preferido. Por exemplo:

-porta 8080

- Edite o ficheiro ES_INSTALL_ROOT/nodeinfo/es.cfg.Especifique o novo caminho de raiz de contexto e porta na propriedade DB2ICDocs4ES, e especifique o novo número da porta na propriedade DB2ICServerPort. Por exemplo: DB2ICDocs4ES=http\://example.server.com\:8080/myHelp/topic DB2ICServerPort=8080
- 6. Reinicie o sistema IBM Content Analytics with Enterprise Search: esadmin system startall
- Confirme que é possível aceder ao Centro de Informações com a nova raiz de contexto e porta. Nos seguintes exemplos, a raiz de contexto foi alterada para myHelp e a porta foi alterada para 8080.
 - a. Aceda directamente ao Centro de Informações ao introduzir directamente este URL:

http://example.server.com:8080/myHelp

b. Aceda ao Centro de Informações a partir da aplicação de procura empresarial ao introduzir este URL e fazendo clique em **Ajuda**:

http://example.server.com:8080/search

Referências relacionadas:

"Configuração do número da porta" na página 31

E Configurar o Centro de Informações para aceder ao sítio da Web público

Alterar nomes de sistema central do servidor ou endereços de IP

Só pode alterar o nome de sistema central ou o endereço de IP em servidores que tenham várias placas de interface de rede, onde o servidor poderá ter mais do que um nome ou endereço de IP.

Por exemplo, poderá querer alterar o endereço de IP se possuir várias placas de interface de rede (NIC) em cada servidor e descobrir que o servidor do controlador

está configurado para utilizar uma rede lenta. Pode alterar as definições nos ficheiros de configuração para permitir que o servidor do controlador utilize uma rede mais rápida.

Sugestão: Se não pretender editar os ficheiros de configuração, pode reinstalar o IBM Content Analytics with Enterprise Search e especificar os novos nomes do sistema central ou endereços de IP durante a execução do programa de instalação.

Restrição: Se alterar o nome do sistema central do computador, terá de reinstalar o IBM Content Analytics with Enterprise Search e os respectivos produtos de pré-requisito.

Para alterar um nome do sistema central ou endereço de IP num servidor que tem várias placas de rede:

- 1. Inicie sessão como administrador predefinido do IBM Content Analytics with Enterprise Search. Se tiver uma configuração múltipla de servidor, inicie sessão em qualquer servidor.
- 2. Pare as sessões do sistema:
 - esadmin system stopall
- **3**. Pare o nível de comunicação comum (CCL, Common Communication Layer). Se tiver uma configuração múltipla do servidor, utilize um dos seguintes métodos para parar o CCL em cada servidor:

AIX ou Linux

stopccl.sh

Linha de comandos do Windows stopcc]

Ferramenta administrativa de Serviços Windows

- a. Inicie os Serviços do Windows: Iniciar > Programas > Ferramentas Administrativas > Serviços.
- b. Faça clique com o botão direito do rato em **IBM Content Analytics** with Enterprise Search e faça clique em Parar.
- 4. Execute os seguintes passos em cada servidor:
 - a. Edite o ficheiro ES_INSTALL_ROOT/configurations/ccl.properties e especifique o nome do sistema central ou o endereço de IP que pretende utilizar para este servidor no parâmetro **es_server_hostName**.
 - b. Edite o ficheiro ES_INSTALL_ROOT/nodeinfo/es.cfg e especifique o nome de sistema central ou o endereço de IP que pretende usar para este servidor no parâmetro LocalHostName.
 - c. Edite o ficheiro ES_NODE_ROOT/master_config/nodes.ini e substitua todas as ocorrências do parâmetro nodeN.destination com o nome do sistema central ou o endereço de IP que pretende utilizar para este servidor.
 - d. Edite o ficheiro ES_NODE_ROOT/config/nodes.ini e substitua todas as ocorrências do parâmetro nodeN.destination com o nome do sistema central ou o endereço de IP que pretende utilizar para este servidor.
- 5. Reinicie o CCL. Se tiver uma configuração múltipla do servidor, utilize um dos métodos seguintes para reiniciar o CCL em cada servidor:

AIX ou Linux startccl.sh -bg

Linha de comandos do Windows startccl

Ferramenta administrativa de Serviços Windows

Para iniciar o CCL em segundo plano:

- a. Inicie os Serviços do Windows: Iniciar > Programas > Ferramentas Administrativas > Serviços.
- b. Faça clique com o botão direito do rato em **IBM Content Analytics** with Enterprise Search e faça clique em Iniciar.
- 6. Reinicie o sistema IBM Content Analytics with Enterprise Search: esadmin system startall

Configuração do suporte para endereços de IP duplos

Se os servidores onde instalou o IBM Content Analytics with Enterprise Search estiverem configurados para suportar endereços de IP duplos, terá de configurar manualmente os servidores para ser executado nesse ambiente.

Por exemplo, um administrador do ambiente de trabalho poderá instalar um Microsoft Loopback Adapter para criar uma rede virtual que suporta os requisitos de funcionamento em rede de certos produtos, tais como o Microsoft SQL Server.

Para configurar o sistema de forma a que possa ser executado em servidores que suportam endereços de IP duplos:

- Inicie sessão como administrador predefinido do IBM Content Analytics with Enterprise Search. Se tiver uma configuração múltipla de servidor, inicie sessão em qualquer servidor.
- 2. Pare as sessões do sistema:

esadmin system stopall

3. Pare o nível de comunicação comum (CCL, Common Communication Layer). Se tiver uma configuração múltipla do servidor, utilize um dos seguintes métodos para parar o CCL em cada servidor:

AIX ou Linux

stopccl.sh

Linha de comandos do Windows stopccl

Ferramenta administrativa de Serviços Windows

- a. Inicie os Serviços Windows: Iniciar > Programas > Ferramentas Administrativas > Serviços.
- b. Faça clique com o botão direito do rato em **IBM Content Analytics** with Enterprise Search e faça clique em Parar.
- 4. Em cada servidor que suporte endereços de IP duplos, edite o ficheiro ES_INSTALL_ROOT/nodeinfo/es.cfg e adicione um parâmetro denominado LocalIPAddress. Para o valor, atribua um endereço de IP que possa ser processado por DNS.
- 5. Reinicie o CCL. Se tiver uma configuração múltipla do servidor, utilize um dos métodos seguintes para reiniciar o CCL em cada servidor:

AIX ou Linux

startccl.sh -bg

Linha de comandos do Windows startccl

Ferramenta administrativa de Serviços Windows Para iniciar o CCL em segundo plano:

- a. Inicie os Serviços Windows: Iniciar > Programas > Ferramentas Administrativas > Serviços.
- b. Faça clique com o botão direito do rato em **IBM Content Analytics** with Enterprise Search e faça clique em Iniciar.
- 6. Reinicie o sistema IBM Content Analytics with Enterprise Search: esadmin system startall

Activar o suporte para o protocolo IPv6 em sistemas Windows

Para activar o suporte para endereços e URLs que aderem ao protocolo IP versão 6 (IPv6), tem de definir uma variável de ambiente que instrua o sistema a utilizar apenas endereços de socket IPv6.

Complete os seguintes passos para preparar o sistema:

- Verifique se os valores para a propriedade LocalHostName e a propriedade LocalIPAddress (se fornecida) no ficheiro ES_INSTALL_ROOT/nodeinfo/es.cfg contêm um nome do sistema central ou um endereço de IPv6 válido para a máquina local. Se alterar este ficheiro, guarde as alterações.
- Verifique se o valor para a propriedade server.host no ficheiro ES_NODE_ROOT/master_config/database/db.properties é um nome do sistema central. Este valor não pode ser um endereço de IPv4 ou um endereço de IPv6. Se alterar esta propriedade, guarde o ficheiro.
- 3. Verifique que os valores para a propriedade destination e a propriedade serverhost no ficheiro ES_NODE_ROOT/master_config/nodes.ini no servidor de controlador contêm um endereço IPv6 válido ou um nome do sistema central válido. Estes valores não podem ser um endereço de IPv4. Se alterar este ficheiro, guarde as alterações.

O suporte para o protocolo IPv6 apenas está disponível em sistemas IBM Content Analytics with Enterprise Search que forem instalados nos servidores do Windows 2003 ou Windows 2008. Depois de activar o suporte para endereços de IPv6, o sistema já não vai utilizar endereços de IPv4 para comunicações de socket.

Os navegadores Internet Explorer e Mozilla Firefox processam os endereços IPv6 de forma diferente:

Mozilla Firefox

Para executar a consola administrativa, a aplicação de procura empresarial ou o prospecção de análise de conteúdo, pode especificar o endereço IPv6 ou o nome do sistema central no URL. Por exemplo:

http://[2001::db8]/ESAdmin
http://SearchServer.com/search/

Internet Explorer

Para executar a consola administrativa, a aplicação de procura empresarial ou o prospecção de análise de conteúdo, não pode especificar o endereço IPv6 no URL. Deve utilizar o formato seguinte e assegurar que o nome do sistema central é correlacionado com o endereço de IPv6 no DNS ou no ficheiro c:\windows\system32\drivers\etc\hosts:

http://SearchServer.com/ESAdmin/
http://SearchServer.com/search/

Se especificar um endereço de IPv6 no URL, é apresentada a mensagem Erro de sintaxe inválido (Invalid syntax error). Para obter mais informações sobre esta restrição, consulte http://support.microsoft.com/ kb/325414.

Inicie um formato URL: As origens de dados que são executadas num servidor IPv4 são suportadas por predefinição pelos pesquisadores. Ao configurar o pesquisador da Web e ao especificar os URLs iniciais no formato do endereço de IPv6, certifique-se de que os URLs estão entre parêntesis. Por exemplo: http://[2001:db8:0:1:0:0:0:1] http://[2001:db8:0:1::1]

Para activar o suporte para o protocolo IPv6:

 No servidor do controlador, inicie a sessão como o administrador predefinido do IBM Content Analytics with Enterprise Search e pare o sistema: esadmin system stopall

Importante: Quando o sistema é interrompido, os utilizadores não podem consultar as colecções.

- **2.** Pare o servidor de common communication layer (CCL) em todos os servidores:
 - a. Inicie os Serviços do Windows: Iniciar > Programas > Ferramentas Administrativas > Serviços.
 - b. Faça clique com o botão direito do rato em **IBM Content Analytics with Enterprise Search** e faça clique em **Parar**.
- **3**. Defina **ES_IPV6=1** nas variáveis de ambiente do sistema. Para uma instalação do servidor múltiplo, execute este passo nos servidores de pesquisa.
- 4. Se utilizar o WebSphere Application Server em vez de o servidor de aplicação da Web incorporado, execute os passos seguintes nos servidores de procura:
 - a. Actualize as propriedades personalizadas da Java Virtual Machine no WebSphere Application Server.
 - 1) Se estiver em execução, pare o servidor da aplicação ESSearchServer no WebSphere Application Server.
 - 2) Abra a consola de administração WebSphere Application Server e navegue até ao painel das propriedades personalizadas da Java Virtual Machine. Seleccione Servidores > Servidores de Aplicação > ESSearchServer > Java e Gestão de Processos > Definição de Processos > Máquina Virtual Java > Propriedades Personalizadas.
 - 3) Configure as propriedades seguintes:
 - java.net.preferIPv4Stack=false java.net.preferIPv6Addresses=true

Se a propriedade personalizada não estiver já na lista, crie-a. Insira o nome da propriedade no campo **Nome** e um nome válido no campo **Valor**.

- 4) Se estiver a executar o WebSphere Application Server Versão 6.1, adicione a propriedade LocalIPAddress ao ficheiro ES_INSTALL_ROOT/ nodeinfo/es.cfg para conter o endereço de IPv6 para esse servidor de pesquisa.
- 5) Reinicie a aplicação ESSearchServer.
- 5. Reinicie o CCL em todos os servidores:

- a. Inicie os Serviços do Windows: Iniciar > Programas > Ferramentas Administrativas > Serviços.
- b. Faça clique com o botão direito do rato em **IBM Content Analytics with Enterprise Search** e faça clique em **Iniciar**.
- 6. Reinicie o sistema IBM Content Analytics with Enterprise Search:

```
esadmin system startall
```

7. Verifique o ficheiro ES_NODE_ROOT/logs/ccl_*hostname_*0.log e verifique se o servidor CCL iniciou sem erros. Se o IPv6 for correctamente activado, serão registadas mensagens semelhantes às seguintes:

```
INFO: CCL server ready for business. Now, waiting for external requests.
CCL host name is fe80::250:56ff:feb4:27d1
CCL host dot.ip address is fe80:0:0:0:250:56ff:feb4:27d1
CCL server port is 6002
Total JVM Runtime memory is 33022Kb
Current free memory is 21346Kb
java.net.preferIPv6Addresses=true
java.net.preferIPv4Stack=false
INFO: Session "TraceDaemonSession" was attached PID=4992
```

- 8. Siga os seguintes passos para verificar se a aplicação de procura empresarial e a consola de administração estão a funcionar correctamente:
 - a. Inicie os servidores de pesquisa para qualquer colecção e depois abra a procura empresarial e submeta uma consulta. Este passo verifica se o processamento de envio da consulta está a funcionar, mesmo que não haja documentos no índice para procurar.
 - b. Abra a consola de administração e verifique se é capaz de iniciar sessão.

Conceitos relacionados:

"Regras para limitar o espaço de pesquisa da Web" na página 138

Desactivar suporte para o protocolo IPv6

Para desactivar o suporte para o protocolo IPv6:

- Veja os passo anteriores para preparar o sistema para a utilização do protocolo IPv6. Para todas as ocorrências de um endereço IPv6, substitua o valor com um nome do sistema central ou um endereço IPv4.
- Siga o procedimento para activar o suporte para o protocolo IPv6 com as seguintes excepções:
 - No passo 3 na página 39, remova a variável de ambiente ES_IPV6.
 - Se utilizar WebSphere Application Server, remova as propriedades de JVM personalizadas que foram definidas para a aplicação ESSearchServer no passo 4 na página 39.
- 3. Remova a propriedade LocalIPAddress do ficheiro ES_INSTALL_ROOT/nodeinfo/ es.cfg caso ele contenha um endereço de IPv6.
- 4. Reinicie o sistema IBM Content Analytics with Enterprise Search:

```
esadmin system
stopall
esadmin system startall
```

5. Siga o procedimento no passo 8 para verificar a aplicação de procura empresarial e a consola de administração.

Activar o suporte para o protocolo IPv6 em sistemas AIX ou Linux

Para activar o suporte para endereços e URLs que aderem ao protocolo IP versão 6 (IPv6), tem de definir uma variável de ambiente que instrua o sistema a utilizar apenas endereços de socket IPv6.

Complete os seguintes passos para preparar o sistema:

- Verifique se os valores para a propriedade LocalHostName e a propriedade LocalIPAddress (se fornecida) no ficheiro ES_INSTALL_ROOT/nodeinfo/es.cfg contêm um nome do sistema central ou um endereço de IPv6 válido para a máquina local. Se alterar este ficheiro, guarde as alterações.
- Verifique se o valor para a propriedade server.host no ficheiro ES_NODE_ROOT/master_config/database/db.properties é um nome do sistema central. Este valor não pode ser um endereço de IPv4 ou um endereço de IPv6. Se alterar esta propriedade, guarde o ficheiro.
- 3. Verifique que os valores para a propriedade destination e a propriedade serverhost no ficheiro ES_NODE_ROOT/master_config/nodes.ini no servidor de controlador contêm um endereço IPv6 válido ou um nome do sistema central válido. Estes valores não podem ser um endereço de IPv4. Se alterar este ficheiro, guarde as alterações.

Os navegadores Internet Explorer e Mozilla Firefox processam os endereços IPv6 de forma diferente:

Mozilla Firefox

Para executar a consola administrativa, a aplicação de procura empresarial ou o prospecção de análise de conteúdo, pode especificar o endereço IPv6 ou o nome do sistema central no URL. Por exemplo:

http://[2001::db8]/ESAdmin
http://SearchServer.com/search/

Internet Explorer

Para executar a consola administrativa, a aplicação de procura empresarial ou o prospecção de análise de conteúdo, não pode especificar o endereço IPv6 no URL. Deve utilizar o formato seguinte e assegurar que o nome do sistema central é correlacionado com o endereço de IPv6 no DNS ou no ficheiro c:\windows\system32\drivers\etc\hosts:

http://SearchServer.com/ESAdmin/ http://SearchServer.com/search/

Se especificar um endereço de IPv6 no URL, é apresentada a mensagem Erro de sintaxe inválido (Invalid syntax error). Para obter mais informações sobre esta restrição, consulte http://support.microsoft.com/ kb/325414.

Inicie um formato URL: As origens de dados que são executadas num servidor IPv4 são suportadas por predefinição pelos pesquisadores. Ao configurar o pesquisador da Web e ao especificar os URLs iniciais no formato do endereço de IPv6, certifique-se de que os URLs estão entre parêntesis. Por exemplo: http://[2001:db8:0:1:0:0:0:1] http://[2001:db8:0:1::1]

Para activar o suporte para o protocolo IPv6:

 No servidor do controlador, inicie a sessão como o administrador predefinido do IBM Content Analytics with Enterprise Search e pare o sistema: esadmin system stopall **Importante:** Quando o sistema é interrompido, os utilizadores não podem consultar as colecções.

2. Pare o servidor de common communication layer (CCL) em todos os servidores:

stopccl.sh

3. Para definir e exportar o parâmetro de ambiente ES_IPV6, insira o comando \$ EXPORT ES_IPV6=1, e depois edite o ficheiro ES_NODE_ROOT/nodeinfo/esprofile e adicione as linhas seguintes. Para uma instalação do servidor múltiplo, execute este passo em todos os servidores.

ES_IPV6=1 EXPORT ES_IPV6

- 4. Se utilizar o WebSphere Application Server em vez de o servidor de aplicação da Web incorporado, execute os passos seguintes nos servidores de procura:
 - a. Actualize as propriedades personalizadas da Java Virtual Machine no WebSphere Application Server.
 - 1) Se estiver em execução, pare o servidor da aplicação ESSearchServer no WebSphere Application Server.
 - 2) Abra a consola de administração WebSphere Application Server e navegue até ao painel das propriedades personalizadas da Java Virtual Machine. Seleccione Servidores > Servidores de Aplicação > ESSearchServer > Java e Gestão de Processos > Definição de Processos > Máquina Virtual Java > Propriedades Personalizadas.
 - 3) Configure as propriedades seguintes:

java.net.preferIPv4Stack=false
java.net.preferIPv6Addresses=true

Se a propriedade personalizada não estiver já na lista, crie-a. Insira o nome da propriedade no campo **Nome** e um nome válido no campo **Valor**.

- 4) Se estiver a executar o WebSphere Application Server Versão 6.1, adicione a propriedade LocalIPAddress ao ficheiro ES_INSTALL_ROOT/ nodeinfo/es.cfg para conter o endereço de IPv6 para esse servidor de pesquisa.
- 5) Reinicie a aplicação ESSearchServer.
- 5. Reinicie o CCL em todos os servidores:

startccl.sh -bg

- 6. Reinicie o sistema IBM Content Analytics with Enterprise Search: esadmin system startall
- 7. Verifique o ficheiro ES_NODE_ROOT/logs/ccl_*hostname_*0.log e verifique se o servidor CCL iniciou sem erros. Se o IPv6 for correctamente activado, serão registadas mensagens semelhantes às seguintes:

```
INFO: CCL server ready for business. Now, waiting for external requests.
CCL host name is fe80::250:56ff:feb4:27d1
CCL host dot.ip address is fe80:0:0:0:250:56ff:feb4:27d1
CCL server port is 6002
Total JVM Runtime memory is 33022Kb
Current free memory is 21346Kb
java.net.preferIPv6Addresses=true
java.net.preferIPv4Stack=false
INFO: Session "TraceDaemonSession" was attached PID=4992
```

8. Siga os seguintes passos para verificar se a aplicação de procura empresarial e a consola de administração estão a funcionar correctamente:

- a. Inicie os servidores de pesquisa para qualquer colecção e depois abra a procura empresarial e submeta uma consulta. Este passo verifica se o processamento de envio da consulta está a funcionar, mesmo que não haja documentos no índice para procurar.
- b. Abra a consola de administração e verifique se é capaz de iniciar sessão.

Desactivar suporte para o protocolo IPv6

Para desactivar o suporte para o protocolo IPv6:

- 1. Consulte os passos anteriores para preparar o sistema para a utilização do protocolo IPv6 e, para cada ocorrência de um endereço IPv6, substitua o valor com um nome do sistema central ou endereço IPv4.
- 2. Siga o procedimento para activar o suporte para o protocolo IPv6 com as seguintes excepções:
 - No passo 3 na página 42, remova as duas linhas que adicionou à variável de ambiente **ES_IPV6**.
 - Se utilizar WebSphere Application Server, remova as propriedades de JVM personalizadas que foram definidas para a aplicação ESSearchServer no passo 4 na página 42.
- 3. Remova a propriedade LocalIPAddress do ficheiro ES_INSTALL_ROOT/nodeinfo/ es.cfg caso ele contenha um endereço de IPv6.
- 4. Reinicie o sistema IBM Content Analytics with Enterprise Search:

```
esadmin system
stopall
esadmin system startall
```

5. Siga o procedimento no passo 8 na página 42 para verificar a aplicação de procura empresarial e a consola de administração.

Administração da colecção

Uma colecção contém o conjunto completo de origens que os utilizadores podem consultar de cada vez.

Ao criar uma colecção, especifique as opções aplicadas a toda a colecção. A colecção está vazia até o conteúdo ser adicionado à mesma.

Federação da colecção

Se o suporte da federação for criado na aplicação da procura empresarial, os utilizadores podem pesquisar várias colecções em simultâneo. A federação também permite dimensionar para além da limitação de tamanho para uma colecção. Por exemplo, os utilizadores podem pesquisar duas colecções em que cada uma contenha 1,000,000 documentos.

A qualidade da procura depende das classificações geradas por colecções individuais, que estão intercaladas para produzir o conjunto de resultados final. Os resultados são iguais aos da submissão de duas consultas em separado e posterior intercalação e classificação dos resultados.

A federação da colecção não é suportada para colecções de análise de conteúdo. No prospecção de análise de conteúdo, pode consultar uma colecção de cada vez.

Diferenças entre tipos de colecções

Estão disponíveis diferentes funções para interpretar, analisar, indexar e procurar conteúdo para colecções de procura empresarial e colecções de análise de conteúdo.

Se criar colecções de procura empresarial, pode utilizar uma aplicação de procura empresarial para consultar várias colecções de uma só vez. Se criar colecções de análise de conteúdo, pode utilizar o prospecção de análise de conteúdo para explorar uma única colecção de cada vez. Algumas das diferenças chave entre os tipos de colecção estão aqui resumidas, seguidas de tabelas que comparam as funções que estão disponíveis para cada tipo de colecção.

Facetas

Numa aplicação de procura empresarial, as facetas são devolvidas, de modo dinâmico, com base em documentos que correspondem à consulta. Para que uma faceta seja apresentada, um valor da faceta deve ocorrer em, pelo menos, um documento que corresponda às condições de consulta. Os administradores podem determinar o modo como as facetas são associadas a documentos no índice ao criar campos de índice ou ao permitir campos de origem para procura de facetas (como, por exemplo, campos pesquisados por pesquisadores, colunas importadas por ficheiros CSV e elementos em ficheiros XML e HTML). As facetas também podem ser associadas a documentos por anotadores que analisem o conteúdo quando os documentos são adicionados ao índice.

No prospecção de análise de conteúdo, todas as facetas que são definidas para a colecção de análise de conteúdo estão sempre disponíveis para que possa explorar e comparar todos os dados na colecção. Em tempo real, pode procurar em todas as facetas para explorar a correlação e estatísticas de frequência, tendências, relações e o modo como as diferentes facetas do seu conteúdo mudam com o tempo. Numa colecção de análise de conteúdo, os valores da faceta derivam do processamento linguístico de conteúdo não estruturado, campos e metadados nos documentos de origem, anotadores e regras e dicionários de análise de texto personalizado.

Para uma colecção de análise de conteúdo, as facetas para os metadados do documento podem ser criadas automaticamente. Quando cria a colecção, especifica se a colecção deve gerar automaticamente facetas a partir do tamanho do ficheiro, extensão do ficheiro e campos de metadados de data de última modificação.

Anotadores

Para ambos os tipos de colecções, os anotadores incorporados melhoram a capacidade de obtenção de conteúdo ao determinar, automaticamente, o idioma de um documento e ao extrair conceitos e frases de conteúdo não estruturado. Se utilizar o IBM Content Classification, um anotador pode associar metadados a documentos no índice com base em categorias e campos de Content Classification.

É possível permitir que uma colecção utilize o anotador de reconhecimento da entidade designada, o qual extrai nomes adequados de texto e correlaciona os nomes com facetas predefinidas para pessoas, localizações e organizações. É, também, possível criar um anotador personalizado e definir as suas próprias regras para analisar e anotar documentos (nem todos os idiomas são suportados).

São facultados anotadores adicionais para colecções de análise de conteúdo. Por exemplo, os anotadores analíticos podem procurar palavras definidas pelo utilizador e marcar as palavras com um caminho de faceta especificado. Um anotador pode, também, extrair padrões de texto e correlacionar padrões com facetas predefinidas para as partes da categoria gramatical.

Segmentação n-gram

Para uma colecção de procura empresarial, é possível permitir a segmentação n-gram para idiomas que não utilizem espaços em branco como delimitadores de palavras. Por exemplo, se a colecção incluir idiomas como chinês, japonês ou coreano, nos quais os espaços em branco não delimitam a palavra, é possível utilizar o método de segmentação n-gram para análise lexical em vez de segmentação de espaços em branco com base em Unicode.

Para ambos os tipos de colecções, é possível activar a segmentação híbrida, uma abordagem que melhora a qualidade da procura ao processar documentos com uma combinação de segmentação baseada no dicionário específico de idioma e com uma segmentação n-gram baseada em regras.

Dicionários e ficheiro de regras

Para ambos os tipos de colecções, é possível associar dicionários de sinónimos personalizados, dicionários de palavras de paragem e dicionários de palavras hierárquicas para influenciar o modo como as consultas são processadas e os resultados classificados.

Para uma colecção de procura empresarial, é possível configurar dicionários e regras personalizadas para expandir as consultas durante o processamento de consultas. Se os termos de consulta corresponderem a um padrão que especificar na regra, a regra é aplicada antes da devolução de resultados. Por exemplo, os termos de consulta podem ser aumentados ou substituídos por palavras na regra. A regra pode, também, fazer com que os documentos que correspondam aos termos de consulta expandida tenham uma classificação mais alta ou baixa nos resultados.

Para uma colecção de análise de conteúdo, é possível configurar dicionários e regras personalizadas para a análise de conteúdo. Quando os processos de análise são executados, as definições nos dicionários são utilizadas para extrair e anotar texto, como, por exemplo, identificação de termos que devem estar presentes como facetas no prospecção de análise de conteúdo. As regras personalizadas são utilizadas para extrair padrões do texto, o que permite que os conceitos sejam analisados para uma melhor percepção do conteúdo.

Análise em tempo real

Ambos os tipos de colecções suportam a API de processamento da linguagem natural em tempo real (NLP). É possível executar a análise de texto ad-hoc em documentos e observar as mesmas informações do que a janela de pré-visualização na aplicação. A análise de texto em tempo real utiliza os recursos de análise textual existentes que estão definidos para uma colecção, mas analisa o conteúdo sem adicionar os documentos ao índice. É possível verificar os resultados da análise sem aguardar pela construção ou actualização do índice.

Exportar documentos e resultados da análise

Para ambos os tipos de colecções, é possível exportar documentos após serem pesquisados e após serem interpretados e analisados. Após os documentos serem indexados, os utilizadores de aplicação podem submeter pedidos para exportar documentos que correspondam às suas condições de consulta.

Para uma colecção de procura empresarial, é possível configurar o suporte para exportação de documentos a partir da aplicação de procura empresarial para uma colecção de análise de conteúdo. Neste cenário, os utilizadores assinalam os documentos que pretendem exportar para que possam efectuar uma análise mais profunda com o prospecção de análise de conteúdo.

Se tiver instalado o IBM Cognos Business Intelligence, é possível exportar documentos a partir de uma colecção de análise de conteúdo directamente para uma base de dados relacional de IBM Cognos BI. É, então, possível utilizar as ferramentas IBM Cognos BI para uma análise mais profunda ao conteúdo, como, por exemplo, modificar o modelo de dados para adicionar campos estruturados de outra base de dados.

Administração de análise e indexação

Tabela 7. Comparação de funções de análise e indexação por tipo de colecção

	Colecção de Procura Empresarial	Colecção de Análise de Conteúdo
Serviços de análise e indexação . Pode ver ou configurar opções para utilizar segmentação n-gram, activar suporte para pesquisa XML nativa e especificar opções para analisar termos compostos.	Х	Х
Campos de índice . É possível especificar opções sobre como os campos podem ser procurados e devolvidos em resultados. Por exemplo, pode especificar se um campo pode ser pesquisado com uma consulta de texto livre ou por nome de campo, e se o valor do campo pode ser consultado mas não apresentado nos resultados.	X	Х

	Colecção de Procura Empresarial	Colecção de Análise de Conteúdo
Vistas contextuais . É possível permitir aos utilizadores explorar conteúdo num contexto específico ao agrupar campos numa vista contextual. Por exemplo, se configurar vistas contextuais de Pergunta e Resposta, os utilizadores podem limitar os resultados às partes dos documentos que correspondem a um ou a ambos os contextos.		X
Facetas . É possível configurar facetas para procura, de modo a que os utilizadores possam filtrar e restringir resultados para encontrar documentos específicos.	Х	
Recursos de análise . É possível configurar facetas de análise para que os utilizadores possam descobrir correlações e anomalias quando exploram a colecção. É possível criar e associar dicionários personalizados e ficheiros de regras para garantir que o analisador aplica as suas preferências de análise quando analisa o conteúdo. É, também, possível permitir que a colecção utilize um índice de faceta separado.		Х
Processos globais . É possível activar e agendar a detecção de documentos duplicados e definir agendamentos para executar a análise de ligação e gerar imagens de miniatura. Também é possível configurar categorias ao agrupar documentos que partilhem padrões URI ou palavras semelhantes, ou ao executar a análise de agrupamento em relação ao índice.	х	Х
Se utilizar um servidor IBM InfoSphere BigInsights, é possível executar tarefas de análise globais personalizadas, como, por exemplo, contar o número de vezes que um documento é mencionado por outros documentos.		
Anotadores . É possível seleccionar os anotadores que pretende activar para a colecção. Os tipos de anotadores que pode activar diferem entre tipos de colecção.	Х	Х
Termos de interesse . É possível permitir que o sistema detecte, automaticamente, relações entre substantivos e verbos e advérbios próximos no texto, identifique termos altamente significativos com base nessa análise e correlacione esses termos com facetas de análise.		Х
Dicionários lexicais . É possível associar um dicionário personalizado a uma colecção para assegurar que o analisador aplica o vocabulário específico da empresa e construções gramaticais ao preparar conteúdo para o índice.	Х	
Correlações de campo de índice para elementos XML . É possível correlacionar elementos em documentos XML com campos de índice predefinidos, o que permite aos utilizadores consultar partes específicas de documentos XML e melhorar a precisão dos resultados.	Х	X
Correlações de campo de índice para metadados HTML . É possível correlacionar elementos de metadados em documentos HTML com campos de índice predefinidos, o que permite aos utilizadores consultar partes específicas de documentos HTML e melhorar a precisão dos resultados.	Х	X

Tabela 7. Comparação de funções de análise e indexação por tipo de colecção (continuação)

	Colecção de Procura Empresarial	Colecção de Análise de Conteúdo
Correlações de campo de índice para metadados extraídos . Caso não pretenda utilizar as correlações predefinidas entre metadados extraídos e campos de índice, é possível configurar o extractor de texto para correlacionar elementos de metadados com diferentes campos de índice.	Х	X
Opções de processamento de texto . Se os motores de análise de texto personalizados foram adicionados ao sistema, pode seleccionar um para utilizar com uma colecção. Depois de utilizar um motor de análise com uma colecção, pode especificar as opções para correlacionar o conteúdo de modo a poder ser linguisticamente analisado e anotado. Também pode especificar a forma como os resultados da análise serão correlacionados para o índice ou para tabelas da base de dados JDBC.	Х	Х
Análise de postura. É possível configurar o analisador para detectar expressões positivas, expressões negativas e expressões sem postura. A análise pode explorar a postura geral para um conjunto de documentos e ver como a postura pode mudar com o tempo.		X
Filtros de campo. É possível especificar opções para filtrar campos de índice e acções de associação. Quando o valor do campo de índice corresponde às condições de filtro, a acção associada, como, por exemplo, ajustar o valor do campo, é aplicada.	Х	Х
Formatos de data personalizados. Para uma análise mais exacta de dados de data, é possível configurar os formatos de data para configurações regionais e fusos horários específicos.	Х	Х
IBM Content Classification . É possível especificar opções que activem IBM Content Analytics with Enterprise Search para categorizar e classificar conteúdo baseado em resultados de campos e categoria que importa de IBM Content Classification.	Х	Х
URIs . É possível remover URIs do índice caso já não pretenda disponibilizá-los para procura. É, também, possível consultar informações actuais e históricas sobre o modo como um documento está indexado e sobre como pode ser procurado.	Х	Х
Memória para análise e indexação . É possível configurar a quantidade máxima de memória a utilizar para análise e indexação de documentos, e a quantidade de memória a ser utilizada por cada processador de documentos.	X	Х

Tabela 7. Comparação de funções de análise e indexação por tipo de colecção (continuação)

Administração do servidor de procura

Tabela 8. Comparação de funções do servidor de procura por tipo de colecção

	Colecção de Procura Empresarial	Colecção de Análise de Conteúdo
Opções do servidor de procura . Existem várias opções de configuração possível para influenciar o processamento de consultas, como, por exemplo, a associação de dicionários personalizados, permitindo uma memória cache de procura e controlando o comprimento dos resumos do documento. Algumas opções, como, por exemplo, limitar o número de documentos nos resultados ou definir um intervalo para detectar actualizações para o índice, estão disponíveis apenas para colecções de procura empresarial.	Х	X
Ligações rápidas . É possível fazer com que documentos específicos sejam automaticamente incluídos nos resultados sempre que um utilizador submeta uma consulta que inclua certas palavras e frases.	Х	Х
Caracteres globais . É possível configurar suporte para caracteres globais. Quando um utilizador submete uma consulta que inclui um carácter global, os resultados incluem todos os documentos que correspondem aos termos de consulta mais todos os documentos que correspondem ao padrão que o carácter global representa.	х	Х
Âmbitos. É possível configurar vistas lógicas do índice. Quando os utilizadores consultam a colecção, procuram apenas os documentos que correspondem à extensão do âmbito.	Х	Х
Suporte de memorização de teclas . É possível especificar opções para mostrar potenciais correspondências para termos de consulta à medida que um utilizador introduz uma consulta. O utilizador pode seleccionar uma das correspondências sugeridas para executar a consulta.	Х	Х
Padrões URI para influenciar pontuações . É possível atribuir factores hierárquicos a padrões URI para influenciar o modo como os documentos que correspondem ao padrão são classificados nos resultados.	Х	Х
Resultados da procura contraídos . É possível especificar opções para resumir documentos nos resultados com o mesmo prefixo URI. Por exemplo, os dois documentos mais relevantes que correspondam a um padrão URI podem aparecer nos resultados. Os utilizadores podem seleccionar ver outros documentos que correspondam ao padrão URI ao expandir os resultados contraídos.	Х	X
Sinalizadores de documentos . É possível definir sinalizadores para ajudar os utilizadores a classificar documentos. Por exemplo, pode criar sinalizadores denominados Arquivo ou Obsoleto para que os utilizadores possam sinalizar documentos candidatos a serem arquivados ou eliminados da origem de dados quando explorarem uma colecção.	х	Х
Classificação de documentos através de expansão de consulta . É possível adicionar regras personalizadas e dicionários para expandir, de forma dinâmica, consultas quando os utilizadores procuram a colecção. Se uma consulta corresponder a um padrão que especificar numa regra, a consulta é automaticamente expandida para aplicar a regra.	х	

	Colecção de Procura Empresarial	Colecção de Análise de Conteúdo
Exportação para uma colecção de análise de conteúdo . É possível configurar uma colecção de análise de conteúdo, nova ou existente, para receber documentos que os utilizadores exportam a partir da aplicação de procura empresarial.	Х	
Memória para processamento de procura . É possível configurar a quantidade máxima de memória a ser utilizada por cada servidor de procura.	Х	Х
Relatórios de inspecção profunda . É possível permitir aos utilizadores obter um relatório de inspecção profunda quando consultam uma colecção com o prospecção de análise de conteúdo. As inspecções profundas são úteis para explorar valores de correlação ou o índice de desvio quando a consulta envolve um grande número de valores da faceta que possam ser difíceis de navegar interactivamente.		Х
Relatórios de IBM Cognos Business Intelligence . É possível permitir aos analistas criar relatórios estatísticos no prospecção de análise de conteúdo. É possível ver os relatórios na aplicação. É, também, possível abrir os relatórios com as ferramentas de análise IBM Cognos BI e partilhá-los com outros analistas.		Х

Tabela 8. Comparação de funções do servidor de procura por tipo de colecção (continuação)

Criar uma colecção

Uma nova colecção está vazia até que lhe adicione conteúdo.

Depois de criar uma colecção, pode alterar algumas das definições seleccionando a colecção na vista Colecções e fazendo clique em **Acções** > **Definições** > **Editar definições de colecções**. Pode ver as seguintes definições, mas não alterá-las:

- O tipo da colecção.
- Se a colecção utiliza um cache de documento.
- Se as imagens em miniatura dos documentos podem ser mostradas nos resultados.
- Se a segurança é activada para a colecção. Se activar a segurança da colecção, pode alterar definições para activar ou desactivar a segurança ao nível do documento, mas não pode alterar a definição ao nível da colecção.
- Se a colecção utiliza um índice de registo de consulta para fornecer assistência de memorização de teclas para introduzir consultas.
- A localização do directório que é utilizado para os dados da colecção.
- O ID da colecção.
- O fuso horário da colecção.

Para criar uma colecção:

- 1. Na vista Colecções, faça clique em Criar colecção.
- 2. Na página Criar uma colecção, forneça informações ou efectue selecções nos seguintes campos necessários:
 - Nome da colecção. Especifique um nome descritivo, tendo em conta o contexto ou a finalidade da colecção.

- **Tipo de colecção**. Especifique o tipo de colecção que pretende criar: uma colecção de procura ou uma colecção de análise de conteúdo.
- Se, anteriormente, configurou o sistema para utilizar o IBM InfoSphere BigInsights e pretende que esta colecção utilize um servidor BigInsights para processar grandes quantidades de dados, faça clique na caixa de verificação **Utilizar IBM InfoSphere BigInsights**.
- Decida se quer que esta colecção utilize um cache de documento. A cache de documento, que é activado por predefinição para colecções de análise de conteúdo, é necessário para gerar miniaturas, exportar documentos pesquisados e exportar documentos a partir de uma aplicação de procura empresarial para uma colecção de análise de conteúdo. A cache de documento também permite reconstruir o índice sem voltar a pesquisar conteúdo. Se planear exportar projectos a partir de Content Analytics Studio para a colecção, garanta que a cache de documento está activo para que actualizações da lógica do anotador fiquem disponíveis sem voltar a pesquisar todos os documentos.
- Decida se quer permitir que os utilizadores vejam imagens em miniatura quando consultam a colecção. Imagens em miniatura podem ser geradas para certos tipos de documentos, tais como documentos Microsoft Office e PDF. Quando agenda opções de processamento global para a colecção, pode especificar uma agenda para gerar imagens em miniatura de cache de documentos. Pode também editar um ficheiro de configuração para alterar o tamanho da apresentação das imagens em miniatura nos resultados da pesquisa.
- Aceite os valores predefinidos para todas as opções desta página ou faça clique em Opções avançadas para especificar preferências para a colecção. As opções que pode configurar e os valores predefinidos dependem do tipo de colecção que criou.
- Faça clique em OK. A vista Colecções enumera a nova colecção com outras colecções no sistema. Se activou a segurança para a colecção, o ícone
 Segurança da colecção está activada é apresentado ao lado do nome da colecção.

A colecção está vazia até o conteúdo ser adicionado à mesma. Expanda a colecção na vista Colecções, adicione conteúdo criando pelo menos um pesquisador ou importando ficheiros CSV e, depois, especifique opções para como pretende que os dados sejam analisados, indexados e procurados.

Conceitos relacionados:

Classificação de documentos

Análise linguística de documentos em chinês, japonês e coreano

Tarefas relacionadas:

"Determinar o ID de colecção" na página 55

"Clonar uma colecção"

Categorizar caracteres SBCS e numéricos como categorias n-gram

"Exportar documentos de um colecção de procura empresarial para uma colecção de análise de conteúdos" na página 304

Clonar uma colecção

Pode criar uma nova colecção que é baseada nas definições que configurou para outra colecção.

Clonar uma colecção copia todos os dados de configuração de uma colecção para uma colecção nova, incluindo todos os pesquisadores, opções para pesquisa, análise e indexação de conteúdos, anotadores e opções para procurar a colecção. Depois de a nova colecção ser criada, pode modificar apenas as definições que pretende alterar para esta colecção, tais como pesquisar origens de dados diferentes ou activar um anotador personalizado diferente.

Para clonar uma colecção:

- 1. Na vista Colecções, faça clique em **Acções** > **Clonar esta colecção** para a colecção que pretende clonar.
- 2. Na página Clonar uma Colecção especifique um nome para a nova colecção. Não pode alterar o tipo da colecção.
- 3. Aceite os valores predefinidos ou valores clonados para todas as restantes opções nesta página e faça clique em Opções avançadas para especificar diferentes preferências para essa colecção. Os valores predefinidos e as opções que pode configurar dependem do tipo de colecção que clonou.
- Faça clique em OK. A vista Colecções enumera a nova colecção com outras colecções no sistema. Se activou a segurança para a colecção, o ícone
 Segurança da colecção está activada é apresentado ao lado do nome da colecção.

A colecção está vazia até o conteúdo ser adicionado à mesma. Para adicionar conteúdo, expanda a colecção na vista Colecções. Pode reconfigurar os pesquisadores que foram clonados para essa colecção ou adicionar conteúdo importando ficheiros CSV. Também pode configurar diferentes opções para analisar, indexar e procurar esta colecção.

Para procurar a colecção clonada, tem de pesquisar ou importar dados para que o índice desta nova colecção possa ser construído.

Tarefas relacionadas:

"Criar uma colecção" na página 51

Editar uma colecção

Quando edita uma colecção, especifica as opções acrescentado conteúdo, analisando e indexando documentos, procurando e indexando conteúdo e registando mensagens de erro.

Para editar uma colecção:

- 1. Na vista Colecções, expanda a colecção que pretende editar.
- 2. Efectue alterações em qualquer das seguintes áreas:

Menu de Acções

Alterar definições gerais para a colecção e definições de visualização que não pode alterar. Outras acções incluem especificar opções para registar ficheiros, clonar a colecção e eliminar a colecção.

Painel Pesquisar e Importar

Adicione conteúdo à colecção configurando pesquisadores e importando ficheiros CSV:

 Quando configura pesquisadores, especifica a origem de dados que pretende pesquisar e especifica opções de como o conteúdo deve ser pesquisado. Uma única colecção pode conter dados a partir de vários tipos de origens de dados. Configure pelo menos um pesquisador para cada tipo de origens de dados que pretende incluir na colecção. Quando importa ficheiros CSV, especifica os caminhos e nomes dos ficheiros CSV que pretende adicionar ao índice da colecção. Importar ficheiros CSV permite ao utilizador adicionar conteúdo à colecção sem configurar um pesquisador.

Também pode parar e iniciar pesquisadores, supervisionar o progresso do pesquisador, supervisionar pedidos de importação, configurar opções para exportar documentos pesquisados e supervisionar o estado dos pedidos de exportação.

PainelAnalisar e Indexar

Especifique as opções sobre como pretende que os documentos que foram pesquisados sejam interpretados, analisados e indexados. Também pode iniciar e parar tarefas de processamento de tarefas e indexar compilações, configurar e supervisionar processos globais, activar e desactivar anotadores, configurar opções para exportar documentos pesquisados ou analisados e supervisionar o estado dos pedidos de exportação.

Para uma colecção de análise de conteúdos, pode configurar a árvore de facetas correlacionando palavras-chave e sinónimos a facetas, adicionar dicionários personalizados e definir regras para padrões extraídos de texto de conteúdo não estruturado. Também pode supervisionar o índice da faceta opcional e implementar novamente a análise de recursos.

Painel Procurar e painel Análise

Especifique opções para os processos de procura, tal como configurar um cache de procura e opções de utilização de caracteres globais em consultas. Também pode configurar opções para exportar documentos a partir de resultados de procura e supervisionar o estado dos pedidos de exportação. Para supervisionar o desempenho de procura, pode ver estatísticas detalhadas sobre o processamento de consulta.

Para uma colecção de análise de conteúdo, pode configurar opções de inspecção profunda se quiser exportar resultados de análises para grandes conjuntos de documentos em vez de ver os resultados na interface do utilizador deprospecção de análise de conteúdo. Também pode configurar opções para gerar relatórios que podem ser vistas com ferramentas IBM Cognos Business Intelligence.

Eliminar uma colecção

Eliminar uma colecção por completo remove toda a informação acerca da colecção do sistema.

Tem de parar todos os processos associados à colecção antes de a poder eliminar.

A eliminação duma colecção pode ser um processo longo. Depois de confirmar que pretende eliminar a colecção, o sistema elimina todos os dados no sistema relacionado com a colecção.O processo de eliminação é terminado quando o nome da colecção já não aparecer na lista das colecções.

Para eliminar uma colecção:

- 1. Na vista Collections, localize a colecção que pretende eliminar.
- 2. Faça clique em Acções > Eliminar esta colecção.

Determinar o ID de colecção

Para muitas tarefas administrativas, necessita de saber o ID da colecção.

Ao criar uma colecção, pode especificar um valor para o ID ou permitir que o sistema atribua um ID automaticamente. Para determinar o ID da colecção depois de uma colecção ser criada, pode utilizar a consola de administração, executar um comando ou ver um ficheiro de configuração.

Para determinar o ID da colecção:

Utilize qualquer uma das seguintes formas:

- Na vista Colecções, localize a colecção que pretende ver e faça clique em Acções
 > Definições > Ver definições de colecção. Pode ver o ID da colecção, o caminho totalmente qualificado em que os dados da colecção estão armazenados e outra informação sobre a colecção que não pode ser alterada.
- Inicie sessão como o administradorIBM Content Analytics with Enterprise Search predefinido e execute o comando **colecções de relatório esadmin**.
- Inicie sessão como administrador IBM Content Analytics with Enterprise Search predefinido e abra o ficheiro ES_NODE_ROOT/master_config/collections.ini.

Para uma visualização mais fácil, ordene este ficheiro e veja o nome de apresentação que corresponde à colecção. As propriedades para a colecção tem o mesmo número da colecção. Procure a propriedade colecçãonúmero.id para determinar o ID da colecção. Na seguinte saída exemplo, coll é o ID da colecção para uma colecção com o nome de apresentação Collection1:

% sort \$ES_NODE_ROOT/master_config/collections.ini | more

```
collection1.configfile=col1_config.ini
collection1.datadir=/home/esearch/node/data/col1
collection1.description=
collection1.displayname=Collection1
collection1.flags=0
collection1.id=col1
collection1.sectiontype=collection
collection1.type=1
...
```

Tarefas relacionadas:

"Criar uma colecção" na página 51

Administração do pesquisador

Configure os pesquisadores para tipos diferentes de dados que pretende para incluir numa colecção. Uma única colecção pode conter qualquer número de pesquisadores.

Um pesquisador tem duas funções principais. Ao configurar um pesquisador, os processos de *identificação* determinam quais as origens disponíveis na origem de dados, tal como os nomes de todas as vistas e pastas numa base de dados Lotus Notes. Após iniciar um pesquisador, o pesquisador processa os dados de cópia a partir das fontes que selecciona para um arquivo de dados, de modo a que o pesquisador possa preparar o conteúdo para indexação e procura.

Configurar pesquisadores

Utilize a consola de administração para criar, editar e eliminar pesquisadores. Normalmente, o pesquisador é configurado por um especialista nos tipos de dados a pesquisar. Por exemplo, para configurar um pesquisador para pesquisar origens de dados do Lotus Notes, o administrador da colecção deve ser um administrador do Notes ou trabalhar próximo de alguém que tenha conhecimentos acerca das bases de dados que estão a ser pesquisadas.

Quando cria um pesquisador, um assistente ajuda-o a especificar as *propriedades do pesquisador* (opções que descrevem o pesquisador e definem limites sobre como este utiliza recursos do sistema) e adiciona origens ao *espaço de pesquisa* (um conjunto de origens de dados que um pesquisador específico deve pesquisar).

Pode também alterar os pesquisadores já existentes em qualquer momento. Pode editar as propriedades do pesquisador ou partes do espaço do pesquisador conforme necessário. Os assistentes do pesquisador ajudam-no a efectuar essas alterações.

Preencher um pesquisador novo com valores base

Pode criar um pesquisador utilizando os valores predefinidos do sistema, ou copiando os valores especificados para um pesquisador do mesmo tipo. Se utilizar um pesquisador existente como base para um novo pesquisador, pode criar rapidamente pesquisadores múltiplos que têm propriedades semelhantes e então configurá-las, por exemplo, para preencher diferentes recursos ou operar em diferentes agendamentos de pesquisa.

Ao copiar um pesquisador, pode dividir o volume de trabalho pesquisado entre os múltiplos pesquisadores que utilizam as mesmas regras de pesquisa. Por exemplo, pode copiar um pesquisador do Notes porque pretende utilizar as mesmas propriedades e as mesmas regras de pesquisa de campo com um servidor de Lotus Notes diferente. As únicas diferenças podem ser as bases de dados que cada pesquisador pesquisa e as definições de segurança ao nível do documento.

Associar tipos de pesquisadores numa colecção

Os pesquisadores são concebidos para reunir informações de tipos específicos de origens de dados. Ao configurar os pesquisadores para uma colecção, tem de decidir o procedimento para associar esses diferentes tipos de origens de dados

para que os utilizadores possam procurar facilmente os dados da empresa. Por exemplo, se pretender que os utilizadores consigam procurar sistemas de ficheiros e pastas públicas do Microsoft Exchange Server Microsoft Windows com uma única consulta, criar uma colecção que inclua os pesquisadores do Sistema de ficheiros Windows e Exchange Server.

Ao associar vários tipos de pesquisadores numa única colecção, certifique-se de que todos os pesquisadores utilizam o mesmo método de classificação estático. (Especifique o método de classificação estático ao criar a colecção.) Por exemplo, se associar origens da Web (que utilizam ligações de documentos como factor de ordenação) e origens NNTP (que normalmente utilizam a data dos documentos como factor de ordenação), a qualidade dos resultados da procura poderá decrescer.

Configurar a segurança de nível de documento

Se activar a segurança para uma colecção ao criá-la, pode configurar as opções de segurança ao nível do documento. Cada pesquisador pode utilizar tokens de segurança com os documentos que pesquisa. Se especificar que pretende utilizar a segurança ao nível do documento quando configurar o pesquisador, este utiliza os tokens de segurança que especifica com cada documento, e esses tokens são adicionados ao índice com os documentos.

Se activar a segurança nas respectivas aplicações personalizadas da procura empresarial, as aplicações podem utilizar os tokens de segurança que os pesquisadores utilizaram com os documentos para autenticar os utilizadores. Essa função permite-lhe restringir o acesso a alguns documentos numa colecção e consentir outros documentos a serem pesquisados por todos os utilizadores. Por exemplo, numa colecção todos os utilizadores podem aceder a todos os documentos nas respectivas pastas públicas do Microsoft Exchange Server, mas só são permitidos utilizadores com IDs de utilizador específicos para aceder aos documentos nas respectivas bases de dados do Lotus Notes.

Pode aplicar regras de negócio personalizadas para determinar o valor dos tokens de segurança, codificando as regras numa classe Java. Quando configurar as propriedades do pesquisador, especifique o nome do suplemento que pretende que o pesquisador utilize quando pesquisar documentos. Os tokens de segurança que o respectivo suplemento adiciona estão armazenados no índice e podem ser utilizados para controlar o acesso aos documentos.

Ao configurar certos tipos de pesquisadores, pode especificar controlos de segurança adicionais. Por exemplo, pode especificar que pretende validar os utilizadores durante o processamento da consulta. Se activar essa opção, as credenciais do utilizador serão comparadas com as listas de controlo de acesso actuais actualizadas pelas origens de dados a pesquisar. Esta validação de credenciais actuais pode ser efectuada em vez de ou além da validação que se baseia em tokens de segurança no índice.

Referências relacionadas:

🖙 Origens de dados suportadas

Criar um pesquisador

Quando criar um pesquisador, especifique o tipo de pesquisador que pretende criar. Um assistente ajudá-lo-á a especificar a informação acerca dos dados que pretende incluir na colecção e como o pesquisador deverá utilizar os recursos do sistema.

Antes de criar muitos tipos de pesquisadores, deve executar scripts de configuração ou implementar serviços da Web no servidor do pesquisador, ou o servidor de origem de dados antes de criar o pesquisador. Consulte os requisitos para o tipo de pesquisador que pretende criar antes de executar o assistente do pesquisador.

Para criar um pesquisador:

- 1. Na vista Colecções, expanda a colecção na qual pretende adicionar um pesquisador.
- 2. No painel Pesquisar e Importar, faça clique em Adicionar um novo pesquisador.
- 3. Seleccione o tipo de pesquisador e os valores base do pesquisador:
 - a. Seleccione o tipo de pesquisador que suporta o tipo de dados que pretende pesquisar, tal como os sítios da Web, as bases de dados do Lotus Notes, ou sistemas de ficheiros UNIX.
 - b. Seleccione os valores base para o pesquisador. Pode alterar os valores iniciais para especificar diferenças para este pesquisador mais à frente no assistente.

Utilizar os valores predefinidos do sistema para o novo pesquisador

Aplica os valores predefinidos da instalação às definições iniciais do pesquisador. Se seleccionar esta opção, faça clique em **Seguinte** para iniciar a configuração do novo pesquisador.

Clonar os valores aplicados a um pesquisador já existente para o novo pesquisador

Preenche as definições iniciais do pesquisador ao copiar valores que foram especificados para outro pesquisador deste tipo. Se seleccionou esta opção, será apresentada uma lista com todos os pesquisadores no sistema que correspondem a esse pesquisador. Seleccione o pesquisador que pretende clonar por esse pesquisador, e depois faça clique **Seguinte** para iniciar a configuração do novo pesquisador.

O ícone A segurança da colecção está activa indica se o pesquisador que seleccionou para clonar pertence a uma colecção segura.

Depois de criar um pesquisador, é listado no painel Pesquisar e Importar com outros pesquisadores que pertencem à colecção. Pode editar as propriedades do pesquisador e do espaço de pesquisa em qualquer momento que precise de fazer alterações ao pesquisador. Para editar um pesquisador:

- 1. Localize o pesquisador no painel Pesquisar e Importar.
- 2. Faça clique em Configurar > Editar propriedades do pesquisador ou Configurar > Editar espaço de pesquisa.

 Para que as alterações se tornem efectivas, pare e reinicie o pesquisador. Se alterar apenas a descrição do pesquisador, não necessita de reiniciar o pesquisador.

Eliminar um pesquisador

A eliminação de um pesquisador implica a remoção de toda a informação acerca do pesquisador do respectivo sistema. As informações anteriormente pesquisadas pelo pesquisador permanece no índice até ocorrer a próxima criação do índice principal.

A eliminação dum pesquisador pode ser um processo longo. Depois de confirmar que pretende eliminar o pesquisador, o sistema elimina todos os dados no sistema relacionados com o pesquisador.

Sugestão: Como esta tarefa leva tempo até ser concluída, pode ver uma mensagem a indicar que a operação solicitada esgotou o tempo apesar de o processo ainda estar a ser executado em segundo plano. O processo de eliminação é terminado quando o nome do pesquisador já não aparecer na lista de pesquisadores.

Para eliminar um pesquisador:

- Na vista Colecções, expanda a colecção do qual pretende eliminar um pesquisador.
- No painel Pesquisador e Importar, expanda a lista de pesquisadores na colecção e faça clique no ícone Apagar para o pesquisador que pretende eliminar.

Agendamentos de pesquisadores

Ao configurar pesquisadores, pode especificar agendamentos para quando cada pesquisador for executado.

Para alguns tipos de origens de dados, um único agendamento controla quando o pesquisador visita todas as origens de dados no espaço de pesquisa. Para outros tipos de origens de dados, pode especificar diferentes agendamentos para diferentes repositórios. Por exemplo, pode especificar diferentes agendamentos para pesquisar cada base de dados do Lotus Notes que o pesquisador do Notes pesquisa.

Os pesquisadores que criar para origens Web podem ser executados de forma contínua. Após iniciar um pesquisador, normalmente, não é necessário pará-lo, a menos que altere a respectiva configuração. Em alguns casos, poderá pretender especificar agendamentos para pesquisar servidor Web específicos em momentos específicos. Por exemplo, poderá criar um agendamento para pesquisar determinados servidores Web durante períodos de baixa utilização como, por exemplo, durante a noite quando a maioria dos utilizadores não está ligada.

Ao configurar o agendamento, especifique o tipo de pesquisa que deve ser efectuado. Pode agendar uma pesquisa completa de todos os documentos no espaço de pesquisa, agendar uma pesquisa que inclua todas as actualizações no espaço de pesquisa (documentos novos, documentos modificados e documentos eliminados) ou agendar uma pesquisa que inclua apenas documentos novos e documentos modificados. Uma pesquisa completa leva mais tempo a concluir. Uma pesquisa que remove os documentos eliminados leva mais tempo do que uma pesquisa que ignora os documentos eliminados.
Se seleccionar a opção de pesquisar apenas documentos novos e modificados, o pesquisador procura documentos com datas de modificação que são posteriores à hora da pesquisa anterior. Se copiar ficheiros para um recurso, a data de modificação pode não se alterar, o que significa que o pesquisador pode não detectar que os ficheiros foram adicionados ao recurso. Por exemplo, se copiar os ficheiros para uma pasta Windows, o Windows não muda automaticamente a data de modificação dos ficheiros. Para assegurar que esses ficheiros são pesquisados, seleccione a opção para pesquisar todas as actualizações ou uma pesquisa completa.

Ao editar um espaço de pesquisa do pesquisador, pode especificar um segundo agendamento de pesquisa. Por exemplo, pode pretender configurar um agendamento para pesquisar todos os documentos no espaço de pesquisa todos os Sábados à noite, e configurar um segundo agendamento que é executado com mais frequência para pesquisar documentos novos e modificados. Para outro exemplo, poderá agendar o pesquisador da Web a efectuar uma pesquisa completa de determinados sítios da Web cada semana, e configurar um segundo agendamento para procurar documentos novos e modificados três ou quatro vezes por dia.

Ao criar múltiplos agendamentos do pesquisador, pode controlar melhor a visita do pesquisador às origens de destino. Por exemplo, para pesquisar bases de dados em diferentes fusos horários, pode agendar o pesquisador para determinadas horas em que os utilizadores estão mais aptos a terminar o seu dia de trabalho.

Pesquisadores do Agente para sistemas de ficheiros Windows

Se instalar o servidor do pesquisador em AIX ou Linux, pode utilizar o pesquisador do Agente para sistemas de ficheiros Windows para adicionar documentos que são armazenados em sistemas de ficheiros remotos Microsoft Windows para uma colecção.

Compare este pesquisador com o pesquisador do Sistema de ficheiros Windows, que exige que o servidor seja instalado em Windows. Como o pesquisador do Sistema de ficheiros Windows, pode utilizar o pesquisador do Agente para sistemas de ficheiros Windows para pesquisar um número qualquer de sistemas de ficheiros do Windows. Quando configurar o pesquisador, seleccione as pastas que pretende pesquisar.

Requisitos de configuração

Para utilizar o pesquisador, tem de instalar um servidor do agente num servidor de ficheiros Windows ou num servidor Windows que instala um ou mais servidores de ficheiros. O servidor do agente é um serviço do Windows que obtém dados de servidores de origens de dados e os envia para o cliente do pesquisador que está em execução no servidor do pesquisador do IBM Content Analytics with Enterprise Search.

O servidor do agente tem um ficheiro de configuração que define a correlação de pastas locais ou de rede como pastas de rede partilhadas. Quando configurar o pesquisador, especifique um ID de utilizador e palavra-passe que permitam ao pesquisador ligar às pastas de rede partilhadas e pesquisar conteúdo.

O servidor do agente é iniciado quando o sistema Windows no computador em que o agente está instalado for iniciado. O agente aguarda numa porta de TCP/IP por pedidos de ligação de clientes do pesquisador. Quando um cliente de pesquisador começar a ser executado, estabelece ligação à porta do servidor do agente. Quando configurar o pesquisador, especifique o nome do sistema central do servidor do agente e as portas a serem utilizadas para transferir dados e autenticar com o servidor. Os números das portas são configurados quando é instalado o servidor do agente.

O servidor do agente remoto e os servidores de ficheiros a serem pesquisados têm de estar no mesmo domínio do Windows ou num grupo de trabalho. Um pesquisador pode recolher dados de controlo de acesso (ACL) de um domínio Windows único. Deste modo, o servidor do agente tem de pertencer ao domínio de que são recolhidos os dados de ACL ou o domínio que se encontra num relacionamento fidedigno com esse domínio.

Credenciais de ligação do pesquisador

Quando cria o pesquisador, pode especificar as credenciais que permitem que o pesquisador se ligue às fontes a pesquisar. Também pode configurar as credenciais de ligação quando especifica as definições gerais de segurança do sistema. Se utilizar a última abordagem, vários pesquisadores e outros componentes do sistema podem utilizar as mesmas credenciais. Por exemplo, os servidores de procura podem utilizar as credenciais para determinar se um utilizador possui autorização para aceder ao conteúdo.

Descrição geral da configuração

Ao criar o pesquisador, um assistente ajudá-lo-á nessas tarefas:

- Especifique as propriedades que controlam o modo como o pesquisador opera e utiliza os recursos do sistema. As propriedades do pesquisador controlam a forma como o pesquisador pesquisa todas as pastas no espaço de pesquisa.
- Defina um agendamento para a pesquisa de sistemas de ficheiros.
- Especifique informações que permitem ao pesquisador ligar ao servidor do agente e aos sistemas de ficheiros a serem pesquisados.
- Seleccione pastas a pesquisar.

Pode especificar quantos mais níveis de subpastas pretende que o pesquisador pesquise.

- Especifique opções para tornar os documentos em pastas pesquisáveis. Por exemplo, pode excluir certos tipos de documentos do espaço de pesquisa.
- Configure as opções de segurança ao nível do documento. Se a segurança tiver sido activada quando a colecção foi criada, o pesquisador pode associar dados de segurança aos documentos existentes no índice. Esses dados permitem às aplicações fazer cumprir os controlos de acesso baseados nas listas de controlo de acesso armazenadas ou nos tokens de segurança.

Pode seleccionar uma opção para validar as credenciais do utilizador no momento em que um utilizador permitir uma consulta. Nesse caso, em vez de comparar as credenciais do utilizador com os dados de segurança indexados, o sistema compara as credenciais com as listas de controlo de acesso actuais que são actualizadas pela origem de dados original.

Para reforçar a segurança ao nível do documento, tem de garantir que o utilizador a informação de conta do domínio seja configurada correctamente no servidor do pesquisador.

Para maximizar o rendimento, o pesquisador do Agente para sistemas de ficheiros Windows pode detectar actualizações a ficheiros e determinar se um ficheiro necessita de ser pesquisado novamente sem o abrir.

Conceitos relacionados:

- 🕒 Instalação do servidor de agente
- 🕩 Configurar directórios partilhados num servidor de agente

Comandos de administração do servidor de agente

Depois de instalar um servidor de agente, pode inserir comandos para iniciar, parar e ver o estado do servidor.

O servidor de agente permite ao pesquisador do Agente para sistemas de ficheiros Windows para pesquisar sistemas de ficheiros remotos do Windows quando o servidor do pesquisador estiver instalado em AIX ou Linux.

Para administrar o servidor de agente, execute os comandos seguinte no servidor de agente.

Iniciar ou parar o servidor de agente

Utilize estes comandos para iniciar ou parar o servidor de agente. Parar o servidor de agente também pára o pesquisador. Por exemplo, se o cliente do pesquisador parar inesperadamente, pode fechar as ligações e libertar recursos para esse pesquisador.

esagent start esagent stop

Obter o estado do servidor de agente

Utilize este comando para obter o estado do servidor agente:

esagent getStatus

A saída do comando getStatus é um ficheiro XML no formato seguinte:

```
<AgentStatus>
  <SpaceStatus>
    <SpaceId>012</SpaceId>
   <RootFolder>E:\\Projects\Analytics\\data\test1</RootFolder>
   <ConnectionNumber>9</ConnectionNumber>
   <StartTime>1244709336093</StartTime>
   <LastTime>1244709385843</LastTime>
   <IdlePeriod>219</IdlePeriod>
  </SpaceStatus>
  <SpaceStatus>
    <SpaceId>013</SpaceId>
   <RootFolder>E:\\Projects\Analytics\\data\test2</RootFolder>
   <ConnectionNumber>10</ConnectionNumber>
   <StartTime>1244709336093</StartTime>
   <LastTime>1244709385843</LastTime>
   <IdlePeriod>219</IdlePeriod>
  </SpaceStatus>
```

Conceitos relacionados:

🕒 Instalação do servidor de agente

Configurar directórios partilhados num servidor de agente

Pesquisadores do Case Manager

Pode configurar um pesquisador de Case Manager para recolher documentos e metadados de soluções de IBM Case Manager e, de seguida, utilizar os dados recolhidos para a análise de texto.

Pode configurar um pesquisador de Case Manager para pesquisar a árvore de pastas de uma solução de caso. O pesquisador começa pela pasta de solução de implementação de nível superior e depois pesquisa as subpastas de tipo caso.

Após recolher documentos e metadados de caso de pastas, o IBM Content Analytics with Enterprise Search pode ser utilizado para analisar casos de IBM Case Manager. Por exemplo, o IBM Content Analytics with Enterprise Search consegue detectar palavras-chave com elevada correlação ou anómalas em casos.

Credenciais de ligação do pesquisador

Quando cria o pesquisador, pode especificar as credenciais que permitem que o pesquisador se ligue às fontes a pesquisar. Também pode configurar as credenciais de ligação quando especifica as definições gerais de segurança do sistema. Se utilizar a última abordagem, vários pesquisadores e outros componentes do sistema podem utilizar as mesmas credenciais. Por exemplo, os servidores de procura podem utilizar as credenciais para determinar se um utilizador possui autorização para aceder ao conteúdo.

Configuração do servidor do pesquisador

Para integrar o IBM Case Manager com o IBM Content Analytics with Enterprise Search, tem de instalar o IBM FileNet Content Engine Client no servidor do pesquisador. Antes de utilizar a consola de administração para configurar um pesquisador do Case Manager, conclua a tarefa seguinte.

• "Configurar o servidor de pesquisadores para suportar pesquisadores de Case Manager e FileNet P8" na página 65.

Descrição geral da configuração

Ao criar o pesquisador, um assistente ajudá-lo-á nessas tarefas:

- Especifique as propriedades que controlam o modo como o pesquisador opera e utiliza os recursos do sistema. As propriedades do pesquisador controlam a forma como o pesquisador pesquisa todos os documentos que adicionar ao espaço de pesquisa.
- Especifique informações que permitem ao pesquisador aceder ao servidor do Case Manager. Ao configurar a ligação, especifique o URL de Content Engine Web Service e um nome de utilizador e palavra-passe válidos. Também pode especificar a localização de um ficheiro de arquivo de chaves SSL no sistema de ficheiros local do servidor de pesquisador. Este valor é definido como a propriedade javax.net.ssl.trustStore para a ligação SSL ao servidor de Content Engine Web Service e é utilizado para autenticar um servidor utilizando uma entidade certificadora fidedigna (CA).
- Seleccione a pasta de solução de implementação de nível superior ou as pastas de caso que pretende incluir no espaço de pesquisa. Ao seleccionar os destinos a pesquisar, o nível superior da árvore hierárquica mostra os armazenamentos de objectos no servidor. No segundo nível, é apresentada a Pasta Raiz. Caso seleccione a Pasta Raiz, é apresentada uma lista de pastas raiz no armazenamento de objectos no terceiro nível. Após expandir as pastas raiz, seleccione uma solução de caso e, de seguida, seleccione as pastas de caso a pesquisar.
- Especifique opções para incluir e excluir documentos específicos e opções para a forma como o conteúdo pode ser pesquisado.

- Seleccione os campos de metadados que pretende incluir e correlacione-os com os campos de índice. O pesquisador adiciona automaticamente propriedades comuns como campos de metadados como, por exemplo, nome de campo, nome de classe, data, estado de caso, ID de caso, tipo de caso, nome do caminho e título. Para adicionar outras propriedades do caso, como os metadados, edite o espaço do pesquisador e especifique os nomes das propriedades como campos de metadados do pesquisador.
- Configure um agendamento para pesquisar pastas individuais.

Tarefas relacionadas:

"Configurar credenciais de ligação" na página 380

Referências relacionadas:

"Formatos de URI no índice" na página 164

"Nomes de campos de índice reservados" na página 199

Configurar o servidor de pesquisadores para suportar pesquisadores de Case Manager e FileNet P8

Antes de configurar um pesquisador de FileNet P8 ou de Case Manager, tem de instalar o IBM FileNet Content Engine Client no servidor de pesquisadores e, de seguida, executar um script de configuração.

A versão do software cliente que instalar no servidor do pesquisador deve corresponder à versão do software do servidor que está instalado nos servidores que tenciona pesquisar.

Para configurar o servidor do pesquisador para suportar pesquisadores de FileNet P8 eCase Manager:

- 1. Instale o IBM FileNet Content Engine Client no servidor do pesquisador.
- 2. Inicie sessão como administrador predefinido do IBM Content Analytics with Enterprise Search no servidor do pesquisador. Num servidor Windows 2008, deve iniciar sessão como Administrador do Windows.
- **3**. Execute o script **escrfilenet**. Quando executar o script, especifique o caminho no servidor do pesquisador onde instalou o IBM FileNet Content Engine Client:

```
AIX ou Linux
```

```
$ES_INSTALL_ROOT/bin/escrfilenet.sh
```

Windows

%ES_INSTALL_ROOT%\bin\escrfilenet.vbs

4. Reinicie o sistema IBM Content Analytics with Enterprise Search:

```
esadmin system
stopall
esadmin system startall
```

Pesquisadores do Content Integrator

Para incluir repositórios do IBM Content Integrator numa colecção, tem de configurar um pesquisador do Content Integrator.

Utilize o pesquisador Content Integrator para pesquisar os seguintes tipos de repositório:

- Documentum
- Hummingbird Document Management (DM)

• OpenText Livelink

Embora o IBM Content Integrator suporte o armazenamento de objectos do IBM FileNet P8 Content Manager e as origens do Microsoft SharePoint, utilize os pesquisadores FileNet P8 e SharePoint para pesquisar directamente estes repositórios em vez de os aceder através do pesquisador Content Integrator.

Credenciais de ligação do pesquisador

Quando cria o pesquisador, pode especificar as credenciais que permitem que o pesquisador se ligue às fontes a pesquisar. Também pode configurar as credenciais de ligação quando especifica as definições gerais de segurança do sistema. Se utilizar a última abordagem, vários pesquisadores e outros componentes do sistema podem utilizar as mesmas credenciais. Por exemplo, os servidores de procura podem utilizar as credenciais para determinar se um utilizador possui autorização para aceder ao conteúdo.

Configuração do servidor do pesquisador

Antes de criar um pesquisador Content Integrator, tem de configurar um conector em IBM Content Integrator para permitir acesso aos repositórios. Também tem de executar o script **escrvbr** no servidor do pesquisador. Este script, que é fornecido com o IBM Content Analytics with Enterprise Search, permite ao pesquisador do Content Integrator comunicar com os servidores do Content Integrator.

Complete as seguintes tarefas:

- "Configurar um conector para aceder a repositórios do Content Integrator" na página 67.
- "Configurar o servidor do pesquisador para pesquisadores do Content Integrator" na página 67.

Descrição geral da configuração

Ao criar o pesquisador, um assistente ajudá-lo-á nessas tarefas:

- Especifique as propriedades que controlam o modo como o pesquisador opera e utiliza os recursos do sistema. As propriedades do pesquisador controlam o modo como este último pesquisa todos os repositórios no espaço de pesquisa.
- Seleccione os repositórios que pretende pesquisar.
- Especifique os IDs de utilizador que permite ao pesquisador aceder ao conteúdo nos repositórios seleccionados.
- Defina um agendamento para a pesquisa de repositórios.
- Seleccione as classes de itens que pretende para pesquisar em cada repositório.
- Especifique opções para fazer com que as características das classes de itens possam ser procuradas. Por exemplo, pode excluir certos tipos de documentos do espaço de pesquisa ou especificar que pretende pesquisar uma versão particular de um repositório.

Também pode especificar uma consulta do Content Integrator ou uma *instrução da condição* para limitar o espaço da pesquisa, como incluir apenas documentos que pertençam a uma determinada categoria ou documentos que contenham determinadas palavras nos seus títulos. Pode especificar uma única instrução da condição por pesquisador. Caso necessite de limitar o espaço de pesquisa para determinadas pastas, utilize o campo **Pastas** para especificar as pastas que pretende excluir em vez de tentar limitar pastas destino através de uma consulta.

 Configure as opções de segurança ao nível do documento. Se a segurança tiver sido activada quando a colecção foi criada, o pesquisador pode associar dados de segurança aos documentos existentes no índice. A pré-filtragem dos documentos desta forma permite às aplicações reforçar os controlos de acesso com base nas listas de controlo de acesso ou em tokens de segurança armazenados no índice.

Para as classes de artigo Documentum e Hummingbird DM,também pode seleccionar uma opção para validar as credenciais do utilizador quando um utilizador submete uma consulta. Nesse caso, em vez de comparar as credenciais do utilizador com os dados de segurança indexados, o sistema compara as credenciais com as listas de controlo de acesso actuais que são actualizadas pela origem de dados original. Ao efectuar a pós-filtragem de documentos desta forma, os documentos não autorizados são removidos dos resultados.

Quando a segurança estiver activada, pesquisador pesquisa os documentos que permitem acesso de leitura, não acesso de procura. Quando um utilizador utiliza a aplicação para consultar uma colecção protegida, apenas o conteúdo que o utilizador tem autoridade para ler é devolvido nos resultados, e não o conteúdo que o utilizador tem autoridade para procurar.

Referências relacionadas:

Configurar suporte para autenticação de início de sessão único (SSO)

Configurar um conector para aceder a repositórios do Content Integrator

O pesquisador Content Integrator utiliza APIs de integração de conteúdo para ligar aos repositórios a serem pesquisados. Ao configurar um conector em IBM Content Integrator, permite ao servidor do pesquisador aceder a origens de conteúdo a serem pesquisadas.

Para configurar um conector de modo a que o pesquisador possa aceder a repositórios do Content Integrator:

- Confirme se as variáveis do ambiente VBR_HOME e JAVA_HOME do ficheiro *iice_install_root*/bin/config.sh (em AIX ou Linux) ou ficheiro *iice_install_root*\bin\config.bat (em Microsoft Windows) especificam os directórios correctos.
- 2. Inicie a consola administrativa Content Integrator e configure um conector para o pesquisador do servidor IBM Content Analytics with Enterprise Search. Consulte o Centro de Informações do IBM Content Integrator para obter mais instruções sobre como criar um conector.

Tarefas relacionadas:

"Pesquisar origens do Documentum no Linux no System z" na página 68

Referências relacionadas:

Li Centro de Informações do Content Integrator

Configurar o servidor do pesquisador para pesquisadores do Content Integrator

Antes de criar um pesquisador Content Integrator, tem de executar um script para configurar o servidor do pesquisador. O script permite que o pesquisador aceda aos repositórios do IBM Content Integrator.

Para configurar o servidor do pesquisador de modo a poder pesquisar os repositórios Content Integrator:

- 1. Inicie sessão como administrador predefinido do IBM Content Analytics with Enterprise Search no servidor do pesquisador. Num servidor do Windows 2008, tem de iniciar sessão como Administrador do Windows.
- Altere para o directório ES_INSTALL_ROOT/bin, execute o script de configuração e responda aos pedidos:

```
AIX ou Linux
cd $ES_INSTALL_ROOT/bin
escrvbr.sh
Windows
```

cd %ES_INSTALL_ROOT%\bin
escrvbr.vbs

3. Reinicie o sistema IBM Content Integrator:

```
esadmin system stopall
esadmin system startall
```

Pesquisar origens do Documentum no Linux no System z

Para pesquisar origens do Documentum, as classes de fundação do Documentum (DFC) têm de estar instaladas no servidor do IBM Content Integrator.

Como o DFC não é suportado na plataforma Linux no System *z*, tem de seguir estes passos para definir a configuração:

- 1. Numa máquina Linux no System *z*, instale o IBM Content Integrator no servidor IBM Content Analytics with Enterprise Search.
- 2. Defina um conector para permitir que o servidor do pesquisador aceda a repositórios do Content Integrator.
- 3. Num máquina Windows separada, instale o DFC (servidor que remete para Documentum) e instale o IBM Content Integrator no modo apenas conectores. Instale ambos os produtos com o mesmo ID de utilizador. Para obter assistência, consulte a documentação para o Documentum e o IBM Content Integrator.
- 4. No servidor do IBM Content Analytics with Enterprise Search, abra a consola de administração do Content Integrator e configure o DocumentConnector para activar a ponte do Remote Method Invocation (RMI) para comunicar com o DFC e o servidor do Content Integrator que instalou no passo anterior. Para obter assistência, consulte o Centro de Informações do Content Integrator para obter mais informações sobre como configurar conectores.
- 5. Abra a consola de administração do IBM Content Analytics with Enterprise Search e configure um pesquisador do Content Integrator para pesquisa as origens do Documentum

Tarefas relacionadas:

"Configurar um conector para aceder a repositórios do Content Integrator" na página 67

Referências relacionadas:

🕩 Centro de Informações do Content Integrator

Pesquisadores do Content Manager

Para incluir os tipos de artigos do IBM Content Manager Enterprise Edition numa colecção, tem de configurar um pesquisador do Content Manager.

Credenciais de ligação do pesquisador

Quando cria o pesquisador, pode especificar as credenciais que permitem que o pesquisador se ligue às fontes a pesquisar. Também pode configurar as credenciais de ligação quando especifica as definições gerais de segurança do sistema. Se utilizar a última abordagem, vários pesquisadores e outros componentes do sistema podem utilizar as mesmas credenciais. Por exemplo, os servidores de procura podem utilizar as credenciais para determinar se um utilizador possui autorização para aceder ao conteúdo.

Configuração do servidor do pesquisador

Antes de poder pesquisar um servidor Content Manager EE, deve executar o script **escrcm** no servidor do pesquisador. Este script, que é fornecido com o IBM Content Analytics with Enterprise Search, permite ao pesquisador do Content Manager comunicar com os servidores do Content Manager EE.

Complete a tarefa adequada ao seu ambiente:

- "Configurar o servidor do pesquisador no UNIX para pesquisadores Content Manager" na página 70.
- "Configurar o servidor do pesquisador no Windows para pesquisadores Content Manager" na página 71.

Descrição geral da configuração

Pode utilizar o pesquisador do Content Manager para pesquisar qualquer número de servidores Content Manager EE. Ao configurar o pesquisador, especifique opções para como o pesquisador deve pesquisar todos os servidores Content Manager EE no espaço do pesquisador. Também pode seleccionar os tipos de itens específicos que pretende pesquisar em cada servidor.

Ao criar o pesquisador, um assistente ajudá-lo-á nessas tarefas:

- Especifique as propriedades que controlam o modo como o pesquisador opera e utiliza os recursos do sistema. As propriedades do pesquisador controlam o modo como este último pesquisa todos os tipos de item de todos os servidores do Content Manager EE no espaço de pesquisa.
- Seleccione os servidores Content Manager EE que pretende pesquisar.
- Especifique os IDs do utilizador e as palavras-passe que permitem o pesquisador aceder ao conteúdo nos servidores Content Manager EE.
- Defina um agendamento para a pesquisa de servidores.
- Seleccione os tipos de itens que pretende pesquisar em cada servidor de Content Manager EE.
- Especifique as opções para fazer com que os atributos possam ser procurados em alguns tipos de itens. Por exemplo, pode excluir certos tipos de documentos do espaço de pesquisa e especificar os atributos que podem regressar nos resultados da procura.
- Configure as opções de segurança ao nível do documento. Se a segurança tiver sido activada quando a colecção foi criada, o pesquisador pode associar dados de segurança aos documentos existentes no índice. Esses dados permitem às aplicações fazer cumprir os controlos de acesso baseados nas listas de controlo de acesso armazenadas ou nos tokens de segurança.

Pode seleccionar uma opção para validar as credenciais do utilizador no momento em que um utilizador permitir uma consulta. Nesse caso, em vez de

comparar as credenciais do utilizador com os dados de segurança indexados, o sistema compara as credenciais com as listas de controlo de acesso actuais que são actualizadas pela origem de dados original. Ao efectuar a pós-filtragem de documentos desta forma, os documentos não autorizados são removidos dos resultados.

Configurar o servidor do pesquisador no UNIX para pesquisadores Content Manager

Se instalar o IBM Content Analytics with Enterprise Search num computador /que esteja a executar o AIX ou o Linux, tem de executar um script para configurar o servidor do pesquisador. O script activa o pesquisador do Content Manager para comunicar com servidores IBM Content Manager Enterprise Edition.

O pesquisador do Content Manager utiliza o conector Java para IBM Content Manager Enterprise Edition para aceder aos servidores Content Manager EE. Este conector é instalado através da instalação de um dos seguintes produtos no servidor do pesquisador. Consulte *Origens de dados suportadas* para informações sobre as versões suportadas destes produtos:

- IBM Information Integrator for Content para AIX ou Linux
- IBM Content Manager Toolkit para Linux

Para assegurar que o pesquisador do Content Manager possa trabalhar com Content Manager EE, pode executar um script de configuração que o IBM Content Analytics with Enterprise Search faculta no servidor do pesquisador depois de instalar o conector.

Para configurar o servidor do pesquisador de modo a poder pesquisar os servidores Content Manager EE:

- 1. Instale o conector Java para IBM Content Manager Enterprise Edition no servidor do pesquisador:
 - a. No servidor do pesquisador, inicie sessão como utilizador root.
 su root
 - su 1000
 - b. Execute o ficheiro db2profile. Por exemplo:
 - . /home/db2inst/sqllib/db2profile
 - c. Exporte a variável do ambiente JAVAHOME. Por exemplo: export JAVAHOME=/usr/IBMJava2-141
 - d. Adicione o directório Java à variável do ambiente PATH: export PATH=\$PATH:\$JAVAHOME/bin
 - e. Insira o CD de instalação do IBM Information Integrator for Content e execute o assistente de instalação.
 - f. Na janela de Selecção de Componentes, execute as seguintes acções:
 - Seleccione Conectores locais a partir da lista Componentes, depois seleccione Conector de Content Manager V8 a partir da lista Subcomponentes.
 - Seleccione Caixas de módulos e amostras do conector a partir da lista Componentes, depois seleccione Conector de Content Manager V8 a partir da lista Subcomponentes.
 - g. Especifique um nome da base de dados, um nome do utilizador e uma palavra-passe na biblioteca Content Manager EE, e aceite as definições predefinidas nas restantes janelas.

- 2. No servidor do pesquisador, inicie sessão com um ID de utilizador que esteja no grupo de administração DB2.
- **3**. Catalogue a base de dados do servidor remoto da biblioteca Content Manager EE, e verifique se o servidor do pesquisador pode estabelecer ligação ao servidor Content Manager EE:

nó tcpip do catálogo db2 *nome_nó* servidor *nome do sistema central* remoto *porta* base de dados do catálogo de db2 *nome_base de dados* como *nome alternativo* no nó *nome_nó*

Em que:

nome nó

Este é o nome abreviado do sistema central do servidor Content Manager EE (como ibmes).

nome do sistema central

É o nome do sistema central totalmente qualificado do servidor Content Manager EE (como ibmes.ibm.com).

porta

É o número da porta do servidor Content Manager EE.

nome_base_de_dados

É o nome da base de dados do Content Manager EE (como ICMNLSDB).

nome_alternativo

É o nome alternativo da base de dados do Content Manager EE (como CMSVR)

 Opcional: inicie sessão como o utilizador root e teste a ligação da base de dados:

. directório_instalação_Information_Integrator_for_Content/bin/cmbenv81.sh cd directório_instalação_Information_Integrator_for_Content/samples/java/icm javac *.java

java SConnectDisconnectICM nome_ICMdatabase ID_CMadmin palavra_passe_CMadmin

- No servidor do pesquisador, execute o script de configuração para o pesquisador do Content Manager:
 - a. Mude para o directório ES_INSTALL_ROOT/bin:

cd \$ES_INSTALL_ROOT/bin

b. Inicie o seguinte script e responda aos pedidos de informação:

escrcm.sh

6. No servidor do pesquisador, pare e reinicie o sistema IBM Content Analytics with Enterprise Search

esadmin system stopall esadmin system startall

Referências relacionadas:

🕩 Origens de dados suportadas

Configurar o servidor do pesquisador no Windows para pesquisadores Content Manager

Se instalar o IBM Content Analytics with Enterprise Search num computador Microsoft Windows, tem de executar um script para configurar o servidor do pesquisador. O script activa o pesquisador do Content Manager para comunicar com servidores IBM Content Manager Enterprise Edition.

O pesquisador do Content Manager utiliza o conector Java para IBM Content Manager Enterprise Edition para aceder aos servidores Content Manager EE. Este conector é instalado através da instalação do IBM Information Integrator for Content para Windows no servidor do pesquisador. Consulte *Origens de dados suportadas* para informações sobre as versões suportadas deste produto.

Para assegurar que o pesquisador do Content Manager possa trabalhar com Content Manager EE, pode executar um script de configuração que o IBM Content Analytics with Enterprise Search faculta no servidor do pesquisador depois de instalar o conector.

Para configurar o servidor do pesquisador de modo a poder pesquisar os servidores Content Manager EE:

- 1. Instale o conector Java para Content Manager EE no servidor do pesquisador:
 - a. Insira o CD de instalação do IBM Information Integrator for Content. O programa de instalação começa automaticamente.

O assistente de instalação do Portal Content Manager EE Enterprise Information abre.

- b. Na janela de Selecção de Componentes, execute as seguintes acções:
 - Seleccione Conectores locais a partir da lista Componentes, depois seleccione Conector de Content Manager V8 a partir da lista Subcomponentes.
 - Seleccione Caixas de módulos e amostras do conector a partir da lista Componentes, depois seleccione Conector de Content Manager V8 a partir da lista Subcomponentes.
- c. Especifique um nome da base de dados, um nome do utilizador e uma palavra-passe na biblioteca Content Manager EE, e aceite as definições predefinidas nas restantes janelas.
- 2. Catalogue a base de dados do servidor remoto da biblioteca Content Manager EE, e verifique se o servidor do pesquisador pode estabelecer ligação ao servidor Content Manager EE. Introduza os seguintes comandos numa linha de comandos no servidor do pesquisador:

nó tcpip do catálogo db2 *nome_nó* servidor *nome do sistema central* remoto *porta* base de dados do catálogo de db2 *nome_base de dados* como *nome alternativo* no nó *nome_nó*

Em que:

nome nó

Este é o nome abreviado do sistema central do servidor Content Manager EE (como ibmes).

nome do sistema central

É o nome do sistema central totalmente qualificado do servidor Content Manager EE (como ibmes.ibm.com).

porta

É o número da porta do servidor Content Manager EE.

nome base de dados

É o nome da base de dados do Content Manager EE (como ICMNLSDB).

nome_alternativo

É o nome alternativo da base de dados do Content Manager EE (como CMSVR)

 No servidor do pesquisador, edite o ficheiro IBMCMR00T/cmgmt/connectors/ cmbicmsrvs.ini e crie uma entrada para a base de dados do servidor de bibliotecas catalogadas. Por exemplo:

CMSERVER=CMSVR ICMSERVERREPTYPE=DB2 ICMSCHEMA=icmadmin ICMSSO=FALSE ICMDBAUTH=SERVER ICMREMOTE=TRUE ICMHOSTNAME=hostname ICMPORT=50000 ICMREMOTEDB=icmnlsdb ICMNODENAME= ICMOSTYPE=NT ICMJDBCDRIVER= ICMJDBCURL=

Alternativamente, é possível executar o utilitário seguinte para adicionar automaticamente a entrada ao ficheiro cmbicmsrvs.ini:

java com.ibm.mm.sdk.util.cmbsrvsicm -a add -s icmnlsdb -sm ICMADMIN

4. No servidor do pesquisador, altere para o directório IBMCMR00T\samples\java\ icm e execute o utilitário seguinte. O comando é apresentado em duas linhas por uma questão de legibilidade. É necessário inserir o comando numa linha:

java com.ibm.mm.sdk.util.cmbenvicm -a add -s CMSVR -u icmconct -p CMConnectionIDpassword -d IBMCMROOT\cmgmt\connectors

Para obter mais informações sobre este utilitário, consulte o tópico sobre o ficheiro conector V8 do Content Manager EE (cmbicmenv.ini) no Centro de Informações de Content Management.

5. Opcional: teste a ligação da base de dados, abrindo uma linha de comandos e introduzindo os seguintes comandos:

cmbenv81.bat

cd *Information_Integrator_for_Content_install_directory*\samples\java\icm javac *.java

java SConnectDisconnectICM ICMdatabaseName icmconct CMConnectionIDpassword

- 6. No servidor de pesquisa, inicie sessão como administrador do IBM Content Analytics with Enterprise Search e execute o script de configuração para o pesquisador do Content Manager. Num servidor com o sistema operativo Windows 2008, tem de iniciar sessão como Administrador do Windows.
 - a. Mude para o directório ES_INSTALL_ROOT/bin:

cd %ES_INSTALL_ROOT%\bin

- b. Inicie o seguinte script e responda aos pedidos de informação: escrcm.vbs
- 7. No servidor do pesquisador, pare e reinicie o sistema IBM Content Analytics with Enterprise Search

esadmin system stopall esadmin system startall

Referências relacionadas:

- 🕒 Origens de dados suportadas
- 🖙 Centro de Informações do Content Management

Pesquisadores de DB2

Pode utilizar um pesquisador de DB2 de modo a incluir base de dados de IBM DB2 numa colecção.

Caso utilize o IBM Federation Server para federar e criar tabelas de nomes alternativos para os seguintes tipos de sistema de base de dados, pode utilizar o pesquisador de DB2 para pesquisar as tabelas através dos nomes alternativos:

- CA-Datacom
- IBM DB2 para z/OS
- DB2 para iSeries
- IBM Informix
- IMS
- Oracle
- Microsoft SQL Server
- Software AG Adabas
- Sybase
- VSAM

Tem de configurar um pesquisador separado para cada servidor da base de dados que pretende pesquisar. Ao configurar o pesquisador, especifique as opções para como o pesquisador deve pesquisar todas as bases de dados no mesmo servidor. Também pode seleccionar as tabelas específicas que pretende pesquisar em cada base de dados.

As tabelas que seleccionar para a pesquisa devem ser tabelas da base de dados, tabelas de nomes alternativos ou vistas. O pesquisador do DB2 não suporta tabelas conjuntas.

Credenciais de ligação do pesquisador

Quando cria o pesquisador, pode especificar as credenciais que permitem que o pesquisador se ligue às fontes a pesquisar. Também pode configurar as credenciais de ligação quando especifica as definições gerais de segurança do sistema. Se utilizar a última abordagem, vários pesquisadores e outros componentes do sistema podem utilizar as mesmas credenciais. Por exemplo, os servidores de procura podem utilizar as credenciais para determinar se um utilizador possui autorização para aceder ao conteúdo.

Configuração do servidor do pesquisador

Antes de poder pesquisar tabelas da base de dados, tem de instalar um DB2 Client no servidor do pesquisador. Em seguida, tem de executar o script **escrdb2** no servidor do pesquisador. Este script, que é fornecido com a IBM Content Analytics with Enterprise Search, permite que o pesquisador de DB2 comunique com servidores de bases de dados.

Complete a seguinte tarefa:

• "Configurar o servidor do pesquisador para DB2 pesquisadores" na página 76.

Publicação de eventos

Se utilizar o IBM Data Event Publisher e se associar bases de dados que pretende pesquisar com os mapas de filas de publicação, o pesquisador do DB2 pode utilizar os mapas para pesquisar as actualizações para a tabelas de bases de dados.

Um mapa de filas de publicação identifica uma fila do WebSphere MQ que recebe mensagens XML quando são publicadas actualizações de uma tabela de base de dados. O pesquisador escuta a fila para informação acerca dos eventos publicados e actualiza o espaço de pesquisa quando as tabelas estiverem actualizadas (a primeira vez em que o pesquisador pesquisa uma tabela, o pesquisador pesquisa todos os documentos).

A publicação de dados permite que se tornem disponíveis documentos novos e alterados para a procura numa base mais rápida do que os documentos que o pesquisador pesquisa de acordo com o agendamento do pesquisador.

Se algumas ou todas as tabelas estiverem configuradas para utilizar a publicação de dados, pode especificar a informação que permite ao pesquisador aceder ao WebSphere MQ e aos mapas de filas de publicação quando o pesquisador for configurado.

Também se deve certificar de que o WebSphere MQ e o Data Event Publisher se encontram configurados no servidor a pesquisar e que o módulo de cliente WebSphere MQ se encontra configurado no servidor do pesquisador. Complete as tarefas seguintes para utilizar a publicação de dados com um pesquisador do DB2:

- "Configurar o WebSphere MQ para pesquisadores de DB2" na página 79.
- "Configurar o Data Event Publisher para pesquisadores do DB2" na página 77.

Descrição geral da configuração

Ao criar o pesquisador, um assistente ajudá-lo-á nessas tarefas:

- Especifique as propriedades que controlam o modo como o pesquisador opera e utiliza os recursos do sistema. As propriedades do pesquisador controlam o modo como este último pesquisa todas as bases de dados num determinado servidor da base de dados.
- Especifique a informação acerca dos tipos de bases de dados que pretende pesquisar.

Se pretender pesquisar bases de dados remotas que não estejam catalogadas no servidor da base de dados local, tem de iniciar o Servidor de Administração DB2 no servidor remoto antes de poder utilizar o pesquisador do DB2 para pesquisar essas bases de dados. Também tem de especificar o nome do sistema central e a porta do servidor da base de dados remota ao configurar o pesquisador.

• Especifique as bases de dados que pretende pesquisar.

Se configurar um pesquisador do DB2 com um cliente do IBM DB2 que não suporte o Assistente de Configuração, as bases de dados não podem ser descobertas pelos processos de descoberta. Tem de especificar o nome das bases de dados a pesquisar quando configurar o pesquisador.

- Especifique os IDs do utilizador e as palavras-passe que permitem ao pesquisador aceder às bases de dados que utilizam controlos de acesso.
- Defina um agendamento para a pesquisa de bases de dados.
- Seleccione as tabelas que pretende pesquisar em cada base de dados.

Aviso: Para optimizar o desempenho dos processos de identificação (e para impedir que o processo de configuração do pesquisador expire), opte por pesquisar todas as tabelas apenas se a base de dados não contiver muitas tabelas ou se as tabelas não contiverem muitas colunas. Se optar por seleccionar algumas tabelas a pesquisar, neste momento, pode editar o espaço de pesquisa posteriormente e adicionar mais tabelas à colecção.

- Seleccione as tabelas que devem ser pesquisadas quando as respectivas actualizações forem publicadas numa fila de publicação de dados, e especifique a informação que permite ao pesquisador aceder à fila de publicação de dados.
- Configure as opções de segurança ao nível do documento. Se a segurança tiver sido activada quando a colecção foi criada, o pesquisador pode associar dados de segurança aos documentos existentes no índice. Esses dados permitem às aplicações fazer cumprir os controlos de acesso baseados nas listas de controlo de acesso armazenadas ou nos tokens de segurança.
- Especifique as opções para fazer com que as colunas possam ser procuradas em alguns tipos de tabelas. Por exemplo, pode activar certas colunas a serem utilizadas em consultas paramétricas, ou especifique as colunas que podem regressar nos resultados da procura.
- Especifique opções para procurar o conteúdo do documento. Cada tabela pode ter apenas uma coluna que pode ser tratada como se contivesse o conteúdo principal do documento. Para ajudar as aplicações a encontrar e recuperar este conteúdo, tal como o conteúdo dos documentos .doc ou .pdf, insira a informação sobre o conteúdo no pesquisador.

Por exemplo, se uma coluna da tabela contiver conteúdo grande binário ou de texto, tal como os documentos com dados do tipo BLOB, pode configurar o pesquisador para permitir que os utilizadores consultem o conteúdo através de consultas de texto gratuitas e visualizem o conteúdo do documento em resumos do documento nos resultados. Pode melhorar a recuperabilidade ao especificar o tipo MIME, tal como aplicação/msword ou aplicação/pdf, e ao especificar as opções para determinar a página de código utilizada para codificar o conteúdo.

Configurar o servidor do pesquisador para DB2 pesquisadores

Antes de criar um pesquisadorDB2, deve executar o script de configuração fornecido, escrdb2, no servidor do pesquisador. O script permite ao pesquisador DB2 comunicar com os servidor da base de dados e tabelas de bases de dados do pesquisador. Se utilizar publicação de eventos, o script também permite que o pesquisador aceda aos gestores de fila e às filas do WebSphere MQ.

Deve instalar o DB2 Data Server Client antes de executar o script de configuração escrdb2.vbs.

Se utilizar publicação de eventos, tem de instalar os módulos MQ WebSphere para Java Messaging no servidor do pesquisador para que o pesquisador DB2 possa aceder aos gestores de filas e às filas doWebSphere MQ. Deve executar o script de configuração escrdb2.vbs após instalar os módulos do WebSphere MQ.

Para configurar o servidor do pesquisador para suportar a pesquisa por parte de pesquisadores de DB2:

 Opcional: Se pretende utilizar publicação de eventos, instale os módulos do WebSphere MQ para Java Messaging no servidor do pesquisador. Veja a documentação doWebSphere MQ para mais instruções.

- Inicie sessão como administrador predefinido do IBM Content Analytics with Enterprise Search no servidor do pesquisador. Num servidor do Windows 2008, tem de iniciar sessão como Administrador do Windows.
- 3. Altere para o directório ES_INSTALL_ROOT/bin, execute o script de configuração e responda aos pedidos:

```
AIX ou Linux
cd $ES_INSTALL_ROOT/bin
escrdb2.sh
```

Windows

cd %ES_INSTALL_ROOT%\bin escrdb2.vbs

4. Reinicie o sistema IBM Content Analytics with Enterprise Search:

esadmin system stopall esadmin system startall

Configurar o Data Event Publisher para pesquisadores do DB2

Antes de configurar um pesquisador de DB2 para utilizar publicação de eventos, certifique-se de que o IBM Data Event Publisher está configurado no servidor a ser pesquisado.

Utilize as seguintes directrizes ao configurar o Data Event Publisher para utilizar com o pesquisador DB2:

- Tanto a coluna alterada como a não alterada nas tabelas origem têm de ser seleccionadas para publicação.
- As linhas eliminadas nas tabelas origem têm de ser seleccionadas para publicação.
- Não é possível partilhar uma fila de publicação de eventos entre as múltiplas bases de dados.
- Uma única base de dados pode ter múltiplos mapas de filas e filas.
- Uma tabela devia ter uma publicação XML associada a um mapa de filas de publicação. (Uma tabela não devia ter mais do que uma publicação XML associada a um único mapa de filas de publicação. Uma tabela pode ter mais do que uma publicação XML se cada uma estiver associada a um mapa de filas de publicação diferente.)

Complete os seguintes passos para configurar um servidor da base de dados de modo a que o pesquisador do DB2 possa aceder às actualizações da tabela que são publicadas numa fila de publicação de eventos. Consulte a documentação do Data Event Publisher para ajudar com estes passos:

- 1. Instale o Data Event Publisher no servidor da base de dados a pesquisar.
- 2. Inicie o Replication Center Launchpad:

AIX ou Linux db2rc

Windows

Faça clique em Iniciar > IBM DB2 Replication Center.

- 3. Criar tabelas de controlo Q Capture:
 - a. Seleccione **Publicação de Eventos** como a vista do launchpad, seleccione **Criar Tabelas de Controlo Q Capture**, e depois faça clique em **Seguinte**.

- b. No campo **Servidor Q Capture**, seleccione o servidor que pretende utilizar como o servidor Q Capture a partir da lista de servidores de bases de dados disponíveis, e faça clique em **OK**.
- **c**. Especifique um ID de utilizador e uma palavra-passe autorizados para a aceder ao servidor Q Capture seleccionado. Altere o esquema Q Capture ou aceite o nome do esquema predefinido, e faça clique em **Seguinte**.
- d. Especifique os nomes do gestor de filas, da liga de administração, e reinicie a fila que especificou quando configurou o WebSphere MQ neste servidor da base de dados, e faça clique em **Seguinte**.
- e. Faça clique em **Terminar**. Depois de ser apresentada uma página com mensagens e scripts SQL, faça clique em **Fechar**.
- f. Para a opção de processamento, seleccione **Executar agora** e faça clique em **OK**. Depois de ser apresentada uma mensagem que indica que os scripts SQL estão terminados, faça clique em **Fechar**.
- 4. Criar uma publicação XML:
 - a. No Replication Center Launchpad, seleccione **Publicação de eventos** como a vista do launchpad, seleccione **Criar uma publicação XML**, e depois faça clique em **Seguinte**.
 - b. Na página Iniciar, faça clique em Seguinte.
 - c. Na página Servidor e Mapa de Filas, confirme se o servidor Q Capture e o esquema Q Capture estão correcto, faça clique na opção ao lado do campo Mapa de filas de publicação e faça clique em Novo para criar um mapa de filas de publicação.
 - d. Na página Geral, escreva um nome do mapa de filas.
 - e. Na página Propriedades, especifique o nome da fila de envio (tal como o nome da fila de dados que especificou quando configurou o MQ neste servidor), seleccione Operação em linha ou Transacção para o tipo de conteúdo da mensagem, limpe as caixas de verificação para enviar mensagens de sinais e adicionar cabeçalhos de mensagens JMS, e faça clique em OK.
 - f. Depois de ser apresentada uma página com mensagens e scripts SQL, faça clique em **Fechar**.
 - g. Para a opção de processamento, seleccione Executar agora e faça clique em OK. Depois de ser apresentada uma mensagem que indica que os scripts SQL estão terminados, faça clique em Fechar.
 - h. Na página Seleccionar Mapa de Filas de Publicação, seleccione o mapa de filas que criou e faça clique em **OK**.
 - i. Na página Servidor e Mapa de Filas, confirme que o nome do mapa de filas está correcto, e faça clique em **Seguinte**.
 - j. Na página Tabela Origem, faça clique em **Adicionar**, faça clique em **Obter Todos**, seleccione uma tabela com que pretende activar a publicação de eventos, faça clique em **OK**, e depois faça clique em **Seguinte**.
 - k. Na página Colunas e Linhas, seleccione as colunas que pretende que o pesquisador do DB2 pesquise (ou todas as colunas) e seleccione as colunas chave. Na página onde selecciona as linhas a pesquisar (ou todas as linhas), seleccione a opção para publicar as eliminações da tabela origem. Após terminar a configuração dessas opções, faça clique em Seguinte.
 - I. Na página Conteúdo da Mensagem, seleccione a opção para incluir tanto as colunas alteradas como as não alteradas nos dados da coluna, e seleccione a opção para apenas os valores de dados novos. Certifique-se de que a caixa de verificação para iniciar as publicações XML é automaticamente seleccionada, e faça clique em Seguinte.

- m. Na página Rever e completar publicações de XML, faça clique em **Seguinte**.
- n. Na página Resumo, faça clique em **Terminar**. Depois de ser apresentada uma página com mensagens e scripts SQL, faça clique em **Fechar**.
- Para a opção de processamento, seleccione Executar agora e faça clique em OK. Depois de ser apresentada uma mensagem que indica que os scripts SQL estão terminados, faça clique em Fechar.
- 5. Iniciar o servidor Q Capture:
 - a. Feche o Replication Center Launchpad e inicie o Replication Center.
 - b. Na árvore de objectos, faça clique em Q Replication > Definições > Servidores Q Capture.
 - c. Faça clique com o botão direito do rato no ícone do servidor Q Capture que configurou e seleccione Activar Base de Dados para Replicação Q.
 - d. Depois de ser apresentada uma mensagem de aviso, faça clique em OK.
 - e. Depois de ser apresentada uma página com mensagens da DB2, faça clique em **Fechar**.
 - f. Na árvore de objectos, faça clique com o botão direito do rato no ícone do servidor Q Capture e seleccione **Iniciar programa Q Capture**.
 - g. Para a opção de processamento, seleccione Executar agora, especifique o nome do sistema, o ID de utilizador e a palavra-passe do utilizador da DB2, o caminho do directório onde os registos são arquivados, e o nome da instância DB2, depois faça clique em OK.
 - h. Depois de ser apresentada uma mensagem que indica que o pedido foi submetido, faça clique em **Fechar**.
 - i. Na árvore de objectos, faça clique com o botão direito do rato no ícone do servidor Q Capture e seleccione **Verificar estado**.

É apresentado o estado do servidor Q Capture. Se ocorrerem erros, uma mensagem de estado informa que o servidor está em baixo. Para rever os registos e para determinar a causa de alguns erros, introduza o seguinte comando numa linha de comandos:

asnqcap Capture_Server=capture server name LOGSTDOUT=y

Configurar o WebSphere MQ para pesquisadores de DB2

Antes de configurar um pesquisador de DB2 para utilizar publicação de eventos, certifique-se de que o IBM WebSphere MQ está configurado no servidor onde o pesquisador irá aguardar.

Certifique-se de que o DB2, o InfoSphere Data Event Publisher e o WebSphere MQ estão instalados no servidor da base de dados de destino.

Se o servidor da base de dados destino for instalado num computador Linux, todos os utilizadores do DB2, WebSphere MQ e IBM Content Analytics with Enterprise Search terão de configurar a seguinte variável de ambiente: exportação LD_ASSUME_KERNEL=2.4.19

Esta variável do ambiente permite que as implementações de adaptação do LinuxThread sejam exportadas de qualquer interface onde a instalação é executada, os comandos de controlo do WebSphere MQ sejam emitidos ou as aplicações do WebSphere MQ sejam executadas. O WebSphere MQ requer que esta variável do ambiente seja exportada. Os pesquisadores do DB2 que utilizam publicação de eventos ligam-se a filas do WebSphere MQ com uma ligação de cliente. Para permitir ligações de cliente, inicie sessão como administrador do WebSphere MQ e execute o seguinte comando para definir o CCSID do gestor de filas como 819:

runmqsc nome_gestor_filas
ALTER QMGR CCSID(819)
END

O pesquisador do DB2 suporta o modo de ligação cliente para o servidor do WebSphere MQ. O pesquisador aguarda por mensagens XML que são publicadas numa fila de publicação de eventos. O pesquisador não pode aguardar por mensagens XML que sejam transportadas em mais do que uma fila.

Depois de configurar o WebSphere MQ, o pesquisador do DB2 utiliza o nome do gestor de filas, o nome da fila, o nome do sistema central, o número de porta do servidor, e o nome do canal do servidor para obter mensagens XML a partir duma fila de publicação. O pesquisador analisa as mensagens e actualiza o espaço de pesquisa com informação acerca de tabelas actualizadas.

Siga estes passos para configurar o servidor da base de dados de modo que o pesquisador do DB2 possa aguardar numa fila de publicação de eventos. Consulte a documentação WebSphere MQ para assistência com estes passos:

- 1. Inicie sessão como função de Administrador do WebSphere MQ e introduza os seguintes comandos para criar um gestor de filas e filas.
 - Numa linha de comandos, introduza o seguinte comando: crtmqm QM1
 - Depois de a mensagem Instalação completa for apresentada, introduza o seguinte comando:

strmqm QM1

c. Depois de a mensagem 'QM1' iniciado ser apresentada, introduza o seguinte comando:

runmqsc QM1

d. Depois de a mensagem Iniciar MQSC para QM1 do gestor de filas ser apresentada, introduza o seguinte comando para criar uma fila de administração:

```
DEFINE QLOCAL('ASN.QM1.ADMINQ')
```

- e. Depois de a mensagem Fila do WebSphere MQ criada ser apresentada, introduza o seguinte comando para criar uma fila de reinício.
 DEFINE QLOCAL (' ASN.QM1.RESTARTQ')
- f. Depois de a mensagem Fila do WebSphere MQ criada ser apresentada de novo, introduza o seguinte comando para criar uma fila de dados: DEFINE QLOCAL(' ASN.QM1.DATAQ')
- g. Depois de a mensagem Fila do WebSphere MQ criada ser apresentada de novo, introduza o seguinte comando para sair: fim
- 2. Introduza o seguinte comando para iniciar o MQ Listener no servidor de base de dados (o MQ Listener tem de estar em execução quando criar um pesquisador do DB2 que utilize publicação de eventos). Neste exemplo, 1414 é o número de porta do servidor e o canal predefinido, SYSTEM.DEF.SVRCONN é utilizado:

runmqlsr -m QM1 -t TCP -p 1414 &

3. Introduza os seguintes comandos para autorizar um utilizador da DB2 a aceder ao gestor de filas e às filas através de Message Queuing Interface (MQI) para publicação de eventos (neste exemplo, o ID de utilizador é db2inst1):

setmqaut -m QM1 -t qmgr -p db2inst1 +allmqi setmqaut -m QM1 -t queue -n ASN.QM1.DATAQ -p db2inst1 +allmqi setmqaut -m QM1 -t queue -n ASN.QM1.ADMINQ -p db2inst1 +allmqi setmqaut -m QM1 -t queue -n ASN.QM1.RESTARTQ -p db2inst1 +allmqi

4. Introduza os seguintes comandos para o ID de utilizador que é utilizado para criar e executar o pesquisador do DB2 com publicação de eventos. Estes comandos autorizam o ID de utilizador a aceder ao gestor de filas e às filas através de Message Queuing Interface (MQI) para publicação de eventos. Neste exemplo, o ID de utilizador é esuser:

```
setmqaut -m ASN.QM1.QM2 -t qmgr -p esuser +allmqi
setmqaut -m ASN.QM1.QM2 -t queue -n ASN.QM1.DATAQ -p esuser +allmqi
```

Pesquisar bases de dados DB2 num servidor de origens de dados clássico

O pesquisador do DB2 pode pesquisar uma base de dados DB2 no servidor de origens de dados clássico através do WebSphere Information Integrator Classic Federation.

Para pesquisar uma base de dados DB2 no servidor de origens de dados clássico, a base de dados tem de ser federada com uma base de dados DB2 no servidor de origens de dados não clássico através da utilização do conector de ODBC do WebSphere Information Integrator Classic Federation.

Para federar a base de dados:

- 1. Instale o WebSphere Information Integrator Classic Federation no servidor de origens de dados clássico.
- Instale o módulo de cliente de WebSphere Information Integrator Classic Federation no servidor de origens de dados não clássico cuja base de dados irá federar-se na base de dados do servidor de origens de dados clássico.
- Configure o controlador de ODBC do WebSphere Information Integrator Classic Federation para estabelecer ligação com o servidor de origens de dados clássico.
- 4. Ligue-se à base de dados de federação e crie o conector de ODBC para se federar com a base de dados do servidor de origens de dados clássico.
- 5. Configure o pesquisador do DB2 para pesquisar a base de dados de federação no servidor de origens de dados não clássico. Esta configuração permite que a base de dados do servidor de origens de dados clássicos seja pesquisada através do WebSphere Information Integrator Classic Federation.

Configurar o nível de isolamento para pesquisar bases de dados DB2

O nível de isolamento que está associado a um processo da aplicação define o grau de isolamento desse processo da aplicação dos outros processos de aplicação em execução simultânea.

Por predefinição, os pesquisadores de DB2 e de Base de dados JDBC pesquisam origens de dados DB2 da IBM com um nível de isolamento que é definido como Leitura Não Consolidada. Neste nível, qualquer linha que uma aplicação leia pode ser alterada por outros processos da aplicação, e qualquer linha que seja alterada por outro processo da aplicação pode ser lida, mesmo que a alteração ainda não tenha sido consolidada pelo processo da aplicação.

Para alterar este comportamento, é possível alterar estes pesquisadores para pesquisar com um nível de isolamento que é definido como Leitura Não Consolidada. Neste nível, qualquer linha que tenha sido alterada por outro processo da aplicação não pode ser lida a não ser que seja consolidada por esse processo da aplicação.

Para reconfigurar o nível de isolamento, é necessário criar uma extensão do ficheiro de configuração do pesquisador. A codificação de caracteres no ficheiro XML tem de ser UTF-8.

Para alterar o nível de isolamento:

- 1. Inicie sessão como administrador predefinido do IBM Content Analytics with Enterprise Search no servidor do pesquisador.
- Aceda ao directório ES_NODE_ROOT/master_config/nome-colecção.nomepesquisador do pesquisador que pretende configurar.
- Crie um ficheiro XML de extensão de configuração denominado db2crawler_ext.xml para um pesquisador do DB2 ou um pesquisador do jdbccrawler_ext.xml for a Base de dados JDBC, e inclua o conteúdo seguinte:

```
<?xml version="1.0" encoding="UTF-8"?>
<ExtendedProperties>
<RemoveChild XPath="/Crawler/DataSources/Server/Target[@Name='target_name']
/OptionalClause"/>
<AppendChild XPath="/Crawler/DataSources/Server/Target[@Name='target_name']"
Name="OptionalClause">WITH CS</AppendChild>
</ExtendedProperties>
```

onde *target_name* identifica o destino de origem dos dados que é especificado no ficheiro db2crawler.xml ou no ficheiro jdbccrawler.xml. No exemplo seguinte, o nome de destino é DB2ADMIN.TEST1:

```
<Crawler><DataSources><Server><Target Name="DB2ADMIN.TEST1"
PrimaryKey="I1" ID="t1224822785487">
```

Sugestão: Se pretender aplicar esta definição a todos os destinos do pesquisador, não especifique um destino. Por exemplo:

```
<?xml version="1.0" encoding="UTF-8"?>
<ExtendedProperties>
<RemoveChild XPath="/Crawler/DataSources/Server/Target/OptionalClause"/>
<AppendChild XPath="/Crawler/DataSources/Server/Target"
Name="OptionalClause">WITH CS</AppendChild>
</ExtendedProperties>
```

4. Reinicie o pesquisador que actualizou.

Pesquisadores do Domino Document Manager

Para incluir as bibliotecas e armários do Domino Document Manager numa colecção, deve configurar um pesquisador Domino Document Manager.

Credenciais de ligação do pesquisador

Quando cria o pesquisador, pode especificar as credenciais que permitem que o pesquisador se ligue às fontes a pesquisar. Também pode configurar as credenciais de ligação quando especifica as definições gerais de segurança do sistema. Se

utilizar a última abordagem, vários pesquisadores e outros componentes do sistema podem utilizar as mesmas credenciais. Por exemplo, os servidores de procura podem utilizar as credenciais para determinar se um utilizador possui autorização para aceder ao conteúdo.

Configuração do servidor do pesquisador

Se um servidor Domino Document Manager que pretende pesquisar utiliza o protocolo de chamada de procedimento remoto do Notes (NRPC), deve executar o script de configuração **escrnote** no servidor do pesquisador. Esse script fornecido com IBM Content Analytics with Enterprise Search, activa o pesquisador do Domino Document Manager para comunicar com os servidores que utilizam o NRPC.

Se um servidor Domino Document Manager que pretende pesquisar utiliza o protocolo Domino Internet Inter-ORB Protocol (DIIOP), não necessita de executar um script de configuração no servidor do pesquisador. Contudo, deve configurar o servidor Domino Document Manager para que o pesquisador Domino Document Manager possa aceder ao servidor.

Importante: Se o servidor Domino Document Manager utilizar DIIOP e configurar o pesquisador para utilizar HTTPS ou DIIOP na vez de SSL para que as transmissões entre o pesquisador e o servidor sejam encriptadas, deve copiar o ficheiro TrustedCerts.class (por exemplo, c:\certs ou /data/certs) do servidor Domino Document Manager para o servidor do pesquisador. Numa configuração de vários servidores, também deve copiar o ficheiro TrustedCerts.class para os servidores de procura. Tem de assegurar que ficheiro se encontra na mesma localização no servidor do pesquisador e nos servidores de procura. O utilizador especifica o caminho do directório para o ficheiro TrustedCerts.class quando configura o pesquisador.

Se IBM Content Analytics with Enterprise Search tiver sido instalado em um sistema da IBM AIX, tem de garantir que o módulo E/S Porta de Conclusão esteja instalado e disponível no servidor do pesquisador.

Antes de utilizar a consola de administração para configurar um pesquisador do Domino Document Manager, execute as tarefas seguintes apropriadas para o ambiente:

- "Configurar o servidor do pesquisador no UNIX para pesquisar origens do Lotus Domino" na página 114.
- "Configurar o servidor do pesquisador no Windows para pesquisar origens do Lotus Domino" na página 116.
- "Configurar servidores que utilizam o protocolo DIIOP" na página 117.
- "Configurar a porta de conclusão E/S no AIX para pesquisar origens no Lotus Domino" na página 118.

Segurança ao nível do documento

Se a segurança de colecções estiver activada e um servidor que pretende pesquisar utilizar o protocolo NRPC, tem de configurar um Servidor Fidedigno Lotus Domino no servidor do pesquisador. O Servidor Fidedigno é utilizado para garantir controlos de acesso ao nível do documento. Antes de disponibilizar uma colecção para os utilizadores, complete as tarefas seguintes:

• Configure os Servidores Fidedignos Lotus Domino para validar credenciais de utilizador.

 Active a segurança do início de sessão da aplicação. Se utilizar o servidor de aplicação da Web incorporado, abra a vista Segurança e configure as definições de início de sessão da aplicação. Se utilizar o WebSphere Application Server, active a segurança global e configure a aplicação para aplicar a segurança de início de sessão.

Activar a segurança de início de sessão da aplicação assegura que as credenciais são validadas quando os utilizadores iniciam sessão em colecções de consulta. Os servidores de procura utilizam as credenciais para verificar se cada utilizador tem autoridade para aceder a documentos Lotus Domino.

Descrição geral da configuração

Pode utilizar o pesquisador Domino Document Manager para pesquisar qualquer número de bibliotecas do Domino Document Manager. Ao criar o pesquisador, selecciona as bibliotecas a pesquisar a partir de um só servidor Domino Document Manager. Posteriormente, quando editar a página de pesquisa, pode adicionar documentos de outro servidor Domino Document Manager que pretenda incluir no mesmo espaço de pesquisa. Ao criar ou editar o pesquisador, pode especificar se pretende pesquisar todos os armários nas bibliotecas que selecciona para a pesquisa, ou se pretende pesquisar armários específicos.

Ao criar o pesquisador, um assistente ajudá-lo-á nessas tarefas:

- Especifique as propriedades que controlam o modo como o pesquisador opera e utiliza os recursos do sistema. As propriedades do pesquisador controlam o modo como este último pesquisa todos os documentos no espaço de pesquisa.
- Identificar o servidor Domino Document Manager e o protocolo de comunicações.
- Se o servidor estiver configurado para utilizar o protocolo DIIOP, pode especificar a forma como o pesquisador se pode ligar aos objectos do Domino. Por exemplo, pode especificar opções para utilizar HTTPS ou Secure Sockets Layer (SSL) para codificar comunicações.
- Seleccione as bibliotecas que pretende pesquisar.
- Defina um agendamento para a pesquisa de bibliotecas.
- Seleccione os documentos que pretende pesquisar. O pesquisador pode pesquisar todos os armários numa biblioteca, ou pesquisa apenas os documentos que estão nos armários que selecciona.
- Especifique as opções para que os campos em várias bibliotecas e em vários armários possam ser pesquisados. Por exemplo, pode excluir determinados campos do espaço de pesquisa e especifique as opções para pesquisar ligações.
- Configure as opções de segurança ao nível do documento. Se a segurança tiver sido activada quando a colecção foi criada, o pesquisador pode associar dados de segurança aos documentos existentes no índice. Esses dados permitem às aplicações fazer cumprir os controlos de acesso baseados nas listas de controlo de acesso armazenadas ou nos tokens de segurança.

Pode seleccionar uma opção para validar as credenciais do utilizador no momento em que um utilizador permitir uma consulta. Nesse caso, em vez de comparar as credenciais do utilizador com os dados de segurança indexados, o sistema compara as credenciais com as listas de controlo de acesso actuais que são actualizadas pela origem de dados original.

Referências relacionadas:

Configurar suporte para autenticação de início de sessão único (SSO)

Pesquisadores do Exchange Server

Para recolher conteúdo de pastas públicas e caixas de correio de utilizadores geridas pelo Microsoft Exchange Server 2007, Service Pack 1 ou posterior, ou Exchange Server 2010, configure um pesquisador do Exchange Server.

Credenciais de ligação do pesquisador

Quando cria o pesquisador, pode especificar as credenciais que permitem que o pesquisador se ligue às fontes a pesquisar. Também pode configurar as credenciais de ligação quando especifica as definições gerais de segurança do sistema. Se utilizar a última abordagem, vários pesquisadores e outros componentes do sistema podem utilizar as mesmas credenciais. Por exemplo, os servidores de procura podem utilizar as credenciais para determinar se um utilizador possui autorização para aceder ao conteúdo.

configuração do pesquisador

Antes de configurar um pesquisador Exchange Server, deve configurar o Exchange Web Service (EWS) no servidor Exchange Server para permitir que o pesquisador aceda ao conteúdo.

Quando configurar um pesquisador Exchange Server na consola de administração:

- Especifique as propriedades que controlam o modo como o pesquisador opera e utiliza os recursos do sistema. As propriedades do pesquisador controlam a forma como o pesquisador recolhe conteúdo de todos os servidores no espaço de pesquisa.
- Especifique a informação sobre o servidor Exchange Server que pretende pesquisar.

Tem de especificar um ID do utilizador e uma palavra-passe de modo a que o pesquisador possa aceder ao conteúdo no servidor. A ID do utilizador pode estar no formato nome principal do utilizador (UPN) ou no formato domínio, tal como Domain\ExampleAccountName.

- Seleccione as pastas públicas ou as pastas pessoais a pesquisar. O pesquisador não pode pesquisar ambos os tipos de pastas na mesma sessão de pesquisa. Para incluir pastas públicas e pastas pessoais numa colecção, crie pesquisadores diferentes.
- Especifique as opções para tornar os documentos pesquisáveis. Por exemplo, pode excluir certos tipos de documentos do espaço de pesquisa.
- Definir um agendamento para pesquisar o servidor Exchange Server.

Pastas públicas

O pesquisador Exchange Server pode pesquisar qualquer número de pastas e subpastas nos servidores de pastas públicas do Exchange Server. Quando criar um pesquisador, seleccione o conteúdo que pretende recolher de um servidor de pasta pública. Posteriormente, pode editar o espaço do pesquisador para adicionar conteúdo de outro servidor.

Caixas de correio do utilizador

O pesquisador Exchange Server pode pesquisar qualquer número de pastas pessoais e itens nas caixas de correio de utilizador do Exchange Server. Quando cria um pesquisador, selecciona o conteúdo que pretende recolher de um servidor de caixa de correio. Posteriormente, pode editar o espaço do pesquisador para adicionar conteúdo de outro servidor. O pesquisador só pode recolher conteúdo de caixas de correio de utilizador, mas não outro tipo de caixas de correio do servidor de caixas de correio.

Se planeia recolher conteúdo de um servidor de caixas de correio, deve implementar um serviço da Web fornecido, ESExchangeServices, no servidor de caixas de correio para que o pesquisador possa aceder a pastas, itens e permissões de utilizador.

O pesquisador suporta os seguintes tipos de caixas de correio de utilizador e tipos de itens. Outros tipos de pastas, tais como feed de notícias e feed RSS, não são suportados. As pastas criadas pelo utilizador incluem todas as pastas de nível superior que um utilizador cria.

Tabela 9. Pasta de caixa de correio de utilizador e tipos de item suportados

pastas pessoais	tipos de itens
caixa de entrada	Mensagem (Correio)
Rascunho	Calendário
Itens enviados	Tarefa
Caixa de saída	Contacto
Calendário	NotasPostItem
Contactos	Mensagem_Reunião
Tarefas	Pedido_Reunião
Notas	Resposta_Reunião
Pastas criadas pelo utilizador	Cancelar_Reunião

Filtros de caixas de correio

Quando seleccionar o conteúdo da caixa de correio a pesquisar, pode especificar os filtros para incluir e excluir documentos. Por exemplo, pode seleccionar tipos de pastas pessoais específicos que pretende incluir na colecção. Também pode filtrar conteúdo de acordo com:

- · Controladores de domínio específicos que organizam o conteúdo no servidor
- Servidores de caixa de correio específicos
- Servidores de base de dados específicos
- Unidades organizacionais específicas (OU), que incluem todas as caixas de correio de utilizadores que pertencem ao OU
- Utilizadores específicos

Se especificar mais do que uma condição para um tipo de filtro específico, tal como dois servidores de caixa de correio, a lista de caixas de correio disponíveis para serem pesquisadas inclui caixas de correio que correspondem a cada condição. Se especificar mais do que um tipo de filtro, tal como um servidor de caixa de correio e um servidor de base de dados, a lista de caixas de correio disponíveis para pesquisa incluem as caixas de correio que correspondem a ambas as condições. Após criar o pesquisador, não pode alterar os filtros. Para pesquisar pastas pessoais diferentes ou um servidor de caixa de correio diferente, por exemplo, deve criar um pesquisador separado.

Segurança

Para obter dados de segurança para procurar caixas de correio de utilizador, deve implementar um serviço da Web fornecido, ESCommonServices, no servidor de caixas de correio do Exchange Server. Este serviço permite que o IBM Content Analytics with Enterprise Search obtenha listas de grupos e permissões necessárias para pré-filtragem dos controlos de acesso

Conceitos relacionados:

"Metadados do Exchange Server" na página 92

"Proteger a procura do conteúdo do Exchange Server" na página 93

Tarefas relacionadas:

"Configurar credenciais de ligação" na página 380

"Definir o Exchange Web Service para os pesquisadores do Exchange Server"

"Implementar o serviço da Web ESExchangeServices no servidor Exchange Server" na página 88

"Implementar o serviço Web ESCommonServices no servidor Exchange Server" na página 90

Referências relacionadas:

"Formatos de URI no índice" na página 164

"Nomes de campos de índice reservados" na página 199

Definir o Exchange Web Service para os pesquisadores do Exchange Server

Para obter a lista de pastas, itens e elementos dos itens a pesquisar, o pesquisador do Exchange Server utiliza o protocolo SOAP do Exchange Web Service (EWS). Quando o Exchange Server Client Access Server é instalado num servidor Microsoft Exchange Server 2007 ou 2010, o componente do EWS é instalado automaticamente.

Para aceder às caixas de correio dos utilizadores, o pesquisador do Exchange Server deve ser capaz de personificar outros utilizadores. Para configurar este suporte, deve preparar um utilizador que possua direitos de personificação e utilizar esta conta quando configurar o pesquisador do Exchange Server.

Definir o EWS para o pesquisador do Exchange Server:

- Verifique se o EWS está instalado no servidor Exchange Server acedendo ao sítio da Web do EWS e confirmando que o WSDL do EWS é apresentado. O URL predefinido é https://hostname/EWS/ExchangeService.asmx, em que hostname é o nome do sistema central do servidor Exchange Server.
- 2. Definir o EWS:
 - Num servidor Exchange Server 2007, atribua dois tipos de permissões, ms-Exch-EPI-Impersonation e ms-Exch-EPI-May-Impersonate. Para mais informações, consulte *Configurar Exchange Impersonation (Exchange Web Services)* no sítio Microsoft MSDN.
 - Num servidor Exchange Server 2010:

- a. Abra o Outlook Web App (OWA) com uma conta de administrador, faça clique em Opções-> Ver todas as opções-> Gerir a sua organização-> Papéis & Auditoria-> Papéis do administrador.
- b. Faça clique em Novo, defina o valor do campo Nome ao utilizar a Gestão de personificação e seleccione AplicaçãoPersonificação como o valor do campo Papéis. No campo Membros, seleccione a conta de utilizador que o pesquisador do Exchange Server deve utilizar ao aceder ao servidor Exchange Server.

Conceitos relacionados:

"Pesquisadores do Exchange Server" na página 85

Implementar o serviço da Web ESExchangeServices no servidor Exchange Server

Para recolher pastas, itens e permissões de utilizador para caixas de correio de utilizadores, deve implementar um serviço da Web fornecido com o IBM Content Analytics with Enterprise Search no servidor de caixas de correio do Microsoft Exchange Server 2007 ou 2010 antes de configurar um pesquisador do Exchange Server.

Para implementar o serviço da Web do ESExchangeServices no servidor de caixa de correio do Exchange Server:

- 1. Copie o ficheiro ES_INSTALL_ROOT/bin/ESExchangeServices.zip para o servidor Exchange Server e extraia todos os ficheiros do ficheiro de arquivos.
- 2. Coloque os ficheiros extraídos no directório C:\inetpub\wwwroot\. Este caminho é o sítio da Web IIS predefinido.
- Apenas Exchange Server 2010: Edite o ficheiro Web.config. Localize a seguinte secção e altere o valor do bindingRedirect newVersion atribuindo para 14.0.0.0: <dependentAssembly>

```
<assemblyIdentity name="Microsoft.Exchange.Management"
    publicKeyToken="31bf3856ad364e35" culture="neutral" />
<bindingRedirect oldVersion="8.0.0.0" newVersion="14.0.0.0"/>
</dependentAssembly>
```

4. Abra o gestor IIS e crie um conjunto de aplicações:

Opção	Descrição
Exchange Server 2007	 Faça clique com o botão direito do rato em Conjunto de aplicações e faça clique em Novo ->Conjunto de aplicações.
	 Defina o valor da ID do conjunto da aplicação para ESServicePool.
	 Seleccione a opção para utilizar o conjunto de aplicações existente como modelo e seleccione DefaultAppPool como nome do conjunto de aplicações.

Opção	Descrição	
Exchange Server 2010	 Faça clique em Conjunto de aplicaçõesno painel esquerdo e faça clique em Adicionar conjunto de aplicações no painel direito. 	
	 Defina o valor do nome para ESServicePool, seleccione v2.0 como a versão .NET framework e seleccione Classic como modelo de canal. 	
	 Faça clique em definições avançadas e defina o valor da identidade do modelo de processamento para NetworkService. 	

5. Abra a árvore do sítio do servidor para o sítio da web predefinido. A pasta ESExchangeServices não está concluída na árvore. Converta esta pasta numa aplicação:

Opção	Descrição
Exchange Server 2007	 Faça clique com o botão direito do rato na pasta e seleccione Propriedades.
	 Faça clique em Criar e seleccione ESServicePool como conjunto de aplicações.
	 Faça clique em Configuração. Se não existir correlação .asmx, faça clique em Adicionar e defina o valor como indicado aqui, altere o número de versão para o seu ambiente:
	Executable: C:\windows\microsoft.net \framework64\v.2.0.50727 \aspnet_isapi.dll Extension: .asmx Verbs: All verbs
	 Abra o separador ASP.NET, seleccione versão ASP.NET e faça clique em Aplicar.
Exchange Server 2010	 Faça clique com o botão direito do rato na pasta e seleccione Converter para Aplicação.
	 Faça novamente clique com o botão direito do rato na pasta, seleccione Gerir Aplicação -> Definições Avançadas e seleccione ESServicePool como conjunto da aplicação.

6. Confirme as definições da aplicação ESExchangeServices. Por predefinição, a definição SSL está activada e a autenticação do Basic e do Windows está activada.

Орção	Descrição		
Exchange Server 2007	 Faça clique com o botão direito do rato na pasta, seleccione Propriedades e abra o separador de segurança do directório. 		
	2. Faça clique em Editar em Comunicações seguras e seleccione Requerer canal seguro.		
	 Faça clique em Editar no controlo de autenticação e acesso e limpe a caixa de verificação activar acesso anónimo. Para um controlo autenticado, seleccione autenticação integrada do Windows e as caixas de verificação de autenticação básica. 		
Exchange Server 2010	 Na vista de funcionalidades, faça clique em Definições SSL e seleccione Requer SSL. 		
	 Faça clique em autenticação e active a autenticação básica e do Windows. Confirme que a autenticação anónima se encontra desactivada. 		

7. Confirme que a implementação foi bem sucedida ao aceder ao servidor através de https://Exchange Server_hostname/ESExchangeServices/ ESExchangeServices.asmx. Após fazer a autenticação, a página de serviços predefinidos deve apresentar as operações suportadas, tal como GetGroup, GetMailbox, GetMailboxPermission, GenUserDomain, e WhoAmI.

Conceitos relacionados:

"Pesquisadores do Exchange Server" na página 85

Implementar o serviço Web ESCommonServices no servidor Exchange Server

Para obter dados de segurança para procurar caixas de correio, deve implementar um serviço web fornecido, ESCommonServices, no servidor de caixa de correio do Microsoft Exchange Server 2007 ou 2010. Este serviço permite que o IBM Content Analytics with Enterprise Search obtenha as listas de grupo e as permissões necessárias para os controlos de acesso pré-filtragem.

Para implementar o serviço web ESCommonServices no servidor de caixa de correio Exchange Server:

- 1. Copie o ficheiro ES_INSTALL_ROOT/bin/ESCommonServices.zip para o servidor Exchange Server e extraia todos os ficheiros do ficheiro de arquivo.
- 2. Coloque os ficheiros extraídos no directório C:\inetpub\wwwroot\. Este caminho é o sítio da Web IIS predefinido.
- Abra a árvore do sítio do servidor para o sítio da Web predefinido. A pasta ESCommonServices está incluída na árvore. Converta esta pasta numa aplicação:

Орção	Descrição		
Exchange Server 2007	 Faça clique com o botão direito do rato na pasta e seleccione Propriedades. 		
	 Faça clique em Criar e seleccione ESServicePool como conjunto de aplicações. 		
	 Faça clique em Configuração. Se não existir correlação .asmx, faça clique em Adicionar e defina o valor como indicado aqui, altere o número da versão do seu ambiente: 		
	Executable: C:\windows\microsoft.net \framework64\v.2.0.50727 \aspnet_isapi.dll Extension: .asmx Verbs: All verbs		
	 Abra o separador ASP.NET, seleccione versão ASP.NET e faça clique em Aplicar. 		
Exchange Server 2010	 Faça clique com o botão direito do rato na pasta e seleccione Converter para Aplicação. 		
	 Faça novamente clique com o botão direito do rato na pasta, seleccione Gerir Aplicação -> Definições Avançadas e seleccione ESServicePool como conjunto da aplicação. 		

4. Confirme as definições da aplicação ESCommonServices. Por predefinição, a definição SSL está activada e a autenticação do Basic e do Windows está activada.

Opção	Descrição		
Exchange Server 2007	 Faça clique com o botão direito do rato na pasta, seleccione Propriedades -> AS .NET -> Editar Configurar -> Aplicação -> definições de identidade, e active a Personificação local. Não necessita de especificar o nome de utilizador ou a palavra-passe. 		
Exchange Server 2010	 Na vista de funcionalidades, faça clique em Definições SSL e seleccione Requer SSL. 		
	2. Faça clique em Autenticação e active a Personificação ASP .Net.		

5. Confirme que a implementação foi bem sucedida ao aceder ao servidor https://*Exchange Server_hostname*/ESCommonServices/ ESCommonServices.asmx.

Conceitos relacionados:

"Pesquisadores do Exchange Server" na página 85

Metadados do Exchange Server

Quando editar um espaço do pesquisador do Exchange Server, pode correlacionar campos de metadados de conteúdo do pesquisador para os campos de índice do IBM Content Analytics with Enterprise Search e, em seguida, especifique como pretender disponibilizar o conteúdo desses campos para a suas aplicações.

Os seguintes campos de metadados comuns são automaticamente correlacionados com os campos de índice. Pode alterar essas correlações ou mantê-las como estão:

•		
Campo de metadados	Campo de índice	
anexo_nome	título	
item_assunto título		
dir	directório	
TipoItem	tipoNome	
tamanho	tamanhoFicheiro	

Tabela 10. Correlações de campo de metadados predefinidos do Exchange Server

Para ver todos os metadados do conteúdo pesquisado, faça clique no botão para configurar os metadados avançados. Os metadados são categorizados em oito separadores para tornar mais fácil encontrar e alterar campos de metadados específicos.

Os nomes de metadados são ajustados para reflectir objectos do Exchange Server. O prefixo e cada campo de metadados apresenta o elementos XML do objecto. Para informações sobre os elementos do objecto, consulte *Exchange Web Services XML Elements* no sítio Microsoft MSDN.

Os seguintes metadados não são criados pelos elementos do objecto do Exchange Server, mas são criados pelo pesquisador do Exchange Server:

- uri: O documento URI
- dir: A localização do item no servidor Exchange Server
- owa: O ID exclusivo do Outlook Web App (OWA)
- tamanho: O tamanho do documento
- Nome_caixa_correio: A caixa de correio em que o item é localizado
- Tipoitem: O tipo de objecto do Exchange Server, tal como a mensagem, o calendário, as tarefas ou o contacto
- TipoItemOriginal: Quando o item é um suplemento, o tipo do suplemento

Os seguintes campos de metadados não podem ser pesquisados:

DireitosEfectivos PropriedadeExpandida ObjectoResposta InícioZonaTemporal e FimZonaTemporal (Objecto do calendário)

Devido aos requerimentos do Exchange Web Service (EWS), o pesquisador não pode obter os valores do últimoTempoAlterado e o tamanho do SuplementoBinário

nos documentos do Exchange Server 2007. Neste caso, o pesquisador retorna valores alternativos, com a data pesquisada como a última data alterada e 0 como o tamanho do ficheiro.

Conceitos relacionados:

"Pesquisadores do Exchange Server" na página 85

"Proteger a procura do conteúdo do Exchange Server"

Proteger a procura do conteúdo do Exchange Server

Se a segurança estiver activada para uma colecção, o pesquisador Exchange Server pode obter as listas de controlo de acesso (ACLs) para os itens que pesquisa e associar dados de segurança a documentos no índice. Estes dados permitem que as aplicações imponham controlos de acesso baseados nos ACLs armazenados ou nos tokens de segurança.

Também pode configurar o pesquisador para validar as credenciais de utilizador quando um utilizador submete uma consulta. Neste caso, em vez de comparar as credenciais de utilizador a dados de segurança no índice, pré-filtragem, o sistema compara as credenciais às actuais listas de controlo de acesso mantidas pela fonte de dados original, pós-filtragem.

O pesquisador Exchange Server suporta autenticação BASIC, autenticação Digest e autenticação NT LAN Manager (NTLM) através dos Serviços de Informação da Internet (IIS) versão 6.x e versão 7.x.

Permissões suportadas

Para obter dados de segurança para procurar caixas de correio de utilizador, deve implementar um serviço da Web fornecido, ESCommonServices, no servidor de Caixas de correio do Exchange Server. Este serviço permite que o IBM Content Analytics with Enterprise Search obtenha as listas de grupo e as permissões necessárias para os controlos de acesso pré-filtragem.Os tipos de permissão suportados são:

- MailboxPermission
- MailboxFolderPermission
- Delegação

Os utilizadores que possuem permissão para aceder a caixas de correio e pastas que pertencem a outros utilizadores (autorização FullAccess como autoridade MailboxPermission ou ReadItem=FullDetails como MailboxFolderPermission), pode procurar itens nessas caixas de correio e pastas. Quando as permissões são atribuídas a um grupo de segurança, os utilizadores que pertencem a esses grupos também podem procurar itens nessas caixas de correio e pastas.

Recusar ACL

Os utilizadores do Exchange Server podem definir a ACL para cada pasta como MailboxFolderPermission. Quando um utilizador atribui MailboxFolderPermission a pastas que correspondem a uma das seguintes condições, o utilizador não pode ver os itens nessas pastas através do Outlook:

- A permissão de acesso de leitura para Predefinido e grupos a que o utilizador pertence é definido para FullDetails.
- A permissão de acesso de leitura do utilizador é definida para Nenhum.

No entanto, o IBM Content Analytics with Enterprise Search não pode recusar os utilizadores que procuram os itens como parte do processo de pré-filtragem. Para assegurar que os utilizadores só vêem os itens que vêem através do Outlook, configure as opções de pós-filtragem de segurança quando configura o pesquisador para assegurar que os actuais controlos de acesso do utilizador são válidos.

Esta questão não afecta os ACLs em que a permissão é definida para MailboxPermission.

Itens privados

Os utilizadores do Exchange Server podem definir uma etiqueta privada para itens para prevenir que os outros utilizadores vejam o item. Quando um utilizador possui acesso às caixas de correio ou pastas de outros utilizadores, através de MailboxPermission ou delegação com direitos de acesso privados, o utilizador pode ver itens privados nessas caixas de correio e pastas através do Outlook. Quando um utilizador possui MailboxFolderPermission ou delegação sem direitos de acesso privados, o utilizador não pode ver os itens privados nas caixas de correio e pastas dos outros utilizadores.

O IBM Content Analytics with Enterprise Search pode portar-se da mesma forma que o Outlook e permitir que os utilizadores que têm acesso às caixas de correio ou pastas de outros utilizadores, através de MailboxPermission ou delegação com direitos de acesso privados, para procurar itens privados nessas caixas de correio e pastas. No entanto, quando a delegação com direitos de acesso privados é atribuída a um grupo, os utilizadores que pertencem a esse grupo não podem procurar itens privados porque o pesquisador não pode obter o nome de um grupo que tem definições de delegação. Para obter a lista de utilizadores que pertencem a um grupo, o pesquisador deve conhecer o nome do grupo.

Através dos Exchange Web Services (EWS) 2010, os utilizadores que possuem qualquer permissão para procurar conteúdo podem ver itens privados. Quando o pesquisador Exchange Server utiliza EWS como um processo de pós-filtragem, a pós-filtragem não funciona da mesma forma que no Outlook. Para proteger a procura, deve activar as opções de segurança pré-filtragem quando configurar o pesquisador.

A tabela seguinte resume as diferenças entre as permissões do Exchange Server para aceder a itens privados e proteger a procura do IBM Content Analytics with Enterprise Search.

	Exchange Server: Outlook	Exchange Server: EWS	Procura segura: permissão atribuída a um utilizador	Procura segura: permissão atribuída a um grupo
MailboxPermission	Sim	Sim	Sim	Sim
MailboxFolderPermission	Não	Sim	Não	Não
Delegação com direitos de acesso privados	Sim	Sim	Sim	Não
Delegação sem direitos de acesso privados	Não	Sim	Não	Não

Tabela 11. Permissões para aceder a itens privados

Conceitos relacionados:

"Pesquisadores do Exchange Server" na página 85 "Metadados do Exchange Server" na página 92

Pesquisadores Exchange Server 2000 e 2003

Pode configurar um pesquisador Exchange Server 2000 e 2003 se pretender incluir pastas públicas de uma dessas versões do Microsoft Exchange Server numa colecção. Para recolher conteúdo dos servidores Microsoft Exchange Server 2007 ou 2010, incluindo pastas públicas ou caixas de correio de utilizadores, configure um pesquisador Exchange Server.

Credenciais de ligação do pesquisador

Quando cria o pesquisador, pode especificar as credenciais que permitem que o pesquisador se ligue às fontes a pesquisar. Também pode configurar as credenciais de ligação quando especifica as definições gerais de segurança do sistema. Se utilizar a última abordagem, vários pesquisadores e outros componentes do sistema podem utilizar as mesmas credenciais. Por exemplo, os servidores de procura podem utilizar as credenciais para determinar se um utilizador possui autorização para aceder ao conteúdo.

Pode utilizar o pesquisador Exchange Server 2000 e 2003 para pesquisar qualquer número de pastas e de subpastas em servidores de pastas públicas Exchange Server. Ao criar um pesquisador, seleccione o conteúdo que pretende pesquisar num servidor de pasta pública. Posteriormente, pode editar o espaço de pesquisa para adicionar o conteúdo a partir de outro servidor da pasta pública.

Ao criar o pesquisador, um assistente ajudá-lo-á nessas tarefas:

- Especifique as propriedades que controlam o modo como o pesquisador opera e utiliza os recursos do sistema. As propriedades do pesquisador controlam o modo como este último pesquisa todas as subpastas de todos os servidores no espaço de pesquisa.
- Especifique a informação acerca do servidor da pasta pública do Exchange Server que pretende pesquisar.

Tem de especificar um ID do utilizador e uma palavra-passe de modo a que o pesquisador possa aceder ao conteúdo no servidor. Se o servidor utilizar o protocolo Secure Sockets Layer (SSL), pode especificar as opções que permitem ao pesquisador aceder ao ficheiro de arquivo de chaves no servidor do pesquisador.

- Defina um agendamento para a pesquisa do servidor de pasta pública.
- Seleccione as pastas e as subpastas para pesquisar.
- Especifique as opções para fazer com que os documentos possam ser procurados em algumas subpastas. Por exemplo, pode excluir certos tipos de documentos do espaço de pesquisa.
- Configure as opções de segurança ao nível do documento. Se a segurança tiver sido activada quando a colecção foi criada, o pesquisador pode associar dados de segurança aos documentos existentes no índice. Esses dados permitem às aplicações fazer cumprir os controlos de acesso baseados nas listas de controlo de acesso armazenadas ou nos tokens de segurança.

Tarefas relacionadas:

"Configurar credenciais de ligação" na página 380

"Verificar o acesso aos documentos Exchange Server protegidos"

Referências relacionadas:

"Formatos de URI no índice" na página 164

"Nomes de campos de índice reservados" na página 199

Verificar o acesso aos documentos Exchange Server protegidos

Para usar um pesquisador Exchange Server 2000 e 2003 para pesquisar documentos protegidos por uma firewall, terá de verificar se o pesquisados é capaz de aceder ao servidor de pastas públicas Microsoft Exchange Server.

Se o servidor do pesquisador não for capaz de aceder a um servidor Exchange Server 2000 ou 2003 seguro, receberá uma mensagem Código 501 de HTTP (Não implementado) do servidor. Também pode ver as mensagens que indicam que foi recebida uma resposta HTTP inesperada.

Para assegurar que o servidor do pesquisador possa aceder aos documentos atrás da firewall:

- 1. Inicie um navegador da Web no servidor do pesquisador.
- 2. Avance para o URL no servidor de pastas públicas Exchange Server que pretende pesquisar. Por exemplo: http://exchange.Empresa.com/public/
- 3. Verifique se pode abrir a página Exchange Server.

Se não conseguir aceder ao servidor do Exchange Server, contacte o administrador do Exchange Server da sua empresa.

Conceitos relacionados:

"Pesquisadores Exchange Server 2000 e 2003" na página 95

Referências relacionadas:

"Requisitos de configuração do pesquisador para suporte de segurança" na página 397

Pesquisadores do FileNet P8

Pode configurar um pesquisador do FileNet P8 para incluir pastas e classes de documentos do IBM FileNet P8 numa colecção.

Credenciais de ligação do pesquisador

Quando cria o pesquisador, pode especificar as credenciais que permitem que o pesquisador se ligue às fontes a pesquisar. Também pode configurar as credenciais de ligação quando especifica as definições gerais de segurança do sistema. Se utilizar a última abordagem, vários pesquisadores e outros componentes do sistema podem utilizar as mesmas credenciais. Por exemplo, os servidores de procura podem utilizar as credenciais para determinar se um utilizador possui autorização para aceder ao conteúdo.

Configuração do servidor do pesquisador

Para integrar o IBM FileNet P8 com o IBM Content Analytics with Enterprise Search, tem de instalar o IBM FileNet Content Engine Client no servidor do
pesquisador. Antes de utilizar a consola de administração para configurar um pesquisador do FileNet P8, conclua a tarefa seguinte:

 "Configurar o servidor de pesquisadores para suportar pesquisadores de Case Manager e FileNet P8" na página 65.

Descrição geral da configuração

Quando configurar o pesquisador, especifique opções para pesquisar documentos em pastas e classes de documentos. Para criar ou alterar um pesquisador do FileNet P8, inicie sessão na consola de administração.

Ao criar o pesquisador, um assistente ajudá-lo-á nessas tarefas:

- Especifique as propriedades que controlam o modo como o pesquisador opera e utiliza os recursos do sistema. As propriedades do pesquisador controlam a forma como o pesquisador pesquisa todos os documentos que adicionar ao espaço de pesquisa.
- Especifique informações que permitem ao pesquisador aceder ao servidor do IBM FileNet P8. Ao configurar a ligação, especifique o URL de Content Engine Web Service e um nome de utilizador e palavra-passe válidos. Também pode especificar a localização de um ficheiro de arquivo de chaves SSL no sistema de ficheiros local do servidor de pesquisador. Este valor é definido como a propriedade javax.net.ssl.trustStore para a ligação SSL ao servidor de Content Engine Web Service e é utilizado para autenticar um servidor utilizando entidade certificadora fidedigna (CA).
- Seleccione as pastas ou classes de documentos que pretende incluir no espaço de pesquisa. Ao seleccionar os destinos a pesquisar, o nível superior da árvore hierárquica mostra os armazenamentos de objectos no servidor. No segundo nível, são apresentadas a Pasta Raiz e a Classe Raiz. Caso seleccione a Pasta Raiz, é apresentada uma lista de pastas raiz no armazenamento de objectos no terceiro nível. Caso seleccione uma Classe Raiz, é apresentada uma lista de classes de documentos raiz no armazenamento de objectos no terceiro nível. Após expandir as pastas raiz e as classes raiz, seleccione uma ou mais pastas ou classes de documento para pesquisar.
- Especifique opções para incluir e excluir documentos específicos e opções para a forma como o conteúdo pode ser pesquisado.
- Seleccione os campos de metadados que pretende incluir e correlacione-os a campos de índice:
 - Ao pesquisar classes de documentos, o pesquisador detecta propriedades das classes. Pode preencher estas classes como campos de metadados de pesquisador especificando propriedades de campo.
 - Ao pesquisar pastas, o pesquisador adiciona automaticamente quatro propriedades comuns como campos de metadados: content_size, crawled_date, docname, filename e lastmodified_date. Para adicionar outras propriedades enquanto metadados do pesquisador, edite o espaço de pesquisa e especifique os nomes da propriedade enquanto campos de metadados do pesquisador.
- Configure um agendamento para pesquisar pastas e classes de documentos individuais.
- Configure as opções de segurança ao nível do documento. Se a segurança tiver sido activada quando a colecção foi criada, o pesquisador pode associar dados de segurança aos documentos existentes no índice. Esses dados permitem às aplicações fazer cumprir os controlos de acesso baseados nas listas de controlo de acesso ou tokens de segurança.

Também pode seleccionar uma opção para validar credenciais de utilizador quando um utilizador submeter uma consulta. Nesse caso, em vez de comparar as credenciais do utilizador com os dados de segurança indexados, o sistema compara as credenciais com as listas de controlo de acesso actuais que são actualizadas pela origem de dados original.

Tarefas relacionadas:

"Configurar credenciais de ligação" na página 380

Referências relacionadas:

"Formatos de URI no índice" na página 164

"Nomes de campos de índice reservados" na página 199

Pesquisadores de Base de dados JDBC

Utilize o pesquisador de Base de dados JDBC para incluir bases de dados que podem ser acedidas com um protocolo de Java Database Connectivity (JDBC) numa colecção.

Tem de configurar um pesquisador separado para cada tipo de sistema de base de dados que pretende pesquisar. Quando cria um pesquisador, especifica opções para pesquisar uma base de dados. Posteriormente, pode adicionar bases de dados do mesmo tipo ao espaço de pesquisa.

Cada linha numa tabela de bases de dados é considerada como um documento e os valores das colunas da base de dados são analisados e indexados como campos pesquisáveis. Pode configurar o pesquisador para pesquisar tabelas estruturadas múltiplas através da associação de um suplemento ao pesquisador durante a configuração das propriedades do pesquisador. Com este suplemento, as linhas das tabelas múltiplas numa base de dados relacional que tenham os mesmos campos de chave, podem ser juntas e consideradas como um só documento. Quando um utilizador procura a base de dados, os dados das tabelas conjuntas aparecem como campos adicionais quando o documento é apresentado nos resultados da procura.

Restrição: Se instalar o IBM Content Analytics with Enterprise Search no sistema operativo Linux com System *z*, pode utilizar o pesquisador de Base de dados JDBC para pesquisar bases de dados do IBM DB2. Outros tipos de bases de dados não são suportados.

Sistemas e controladores de base de dados suportados

Para utilizar o protocolo JDBC para pesquisar tabelas numa base de dados, tem de existir o controlador de JDBC apropriado no servidor do pesquisador. O pesquisador de Base de dados JDBC suporta os seguintes sistemas de base de dados e controladores de JDBC de tipo 4:

Sistema de base de dados	Nome do controlador de JDBC de tipo 4	Caminhos de classe do controlador de JDBC padrão
IBM DB2 Enterprise Server Edition para Linux UNIX e	com.ibm.db2.jcc.DB2Driver	<i>db2_install_root/</i> java/db2jcc.jar <i>db2_install_root/</i> java/db2jcc_license_cu.jar
Windows		Para pesquisar bases de dados do DB2 num sistema Linux com System z, utilize o ficheiro db2jcc.jar e o ficheiro db2jcc_license_cisuz.jar que estão associados ao DB2 Connect.
Oracle 9i e 10g	Oracle.jdbc.driver.OracleDriver	<pre>oracle_home/jdbc/lib/ojdbc14.jar</pre>

Sistema de base de dados	Nome do controlador de JDBC de tipo 4	Caminhos de classe do controlador de JDBC padrão
Oracle 11g	Oracle.jdbc.driver.OracleDriver	oracle_home/jdbc/lib/ojdbc6.jar
Microsoft SQL Server 2000	com.microsoft.jdbc.sqlserver. SQLServerDriver	mssql_jdbc_home/lib/mssqlserver.jar mssql_jdbc_home/lib/msbase.jar mssql_jdbc_home/lib/msutil.jar
Microsoft SQL Server 2005	com.microsoft.sqlserver.jdbc. SQLServerDriver	<pre>install_dir/sqljdbc_1.0/loc/sqljdbc.jar em que loc representa o seu locale, como install_dir/sqljdbc_1.0/ enu/sqljdbc.jar</pre>
Microsoft SQL Server 2008	com.microsoft.sqlserver.jdbc. SQLServerDriver	<pre>install_dir/sqljdbc_2.0/loc/sqljdbc4.jar em que loc representa o seu locale, como por exemplo install_dir/sqljdbc_2.0/enu/sqljdbc4.jar</pre>
		Quando configura o pesquisador, seleccione Base de dados JDBC genérica para o tipo de base de dados JDBC antes de poder especificar o nome do controlador JDBC.
MySQL 5.0	com.mysql.jdbc.Driver	install_directory/mysql-connector-java-version- bin.jar

Pesquisadores de Base de dados JDBC em contraste com pesquisadores do DB2

Se estiver presentemente a utilizar o pesquisador de DB2, pode achar conveniente continuar a utilizá-la. Não pode migrar dados armazenados para um pesquisador de DB2 num pesquisador de Base de dados JDBC.

Utilize o pesquisador de DB2 em vez do pesquisador de Base de dados JDBC nas seguintes situações:

- Se pretender pesquisar bases de dados do DB2 com um controlador JDBC de tipo 2.
- Se pretender pesquisar bases de dados Oracle e SQL Server que são federadas com uma base de dados do DB2. Com o pesquisador de DB2, pode aceder a todos estes tipos de bases de dados através de um nome alternativo.
- Se pretender pesquisar bases de dados do DB2 para z/OS, DB2 for iSeries, Informix, Sybase, VSAM, IMS, CA-Datacom ou Software AG Adabas. Federar este tipo de bases de dados com uma base de dados DB2 e acedê-los com o pesquisador do DB2 através de um nome alternativo.
- Pretende utilizar a publicação de eventos para actualizar o índice quando são publicadas actualizações à base de dados.

Credenciais de ligação do pesquisador

Quando cria o pesquisador, pode especificar as credenciais que permitem que o pesquisador se ligue às fontes a pesquisar. Também pode configurar as credenciais de ligação quando especifica as definições gerais de segurança do sistema. Se utilizar a última abordagem, vários pesquisadores e outros componentes do sistema podem utilizar as mesmas credenciais. Por exemplo, os servidores de procura podem utilizar as credenciais para determinar se um utilizador possui autorização para aceder ao conteúdo.

Descrição geral da configuração

Ao criar o pesquisador, um assistente ajudá-lo-á nessas tarefas:

- Especifique as propriedades que controlam o modo como o pesquisador opera e utiliza os recursos do sistema. As propriedades do pesquisador controlam o modo como este último pesquisa todas as bases de dados no espaço de pesquisa.
- Especifique o tipo de base de dados que pretende pesquisar.
- Seleccione a base de dados que pretende pesquisar e, se necessário, especifique um ID de utilizador e palavra-passe que permita que o pesquisador aceda à base de dados.
- Configure um agendamento para pesquisar a base de dados.
- Seleccione as tabelas que pretende pesquisar.

Aviso: Para optimizar o desempenho dos processos de identificação e para impedir que o processo de configuração do pesquisador expire, opte por pesquisar todas as tabelas apenas se a base de dados não contiver muitas tabelas ou se as tabelas não contiverem muitas colunas. Se optar por seleccionar algumas tabelas a pesquisar, neste momento, pode editar o espaço de pesquisa posteriormente e adicionar mais tabelas à colecção.

- Configure as opções de segurança ao nível do documento. Se a segurança tiver sido activada quando a colecção foi criada, o pesquisador pode associar dados de segurança aos documentos existentes no índice. Esses dados permitem às aplicações fazer cumprir os controlos de acesso baseados nas listas de controlo de acesso armazenadas ou nos tokens de segurança.
- Especifique as opções para fazer com que as colunas possam ser procuradas em alguns tipos de tabelas. Por exemplo, pode activar certas colunas a serem utilizadas em consultas paramétricas, ou especifique as colunas que podem regressar nos resultados da procura.
- Especifique opções para procurar o conteúdo do documento. Cada tabela pode ter apenas uma coluna que pode ser tratada como se contivesse o conteúdo principal do documento. Para ajudar as aplicações a encontrar e recuperar este conteúdo, tal como o conteúdo dos documentos .doc ou .pdf, insira a informação sobre o conteúdo no pesquisador.

Por exemplo, se uma coluna da tabela contiver conteúdo grande binário ou de texto, tal como os documentos com dados do tipo BLOB, pode configurar o pesquisador para permitir que os utilizadores consultem o conteúdo através de consultas de texto gratuitas e visualizem o conteúdo do documento em resumos do documento nos resultados. Pode melhorar a recuperabilidade ao especificar o tipo MIME, tal como aplicação/msword ou aplicação/pdf, e ao especificar as opções para determinar a página de código utilizada para codificar o conteúdo.

Mapas de relações para bases de dados de JDBC

Quando cria regras para um suplemento que pesquisa tabelas de bases de dados de JDBC estruturadas múltiplas, especifica informações sobre a tabela principal e como é que as tabelas ascendentes e descendentes são unidas.

Um suplemento facultado permite ao pesquisador de Base de dados JDBC juntar tabelas estruturadas múltiplas. O utilizador cria o suplemento especificando regras no ficheiro ES_INSTALL_ROOT/default_config/crawler_rdb_plugin.xml. Depois de configurar o pesquisador para utilizar o suplemento, as linhas das tabelas que têm os mesmos campos de chave são unidas e consideradas como um só documento. Quando um utilizador procura a base de dados, os dados das tabelas conjuntas aparecem como campos adicionais quando o documento é apresentado nos resultados da procura.

Juntar tabelas através de colunas chave

A figura seguinte apresenta como é que o mapa de relações para tabelas múltiplas é construído. O pesquisador de Base de dados JDBC procura uma tabela principal numa base de dados. Algumas das colunas na tabela são campos de chave que podem ser utilizados para juntar a tabela a outras tabelas. As colunas nas tabelas conjuntas podem assim ser utilizadas como chaves para juntar tabelas adicionais. As linhas nas tabelas múltiplas são consideradas como um só documento no espaço de pesquisa. Os valores da coluna são considerados como metadados do documento. A tabela principal é *o ascendente (the parent)* na relação e a tabela conjunta do primeiro nível é *um descendente (a child)*. As tabelas descendentes do primeiro nível também podem ser ascendentes de tabelas que são unidas num segundo nível.

Neste exemplo, as colunas Chave 1 e Chave 2 na tabela principal (ascendente) são campos de chave que permitem que a tabela se junte a tabelas secundárias que têm também colunas Chave 1 e Chave 2. Uma tabela conjunta a este primeiro nível tem campos de chave, Chave 3 e Chave 4, que permitem que a tabela se junte a tabelas adicionais.



Figura 2. Tabelas de bases de dados de JDBC juntas através de campos de chave

Visualizar resultados da procura

A figura seguinte mostra é que os dados das tabelas estruturadas múltiplas são apresentados como um só documento nos resultados da procura. Sem o suplemento, um utilizador que procure a tabela EMPLOYEE, poderá visualizar uma linha da tabela principal apresentada nos resultados da procura e só visualizar os valores para as colunas da tabela EMPLOYEE (ID, Name e Office).

Com o suplemento, o pesquisador tem a capacidade de utilizar a coluna Office como uma chave para juntar a tabela EMPLOYEE com a tabela OFFICE. A coluna Country na tabela OFFICE serve como uma chave para juntar essa tabela com a tabela COUNTRY. Após a junção das tabelas, os utilizadores que procurem a tabela EMPLOYEE podem visualizar valores das colunas nas tabelas OFFICE e COUNTRY como campos adicionais nos resultados da procura.



Figura 3. Os valores das tabelas de JDBC conjuntas são apresentados nos resultados da procura

Pesquisar tabelas de bases de dados de JDBC estruturadas múltiplas

Pode configurar o pesquisador de Base de dados JDBC para se juntar a tabelas estruturadas múltiplas que têm os mesmos campos de chave.

Quando configura as propriedades do pesquisador para um pesquisador de Base de dados JDBC, pode especificar um suplemento para pesquisar múltiplas tabelas estruturadas que estão relacionadas umas com as outras através de campos de chave. Sem este suplemento, as linhas numa tabela de bases de dados são consideradas como documentos individuais e os valores das colunas das bases de dados são pesquisáveis como campos individuais. Com este suplemento, as linhas das tabelas múltiplas numa base de dados relacional que tenham os mesmos campos de chave são unidas e consideradas como um só documento. O pesquisador adiciona dados que obtém das tabelas conjuntas para os metadados, para a linha original de uma tabela de bases de dados. Quando um utilizador procura a base de dados, os dados adicionais aparecem como campos adicionais quando o documento é apresentado nos resultados da procura.

Tipos de dados que não podem ser pesquisados

O pesquisador não pode pesquisar campos nas tabelas que o utilizador juntou, que contenham estes tipos de dados binários:

BLOB CHARACTER FOR BIT DATA VARCHAR FOR BIT DATA LONG VARCHAR FOR BIT DATA

Limitações no âmbito do espaço de pesquisa

As tabelas a serem unidas têm de estar na mesma base de dados relacional. Não pode juntar tabelas ao longo de bases de dados.

Se uma tabela, numa base de dados, estiver configurada para se juntar a outras tabelas, esta definição é universal para todos os pesquisadores numa colecção, que estejam activados para utilizar o suplemento. No entanto, é possível criar colecções múltiplas e configurar pesquisadores separados para pesquisar tabelas principais diferentes e juntar tabelas diferentes.

Restrições na utilização de outros suplementos

Se configurar o pesquisador para utilizar o suplemento para pesquisar tabelas estruturadas múltiplas, não pode associar outro suplemento ao pesquisador. Por exemplo, não é possível especificar um suplemento personalizado para aplicar regras de negócio e de segurança. Não pode associar mais do que um suplemento a um pesquisador.

Restrições no número de tabelas, linhas, campos e chaves

O número máximo de tabelas conjuntas por base de dados é cinco e a soma das linhas nessas tabelas deve ser inferior a 1 000 000. O número máximo de campos que podem ser lidos a partir de uma tabela é 10. Para juntar tabelas, utiliza-se um par de chaves. Isso significa que não é possível juntar tabelas utilizando chaves múltiplas.

Garantir que as alterações em tabelas conjuntas são pesquisadas

Se as linhas numa tabela principal não alteram entre pesquisas e o pesquisador não está configurado para efectuar uma pesquisa completa, o pesquisador ignora as linhas inalteradas. Se as linhas numa tabela unida à tabela principal se alterarem, mesmo que a tabela principal não o faça, terá de executar uma das seguintes acções para garantir que as alterações são detectadas e pesquisadas:

- Uma tabela principal na base de dados destino deve possuir um campo de marca de tempo. Configure a base de dados destino para ter um campo de marca de tempo que é actualizado quando uma linha na tabela principal se altera ou quando as linhas em qualquer das tabelas secundárias conjuntas se alteram. Quando define o pesquisador de Base de dados JDBC, certifique-se de que especifica o campo de marca de tempo como o campo que o pesquisador utiliza para determinar se ocorreram alterações nas tabelas.
- Especifique que o pesquisador deve efectuar uma pesquisa completa sempre que configurar a agenda do pesquisador. Esta opção garante que todas as tabelas são pesquisadas todas as vezes, independentemente de terem ocorrido quaisquer alterações ou não.

Para configurar o pesquisador de Base de dados JDBC para pesquisar tabelas estruturadas múltiplas:

- Na consola de administração, crie um pesquisador de Base de dados JDBC. Seleccione apenas a tabela principal a pesquisar na configuração do pesquisador.
- 2. Inicie sessão como administrador predefinido do IBM Content Analytics with Enterprise Search no servidor do pesquisador.
- Copie o ficheiro ES_INSTALL_ROOT/default_config/crawler_rdb_plugin.xml para criar o ficheiro ES_NODE_ROOT/master_config/crawler_rdb_plugin.xml.
- 4. Edite o ficheiro ES_NODE_ROOT/master_config/crawler_rdb_plugin.xml através de um editor de texto que suporte ficheiros XML com codificação de caracteres UTF-8, tal como o Notepad. Quando editar o ficheiro de configuração:
 - Introduza as alterações adequadas para as tabelas da base de dados que a Base de dados JDBC deve pesquisar.
 - Introduza todos os nomes de colunas das tabelas da base de dados em maiúsculas.
 - Guarde todas as alterações em formato de codificação UTF-8.
 - a. Edite o elemento <Server DBURL="jdbc:db2://db_server_url:50000/ SAMPLE"> e substitua jdbc:db2://db_server_url:50000/SAMPLE pelo URL da base de dados de JDBC para ser pesquisado.
 - b. Se a base de dados a pesquisar não for uma base de dados DB2, edite o elemento <JDBCDriver>com.ibm.db2.jcc.DB2Driver</JDBCDriver> e substitua com.ibm.db2.jcc.DB2Driver pelo controlador JDBC adequado. Por exemplo, para uma base de dados Oracle, introduza oracle.jdbc.driver.OracleDriver. Certifique-se de que especifica o mesmo controlador que especificou quando criou o pesquisador.
 - **c.** Edite o elemento <User>username</User> e substitua o username por um ID de utilizador que tem autoridade para aceder à base de dados a pesquisar.
 - d. Edite o elemento <Password Encryption="True">encrypted_password</Password> e substitua a encrypted_password com uma palavra-passe codificada para o ID de utilizador especificado. Pode copiar a palavra-passe codificada a partir do ficheiro ES_NODE_ROOT/master_config/col_collection_name.JDBC_crawler_name/jdbccrawler.xml e colá-la aqui. Se a palavra-passe não necessitar de ser codificada, substitua Encryption="True" com Encryption="False" e substitua a encrypted_password com uma palavra-passe de texto simples.
 - e. Se não alterar o elemento <Delimiters Use="True">, vários termos numa coluna são separados pelo carácter vírgula(,) definido no elemento <Delimiter>. Conjuntos de termos por tabela são separados pelo carácter ponto e vírgula (;) definido no elemento <SecondDelimiter>. Se definir <Delimiters Use="True"> para <Delimiters Use="False">, os caracteres de delimitador não são utilizados e os campos de metadados múltiplos, com o mesmo nome do campo, são adicionados como metadados do documento.
 - f. Se utilizar o elemento <Delimiters Use="True">, edite o elemento <Delimiter>,</Delimiter> e o elemento <SecondDelimiter>;</ SecondDelimiter> para especificar os caracteres que pretende utilizar como separadores de valor.
 - g. Edite o elemento <RelationMap Root="DB2INST1.TABLE_0"> e substitua DB2INST1.TABLE_0 com o nome de uma tabela principal que irá ser pesquisada.
 - h. Edite o elemento <Relation Parent="DB2INST1.TABLE_0" ParentAlias="T0" ParentKey="ID" Child="DB2INST1.TABLE_1" ChildAlias="T1" ChildKey="ID"/>.

- Substitua Parent="DB2INST1.TABLE_0" pelo nome de uma tabela que é um ascendente na relação.
- Substitua ParentKey= "T0" pelo nome alternativo da tabela principal. O nome alternativo deve ser exclusivo e não duplicado no ficheiro crawler_rdb_plugin.xml.
- Substitua ParentKey="ID" pelo nome de uma coluna que é utilizada como um campo de chave na relação.
- Substitua Child="DB2INST1.TABLE_1" ChildAlias="T1" ChildKey="ID" pelas informações sobre uma tabela descendente a pesquisar.

Esta estrutura define como é que as tabelas devem ser unidas. Por exemplo, o mapa de relações seguinte especifica que uma tabela principal designada por DB2INST1.TABLE_A deve ser pesquisada. As tabelas DB2INST1.TABLE_B e DB2INST1.TABLE_C são juntas sob a condição

DB2INST1.TABLE_A.ID=DB2INST1.TABLE_B.ID AND

DB2INST1.TABLE_B.ID=DB2INST1.TABLE_C.ID.

<RelationMap Root="DB2INST1.TABLE_A">

```
<Relation Parent="DB2INST1.TABLE_A" ParentAlias="TA" ParentKey="ID"
Child="DB2INST1.TABLE_B" ChildAlias="TB" ChildKey="ID"/>
```

```
<Relation Parent="DB2INST1.TABLE_B" ParentAlias="TB" ParentKey="ID"
Child="DB2INST1.TABLE_C" ChildAlias="TC" ChildKey="ID"/>
```

Restrição: Todas as referências à tabela ou aos nomes de campos na base de dados são sensíveis a maiúsculas. Especifique-os em maiúsculas e minúsculas como definido na base de dados.

- i. Repita o passo 4h na página 104 para criar elementos <Relation> para todas as relações que juntam tabelas a partir de uma tabela principal.
- j. Edite o elemento <Target TableAlias="T1"> e substitua TableAlias="T1" com um valor ChildAlias que definiu no passo 4h na página 104.
- k. Edite o elemento <Field Name="ID" FieldName="ID_1" Enabling="True" Searchable="True" FieldSearchable="True" IsContent="True"/>.
 - Substitua Name="ID" com o nome de uma coluna nos documentos a pesquisar.
 - Substitua FieldName="ID_1" com o nome de um campo de metadados nos documentos a pesquisar. Este valor é utilizado como nome de apresentação para a coluna na consola de administração e os resultados da pesquisa.
 - Substitua Enabling="True" por "False" se esta coluna não for incluída nos metadados do documento.
 - Substitua Searchable="True" com "False" para impedir os utilizadores de procurar esta coluna com uma consulta de textos livre.
 - Substitua FieldSearchable="True" com "False" para impedir os utilizadores de procurar esta coluna pelo nome da coluna.
 - Substitua IsContent="True" com "False" para indicar que a coluna não contém conteúdo pesquisável. Se especificar Searchable="True" e IsContent="True", o valor da coluna é utilizado para detectar documentos duplicados. O valor também se torna parte do resumo do documento dinâmico nos resultados da procura.
- I. Repita o passo 4k para criar elementos <Field> para todas as colunas que são para pesquisar.
- m. Repita os passos 4j e 4k para criar elementos <Target> e <Field> para todas as tabelas descendentes no mapa de relações (<RelationMap>).

- n. Repita os passos 4g na página 104 através de 4m na página 105 para criar mapas de relações múltiplas para tabelas principais múltiplas.
- **o**. Repita os passos 4a na página 104 através de 4n para configurar um mapa de relações para outra base de dados.
- 5. Configure o pesquisador para utilizar o suplemento:
 - a. Abra a consola de administração, edite uma colecção e seleccione a página Pesquisar.
 - b. Na consola de administração, seleccione o pesquisador que criou e faça clique em **Configurar** > **propriedades do pesquisador**.
 - **c**. No campo **Nome de classe do suplemento**, escreva o nome do suplemento para pesquisar múltiplas tabelas estruturadas:

com.ibm.es.plugin.rdb.RDBPlugin

d. No campo **caminho de classe de suplemento**, introduza os caminhos totalmente qualificados para o suplemento e os controladores JDBC utilizados pelo suplemento. Por exemplo, o caminho para o controlador JDBC para a base de dados DB2 num sistema Windows pode ser:

C:\Program Files\IBM\es\lib\plugin_rdb.jar;C:\Program Files\IBM\SQLLIB\java\db2jcc.jar;C:\Program Files\IBM\SQLLIB\java\ db2jcc license cu.jar

Num sistema AIX ou Linux, o caminho pode ser:

/opt/IBM/es/lib/plugin_rdb.jar:IBM/SQLLIB/java/db2jcc.jar:/opt/IBM/ SQLLIB/java/db2jcc_license_cu.jar

- 6. Criar campos de índice para a tabela mais recente:
 - a. No painel Analisar e Indexar, faça clique em **Configurar** > **Campos de Índice**.
 - b. Faça clique em **Criar campo de índice** e crie um campo de índice que definiu no ficheiro crawler_rdb_plugin.xml como parte de uma etapa 4k na página 105. Repita esta etapa para criar outros campos de índice.
- 7. Para implementar o ficheiro crawler_rdb_plugin.xml para a configuração do sistema, reinicie o sistema do IBM Content Analytics with Enterprise Search:

Paragem do sistema esadmin Início do sistema esadmin

Conceitos relacionados:

"Opções para pesquisar várias tabelas de base de dados JDBC estruturadas"

Opções para pesquisar várias tabelas de base de dados JDBC estruturadas

Para definir junções mais complexas, estão disponíveis directrizes adicionais para configurar o pesquisador Base de dados JDBC para juntar várias tabelas estruturadas.

Para configurar opções para o suplemento, edite o ficheiro ES_NODE_ROOT/ master_config/crawler_rdb_plugin.xml através do editor de texto que suporta ficheiros XML com codificação de caracteres UTF-8.

Especificar várias chaves principais de uma tabela principal ascendente

Se uma tabela principal tiver mais do que uma chave primária, deve especificar todas as chaves no ficheiro de configuração. No ficheiro de configuração, adicione

os elementos <RootKey> que possuem valores de atributos ParentKey. Neste exemplo, ID-1 e ID-2 são chaves principais de uma tabela principal com o nome DB2INST1.TABLE_0.

```
<RelationMap Root="DB2INST1.TABLE_0">
<RootKey ParentKey="ID-2"/>
<Relation Parent="DB2INST1.TABLE_0" ParentAlias="T0" ParentKey="ID-1"
Child="DB2INST1.TABLE_1" ChildAlias="T1" ChildKey="ID-A"/>
</RelationMap>
```

Juntar tabelas descendentes com várias chaves

Se as tabelas que foram anexadas umas às outras possuem várias chaves, especifique o conjunto de chaves no ficheiro de configuração. No ficheiro de configuração, defina vários elementos de <Relation> que possuem o valor de atributo Duplicated="True". Neste exemplo, DB2INST1.TABLE_0 e DB2INST1.TABLE_1 são anexadas através de chaves com uma condição de correspondência (ID-1=ID-A e ID-2=ID-B).

```
<RelationMap Root="DB2INST1.TABLE_0">
    <Relation Parent="DB2INST1.TABLE_0" ParentAlias="T0" ParentKey="ID-1"
    Child="DB2INST1.TABLE_1" ChildAlias="T1" ChildKey="ID-A"/>
    <Relation Parent="DB2INST1.TABLE_0" ParentAlias="T0" ParentKey="ID-2"
    Child="DB2INST1.TABLE_1" ChildAlias="T1" ChildKey="ID-B" Duplicated="True"/>
    <Relation Parent="DB2INST1.TABLE_0" ParentAlias="T0" ParentKey="ID-3"
    Child="DB2INST1.TABLE_1" ChildAlias="T1" ChildKey="ID-C" Duplicated="True"/>
    <Relation Parent="DB2INST1.TABLE_1" ChildAlias="T1" ChildKey="ID-C" Duplicated="True"/>
    </RelationParent="DB2INST1.TABLE_1" ChildKey="ID-C" Duplicated="True"/>
    </RelationParent="Comparent="Comparent="Comparent="Comparent="Comparent="Comparent="Comparent="Comparent="Comparent="Comparent="Comparent="Comparent="Comparent="Comparent="Comparent="Comparent="Comparent="Comparent="Comparent="Comparent="Comparent="Comparent="Comparent="Comparent="Comparent="Comparent="Comparent="Comparent="Comparent="Comparent="Comparent="Comparent="Co
```

Listar valores de coluna duplicados num campo de metadados

Se forem obtidos vários valores de colunas para um campo de metadados de tabelas conjuntas, os valores duplicados são eliminados por predefinição. Por exemplo, suponhamos que o ficheiro de configuração foi definido para adicionar um valor de campo de metadados para o campo NAME_1 a partir de valores de coluna de uma tabela descendente, T1, que é anexada a uma tabela ascendente através das chaves ID-1 e ID-2. Se as duas linhas corresponderem a esta condição e as colunas NAME em ambas as linhas possuírem o mesmo valor, tal como "Mr.X", o valor do campo NAME_1 será "Mr.X".

```
<Target TableAlias="T1">

<Field Name="ID-1" FieldName="ID_1" Enabling="False" Searchable="True"

FieldSearchable="True" IsContent="False"/>

<Field Name="ID-2" FieldName="ID_2" Enabling="False" Searchable="True"

FieldSearchable="True" IsContent="False"/>

<Field Name="NAME" FieldName="NAME_1" Enabling="True" Searchable="True"

FieldSearchable="True" IsContent="True"/>

<Field Name="OFFICE" FieldName="OFFICE_1" Enabling="True" Searchable="True"

FieldSearchable="True" IsContent="True"/>

<Field Name="SUBID" FieldName="SUBID_1" Enabling="False" Searchable="True"

FieldSearchable="True" IsContent="False"/>

<Field Name="SUBID" FieldName="SUBID_1" Enabling="False" Searchable="True"

</Face Searchable="True" Searchable="True"

</Face Searchable="True" Sublock Searchable="True"

</Face Searchable="True" Searchable="True" Searchable="True" Searchabl
```

Pode alterar este comportamento predefinido ao adicionar o valor de atributo Primarykey="True" aos elementos <Field> das colunas chave principais numa tabela descendente, como apresentado para as chaves ID-1 e ID-2 no seguinte exemplo:

```
<Target TableAlias="T1">

<Field Name="ID-1" FieldName="ID_1" Enabling="False" Searchable="True"

FieldSearchable="True" IsContent="False" PrimaryKey="True"/>

<Field Name="ID-2" FieldName="ID_2" Enabling="False" Searchable="True"

FieldSearchable="True" IsContent="False" PrimaryKey="True"/>

<Field Name="NAME" FieldName="NAME_1" Enabling="True" Searchable="True"

FieldSearchable="True" IsContent="True"/>
```

```
<Field Name="OFFICE" FieldName="OFFICE_1" Enabling="True" Searchable="True"
FieldSearchable="True" IsContent="True"/>
<Field Name="SUBID" FieldName="SUBID_1" Enabling="False" Searchable="True"
FieldSearchable="True" IsContent="False"/>
</Target>
```

Com esta alteração, o valor do campo NAME_1 será "MrX, Mr.X" em que "," é um carácter de delimitação que especifica no ficheiro de configuração.

Especificar um conjunto total de atributos para um campo de metadados

Ao adicionar valores de atributo nos elementos de <Field>, pode definir explicitamente valores de atributo em vez de utilizar os valores predefinidos.

Atributo do campo	Valor predefinido	Valores alternativos
Activar	False	True
Procura	True	False
Procura_de_campo	True	False
ConteúdoIs	False	True
Procura_paramétrica	False	True
Metadados	True	False
Resolver_conflito	Metadados_pretendidos	Conteúdo_pretendido Coexistir
Correspondência_exacta	False	True
Ordenável	False	True

Tabela 12. Atributos para elementos de campo

Neste exemplo, os atributos do campo ID_1 são explicitamente especificados:

<Field Name="ID" FieldName="ID_1" Enabling="False" Searchable="True"
FieldSearchable="True" IsContent="False" ParametricSearchable="False"
Metadata="True"\@ResolveConflict="MetadataPreferred" ExactMatch="False"
Sortable="False"/>

Se o elemento <Delimiters Use="True"> for definido no ficheiro de configuração, não é possível utilizar a procura paramétrica. Para permitir uma procura paramétrica, defina Use="False". Por exemplo:

```
<Delimiters Use="False"
  <Delimiter>,</Delimiter>
  <SecondDelimiter>;</SecondDelimiter>
</Delimiters>
```

Não adicionar um carácter de delimitação se não existir qualquer valor de coluna a adicionar

Por predefinição, um carácter de delimitação é adicionado a um valor de campo quando nenhum linha de uma tabela descendente corresponde à condição de junção e não existe qualquer valor de coluna a adicionar. Para alterar este comportamento, adicione o atributo IgnoreNull="True" ao elemento <Field>. Com este atributo, não é adicionado qualquer delimitador ao valor de campo se não existir qualquer valor de coluna a adicionar. Neste exemplo, suponhamos que nenhuma das linhas da tabela descendente anexada correspondeu ao valor chave. Por predefinição, o valor do campo ID_1 é ",", um carácter de delimitação, no entanto, ao adicionar IgnoreNull="True" ao elemento <Field>, o valor torna-se vazio.

```
<Field Name="ID" FieldName="ID_1" Enabling="False" Searchable="True"
FieldSearchable="True" IsContent="False" ParametricSearchable="False"
Metadata="True"\carcolectered" ExactMatch="False"
Sortable="False" IgnoreNull="True"/>
```

Configurar o nível de isolamento para pesquisar

Por predefinição, o nível de isolamento és definido para leitura não consolidada. Para alterar este comportamento para leitura consolidada, adicione o elemento <0ptionalClause>WITH CS</0ptionalClause> ao ficheiro de configuração, como indicado neste exemplo:

```
<Crawler>
<DataSources>
<Server>
<RelationMap>
<OptionalClause>WITH CS</OptionalClause>
</RelationMap>
</Server>
</DataSources>
</Crawler>
```

Tarefas relacionadas:

"Pesquisar tabelas de bases de dados de JDBC estruturadas múltiplas" na página 102

NNTP pesquisadores

Para incluir artigos de news groups de NNTP numa colecção, tem de configurar um pesquisador do NNTP.

Pode utilizar o pesquisador do NNTP para pesquisar qualquer número de servidores NNTP. Quando configura o pesquisador, selecciona os newsgroups a pesquisar a partir de um único servidor de NNTP. Posteriormente, quando edita o espaço do pesquisador, pode adicionar outros servidores de NNTP que pretenda que o pesquisador pesquise.

Quando identifica os newsgroups a pesquisar, pode seleccionar grupos a incluir e grupos a excluir do espaço do pesquisador. Com esta concepção, pode permitir facilmente que o pesquisador pesquise a maioria dos novos grupos num servidor e proibir que o pesquisador pesquise alguns grupos novos que não quer que os utilizadores procurem.

Por exemplo, pode especificar as regras para incluir todos os newsgroups num servidor NNTP específico, e depois especificar que pretende excluir newsgroups nesse servidor se os respectivos nomes incluírem a cadeia privado .

Ao criar o pesquisador, um assistente ajudá-lo-á nessas tarefas:

- Especifique as propriedades que controlam o modo como o pesquisador opera e utiliza os recursos do sistema. As propriedades do pesquisador controlam o modo como este último pesquisa todos os newsgroups no espaço de pesquisa.
- Identifique o servidor de NNTP a pesquisar. Se o servidor estiver protegido por palavra-passe, tem de fornecer um ID de utilizador e palavra-passe para o pesquisador utilizar para aceder a newsgroups no servidor.

- Configure um agendamento para pesquisar o servidor.
- Especifique padrões para incluir newsgroups, e especifique padrões para excluir certos newsgroups do espaço de pesquisa.
- Especifique se o pesquisador deve detectar automaticamente o idioma e a página de código dos artigos a pesquisar ou se o pesquisador deve utilizar um idioma e uma página de código específicos.
- Configure as opções de segurança ao nível do documento. Se a segurança tiver sido activada quando a colecção foi criada, o pesquisador pode associar dados de segurança aos documentos existentes no índice. Esses dados permitem às aplicações fazer cumprir os controlos de acesso baseados nas listas de controlo de acesso armazenadas ou nos tokens de segurança.

Credenciais de ligação do pesquisador

Quando cria o pesquisador, pode especificar as credenciais que permitem que o pesquisador se ligue às fontes a pesquisar. Também pode configurar as credenciais de ligação quando especifica as definições gerais de segurança do sistema. Se utilizar a última abordagem, vários pesquisadores e outros componentes do sistema podem utilizar as mesmas credenciais. Por exemplo, os servidores de procura podem utilizar as credenciais para determinar se um utilizador possui autorização para aceder ao conteúdo.

Pesquisadores do Notes

Para incluir conteúdo de IBM Lotus Notes numa colecção, tem de configurar um pesquisador Notes

Credenciais de ligação do pesquisador

Quando cria o pesquisador, pode especificar as credenciais que permitem que o pesquisador se ligue às fontes a pesquisar. Também pode configurar as credenciais de ligação quando especifica as definições gerais de segurança do sistema. Se utilizar a última abordagem, vários pesquisadores e outros componentes do sistema podem utilizar as mesmas credenciais. Por exemplo, os servidores de procura podem utilizar as credenciais para determinar se um utilizador possui autorização para aceder ao conteúdo.

Configuração do servidor do pesquisador

Se um servidor Lotus Notes que tenciona pesquisar utilizar o protocolo de chamada de procedimento remoto (NRPC) Notes, tem de executar o script de configuração **escrnote** no servidor do pesquisador. Esse script fornecido com IBM Content Analytics with Enterprise Search, activa o pesquisador do Notes para comunicar com os servidores que utilizam o NRPC.

Se um servidorLotus Notes que pretende pesquisar utilizar o Domino Internet Inter-ORB Protocol (DIIOP), não necessita de executar um script de configuração no servidor do pesquisador. No entanto, tem de configurar o servidor Lotus Notes de modo a que o pesquisador Notes possa aceder ao servidor.

Importante: Se o servidor Lotus Notes utilizar o DIIOP e o utilizador configurar o pesquisador para utilizar o HTTPS ou DIIOP através de SSL, para que as transmissões entre o pesquisador e o servidor sejam codificadas, é necessário copiar o ficheiro TrustedCerts.class (por exemplo, c:\certs ou /data/certs) do servidor Lotus Notes para o servidor do pesquisador. Numa configuração de

vários servidores, também deve copiar o ficheiro TrustedCerts.class para os servidores de procura. Tem de assegurar que ficheiro se encontra na mesma localização no servidor do pesquisador e nos servidores de procura. O utilizador especifica o caminho do directório para o ficheiro TrustedCerts.class quando configura o pesquisador.

Se IBM Content Analytics with Enterprise Search tiver sido instalado em um sistema da IBM AIX, tem de garantir que o módulo E/S Porta de Conclusão esteja instalado e disponível no servidor do pesquisador.

Antes de utilizar a consola de administração para configurar um pesquisador do Notes, execute as tarefas seguintes apropriadas para o ambiente:

- "Configurar o servidor do pesquisador no UNIX para pesquisar origens do Lotus Domino" na página 114.
- "Configurar o servidor do pesquisador no Windows para pesquisar origens do Lotus Domino" na página 116.
- "Configurar servidores que utilizam o protocolo DIIOP" na página 117.
- "Configurar a porta de conclusão E/S no AIX para pesquisar origens no Lotus Domino" na página 118.

Bases de dados, directórios e vistas de coluna

Quando cria um pesquisador Notes, tem de configurar o pesquisador para pesquisar bases de dados ou directórios. Um pesquisador configurado para pesquisar bases de dados não pode ser configurado para também pesquisar directórios. Se configurar o pesquisador para recolher conteúdos de bases de dados quando cria o pesquisador, não é possível adicionar directórios quando editar o espaço de pesquisa. De igual modo, não é possível adicionar bases de dados a um pesquisador que esteja configurado para pesquisar directórios.

Quando configura o pesquisador Notes para pesquisar bases de dados, é possível especificar se o pesquisador deve pesquisar todos os documentos ou vistas e pastas específicas. Se configurar o pesquisador para pesquisar vistas e pastas, é possível especificar se o pesquisador deve pesquisar os valores dos campos nas vistas (os *valores da coluna*) em acréscimo a pesquisar campos dos documentos. Pesquisar os valores da coluna em vistas da base de dados fornece uma forma eficaz de recolher o conteúdo a partir da base de dados.

Quando especifica opções para pesquisar vistas da base de dados, é possível fazer clique num botão para validar os valores da coluna a serem pesquisados. Após a validação, os nomes de campos candidatos são apresentados na lista de campos a serem pesquisados. Os nomes de campos extraídos de colunas são baseados no título da coluna com o prefixo Coluna_, tal como Coluna_Data ou Coluna_Categoria. Se o título da coluna for nulo, o pesquisador utiliza o nome do item programático para a coluna como nome de campo candidato. Se um título incluir caracteres inválidos, o pesquisador substitui tais caracteres por um carácter de traço sublinhado.

Restrição: O pesquisador do Notes não consegue devolver vistas ou ficheiros que contenham os seguintes tipos de valores:

- Uma fórmula que contenha uma função de apenas interface do utilizador
- Uma constante

Quando configura o pesquisador Notes para pesquisar directórios, é possível especificar padrões para incluir ou excluir bases de dados. Ao filtrar as bases de dados, é possível dividir a tarefa de pesquisar directórios amplos por vários pesquisadores.

Se deseja que o pesquisador recolha valores da coluna a partir das bases de dados num directório, tem de especificar o nome de uma vista que contenha os valores da coluna a serem analisados. O pesquisador aplica a mesma análise de vista a todas as bases de dados sob o directório de destino. Se uma vista não existir numa base de dados, o pesquisador pode ignorar a base de dados ou apenas pesquisar os campos dos documentos nas bases de dados. Devido ao tempo que pode demorar a identificar valores possíveis em todas as bases de dados num directório, não está disponível um botão para validar os valores da coluna.

Para pesquisa de base de dados e de directório, é possível configurar o pesquisador para ignorar colunas ocultas quando configura as vistas a serem pesquisadas.

Segurança ao nível do documento

Se a segurança de colecções estiver activada e um servidor que pretende pesquisar utilizar o protocolo NRPC, tem de configurar um Servidor Fidedigno Lotus Domino no servidor do pesquisador. O Servidor Fidedigno é utilizado para garantir controlos de acesso ao nível do documento. Antes de disponibilizar uma colecção para os utilizadores, complete as tarefas seguintes:

- Configure os Servidores Fidedignos Lotus Domino para validar credenciais de utilizador.
- Active a segurança do início de sessão da aplicação. Se utilizar o servidor de aplicação da Web incorporado, abra a vista Segurança e configure as definições de início de sessão da aplicação. Se utilizar o WebSphere Application Server, active a segurança global e configure a aplicação para aplicar a segurança de início de sessão.

Activar a segurança de início de sessão da aplicação assegura que as credenciais são validadas quando os utilizadores iniciam sessão em colecções de consulta. Os servidores de procura utilizam as credenciais para verificar se cada utilizador tem autoridade para aceder a documentos Lotus Domino.

O pesquisador Notes suporta três níveis de listas de controlo de acesso (ACLs): documento, base de dados e servidor. Em acréscimo, o pesquisador apenas recolhe dados de ACLs positivas. Os utilizadores incluídos na ACL positiva e que também estão na ACL negativa, em cada nível, podem consultar documentos que não estão totalmente autorizados a ver. Para segurança e desempenho máximos:

- Inclua o utilizador na ACL em cada nível (documento, base de dados e servidor).
- Configure o pesquisador para indexar documento, base de dados e ACLs do servidor (active a pré-filtragem de resultados).
- Configure o pesquisador para validar credenciais de utilizador durante o processamento de consulta (active a pós-filtragem de resultados).

Pode configurar o pesquisador apenas para indexar a base de dados e ACLs de servidor. Se seleccionar esta opção, pode melhorar o desempenho ao criar um ficheiro de configuração personalizado para indexar as ACLs de nível de

documento. Para obter detalhes, consulte o procedimento para indexação de segurança de nível de documento Lotus Notes para melhorar o desempenho da procura.

No Lotus Domino, as permissões para documentos e vistas são definidas como listas de controlo de acesso (ACLs) diferentes. Por exemplo, quando utilizar o Lotus Notes, não é possível a um utilizador aceder a documentos numa base de dados caso o utilizador não esteja incluído nas ACLs de todas as vistas na base de dados, ainda que uma ACL de documento inclua o utilizador. O pesquisador Notes, no entanto, apenas se refere a ACLs de documentos. Se um utilizador estiver incluído na ACL de documento, o utilizador pode aceder ao documento mesmo se o utilizador não estiver incluído na ACL de nível de visualização. Para impedir os utilizadores de acederem a documentos específicos, o administrador do Lotus Domino tem de configurar ACLs ao nível do documento, não ao nível de visualização.

Quando existir um ou mais pesquisadores Notes configurados para suportar dois níveis de indexação de ACL, outros pesquisadores Notes que estejam configurados para suportar a indexação de ACL nos três níveis (servidor, base de dados e documento) podem processar apenas a indexação de ACL de dois níveis. Para suportar consultas que tenham como alvo os três níveis de ACLs, todos os pesquisadores Notes têm de ser configurados para indexar os três níveis de ACLs.

Descrição geral da configuração

É possível utilizar o pesquisadorNotes para pesquisar um número de bases de dados padrão Lotus Notes (ficheiros .nsf). Ao criar o pesquisador, selecciona as bases de dados ou directórios a pesquisar a partir de um único servidorLotus Notes. Posteriormente, ao editar o espaço de pesquisa, é possível adicionar documentos de outro servidorLotus Notes que pretenda incluir no mesmo espaço de pesquisa.

Ao criar o pesquisador, um assistente ajudá-lo-á nessas tarefas:

- Especifique as propriedades que controlam o modo como o pesquisador opera e utiliza os recursos do sistema. As propriedades do pesquisador controlam o modo como este último pesquisa todos os documentos no espaço de pesquisa.
- Identifique o nome do sistema central, a porta e o protocolo de comunicação do servidor Lotus Notes.
- Se o servidor estiver configurado para utilizar o protocolo DIIOP, pode especificar a forma como o pesquisador se pode ligar aos objectos do Domino. Por exemplo, pode especificar opções para utilizar HTTPS ou Secure Sockets Layer (SSL) para codificar comunicações.
- Seleccione as bases de dados ou os directórios que pretende pesquisar.
- Defina um agendamento para a pesquisa das bases de dados ou dos directórios.
- Seleccione os documentos que pretende pesquisar. Pode pesquisar todos os documentos num directório, todos os documentos numa base de dados, ou os documentos de vistas e pastas seleccionadas de uma base de dados.
- Especifique opções para pesquisar os valores da coluna nas vistas da base de dados.
- Especifique as opções para fazer com que os campos em várias bases de dados, vistas e pastas possam ser procurados. Por exemplo, pode excluir determinados campos do espaço de pesquisa e especifique as opções para pesquisar ligações.
- Configure as opções de segurança ao nível do documento. Se a segurança tiver sido activada quando a colecção foi criada, o pesquisador pode associar dados

de segurança aos documentos existentes no índice. Esses dados permitem às aplicações fazer cumprir os controlos de acesso baseados nas listas de controlo de acesso armazenadas ou nos tokens de segurança.

Pode seleccionar uma opção para validar as credenciais do utilizador no momento em que um utilizador permitir uma consulta. Nesse caso, em vez de comparar as credenciais do utilizador com os dados de segurança indexados, o sistema compara as credenciais com as listas de controlo de acesso actuais que são actualizadas pela origem de dados original.

Referências relacionadas:

Configurar suporte para autenticação de início de sessão único (SSO)

Configurar o servidor do pesquisador no UNIX para pesquisar origens do Lotus Domino

Se instalar o IBM Content Analytics with Enterprise Search num computador que esteja a executar o IBM AIX ou oLinux, e pretende pesquisar servidores que utilizem o protocolo de chamada de procedimento remoto do Notes (NRPC, Notes Remote Procedure Call), tem de executar um script de configuração no servidor do pesquisador. O script permite aos pesquisadores Notes, Quickr for Domino, e Domino Document Manager comunicar com os servidores de base de dados.

Um Servidor Domino não consegue executar ao mesmo tempo, no mesmo computador, com um pesquisador do Notes, Quickr for Domino, ou Domino Document Manager que esteja configurado para utilizar o protocolo NRPC. Se tentar iniciar um destes pesquisadores enquanto o Servidor Domino estiver em execução, ocorre um erro e o pesquisador pára.

Os pesquisadores que utilizam o protocolo NRPC utilizam as bibliotecas Domino como um cliente. Pode instalar estas bibliotecas instalando o Lotus Domino Server no servidor do pesquisador. Para assegurar que os pesquisadores possam trabalhar com as bibliotecas do Domino, tem de executar um script de configuração que o IBM Content Analytics with Enterprise Search fornece no servidor do pesquisador após a instalação das bibliotecas do Domino.

Para configurar o servidor do pesquisador de forma a que possa pesquisar nos servidores do Notes, do Quickr para Domino, e doDomino Document Manager:

- 1. Crie o utilizador servidor e o grupo notas no servidor do pesquisador:
 - a. Inicie sessão como utilizador root:
 - su root
 - b. Adicione um utilizador:
 - useradd server
 - c. Adicione uma palavra-passe a este utilizador:
 - passwd server

Ser-lhe-á solicitado para alterar a palavra-passe.

- 2. Instale o Servidor Domino do Lotus no servidor do pesquisador:
 - a. Insira o CD de instalação do Servidor Domino. (Se não tiver um CD, pode descarregar a imagem.)
 - b. Mude para a pasta do respectivo sistema operativo.

AIX: cd /mnt/cdrom/aix Linux: cd /mnt/cdrom/linux

c. Inicie o programa de instalação:

./instalação

d. Responda aos pedidos e aceite os valores predefinidos ou especifique as definições de instalação preferidas (tais como os caminhos para o directório da instalação e o directório dos dados).

Consulte a documentação do Domino se tiver necessidade de assistência quando estiver a instalar o Servidor Domino.

- e. Execute o programa de configuração do servidor Domino, o qual regra geral se inicia da primeira vez que iniciar o servidor Domino. Consulte a documentação do servidor Domino ou peça ajuda ao administrador, se tiver necessidade de assistência ao configurar o Servidor Domino.
- f. Certifique-se de que o ID de administrador predefinido IBM Content Analytics with Enterprise Search tem permissão para aceder ao directório home/server.
- **3**. No servidor do pesquisador, execute o script de configuração fornecido pelo pesquisador do IBM Content Analytics with Enterprise Search:
 - a. Inicie sessão como administrador predefinido do IBM Content Analytics with Enterprise Search.
 - b. Inicie o seguinte script instalado no directório \$ES_INSTALL_ROOT/bin: escrnote.sh
 - c. Responda aos pedidos de informação:
 - No pedido seguinte, responda S se o Servidor Domino for instalado no directório predefinido, e N se não for:

Foi encontrado o caminho do directório do Lotus Notes /opt/lotus/notes/latest/linux. Este caminho do directório do Lotus Notes está correcto?

O caminho predefinido para o AIX é /opt/lotus/notes/latest/ibmpow. O caminho predefinido para o Linux é /opt/lotus/notes/latest/linux.

• Se o Servidor Domino não estiver instalado no directório predefinido do servidor do pesquisador, especifique onde foi instalado o Domino em resposta ao pedido seguinte:

Introduza o caminho para o directório do Lotus Notes

Por exemplo, num computador com o sistema operativo Linux pode especificar/opt/lotus/notes/latest/linux.

• Por exemplo, no pedido seguinte, responda S se o directório dos dados do Domino Server for instalado no directório predefinido, e N se não for:

Foi encontrado o caminho do directório dos dados do Lotus Notes /local/notesdata.

É este o caminho correcto do directório do Lotus Notes?

O caminho predefinido é /local/notesdata.

• Se o directório dos dados do Domino Server não for implementado na localização predefinida do servidor do pesquisador, especifique o caminho dos dados do Domino em resposta ao seguinte pedido:

Introduza o caminho do directório dos dados do Lotus Notes.

4. No servidor do pesquisador, reinicie o sistema IBM Content Analytics with Enterprise Search:

esadmin system stopall esadmin system startall

Configurar o servidor do pesquisador no Windows para pesquisar origens do Lotus Domino

Se instalar o IBM Content Analytics with Enterprise Search num computador com o sistema operativo Microsoft Windows, e pretender pesquisar servidores que utilizem o protocolo de chamada de procedimento remoto do Notes (NRPC, Notes Remote Procedure Call), tem de executar um script de configuração no servidor do pesquisador. O script permite aos pesquisadores Notes, Quickr for Domino, e Domino Document Manager comunicar com os servidores de base de dados.

O Servidor Lotus Domino e o cliente do Lotus Notes não conseguem estar a executar ao mesmo tempo no mesmo computador com um pesquisador do Notes, Quickr for Domino, ou Domino Document Manager que esteja configurado para utilizar o protocolo NRPC. Se tentar iniciar um destes pesquisadores enquanto o Servidor Domino estiver em execução, ocorre um erro e o pesquisador pára.

Os pesquisadores que utilizam o protocolo NRPC utilizam as bibliotecas do Lotus Domino como um cliente. Pode instalar estas bibliotecas instalando o Lotus Domino Server no servidor do pesquisador. Para assegurar que os pesquisadores possam trabalhar com as bibliotecas do Domino, tem de executar um script de configuração que o IBM Content Analytics with Enterprise Search fornece no servidor do pesquisador após a instalação das bibliotecas do Domino.

Este procedimento dá instruções para instalar um cliente Notes. Se tenciona filtrar resultados de procura para cumprir a segurança ao nível de documento, tem de instalar um servidor Domino no servidor do pesquisador, e não no cliente Notes.

Para configurar o servidor do pesquisador de forma a que possa pesquisar nos servidores do Notes, do Quickr para Domino, e doDomino Document Manager:

- 1. No servidor do pesquisador, inicie sessão com um ID de utilizador que seja membro do grupo de Administradores Windows. Certifique-se de que o ID de utilizador tem autoridade para instalar o Lotus Notes.
- 2. Instale o Lotus Notes:
 - a. Insira o CD de instalação do Servidor Domino. (Se não tiver um CD, pode descarregar a imagem.)
 - b. Inicie o programa de instalação: setup.exe
 - c. Responda aos pedidos e aceite os valores predefinidos ou especifique as definições de instalação preferidas (tais como os caminhos para o directório da instalação e o directório dos dados).

Consulte a documentação do Lotus Domino se necessitar de assistência.

- d. Certifique-se de que o ID de administrador predefinido IBM Content Analytics with Enterprise Search tem permissão para aceder ao directório de dados Domino.
- **3.** No servidor do pesquisador, execute o script de configuração fornecido pelo pesquisador de IBM Content Analytics with Enterprise Search:
 - a. Inicie sessão como administrador predefinido do IBM Content Analytics with Enterprise Search. Num servidor com o sistema operativo Windows 2008, tem de iniciar sessão como Administrador do Windows.
 - b. Inicie o seguinte script, que está instalado no directório %ES_INSTALL_ROOT%\bin:

escrnote.vbs

c. Responda aos pedidos de informação:

• No pedido seguinte, responda S se o Lotus Notes estiver instalado no directório predefinido, e N se não estiver:

Foi encontrado o caminho do directório do Lotus Notes c:\lotus\notes. Este caminho do directório do Lotus Notes está correcto?

O caminho de instalação comum num computador Windows éc:\lotus\notes ou c:\lotus\domino.

• Se o Lotus Notes não for instalado no directório predefinido no servidor do pesquisador, especifique onde foi instalado o Lotus Notes foi instalado em resposta ao pedido seguinte:

Introduza o caminho para o directório do Lotus Notes

• No pedido seguinte, responda S se o directório de dados do Lotus Notes for instalado na localização predefinida e N se não for:

Foi encontrado o caminho do directório de dados do Lotus Notes c:\lotus\notes\data.

É este o caminho correcto do directório do Lotus Notes?

O caminho comum num computador Windows é c:\lotus\notes\data ou c:\lotus\domino\data.

 Se o directório de dados do Lotus Notes não for implementado na localização predefinida no servidor do pesquisador, especifique o caminho do directório de dados em resposta ao seguinte pedido:

Introduza o caminho do directório dos dados do Lotus Notes.

- 4. Para evitar erros que possam ocorrer quando um utilizador inicia sessão numa aplicação, depois de o utilizador que iniciou o IBM Content Analytics with Enterprise Search terminar a sessão, defina a variável de sistema NOTESNTSERVICE=1 no Windows:
 - a. Abra o Painel de Controlo e depois as definições de Sistema.
 - b. Seleccione o separador Avançadas e faça clique no botão **Variável de Ambiente**.
 - c. Seleccione **Variável de sistema**, adicione uma variável de ambiente nova chamada NOTESNTSERVICE que tenha o valor 1, e guarde as alterações.
- 5. No servidor do pesquisador, reinicie o sistema IBM Content Analytics with Enterprise Search:

```
esadmin system stopall
esadmin system startall
```

Configurar servidores que utilizam o protocolo DIIOP

Para pesquisar os servidores que utilizam o protocolo Domino Internet Inter-ORB Protocol (DIIOP), tem de configurar o servidor de modo que os pesquisadores Notes, Quickr for Domino, e Domino Document Manager possam utilizar o protocolo.

O servidor que pretende para efectuar a pesquisa tem de estar a executar as tarefas DIIOP e HTTP.

Para configurar os servidores que utilizam o protocolo DIIOP:

- 1. Configure o documento do servidor:
 - a. Abra o documento servidor no Lotus Notes, Lotus Quickr for Domino, ou o servidor do Domino Document Manager que pretende pesquisar. Este documento está armazenado no directório Domino.
 - b. Na página Configuração, expanda a secção do servidor.

- c. Na página Segurança, na área **Restrições de programabilidade**, especifique as restrições de segurança apropriadas ao ambiente nos campos seguintes:
 - Executar agentes Lotus Script/Java restritos
 - Executar Java/Javascript/COM restrito
 - Executar Java/Javascript/COM não restrito

Por exemplo, pode especificar um asterisco (*) para permitir o acesso não restrito por agentes Lotus Script/Java, e especifique os nomes de utilizador que são registados no Directório Domino como restrições Java/Javascript/COM.

Importante: O pesquisador que configurar para pesquisar este servidor com o protocolo DIIOP tem de poder utilizar os nomes dos utilizadores que especifica nestes campos.

- d. Abra a página Protocolo Internet, depois abra a página HTTP, e defina a opção **Permitir aos clientes HTTP navegar na base de dados** para **Sim**.
- 2. Configure o documento do utilizador:
 - a. Abra o documento utilizador no servidor que pretende pesquisar. Este documento está armazenado no directório Domino.
 - b. No página Básicas, no campo **Palavra-passe de Internet**, especifique uma palavra-passe.

Quando utilizar a consola de administração para configurar as opções de pesquisa deste servidor, especifique este ID de utilizador e a palavra-passe na página em que identifica o servidor a pesquisar. O pesquisador utiliza estas credenciais para aceder ao servidor.

3. Reinicie a tarefa DIIOP no servidor.

Configurar a porta de conclusão E/S no AIX para pesquisar origens no Lotus Domino

Antes de poder utilizar os pesquisadores Notes, Quickr for Domino, ou Domino Document Manager num sistema IBM AIX, terá de instalar os módulos da porta de conclusão E/S (IOCP) e configurá-los para utilização pelo pesquisador.

Sem o módulo IOCP, o processo de identificação falhará quando tentar criar um pesquisador. É apresentada a seguinte mensagem de erro:

FFQM0105E Erro recebido do servidor -

Mensagem: FFQG0024E Foi detectada uma excepção inesperada: identificar

A mensagem seguinte, que inclui o erro ENOEXEC está gravada no ficheiro \$ES_NODE_ROOT/logs/system_*yyyymmdd*.log. (Algum deste texto da mensagem é dividido ao longo de múltiplas linhas para melhorar a leitura.)

5/20/05 18:08:52.423 JST [Error] [ES_ERR_EXCEPTION_DEFAULT_MESSAGE] [] [identificação] iies10.yamato.ibm.com:0:2108088751:control:ComponentDiscoveryW.java: com.ibm.es.control.discovery.server.ComponentDiscoveryW.discover:86 FFQ00277E Foi detectada uma excepção com o detalhe 'java.lang.UnsatisfiedLinkError: /opt/lotus/notes/65010/ibmpow/liblsxbe_r.a: load ENOEXEC on shared library(s) /opt/lotus/notes/latest/ibmpow/libnotes_r.a' e um rastreio de pilha de 'java.lang.UnsatisfiedLinkError: /opt/lotus/notes/65010/ibmpow/liblsxbe_r.a: load ENOEXEC on shared library(s) /opt/lotus/notes/latest/ibmpow/libnotes_r.a at java.lang.ClassLoader\$NativeLibrary.load(Native Method) at java.lang.ClassLoader.loadLibrary0(ClassLoader.java:2120) at java.lang.Runtime.loadLibrary0(Runtime.java:824) at java.lang.System.loadLibrary(System.java:908)

- at lotus.domino.NotesThread.load(NotesThread.java:306)
- at lotus.domino.NotesThread.checkLoaded(NotesThread.java:327)
- at lotus.domino.NotesThread.sinitThread(NotesThread.java:181)
- at com.ibm.es.crawler.discovery.notes.NotesLibrary\$NotesOperation.discover
 (Unknown Source)
- at com.ibm.es.crawler.discovery.api.DiscoveryAPI.discover(Unknown Source)
- at com.ibm.es.control.discovery.server.ComponentDiscoveryW.discover (ComponentDiscoveryW.java:72)
- at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
- at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:85)
- at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:58)
- at sun.reflect.DelegatingMethodAccessorImpl.invoke
- (DelegatingMethodAccessorImpl.java:60)
 at java.lang.reflect.Method.invoke(Method.java:391)
- at com.ibm.es.ccl.sessionwrapper.CallThread.run(CallThread.java:77)

Para instalar o módulo IOCP e assegurar que o mesmo esteja instalado correctamente no servidor do pesquisador:

1. Instalar o módulo IOCP (bos.iocp.rte) a partir do CD do produto AIX no servidor do pesquisador.

Depois de instalar o módulo IOCP, e antes de criar um pesquisador do Notes, Quickr for Domino, ou Domino Document Manager, aplique uma correcção de software no módulo. Para obter mais instruções, consulte *Resolver erros da porta de conclusão E/S*.

2. Introduza o comando seguinte para assegurar que o módulo IOCP está instalado no servidor do pesquisador:

\$ lslpp -l bos.iocp.rte

A saída de dados do comando lslpp devia ser semelhante ao seguinte exemplo: Descrição do Estado do Nível do Conjunto de Ficheiros

Caminho: /usr/lib/objrepos bos.iocp.rte 5.2.0.10 COMMITTED I/O Completion Ports API

Caminho: /etc/objrepos

bos.iocp.rte 5.2.0.10 COMMITTED I/O Completion Ports API

3. Introduza o comando seguinte para assegurar que o estado da porta IOCP está **Disponível**:

\$ lsdev -Cc iocp

A saída de dados do comando lsdev devia corresponder ao seguinte exemplo: Portas de Conclusão E/S Disponíveis iocp0

- 4. Se o estado da porta IOCP estiver **Definido**, altere o estado para **Disponível**:
 - a. Inicie sessão no servidor do pesquisador como root e emita o seguinte comando:

smit iocp

- b. Seleccione Alterar / Mostrar Características de Portas de Conclusão de E/S e altere STATE a configurar no reinício do sistema a partir de Definido para Disponível.
- c. Reinicialize o servidor do pesquisador.
- d. Introduza o comando 1sdev de novo e confirme que o estado da porta IOCP foi alterado para **Disponível**.

Referências relacionadas:

Resolver erros de porta de conclusão E/S

Melhorar a forma como os documentos Notes são novamente pesquisados

Para melhorar o desempenho dos pesquisadores, pode alterar a forma como os pesquisadores do Notes que utilizam o protocolo a chamada remota de procedimento do Notes (NRPC) pesquisam novamente os documentos.

Por predefinição, o pesquisador verifica a data modificada de todos os documentos ao determinar quais os documentos necessários a pesquisar novamente. Em alguns casos, esta verificação resulta num tempo de nova pesquisa que é não é muito menor do que o tempo que demora a efectuar uma pesquisa completa. Como alternativa, configure o pesquisador para obter uma lista dos documentos que é necessário recuperar nas etapas inicias da nova pesquisa. Esta alteração reduz o tempo que demora a determina quais os documentos que foram modificados e que precisam ser novamente pesquisados.

Restrições:

- Esta função funciona apenas para um pesquisador do Notes que utilize o protocolo de NRPC.
- Esta função funciona apenas para pesquisadores que estejam configurados para fazer cumprir a procura. Antes de actualizar a configuração do pesquisador para utilizar esta função, confirme que a procura protegida através de um pesquisador NRPC do Notes está a funcionar correctamente.
- Neste modo, o pesquisador refere-se aos canhotos de eliminação na base de dados. Os canhotos são removidos em cada intervalo de supressão, um valor que é configurado para a base de dados. Para capturar todos os documentos eliminados desta forma, terá de pesquisar mais frequentemente do que o intervalo de supressão.

Para altera a forma como o pesquisador NRPC do Notes determina quais os documentos que é necessário pesquisar novamente:

- 1. Inicie sessão como administrador predefinido do IBM Content Analytics with Enterprise Search no servidor do pesquisador.
- 2. Com um editor de texto capaz de ler UTF-8, crie um ficheiro de configuração XML denominado notescrawler_ext.xml e adicione o seguinte conteúdo:

```
<?xml version="1.0" encoding="UTF-8"?>
<ExtendedProperties>
<SetAttribute XPath="/Crawler/DefaultSetting" Name="DetectDelete">True
</SetAttribute>
<SetAttribute XPath="/Crawler/DataSources/Server/Target"
Name="DetectDelete">True
</SetAttribute>
</SetAttribute>
</SetAttribute>
</ExtendedProperties>
```

- **3**. Abra a consola de administração e pare o pesquisador do Notes que pretende configurar.
- Copie o ficheiro notescrawler_ext.xml que criou no directório ES_NODE_ROOT/master_config/crawler_session/ para o pesquisador do Notes. O pesquisador utiliza este ficheiro de extensão juntamente com o ficheiro notescrawler.xml.
- 5. Reinicie o pesquisador do Notes.

Sugestões para pesquisar campos nas bases de dados do Lotus Domino

Revejas as directrizes para pesquisar campos nas bases de dados do Lotus Domino antes de configurar um pesquisador Notes.

- As bases de dados do Notes baseadas em modelos padrão (tais como uma base de dados de discussão) são o melhor tipo de base de dados para efectuar a pesquisa.
- O pesquisador do Notes aplica as seguintes regras de mapeamento de campo:
 - Os nomes principais do campo dos modelos padrão do Domino são inicialmente registados.
 - Os valores de campos do Notes que são especificados na tabela de regras de correlação são utilizados como resumos do documento nos resultados da procura.
 - Os valores dos campos do Notes que não são especificados na tabela de regras de correlação não são utilizados nos resumos do documento.
 - Os valores dos campos Notes que são correlacionados com o campo Título são utilizados como o título do documento nos resultados da procura.
 - Os campos na tabela seguinte são correlacionados com campos de índice por predefinição:

Nome do campo de índice
Título
Título
Título
Corpo
Corpo
Categorias
Categorias

Tabela 13. Regras de mapeamento de campos por predefinição

- O pesquisador do Notes pode pesquisar todos os tipos de campos excepto para calculado para campos de visualização.
- O texto estático e as imagens que se encontram num formulário do Notes não são pesquisados.
- Configure o pesquisador para pesquisar todos os campos. Se, primeiro, maximizar os dados do campo a serem pesquisados, é possível, depois, especificar as excepções para excluir os poucos campos que não pretende incluir. Para maximizar a pesquisa de campos desnecessários, exclua os campos que não estão correlacionados com os campos de índice.

Sugestões para pesquisar anexos para documentos do Notes

Se a sua colecção incluir anexos para documentos do Lotus Notes, e se configurar os âmbitos para que apresentem apenas o URL para o anexo nos resultados, o título que é exibido para o anexo pode ser incorrecto ou inconsistente.

Por exemplo, o título pode conter o valor Subtítulo Título, a primeira linha do ficheiro PDF ou Word anexado, ou o nome do ficheiro. Utilize uma das seguintes abordagens para garantir que o título correcto do anexo é apresentado.

Personalize a aplicação de procura empresarial

É possível personalizar o código fonte da aplicação de procura de amostra facultada. Por exemplo, na consola administrativa, correlacione o campo F_Title com um novo campo de índice, como, por exemplo, alternativeTitle. De seguida, personalize a aplicação para que mostre sempre alternativeTitle como o título do documento em que os utilizadores podem fazer clique nos resultados de procura.

Desactive a extracção da propriedade do documento de título

- 1. Na consola administrativa, certifique-se de que o campo que contém o título que deve ser exibido está correlacionado com um campo de índice de título. Por exemplo, correlacione o campo F_Title com o campo de índice de título.
- Na consola administrativa, pare os serviços de análise e indexação para a colecção.
- Inicie sessão como o administrador predefinido de IBM Content Analytics with Enterprise Search no servidor do controlador e edite o ficheiro parser_config.xml no directório ES_NODE_ROOT/master_config/ ID_Colecção.indexservice.
- 4. Adicione o parâmetro field.name.title sob o elemento ParserFactoryClass:

```
<Parser>
<ParserName>stellent</ParserName>
<ParserFactoryClass>com.ibm.es.oze.indexservice.internal.ParserFactory$OutsideIn
</ParserFactoryClass>
<Parameter Name="field.name.title"></Parameter>
</Parser>
```

- 5. Na consola administrativa, reinicie os serviços de análise e indexação para a colecção.
- 6. Volte a pesquisar os documentos para que possam voltar a ser analisados e indexados.

Pesquisadores do Quickr for Domino

Se utilizar o Lotus Quickr o Domino, é possível configurar o pesquisador Quickr for Domino para adicionar documentos que estão armazenados nas bibliotecas a uma colecção.

Se utilizar o Lotus Quickr for WebSphere Portal, utilize o pesquisador Lista de valores geradores para adicionar documentos que estão armazenados nas bibliotecas Lotus Quickr a uma colecção.

Credenciais de ligação do pesquisador

Quando cria o pesquisador, pode especificar as credenciais que permitem que o pesquisador se ligue às fontes a pesquisar. Também pode configurar as credenciais de ligação quando especifica as definições gerais de segurança do sistema. Se utilizar a última abordagem, vários pesquisadores e outros componentes do sistema podem utilizar as mesmas credenciais. Por exemplo, os servidores de procura podem utilizar as credenciais para determinar se um utilizador possui autorização para aceder ao conteúdo.

Configuração do servidor do pesquisador

Se um servidor que tenciona pesquisar utiliza o protocolo de chamada de procedimento remoto (NRPC) Notes, tem de executar o script de configuração **escrnote** no servidor do pesquisador. Esse script fornecido com IBM Content

Analytics with Enterprise Search, activa o pesquisador do Quickr for Domino para comunicar com os servidores que utilizam o NRPC.

Se um servidor que pretende pesquisar utilizar o Domino Internet Inter-ORB Protocol (DIIOP), não necessita de executar um script de configuração no servidor do pesquisador. No entanto, tem de configurar o servidor destino de modo que o pesquisador do Quickr for Domino possa aceder ao servidor.

Se um servidor que pretende pesquisar utilizar um servidor de Lightweight Directory Access Protocol (LDAP), o servidor destino terá de ser configurado para utilizar o protocolo DIIOP (o pesquisador de Quickr for Domino não pode utilizar o protocolo NRPC para pesquisar os dados de LDAP). Também tem de configurar uma base de dados de Assistência do Directório e configurar o servidor de destino para utilizar o servidor LDAP como um servidor Domino secundário.

Importante: Se o servidor destino utilizar o DIIOP e o utilizador configurar o pesquisador para utilizar o HTTPS ou DIIOP através de SSL, para que as transmissões entre o pesquisador e o servidor sejam codificadas, é necessário copiar o ficheiro TrustedCerts.class (por exemplo, c:\certs ou /data/certs) do servidor destino para o servidor do pesquisador. Numa configuração de vários servidores, também deve copiar o ficheiro TrustedCerts.class para os servidores de procura. Tem de assegurar que ficheiro se encontra na mesma localização no servidor do pesquisador e nos servidores de procura. O utilizador especifica o caminho do directório para o ficheiro TrustedCerts.class quando configura o pesquisador.

Se IBM Content Analytics with Enterprise Search tiver sido instalado em um sistema da IBM AIX, tem de garantir que o módulo E/S Porta de Conclusão esteja instalado e disponível no servidor do pesquisador.

Antes de utilizar a consola de administração para configurar um pesquisador do Quickr for Domino, execute as tarefas seguintes apropriadas para o ambiente:

- "Configurar o servidor do pesquisador no UNIX para pesquisar origens do Lotus Domino" na página 114.
- "Configurar o servidor do pesquisador no Windows para pesquisar origens do Lotus Domino" na página 116.
- "Configurar servidores que utilizam o protocolo DIIOP" na página 117.
- "Configurar o Quickr para o servidor do Domino para utilizar a segurança Utilizador Local" na página 407.
- "Configurar Assistência de Directório no Quickr para servidor do Domino" na página 408.
- "Configurar a porta de conclusão E/S no AIX para pesquisar origens no Lotus Domino" na página 118.

Segurança ao nível do documento

Se a segurança de colecções estiver activada e um servidor que pretende pesquisar utilizar o protocolo NRPC, tem de configurar um Servidor Fidedigno Lotus Domino no servidor do pesquisador. O Servidor Fidedigno é utilizado para garantir controlos de acesso ao nível do documento. Antes de disponibilizar uma colecção para os utilizadores, complete as tarefas seguintes:

• Configure os Servidores Fidedignos Lotus Domino para validar credenciais de utilizador.

 Active a segurança do início de sessão da aplicação. Se utilizar o servidor de aplicação da Web incorporado, abra a vista Segurança e configure as definições de início de sessão da aplicação. Se utilizar o WebSphere Application Server, active a segurança global e configure a aplicação para aplicar a segurança de início de sessão.

Activar a segurança de início de sessão da aplicação assegura que as credenciais são validadas quando os utilizadores iniciam sessão em colecções de consulta. Os servidores de procura utilizam as credenciais para verificar se cada utilizador tem autoridade para aceder a documentos Lotus Domino.

Pesquisa de anexos

No Lotus Quickr for Domino, é possível importar e publicar documentos do Microsoft Office (as opções que selecciona durante a importação incluem a Folha Importada, a Folha do Microsoft Word, Folha de Cálculo do Microsoft, o Diapositivo do Microsoft PowerPoint e Múltiplas Páginas Importadas). O pesquisador do Quickr for Domino pode pesquisar estes tipos de documentos importados como anexos apenas sob as seguintes condições:

- Se o servidor utilizar o protocolo DIIOP.
- Se activar a pesquisa de anexos quando configurar opções de pesquisa para o pesquisador.
- Se configurar o pesquisador para pesquisar o campo "\$FILE" ou todos os campos.

Descrição geral da configuração

É possível utilizar o pesquisador Quickr for Domino para pesquisar qualquer número de Quickr para locais Domino. Quando cria o pesquisador, selecciona os locais a pesquisar a partir de um único Quickr para o servidor Domino. Posteriormente, quando edita o espaço de pesquisa, é possível adicionar documentos de outro servidor Quickr for Domino que pretende incluir no mesmo espaço de pesquisa. Ao criar ou editar o pesquisador, é possível especificar se pretende pesquisar todos os locais no servidor ou apenas os locais que especifica, ou se pretende pesquisar todas as salas nos locais a pesquisar ou apenas as salas que especifica.

Restrição: Quando especifica um ID de utilizador para o pesquisador a utilizar, certifique-se de especificar um ID com autoridade suficiente para aceder a todos os Quickr para locais Domino no domínio Domino. É possível fazer isto ao configurar um servidor do Quickr for Domino e atribuir um ID de utilizador ao grupo reservado designado QuickPlaceAdministratorsSUGroup. Note que o nome do grupo não contém espaços incorporados e é sensível a maiúsculas e minúsculas.

Ao criar o pesquisador, um assistente ajudá-lo-á nessas tarefas:

- Especifique as propriedades que controlam o modo como o pesquisador opera e utiliza os recursos do sistema. As propriedades do pesquisador controlam o modo como este último pesquisa todos os documentos no espaço de pesquisa.
- · Identifique o Quickr para o servidor Domino e o protocolo de comunicação.
- Se o servidor estiver configurado para utilizar o protocolo DIIOP, pode especificar a forma como o pesquisador se pode ligar aos objectos do Domino. Por exemplo, pode especificar opções para utilizar HTTPS ou Secure Sockets Layer (SSL) para codificar comunicações.

- Especifique a informação acerca do directório do utilizador que é utilizada com o servidor (o pesquisador necessita dessa informação de modo a fazerem-se cumprir os controlos de acesso quando os utilizadores procurarem a colecção).
- Seleccione os locais que pretende pesquisar.
- Defina um agendamento para a pesquisa de locais.
- Seleccione os documentos que pretende pesquisar. O pesquisador pode pesquisar todas as salas num local, ou pesquisa apenas os documentos que estão nas salas que selecciona.
- Especifique as opções para fazer com que os campos possam ser procurados em vários locais e em várias salas. Por exemplo, pode excluir determinados campos do espaço de pesquisa e especifique as opções para pesquisar ligações.
- Configure as opções de segurança ao nível do documento. Se a segurança tiver sido activada quando a colecção foi criada, o pesquisador pode associar dados de segurança aos documentos existentes no índice. Esses dados permitem às aplicações fazer cumprir os controlos de acesso baseados nas listas de controlo de acesso armazenadas ou nos tokens de segurança.

Pode seleccionar uma opção para validar as credenciais do utilizador no momento em que um utilizador permitir uma consulta. Nesse caso, em vez de comparar as credenciais do utilizador com os dados de segurança indexados, o sistema compara as credenciais com as listas de controlo de acesso actuais que são actualizadas pela origem de dados original.

Referências relacionadas:

Configurar suporte para autenticação de início de sessão único (SSO)

Pesquisadores Lista de valores geradores

Configure um pesquisador Lista de valores geradores caso pretenda pesquisar documentos em servidores WebSphere Portal, documentos em Quickr para bibliotecas WebSphere Portal ou documentos IBM Connections.

É possível configurar um pesquisador Lista de valores geradores para o conteúdo da colecção a partir das seguintes origens. Para informações actuais sobre as versões do produto que podem ser pesquisadas, incluindo todos os iFixes necessários, consulte as informações relacionadas sobre as origens de dados suportadas.

- IBM Connections (IBM Connections tinha a denominação anterior de Lotus Connections.) A procura protegida não é suportada para IBM Connections Versão 2.0.
- IBM Web Content Manager
- IBM WebSphere Portal
- Lotus Quickr para WebSphere Portal

Directrizes para IBM Connections

Se utilizar o IBM Connections, é possível utilizar o pesquisador Lista de valores geradores para adicionar actividades de rede social, perfis, blogues, comunidades, marcadores de canto dobrado e outro conteúdo baseado na Web a uma colecção. Os tipos de conteúdo que pode adicionar dependem da versão de IBM Connections que utiliza.

IBM Connections necessita WebSphere Application Server. Para configurar suporte para pesquisar conteúdo, tem de configurar um utilizador de administrador de procura para o IBM Connections no WebSphere Application Server antes de criar um pesquisador Lista de valores geradores.

Directrizes para o Lotus Quickr para WebSphere Portal

Uma biblioteca Quickr é um contentor para ficheiros de documentos. O pesquisador de Lista de valores geradores não suporta pesquisar conteúdo com base na Web, tais como wikis e blogues.

Quando configura o pesquisador, especifica o servidor a pesquisar. O pesquisador pesquisa assim todos os documentos nas bibliotecas do Lotus Quickr nesse servidor.

Os documentos a serem pesquisados têm de ser acessíveis pelo mesmo ID e palavra-passe de administrador do Quickr. Para pesquisar sítios que utilizem credenciais diferentes, tem de configurar um pesquisador do Lista de valores geradores separada.

Uma cadeia de agentes do utilizador identifica que navegador ou mecanismo está a aceder a um servidor. Quando pesquisar documentos Quickr, o pesquisador Lista de valores geradores utiliza a cadeia de agente do utilizador OmniFind SeedlistCrawler/1.0.

Se utilizar o Quickr para Domino, utilize o pesquisadorQuickr for Domino para adicionar documentos Quickr a uma colecção.

Credenciais de ligação do pesquisador

Quando cria o pesquisador, pode especificar as credenciais que permitem que o pesquisador se ligue às fontes a pesquisar. Também pode configurar as credenciais de ligação quando especifica as definições gerais de segurança do sistema. Se utilizar a última abordagem, vários pesquisadores e outros componentes do sistema podem utilizar as mesmas credenciais. Por exemplo, os servidores de procura podem utilizar as credenciais para determinar se um utilizador possui autorização para aceder ao conteúdo.

Pesquisar conteúdo

O pesquisador Lista de valores geradores recolhe todos os documentos que estão incluídos na lista de valores geradores. O pesquisador recolhe todos os campos numa entrada da lista de valores geradores para que os campos possam ser correlacionados com campos de índice e facetas. É possível criar campos de índice e facetas baseados em valores de campo da origem de dados.

Caso necessite de pesquisar determinadas páginas e subpáginas em vez de pesquisar todas as páginas no sítio do WebSphere Portal, utilize uma das abordagens seguintes:

- Utilize o pesquisador da Web. Uma vez que o pesquisador da Web é concebido para trabalhar com aplicações da Web, pode pesquisar portlets da mesma forma que pesquisa outros sítios da Web.
- Em vez de pesquisar documentos, utilize APIs para adicionar documentos específicos ao índice.

Descrição geral da configuração

Ao criar o pesquisador, um assistente ajudá-lo-á nessas tarefas:

- Especifique as propriedades que controlam o modo como o pesquisador opera e utiliza os recursos do sistema. As propriedades do pesquisador controlam a forma como o pesquisador pesquisa o servidor.
- Especifique o servidor a pesquisar e informações que permitam que o pesquisador se ligue ao servidor.

Quando cria ou edita o pesquisador, é possível testar a respectiva capacidade de estabelecer ligação com os documentos a pesquisar. As mensagens indicam-lhe se o pesquisador pode ou não aceder aos documentos a pesquisar antes de a iniciar.

- Configurar opções de segurança ao nível de documentos:
 - Para activar a segurança para locais WebSphere Portal ou Quickr para documentos WebSphere Portal, tem de executar um script de configuração que é facultado com IBM Content Analytics with Enterprise Search no servidorWebSphere Portal antes de configurar o pesquisador. São facultados ficheiros JAR distintos para diferentes versões de WebSphere Portal. O suporte para procura protegida está disponível apenas para consultas processadas através da portlet de Procura para colecções de procura empresarial no WebSphere Portal (ESSearchPortlet), não para aplicações autónomas de procura empresarial.
 - Para activar a segurança para o IBM Connections, não é necessário executar o script de configuração facultado no servidor WebSphere Portal. Pode configurar suporte apenas para controlos de acesso pré-filtragem; a validação de credenciais de utilizadores (pós-filtragem) actual não é suportada. Consulte as origens de dados suportadas para o IBM Content Analytics with Enterprise Search para obter detalhes sobre as versões do IBM Connections que suportam a segurança ao nível de documentos.
- Especifique informações que permitam que o pesquisador comunique com um servidor proxy, caso o servidor proxy seja utilizado para assistir páginas.
- Se o servidor de origem de dados estiver protegido por um produto que suporte autenticação de início de sessão único (SSO), especifique as informações que permitem ao pesquisador obter autenticação de início de sessão único e recolher o conteúdo que pretende incluir no espaço de pesquisa. Quando o pesquisador necessitar de autenticação de início de sessão único, instale os servidores IBM Content Analytics with Enterprise Search na mesma rede com os servidores a serem pesquisados sem firewalls ou routers entre os servidores.

Para suportar autenticação de início de sessão único para procura protegida, em vez de permitir que o pesquisador especifique credenciais de início de sessão único ao pesquisar conteúdo, tem de activar o suporte de início de sessão único no WebSphere Application Server.

- Especifique informações sobre um ficheiro de arquivo de chaves para que o pesquisador possa utilizar o protocolo Secure Sockets Layer (SSL) para ligar ao servidor.
- Especifique o idioma e a página de códigos dos documentos a pesquisar.
- Especifique as opções para pesquisar e procurar metadados em documentos.
- Especifique agendamentos para pesquisar documentos.

Conceitos relacionados:

Adicionar documentos a uma colecção

Referências relacionadas:

🕩 Origens de dados suportadas

Como pesquisar conteúdo de IBM Web Content Manager através da interface seedlist 1.0

Integração do WebSphere Portal com IBM Content Analytics with Enterprise Search

Configurar suporte para autenticação de início de sessão único (SSO)

Configurar um utilizador de administrador de procura para o IBM Connections

O IBM Connections (anteriormente denominado de IBM Lotus Connections) necessita de um WebSphere Application Server. Para configurar o suporte pata o IBM Content Analytics with Enterprise Search, tem de preparar um utilizador de administrador de procura para o IBM Connections antes de criar um pesquisador da Lista de valores geradores.

Para configurar um utilizador de administrador de procura para utilizar para pesquisar origens do IBM Connections:

- Inicie sessão no WebSphere Application Server Integrated Solutions Console no servidor do IBM Connections como um utilizador administrativo do WebSphere Application Server.
- 2. Seleccione Aplicações > Enterprise Application.
- **3**. Quando a lista de aplicações empresariais for apresentada, seleccione uma aplicação e faça clique em **Correlação de funções de segurança com utilizador/grupo**.
- 4. Atribua um utilizador administrativo à função denominada **search-admin**. Quando configura o pesquisador da Lista de valores geradores para pesquisar origens do IBM Connections, tem de especificar o utilizador que atribui aqui.
- 5. Faça clique em OK duas vezes e guarde as alterações na configuração principal.
- **6**. Repita os passos 3, 4 e 5 para todas as aplicações empresariais no IBM Connections.

Tarefas relacionadas:

Configurar autenticação J2C para pesquisa

Configurar campos de índice para valores de campo que estão publicados numa lista de valores gerados do IBM Connections

Pode atribuir quaisquer campos publicados na lista de valores geradores aos campos de índice do IBM Content Analytics with Enterprise Search.

Para configurar correlações de campo:

 Obtenha a lista de valores geradores ao aceder ao seguinte URL do servidor do IBM Connections com um utilizador que tenha a função de search-admin: http://host_name:port/feature_name/seedlist/myserver

Por exemplo: http://ica.example.com:9081/blogs/seedlist/myserver

2. No conteúdo da lista de valores geradores, localize o ID do campo de onde pretende atribuir um valor. Em seguida localize o elemento fieldInfo que tem o mesmo ID. Lembre-se do valor do atributo de nome do elemento fieldInfo. Pode ver este valor como o nome nativo (original) do campo quando configurar opções de metadados para o pesquisador. Por exemplo, o elemento seguinte está na lista de valores geradores e vai ser atribuído a um campo de índice designado lcfield_tag. Neste exemplo, o nome do campo nativo é Tag.

<wplc:field id="FIELD_TAG">test</wplc:field> <wplc:fieldInfo id="FIELD_TAG" name="Tag" description="Tag associated with Blogs Entry" type="string" contentSearchable="true" fieldSearchable="true" parametric="false" returnable="false" sortable="false" supportsExactMatch="false"/>

- Inicie sessão na consola de administração. Localize o pesquisador Lista de valores geradores que quer editar e,de seguida, faça clique em Configurar > Espaço de pesquisa.
- 4. Na página do espaço de pesquisa, faça clique na ligação Editar metadados.
- 5. Adicione um novo campo e seleccione a caixa de verificação **Pesquisável** para o campo.
- 6. Escreva o nome do campo nativo no campo **Nome** e escreva o nome do campo correlacionado no campo **Apresentar nome**.
- 7. Seleccione as caixas de verificação para outras opções que pretende activar para o campo correlacionado.
- 8. Repita os passos anteriores até terminar a configuração de todos os campos de índice que pretende activar.
- **9**. Conclua os passos restantes no assistente de configuração do pesquisador e guarde a configuração.

Pesquisadores do SharePoint

É possível configurar um pesquisador SharePoint para incluir o conteúdo do servidorWindows SharePoint numa colecção.

Antes de configurar um pesquisador SharePoint, tem de implementar os serviços da Web facultados no servidor SharePoint. Estes serviços da Web activam o suporte para várias funções, incluindo pesquisa consciente de farm e procura protegida.

Credenciais de ligação do pesquisador

Quando cria o pesquisador, pode especificar as credenciais que permitem que o pesquisador se ligue às fontes a pesquisar. Também pode configurar as credenciais de ligação quando especifica as definições gerais de segurança do sistema. Se utilizar a última abordagem, vários pesquisadores e outros componentes do sistema podem utilizar as mesmas credenciais. Por exemplo, os servidores de procura podem utilizar as credenciais para determinar se um utilizador possui autorização para aceder ao conteúdo.

Restrição:

• Para pesquisar um servidor SharePoint, o ID do utilizador da ligação que o pesquisador utiliza tem de ser capaz de aceder ao URL do servidor SharePoint de destino.

- O ID do utilizador da ligação que o pesquisador utiliza tem de ter permissão de Administrador de Colecção de Sítio para que as colecções do sítio sejam pesquisadas.
- Para pesquisar elementos sociais, o ID de ligação utilizado pelo pesquisador tem de ser adicionado como Administrador nas definições da Aplicação de Serviço de Perfil de Utilizador e receber permissão de Gestão de Dados Sociais.
- Quando cria um pesquisador SharePoint, apenas os sítios que podem ser pesquisados com as credenciais de ligação especificadas são listados como candidatos.

Pesquisa consciente de farm

É possível configurar um pesquisador SharePoint para pesquisar os sítios de nível superior num servidor SharePoint e especificar opções para pesquisar locais, listas ou itens da lista que pertençam ao local de nível superior. Para pesquisar um sítio de nível superior diferente, é possível configurar um pesquisador em separado.

No entanto, num ambiente de SharePoint, o número de locais de nível superior pode aumentar bastante rapidamente. Para ajudar o utilizador a incluir vários sítios de nível superior no espaço de pesquisa, todos os sítios de nível superior que pertençam a um farm num servidor SharePoint podem ser automaticamente detectados. É possível configurar um pesquisador para recolher conteúdo de sítios de nível superior. Para activar o pesquisador para recolher conteúdo de um farm de SharePoint, tem de implementar os serviços da Web facultados no servidor SharePoint antes de criar um pesquisador.

Quando selecciona o conteúdo do Sharepoint a pesquisar, especifica os filtros para incluir ou excluir conteúdo do espaço de pesquisa. As regras de filtragem podem incluir caracteres globais para nomes de recolha do sítio, nomes do sítio e URLs. Utilize caracteres globais para incluir e excluir conteúdo que corresponda a um padrão.

Elementos sociais

O SharePoint 2010 expande o suporte a vários elementos sociais, comentários, médias de classificação e etiquetas. Se o pesquisador tiver permissão para aceder a elementos sociais, estes elementos sociais são correlacionados com metadados de campos quando os documentos forem pesquisados. Tem de adicionar a permissão de Gestão de Dados Sociais para o utilizador do pesquisador, no SharePoint, antes de criar o pesquisador SharePoint.

Segurança

Se a segurança estiver activada quando uma colecção for criada, o pesquisador pode associar dados de segurança aos documentos no índice. Estes dados permitem que as aplicações efectuem a pré-filtragem dos resultados de procura com base em listas de controlo de acesso ou tokens de segurança. Para suportar a pré-filtragem, para que os utilizadores apenas procurem os documentos que estão autorizados a ver, assegure-se de que os utilizadores têm as seguintes permissões de SharePoint, no mínimo:

- Ver Itens
- Ver Páginas de Aplicação
- Ver Páginas
- Abrir

É, também, possível configurar as definições de segurança para validar credenciais de utilizador quando um utilizador submeter uma consulta. Neste caso, em vez de comparar credenciais de utilizador com dados de segurança indexados, o sistema efectua uma pós-filtragem aos resultados de procura ao comparar as credenciais de utilizador com as listas de controlo de acesso actuais que são mantidas pela origem de dados original. O pesquisador SharePoint suporta a autenticação BASIC, autenticação Digest e autenticação NT LAN Manager (NTLM) através dos Serviços de Informação da Internet (IIS) Versão 7.x.

As seguintes tabelas resumem o suporte para procurar conteúdo de SharePoint mediante se a segurança estiver ou não activada na configuração do pesquisador SharePoint.

Tabela 14. Suporte para procurar conteúdo de SharePoint quando a segurança está activada

SharePoint 2007	SharePoint 2010	SharePoint 2007	SharePoint 2010
Não-Farm	Não-Farm	Farm	Farm
Suportado ^{1,3}	Não suportado	Suportado ^{2,3}	Suportado ²

Notas de tabela:

- 1. A procura protegida de servidores não conscientes de farm não é recomendada.
- 2. Para suportar a pesquisa consciente de farm e a procura protegida, tem de implementar os serviços da Web facultados no servidor SharePoint.
- 3. Se actualizar para IBM Content Analytics with Enterprise Search Versão 3.0 a partir de OmniFind Enterprise Edition Versão 9.1 ou IBM Content Analytics Versão 2.2, é possível continuar a utilizar o servidor de agente que instalou no servidor SharePoint para suportar a procura protegida nos seus pesquisadores existentes. Para pesquisadores novos, tem de implementar os serviços da Web facultados no servidor SharePoint.

Tabela 15. Suporte para procurar conteúdo de SharePoint quando a segurança não está activada

SharePoint 2007	SharePoint 2010	SharePoint 2007	SharePoint 2010
Não-Farm	Não-Farm	Farm	Farm
Suportado	Suportado	Suportado ¹	Suportado ¹

Notas de tabela:

1. Para suportar a pesquisa consciente de farm, tem de implementar os serviços da Web facultados no servidor SharePoint.

Tarefas relacionadas:

- "Configurar credenciais de ligação" na página 380
- "Implementar serviços da Web para pesquisadores SharePoint"

Referências relacionadas:

"Formatos de URI no índice" na página 164

"Nomes de campos de índice reservados" na página 199

Implementar serviços da Web para pesquisadores SharePoint

Para suportar a procura protegida e a pesquisa consciente de farm, tem de implementar serviços da Web que são facultados com o IBM Content Analytics with Enterprise Search no servidor do Sharepoint do Windows antes de configurar um pesquisador SharePoint.

Mediante as suas preferências, é possível implementar, manualmente, os serviços da Web ou executar um script que as implemente automaticamente:

- 1. Siga estes passos caso prefira aplicar, manualmente, as alterações ao servidor SharePoint:
 - a. Copie o ficheiro ES_INSTALL_ROOT/bin/ESSPServices.zip e o ficheiro ESCommonServices.zip para o servidor SharePoint.
 - b. Extraia ficheiros dos ficheiros de arquivo. Cada arquivo contém um ficheiro DLL e um ficheiro ASMX.
 - c. Coloque os ficheiros DLL no directório C:\inetpuwwwroot\wss\ VirtualDirectories\80\bin\. Este caminho é o sítio da Web IIS predefinido. Se o directório \bin não existir, crie-o e coloque ficheiros nele. O número da porta predefinido do servidor SharePoint neste caminho, 80, pode ser diferente no seu ambiente SharePoint.
 - d. Coloque os ficheiros ASMX no directório C:\Program Files\Common Files\Microsoft Shared\Web Server Extensions\14\ISAPI\. Este caminho é o directório predefinido da aplicação do SharePoint. O número no caminho especificado pode ser diferente, mediante a versão do servidor do SharePoint. Para o SharePoint 2010, o número é, geralmente, 14. Para o SharePoint 2007, o número é, geralmente, 12.
 - e. Aceda aos URLs dos serviços da Web, com o seu navegador, para verificar o sucesso da implementação:

http://servidor_SharePoint:porta/_vti_bin/ESSPServices.asmx
http://servidor_SharePoint:porta/_vti_bin/ESCommonServices.asmx

2. Se preferir automatizar a implementação dos serviços da Web, execute apenas um dos seguintes scripts no servidor SharePoint. ESSPSolution.wsp é facultado com IBM Content Analytics with Enterprise Search.

stsadm

Se utilizar o SharePoint 2007, tem de utilizar o comando **stsadm** para implementar os serviços da Web:

stsadm -o addsolution -filename ESSPSolution.wsp

Add-SPSolution

Se utilizar o SharePoint 2010, é possível utilizar o comando **stsadm** ou executar o seguinte Windows PowerShell cmdlet:

Add-SPSolution -LiteralPath C:\files\ESSPSolution.wsp

Após a conclusão da implementação, a solução farm é listada na consola administrativa do SharePoint. Um administrador pode activar ou desactivar a solução e agendar activadores.

- 3. Opcional: Para completar a implementação em alguns ambientes (independentemente da abordagem que utiliza para implementar os serviços da Web), pode necessitar de aplicar as seguintes configurações ao servidor IIS que alojar o servidor SharePoint e os serviços da Web:
 - Permita a representação de .NET em IIS.
 - Altere o nível de confiança de ASP.NET para WSS_Medium.

No SharePoint 2010, é possível aplicar estas configurações através do Gestor de Serviços de Informação da Internet.

Para alterar o nível de confiança de ASP.NET no SharePoint 2007, modifique o ficheiro web.config e altere o nível de confiança de mínimo para médio:
```
De: <trust level="WSS_Minimum" originUrl="" />
Para: <trust level="WSS_Medium" originUrl="" />
```

O caminho predefinido para este ficheiro é C:\Inetpub\wwwroot\wss\ VirtualDirectories\80\web.config.

Conceitos relacionados:

"Pesquisadores do SharePoint" na página 129

Pesquisadores do Sistema de ficheiros UNIX

Para incluir documentos que estão armazenados nos sistemas de ficheiros AIX ou Linux numa colecção, tem de configurar um pesquisador Sistema de ficheiros UNIX.

Pode utilizar o pesquisador do Sistema de ficheiros UNIX para pesquisar qualquer número de sistemas de ficheiros. Ao configurar o pesquisador, seleccione os directórios locais e remotos e os subdirectórios que pretenda pesquisar.

Se instalar o servidor do pesquisador num computador Windows, não é possível pesquisar origens do sistema de ficheiros AIX ou Linux (o pesquisador Sistema de ficheiros UNIX não surge na lista de tipos de pesquisador disponíveis).

O pesquisador do Sistema de ficheiros UNIX pesquisa documentos de acordo com permissões de leitura que são especificadas para o administrador.

Ao criar o pesquisador, um assistente ajudá-lo-á nessas tarefas:

- Especifique as propriedades que controlam o modo como o pesquisador opera e utiliza os recursos do sistema. As propriedades do pesquisador controlam o procedimento do pesquisador para pesquisar todos os subdirectórios no espaço de pesquisa.
- Defina um agendamento para a pesquisa de sistemas de ficheiros.
- Seleccione os subdirectórios e os níveis de subdirectórios que pretende que o pesquisador pesquise.
- Especifique as opções para fazer com que os documentos possam ser procurados em subdirectórios. Por exemplo, pode excluir certos tipos de documentos do espaço de pesquisa.
- Configure as opções de segurança ao nível do documento. Se a segurança tiver sido activada quando a colecção foi criada, o pesquisador pode associar dados de segurança aos documentos existentes no índice. Esses dados permitem às aplicações fazer cumprir os controlos de acesso baseados nas listas de controlo de acesso armazenadas ou nos tokens de segurança.

Credenciais de ligação do pesquisador

Quando cria o pesquisador, pode especificar as credenciais que permitem que o pesquisador se ligue às fontes a pesquisar. Também pode configurar as credenciais de ligação quando especifica as definições gerais de segurança do sistema. Se utilizar a última abordagem, vários pesquisadores e outros componentes do sistema podem utilizar as mesmas credenciais. Por exemplo, os servidores de procura podem utilizar as credenciais para determinar se um utilizador possui autorização para aceder ao conteúdo.

Pesquisadores da Web

Para incluir páginas de sítios da Web numa colecção, tem de configurar um pesquisador da Web.

Pode utilizar o pesquisador da Web para pesquisar um número de servidores Hypertext Transfer Protocol (HTTP) e proteger os servidores HTTP (HTTPS). O pesquisador visita um sítio da Web e lê os dados no sítio. Depois, segue ligações nos documentos para pesquisar documentos adicionais. O pesquisador da Web pode pesquisar e extrair ligações de páginas individuais ou *conjuntos de segmentos* (páginas que são criadas com segmentos HTML).

Os dados pesquisados podem estar num dos muitos formatos, e têm várias origens na intranet ou na Internet. Os formatos comuns incluem o HTML, PDF, Microsoft Word, Lotus WordPro, Extensible Markup Language (XML), etc.

Ao criar o pesquisador, um assistente ajudá-lo-á nessas tarefas:

- Especifique as propriedades que controlam o modo como o pesquisador opera e utiliza os recursos do sistema. As propriedades do pesquisador controlam o modo como este último pesquisa todas as páginas da Web no espaço de pesquisa.
- Especifique as regras para permitir e proibir visitas ao sítio da Web. Quando especificar as regras de pesquisa, pode testar as regras e verificar que o pesquisador pode aceder aos sítios que pretende incluir no espaço de pesquisa.
- Especifique opções para incluir certos tipos de ficheiros e excluir ficheiros com certos tipos do nome de ficheiros.
- Especifique regras para como o pesquisador pode processar páginas de erros esporádicos.
- Configure as opções de segurança ao nível do documento. Se a segurança tiver sido activada quando a colecção foi criada, o pesquisador pode associar dados de segurança aos documentos existentes no índice. Esses dados permitem às aplicações da procura empresarial fazer cumprir os controlos de acesso baseados nas listas de controlo de acesso armazenadas ou nos tokens de segurança.
- Especifique opções para pesquisar sítios da Web protegidos contra palavra-passe (os servidores da Web a pesquisar têm de utilizar uma autenticação base HTTP ou formas HTTP para pedir palavras-passe).
- Especifique opções para pesquisar sítios da Web servidos por um servidor proxy.
- Especifique agendamentos para pesquisar servidores Web específicos em momentos específicos.

Credenciais de ligação do pesquisador

Quando cria o pesquisador, pode especificar as credenciais que permitem que o pesquisador se ligue às fontes a pesquisar. Também pode configurar as credenciais de ligação quando especifica as definições gerais de segurança do sistema. Se utilizar a última abordagem, vários pesquisadores e outros componentes do sistema podem utilizar as mesmas credenciais. Por exemplo, os servidores de procura podem utilizar as credenciais para determinar se um utilizador possui autorização para aceder ao conteúdo.

Configuração do agente do utilizador

Para pesquisar um sítio da Web que utilize o protocolo Robots Exclusion, certifique-se de que o ficheiro robots.txt no sítio da Web permite que o nome de agente do utilizador que configurar para o pesquisador da Web aceda ao sítio da Web.

Quando o sistema for iniciado, o pesquisador da Web carrega o nome do agente do utilizador que configurar para ele. Antes de o pesquisador descarregar uma página do sítio da Web não visitada previamente (ou que nunca foi visitada), primeiro o pesquisador tenta descarregar um ficheiro chamado robots.txt. O ficheiro encontra-se no directório raiz do sítio da Web.

Se o ficheiro robots.txt não existir, o sítio da Web está aberto à pesquisa não restrita. Se o ficheiro não existir, especifica as áreas do sítio (directórios) que estão fora dos limites dos pesquisadores. O ficheiro robots.txt especifica as permissões para os pesquisadores, identificando o respectivo nome do utilizador-agente.

O protocolo Robots Exclusion é voluntário, mas o pesquisador da Web tenta entrar em conformidade com ele:

- Se um ficheiro robots.txt contiver uma entrada do nome do agente do utilizador que esteja configurada para o pesquisador da Web, este é compelido com as restrições ao agente do utilizador.
- Se o nome do agente do utilizador não aparecer no ficheiro robots.txt, mas a última entrada especificar Utilizador-agente: * (o que significa qualquer agente do utilizador) e a restrição seja Disallow: / (o que não significa que a pesquisa não é permitida, iniciando na raiz do sítio da Web), o pesquisador não pode pesquisar esse sítio.
- Se o nome do agente do utilizador não aparecer no ficheiro robots.txt, mas a última entrada especificar Utilizador-agente: * (o que significa qualquer agente do utilizador) e a restrição seja Allow: /, o pesquisador é permitido para pesquisar esse sítio.

Os administradores do sítio da Web especificam com frequência uma entrada final que proíbe o acesso a todos os pesquisadores que não concedem explicitamente o acesso. Se estiver a configurar um novo pesquisador da Web e souber que alguns dos sítios da Web que pretende pesquisar utilizam o protocolo Robots Exclusion, peça aos administradores do sítio da Web para adicionar uma entrada para o respectivo pesquisador aos respectivos ficheiros robots.txt.

Tenha a certeza ao especificar o mesmo nome do agente do utilizador nas características do pesquisador da Web e em todos os ficheiros robots.txt que pertencem aos sítios da Web de interesse.

Se nenhum dos sítios da Web a pesquisar utilizar o protocolo Robots Exclusion, o valor que especifica para a característica do agente do utilizador não interessa. Todavia, alguns servidores de aplicação, JSPs, e servlets personalizam as respectivas respostas para o nome do agente do utilizador. Por exemplo, existem respostas diferentes para processar incompatibilidades de navegador. O nome do agente do utilizador que especifica para o pesquisador pode interessar nestas situações, tendo em conta do protocolo Robots Exclusion. Se tiver a necessidade de pesquisar esses tipos de sítios, consulte os administradores do sítio da Web para garantir que o pesquisador permita o acesso.

Como o pesquisador da Web utiliza o protocolo Robots Exclusion

A não ser que configure as propriedades do pesquisador da Web para ignorar um ficheiro robots.txt do servidor da Web, o pesquisador tenta estar em conformidade com o protocolo Robots Exclusion e não pesquisa os sítios da Web se as regras no ficheiro robots.txt não permitirem a pesquisa.

Quando o pesquisador está configurado para seguir os ficheiros robots.txt, uma transferência bem-sucedida é quando o pesquisador consegue recuperar o ficheiro robots.txt de um servidor da Web ou confirmar que um ficheiro robots.txt não existe. A transferência é considerada como falhada quando o pesquisador não consegue obter as regras ou confirmar a existência de um ficheiro robots.txt.

Uma transferência bem-sucedida não significa que o pesquisador tenha permissão para pesquisar porque as regras existentes no ficheiro robots.txt podem não permitir a pesquisa. Uma falha de transferência proíbe temporariamente a pesquisa porque o pesquisador não consegue determinar quais são as regras.

Estes são os passos que o pesquisador executa quando tenta transferir o ficheiro robots.txt:

- 1. Quando o pesquisador descobre um novo sítio, tenta obter o endereço de IP do servidor. Se esta tentativa falhar, a pesquisa não é possível.
- 2. Quando está disponível, pelo menos, um endereço de IP, o pesquisador tenta transferir o ficheiro robots.txt utilizando HTTP (ou HTTPS) GET.
- **3**. Se o tempo de espera da ligação de encaixe se esgotar, se esta for interrompida ou se ocorrer outro erro de nível inferior (como um problema com um certificado de SSL), o pesquisador regista o problema e repete a tentativa em cada endereço de IP conhecido pelo servidor destino.
- 4. Se não for estabelecida nenhuma ligação após o pesquisador tentar contactar todos os endereços, esta aguarda dois segundos e, em seguida, tenta todos os endereços mais uma vez.
- 5. Se for estabelecida uma ligação, e se forem trocados cabeçalhos de HTTP, o estado de retorno é examinado. Se o código de estado for 500 ou superior, o pesquisador interpreta-o como uma ligação falhada e continua a tentar outros endereços de IP. Para qualquer outro estado, o pesquisador pára de tentar contactar endereços de IP alternativos e procede de acordo com o código de estado.

Após o pesquisador receber um código de estado de HTTP inferior a 500, ou após tentar contactar pela segunda vez todos os endereços de IP, ela procede do seguinte modo:

- 1. Se não tiver sido recebido nenhum estado de HTTP inferior a 500, o sítio fica desqualificado temporariamente.
- 2. Se tiver sido recebido um estado de HTTP de 400, 404 ou 410, o sítio é qualificado para a pesquisa sem regras.
- **3**. Se tiver sido recebido um estado de HTTP de 200 a 299, as seguintes condições determinam a acção a executar em seguida:
 - Se o conteúdo estiver truncado, o sítio fica desqualificado temporariamente.
 - Se o conteúdo analisado não tiver erros, o sítio é qualificado para a pesquisa com as regras que foram encontradas.
 - Se o conteúdo analisado tiver erros, o sítio é qualificado para a pesquisa sem regras.

4. Se tiver sido devolvido qualquer outro estado de HTTP, o sítio fica desqualificado temporariamente.

Quando o pesquisador tenta transferir o ficheiro robots.txt para um sítio, ela actualiza uma marca de tempo persistente para esse sítio, chamada a "data de robots". Se um sítio for desqualificado por as informações do ficheiro robots.txt não estarem disponíveis, a contagem de falhas persistentes de robots é incrementada.

Quando o intervalo de reenvio for atingido, o pesquisador tenta de novo obter informações sobre o robots.txt para o sítio que falhou. Se o número de falhas sucessivas atingir o número máximo de falhas permitidas, o pesquisador pára de tentar obter o ficheiro robots.txt para o sítio e desqualifica-o para a pesquisa.

Depois de um sítio ser qualificado para pesquisa (procura das regras do ficheiro robots.txt com êxito), a contagem de falhas é reposta a zeros. O pesquisador utiliza os resultados da transferência até expirar o intervalo para verificação de regras. Nesse momento, o sítio tem de ser qualificado de novo.

Sugestão:

- Se um servidor devolver conteúdo, mas contiver erros de sintaxe, ou se o servidor utilizar um protocolo robots diferente da versão 1994, ou se o conteúdo tiver algo diferente de regras de robots (como uma página de erros esporádicos), o pesquisador age como se o servidor não tivesse um ficheiro de regras aplicável e pesquisa o sítio. Esta acção está, normalmente, correcta porque os administradores de colecções não controlam o conteúdo dos sítios ou o comportamento do servidor predefinido. Se um administrador de servidor da Web não pretender que sejam pesquisados para um sítio e não pretender instalar um ficheiro de regras relacionadas, o administrador de colecções pode bloquear o sítio a partir do pesquisador especificando o domínio do sítio, o endereço de IP ou o prefixo de HTTP nas regras do pesquisador.
- Se um servidor devolver um código de estado 302 ou outros códigos de redireccionamento, o pesquisador interpreta o código para que este signifique que o sítio tem um ficheiro robots.txt que deve ser utilizado, mas que o ficheiro não se encontra no local adequado (na raiz do sítio). O administrador do servidor da Web tem de mover o ficheiro para a localização correcta de modo a que o pesquisador se possa orientar pelas regras existentes no ficheiro.
- Se existirem problemas relacionados com certificados (por exemplo, o certificado pode estar desactualizado, a respectiva autoridade pode não ser fidedigna ou o certificado pode ser de autenticação própria e o pesquisador não está configurado para aceitar certificados de autenticação própria), o pesquisador interpreta o problema como uma falha na ligação ao sítio e desqualifica-o. De qualquer maneira, o mesmo problema impediria provavelmente a pesquisa noutras páginas do sítio. Para permitir a pesquisa no sítio, o administrador da colecção tem de permitir certificados de autenticação própria, adicionar a autoridade do sítio ao ficheiro de arquivo de chaves fidedigno ou pedir ao administrador do servidor da Web que obtenha um certificado actualizado.
- O pesquisador da Web pode ser configurado para utilizar a autenticação base de HTTP (incluindo a autenticação de proxy base de HTTP). Se for devidamente configurada, a autenticação também é necessária para transferências de ficheiros robots.txt. Um código de estado de 403, 407 ou outras respostas relacionadas com a autenticação, indicam problemas de autorização, caso em que o pesquisador desqualifica o sítio. (Só é suportada a autenticação base de HTTP.)

 Se o ficheiro robots.txt para um sítio exceder o comprimento máximo para uma página de robots, o administrador da colecção pode aumentar o valor máximo configurado (o valor predefinido de um milhão de bytes deverá ser suficiente).

Para ajudar na resolução de problemas, pode pedir um relatório de sítio quando supervisiona o pesquisador da Web. Seleccione opções para visualizar o conteúdo do ficheiro robots.txt (para ver se as regras proíbem o acesso do pesquisador ao sítio), ver a data e hora em que o pesquisador tentou pela última vez transferir o ficheiro robots.txt (o pesquisador não tentará de novo enquanto o intervalo de reenvio não expirar) e ver quantas tentativas falhadas consecutivas o pesquisador efectuou para transferir o ficheiro robots.txt.

Suporte para JavaScript

O pesquisador da Web pode encontrar algumas ligações (URLs) que estão contidas nas partes de JavaScript dos documentos da Web. Se determinar que um número elevado de URLs incorporado no texto é de baixa relevância, pode desactivar a análise de ligações no texto configurando as propriedades avançadas do pesquisador Web.

O pesquisador da Web pode encontrar ambas as ligações relativas e absolutas. Se um documento HTML contiver um elemento BASE, o pesquisador utilizará esse elemento para solucionar ligações relativas. Por outro lado, o pesquisador utiliza o URL próprio do documento.

O Suporte de JavaScript é limitado à extracção de ligação. O pesquisador não analisa JavaScript, não cria um DOM (Document Object Model), e não interpreta ou executa instruções JavaScript. O pesquisador procura cadeias no conteúdo do documento (incluindo, mas não limitado às partes JavaScript), que pretendem ser URLs em instruções JavaScript. Isto tem dois significados:

- Alguns URLs que serão encontrados, serão ignorados pelo analisador HTML estrito. O pesquisador rejeitará tudo o que não for um URL sintacticamente válido, mas alguns dos URLs válidos devolvidos pela fase de varrimento podem ter pouco interesse para a procura.
- O conteúdo do documento é gerado por JavaScript, tal como quando um utilizador vê uma página com um navegador e o navegador executa qualquer JavaScript, não pode ser detectado pelo pesquisador da Web e, por isso, não será indexado.

Porque o pesquisador da Web não analisa JavaScript em ficheiros de HTML, os URLs em JavaScript não são pesquisados. Para permitir que o pesquisador da Web pesquise URLs em JavaScript, pode executar uma das seguintes acções:

- Na consola de administração, edite o pesquisador da Web e, na página Espaço de Pesquisa da Web, adicione os URLs à lista de URLs que o pesquisador tem de utilizar como ponto de partida para adicionar os URLs à colecção (Iniciar URLs). Para que as alterações se tornem efectivas, reinicie o pesquisador da Web (não é necessário iniciar uma pesquisa completa).
- Utilize a etiqueta de ancoragem () para especificar os URLs como ligações de hipertexto no ficheiro de HTML.

Regras para limitar o espaço de pesquisa da Web

Para garantir que os utilizadores acedam apenas às páginas da Web que pretende que eles procurem, especifique as regras para limitar o que o pesquisador da Web pode pesquisar. Quando um pesquisador da Web pesquisar uma página da Web, identifica ligações às outras páginas e coloca aquelas numa fila a pesquisar de seguida. A pesquisa e a identificação podem ser repetidas enquanto o tempo e os recursos da memória permitirem. Ao configurar um pesquisador da Web, especifique onde é que o pesquisador deve iniciar a pesquisa. Desses URLs iniciais (que são chamados *iniciar URLs*), os pesquisadores da Web podem alcançar qualquer documento na Web que esteja ligado directa ou indirectamente.

Para limitar o espaço de pesquisa, configurar o pesquisador da Web para pesquisar determinados URLs e ignorar as ligações que apontam para fora da área de interesse. Como o pesquisador, que por predefinição, aceita qualquer URL que localize, tem de especificar as regras que identifiquem os URLs que pretende incluir na colecção e eliminar o resto das páginas.

Pode especificar em várias maneiras o que pretende que o pesquisador pesquise ou não pesquise. Pode especificar:

- Uma lista de URLs iniciais onde o pesquisador deve iniciar a pesquisa
- Três tipos de regras de pesquisa: domínio, endereço (IP) de Internet Protocol e prefixo de URL
- Uma lista de tipos MIME para documentos que pretende incluir
- Uma lista de extensões do nome do ficheiro que pretende excluir
- O número máximo de directórios num caminho URL
- O número máximo de ligações que o pesquisador atravessa quando segue as ligações de um documento para outro

As regras de pesquisa têm o formato: destino do tipo de acção

acção proibida ou permitida; tipo é domínio, endereço IP ou prefixo URL (HTTP ou HTTPS); e destino depende do valor do tipo. Pode especificar um asterisco (*) como carácter global, de maneira limitada, para especificar os destinos que correspondem a um padrão.

Regras de domínio

O destino de uma regra de domínio é um nome do domínio DNS. Por exemplo, pode especificar que todo o domínio www.IBM.com deve ser pesquisado: permite o domínio www.IBM.com

Pode especificar um asterisco como carácter global, o que faz com que a regra seja aplicada a qualquer nome de sistema central que corresponda ao resto do padrão. Por exemplo, pode especificar que nenhum domínio que comece por servidor e termine com ibm.com deverá ser pesquisado.

forbid domain server*.ibm.com

A correspondência do nome do sistema central é sensível a maiúsculas e minúsculas, se especificar um nome do domínio explícito ou um padrão do nome do domínio. Por exemplo, *.utilizador.IBM.com corresponde a joão.utilizador.IBM.com e maria.silva.utilizador.IBM.com, mas não a joão.utilizador.IBM.com.

Uma regra de domínio que não especifique uma porta aplica-se a todas as portas desse domínio. No exemplo seguinte, todas as portas no domínio sales são permitidas:

allow domain sales.IBM.com

Se uma regra de domínio especificar um número de porta, a regra aplica-se apenas a essa porta. No exemplo seguinte, apenas a porta 443 no domínio sales é permitida:

allow domain sales.IBM.com:443

Regras de prefixo

Uma regra de prefixo controla a pesquisa de URLs que iniciam com uma cadeia correspondente. O destino é um único URL, que normalmente contém um ou mais asteriscos para assinalar um padrão. Por exemplo, um asterisco é frequentemente especificado como carácter final na cadeia de prefixo.

Uma regra de prefixo permite pesquisar a totalidade ou parte de um sítio da Web. Pode especificar um caminho ou padrão, e depois permitir ou proibir tudo desse ponto em diante na árvore de directórios. Por exemplo, as regras seguintes permitem ao pesquisador pesquisar tudo no directório público em sales.IBM.com, mas proíbem o pesquisador de aceder a outras páginas desse sítio:

```
allow prefix http://sales.IBM.com/public/*
forbid prefix http://sales.IBM.com/*
```

Ao especificar regras de prefixo, pode especificar mais do que um asterisco e pode especificá-los em qualquer parte da cadeia de prefixo, não só na última posição. Por exemplo, a regra seguinte proíbe o pesquisador de pesquisar documentos num directório de nível superior do sítio sales.IBM.com se o nome e do directório terminar em fs. (Por exemplo, pode ter montagens de sistema de ficheiros que não contêm informação que seria útil no índice da procura.)

forbid http://sales.IBM.com/*fs/*

Regras do endereço

Uma regra do endereço permite-lhe controlar a pesquisa de todos os sistemas centrais ou de todas as redes, especificando um endereço de IP e uma máscara de rede como o destino. Por exemplo:

- IPv4 allow address 9.0.0.0 255.0.0.0
- IPv6 Se executar o sistema num servidor Windows 2003, e tiver activado o sistema para utilizar o protocolo de IP versão 6 (IPv6), deve colocar o endereço entre parêntesis.

allow address [2001:db8:0:1:0:0:0:1]

A máscara de rede permite-lhe especificar a correspondência do padrão. Uma regra de endereço só se aplica a possíveis endereços de IP se o endereço de IP da regra e o possível endereço de IP forem idênticos, excepto quando encobertos por zeros na máscara de rede. Uma regra de endereço define um padrão e uma máscara de rede define os bits significantes no padrão de endereço. Um zero numa máscara de rede funciona como carácter global e significa que qualquer valor especificado naquela posição binária do endereço terá correspondência.

No exemplo anterior, a regra seguinte aplica-se a qualquer endereço de IP com 9 no primeiro octeto, e qualquer valor nos últimos três octetos.

A regra seguinte é uma regra útil a incluir como o endereço final na respectiva lista de regras de endereço. Essa regra corresponde a um endereço de IP porque a

máscara de rede torna todos os bits insignificantes (a regra proíbe todos os endereços que não são permitidos por uma regra anterior na respectiva lista de regras).

IPv4 forbid address 0.0.0.0 0.0.0.0

IPv6

forbid address :: ::

Restrições aos servidores proxy: Se pretender pesquisar sítios da Web fornecidos por um servidor proxy, não especifique as regras de endereço de IP. Normalmente, um servidor proxy é utilizado quando um agente do utilizador (navegador ou pesquisador) não tem acesso directo às redes onde os servidores da Web se encontram. Por exemplo, um servidor proxy HTTP pode retransmitir pedidos HTTP de um pesquisador para um servidor da Web e devolver as respostas para o pesquisador.

Quando um pesquisador da Web utiliza um servidor proxy, o endereço de IP do servidor proxy é apenas o endereço de IP que o pesquisador tem para outro sistema central. Se as regras de endereços de IP forem utilizadas para restringir o pesquisador a uma sub-rede de endereços de IP, a restrição faz com que quase todos os URLs sejam classificados com o código de estado 760 (que indica que são proibidos pelo espaço da Web).

Pesquisar ordem de regra

O pesquisador aplica as regras de pesquisa várias vezes durante o processo de identificação e de pesquisa de URLs. A ordem das regras é importante, mas apenas dentro das regras de cada tipo. O posicionamento da regra de endereço antes ou depois de outra regra de endereço é relevante, mas é indiferente se uma regra de endereço estiver posicionada antes ou depois de uma regra de prefixo, porque o pesquisador não aplica as regras ao mesmo tempo.

Dentro do conjunto de regras para um único tipo, o pesquisador testa um possível domínio, endereço ou URL contra cada regra, da primeira regra especificada à última, até encontrar uma regra que aplique. É utilizada a acção especificada para a primeira regra que se aplica.

A dependência da ordem conduz a uma estrutura típica para a maioria das regras de pesquisa:

- Normalmente, um conjunto de regras de domínio inicia com regras de proibição que eliminam domínios únicos do espaço de pesquisa. Por exemplo, o administrador da colecção pode determinar que certos domínios não contêm informação útil.
- A lista de regras de proibição é normalmente seguida por uma série de regras de permissão (com caracteres globais) que permitem ao pesquisador visitar qualquer domínio que termine num dos nomes do domínio de nível superior que defina uma intranet da empresa (tal como *.IBM.com e *.lotus.com).

Termine o conjunto das regras de domínio com regra predefinida seguinte, que elimina os domínios que não são permitidos por uma regra anterior:

forbid domain *

Essa regra final é crítica, porque impede o espaço de pesquisa de incluir toda a Internet.

 O conjunto de regras de endereço inicia normalmente com um número reduzido de regras de permissão que permite ao pesquisador pesquisar as redes de nível superior (classe - A, classe - B ou classe - C), que ampliam uma intranet da empresa.

Consulte a discussão precedente sobre as regras de endereço para exemplos de como especificar a regra final na respectiva lista de regras de endereço, para prevenir o pesquisador de pesquisar sítios da Web que estejam fora da rede corporativa.

 O conjunto de regras de prefixo é normalmente o maior, porque contém de forma arbitrária especificações detalhadas de regiões permitidas ou proibidas especificadas como árvores ou sub-árvores. Uma boa abordagem é a de primeiro permitir ou proibir mais regiões localizadas de forma exacta, e depois especificar a regra oposta, num padrão mais geral, para permitir ou proibir mais alguma coisa.

Normalmente, a secção de prefixo não termina com uma regra comum. As regras do domínio e do endereço sugeridas no final podem garantir que o pesquisador não pesquise para além da rede da empresa de forma mais eficiente que por meio da verificação de prefixos de URL.

O pesquisador pode aplicar regras de prefixo de forma mais eficiente se agrupar as regras por acção (proibir ou permitir). Por exemplo, em vez de especificar sequências curtas de regras de permissão e de proibição que se alternam entre si, especifique uma sequência longa de regras que estipulam uma acção e, então, especifique uma sequência longa de regras que estipulam a outra acção. Pode fazer com que as regras de permissão e proibição atinjam os objectivos do respectivo espaço de pesquisa. Porém, o agrupamento de regras de permissão e de proibição em conjunto pode melhorar o desempenho do pesquisador.

Extensões de ficheiro e tipos MIME

Essas opções fornecem maneiras adicionais de especificar o conteúdo para o espaço de pesquisa. Pode excluir certos tipos de documentos baseados na extensão do nome do ficheiro do documento e pode incluir certos tipos de documentos baseados no tipo MIME do documento. Ao especificar que tipos MIME pretende que o pesquisador pesquise, considere que o tipo MIME é frequentemente definido de forma incorrecta em documentos da Web.

Comprimento máximo do caminho do URL e comprimento máximo da ligação

O comprimento máximo do caminho do URL é o número de barras num URL a partir da raiz do sítio. Essa opção permite-lhe impedir a criação em imagens do pesquisador nas estruturas do sistema de ficheiros recursivas de comprimento infinito. O comprimento da pesquisa não corresponde aos níveis que o pesquisador atravessa quando segue as ligações de um documento para outro.

O comprimento máximo da ligação controla quantos documentos deve o pesquisador incluir quando segue as ligações de um documento para outro. Se uma ligação para um documento exceder o comprimento máximo da ligação, o documento é excluído do espaço de pesquisa.

Se alterar estas propriedades avançadas, tem de reiniciar o pesquisador da Web para que o índice possa ser actualizado. Por exemplo, se especificar um número superior para o comprimento máximo da ligação, as páginas previamente excluídas podem ser incluídas. Se especificar um número inferior para o comprimento máximo da ligação, as páginas previamente incluídas são removidas do índice. Os documentos que excederem o comprimento máximo da ligação serão removidos da próxima vez que forem pesquisados.

URLs iniciais

Os URLs iniciais são os URLs que o pesquisador inicia a pesquisar, e esses URLs são inseridos na pesquisa sempre que o pesquisador é iniciado. Se os URLs iniciais já tiverem sido identificados, não serão pesquisados nem pesquisados de novamente mais cedo do que outros sítios da Web permitidos nas regras de pesquisa.

Um URL inicial é importante na primeira vez em que um pesquisador da Web for iniciado e o espaço de pesquisa estiver vazio. Um URL inicial também é importante ao adicionar um URL não previamente identificado para a lista de URLs iniciais num espaço de pesquisa.

Os URLs iniciais têm de ser URLs completos, não apenas nomes do domínio. Tem de especificar o protocolo e, se a porta não for 80, o número da porta.

Os URLs seguintes são URLs iniciais válidos: http://w3.ibm.com/ http://sales.IBM.com:9080/

O URL seguinte não é um URL inicial válido: www.ibm.com

Tem de incluir os URLs iniciais nas respectivas regras de pesquisa. Por exemplo, o pesquisador não pode iniciar a pesquisa com um URL inicial especificado se as regras de pesquisa não permitirem que o URL seja pesquisado.

Suporte para endereços de IPv6: Se executar o IBM Content Analytics with Enterprise Search num servidor Windows 2003, e tiver activado o sistema para utilizar o protocolo de IP versão 6 (IPv6), tem de colocar os URLs iniciais entre parêntesis. Por exemplo:

http://[2001:db8:0:1:0:0:0:1] http://[2001:db8:0:1::1]

Tarefas relacionadas:

"Activar o suporte para o protocolo IPv6 em sistemas Windows" na página 38

Testar ligações de URL com o pesquisador da Web

Depois de especificar URLs para o pesquisador da Web a pesquisar, pode testar a configuração das regras de pesquisa.

Pode fazer clique em **Testar** quando especificar os domínios, prefixos de HTTP ou endereços de IP a pesquisar, ou seleccionar a página Testar URLs para testar a capacidade do pesquisador para estabelecer ligação com os URLs iniciais para além dos URLs que especificar.

Os resultados do teste mostram se o pesquisador consegue aceder a URLs com o nome de agente utilizador especificado nas propriedades do pesquisador. Os resultados de teste também mostram se um URL não pode ser pesquisado devido às regras de exclusão (por exemplo, um documento pode não ser pesquisado, porque tem uma extensão do nome do ficheiro que corresponde à extensão excluída do espaço de pesquisa). Depois de um sítio ser pesquisado pelo menos uma vez, pode testar os URLs para obter informação adicional. Por exemplo, o relatório de teste pode fornecer o código de estado de HTTP mais recente (que indica se uma pesquisa do URL teve êxito), mostrar quando o URL foi pesquisado pela última vez, quando foi agendado para nova pesquisa e mostrar se o agente do utilizador está a utilizar o ficheiro robots.txt actual do servidor da Web.

Nova pesquisa de definições de intervalo no pesquisador da Web

Para influenciar a frequência com que o pesquisador volta a visitar os URLs, especifique as opções nas características do pesquisador da Web.

A maioria dos outros tipos de pesquisadores são executados segundo os agendamentos especificadas por um administrador. Em contrário, depois de iniciar um pesquisador da Web, normalmente esse é executado de forma contínua. Para controlar a frequência com que volta a visitar os URLs que pesquisou previamente, especifique intervalos mínimos e máximos da nova pesquisa.

Quando utiliza a consola de administração para criar um pesquisador da Web ou para editar propriedades do pesquisador da web, pode seleccionar uma opção para configurar propriedades avançadas. Na página características adicionais do pesquisador da Web, especifique as opções de um intervalo mínimo e máximo da nova pesquisa. O pesquisador da Web utiliza o valor especificado para calcular um intervalo para pesquisar de novo os dados.

A primeira vez que uma página for pesquisada, o pesquisador utiliza a data e hora em que uma página é pesquisada e uma média dos intervalos mínimos e máximos da nova pesquisa para definir uma data da nova pesquisa. A página não será pesquisada de novo antes da data. O momento em que a página é pesquisada de nova após a data depende do carregamento do pesquisador e do equilíbrio entre URLs novos e antigos no espaço de pesquisa.

Sempre que a página for pesquisada de novo, o pesquisador verifica se o conteúdo foi alterado. Se o conteúdo tiver sido alterado, o intervalo seguinte da nova pesquisa será menor que o anterior, mas nunca menor que o intervalo mínimo da nova pesquisa especificado. Se o conteúdo não tiver sido alterado, o intervalo seguinte da nova pesquisa será maior que o anterior, mas nunca maior que o intervalo máximo da nova pesquisa especificado.

Opções para visitar URLs com o pesquisador da Web

Pode forçar o pesquisador para visitar URLs específicos logo que possível.

Se tiver necessidade de renovar o espaço do pesquisador com informações de certos sítios da Web, pode supervisionar o pesquisador, seleccionar a opção **URLs a visitar ou visitar de novo** e, em seguida, especificar os URLs ou padrões de URL das páginas que têm de ser pesquisadas ou pesquisadas de novo.

Por exemplo, se o departamento de Comunicações adicionar uma página da Web à intranet, ou fizer a revisão de uma página para reflectir uma alteração importante de política, pode especificar o URL da página nova ou alterada. Se o pesquisador estiver em execução, o pesquisador coloca o URL específico em fila para pesquisar na vez seguinte que verificar por páginas que estejam a aguardar para serem visitadas (tipicamente a cada dez minutos). Se o pesquisador não estiver em execução, esta coloca o URL especificado em fila de forma a poder ser pesquisado na vez seguinte que o pesquisador for iniciado.

Certifique-se de que as regras de pesquisa incluem a regra que permite ao pesquisador visitar os URLs especificados. O pesquisador pode visitar os URLs especificados mais cedo do que o faria normalmente. Todavia, para um URL a ser pesquisado de todo, tem de existir uma regra de pesquisa que permita ao URL ser pesquisado.

Os dados recém-pesquisados ficam disponíveis para procura da próxima vez que ocorrer a criação de índice principal.

Como o pesquisador da Web processa páginas de erros esporádicos

Pode configurar o pesquisador para processar páginas personalizadas que os administradores do sítio da Web criam quando não pretendem devolver um código do registo de erros comum em resposta aos pedidos de certas páginas.

Se um servidor de HTTP não puder devolver a página que um cliente pede, normalmente o servidor devolve uma resposta que consiste num cabeçalho com um código de estado. O código de estado indica qual é o problema (tal como o erro 404, que indica que não foi possível encontrar o ficheiro). Alguns administradores do sítio da Web criam páginas especiais que explicam o problema de forma mais detalhada e configuram o servidor de HTTP para devolver essas páginas. Essas páginas personalizadas são chamadas de *páginas de erros esporádicos*.

As páginas de erro recuperável podem distorcer os resultados do pesquisador da Web. Por exemplo, em vez de receber um cabeçalho que indica um problema, o pesquisador recebe uma página de erros esporádicos e o código de estado 200, o que indica a transferência bem-sucedida de uma página de HTML válida. Contudo, esta página de erro recuperável transferida não está relacionada com o URL solicitado, e o seu conteúdo é quase idêntico sempre que é devolvida em vez da página solicitada. Estas páginas irrelevantes e quase idênticas distorcem o índice e os resultados da procura.

Para processar essa situação, pode especificar as opções para processar páginas de erro esporádico quando configura o pesquisador da Web. O pesquisador da Web necessita da seguinte informação acerca de cada página da Web que devolve páginas de erro esporádico:

- Um padrão de URL para um sítio da Web que utiliza páginas de erro esporádico. Esse padrão de URL consiste no protocolo (HTTP ou HTTPS), no nome do sistema central, no número da porta (se não for padrão), e no nome do caminho. Pode utilizar um asterisco (*) como carácter global para corresponder a um ou vários caracteres até à ocorrência seguinte de um carácter não global no padrão. O padrão especificado é sensível a maiúsculas e minúsculas.
- Um padrão de título para o texto que corresponde ao controlo <TITLE> de um documento HTML. Pode utilizar o asterisco (*) como carácter global para especificar este padrão. O padrão especificado é sensível a maiúsculas e minúsculas.
- Um padrão de conteúdo para o texto que corresponde ao conteúdo de um documento HTML. O conteúdo não é só o conteúdo do controlo <BODY>, se um controlo<BODY> estiver presente. O conteúdo é tudo o que se segue ao cabeçalho HTTP do ficheiro. Pode utilizar o asterisco (*) como carácter global para especificar este padrão. O padrão especificado é sensível a maiúsculas e minúsculas.

 Um número inteiro que representa o código de estado a utilizar para documentos que correspondam aos padrões de URL, título e conteúdo especificados.

Exemplo

A configuração que se segue indica ao pesquisador da Web que compare todas as páginas de HTML válidas (código de estado 200) que são devolvidas do sítio da Web http://www.mysite.com/hr/* aos padrões de título e conteúdo especificados. Se o controlo <TITLE> de uma página começar por "Lamento, a página" e o conteúdo do documento contiver qualquer coisa (*), o pesquisador trata a página tal como um código de estado 404 (a página não foi encontrada).

Tal	bela	a 1	6.	Exemplo	de	página	de	erro	esporádico	
-----	------	-----	----	---------	----	--------	----	------	------------	--

Padrão de URL	Padrão do título	Padrão de conteúdo	Código de estado de HTTP
http://www.mysite.com/hr/*	Desculpe, a página*	*	404

Pode criar entradas múltiplas para o mesmo sítio da Web para processar códigos de estado diferentes. Cada código de estado do mesmo sítio da Web requer a sua própria entrada na configuração do pesquisador da Web.

Utilizar caracteres globais

Os padrões de URL, título e conteúdo não são expressões globais. O carácter do asterisco corresponde a caracteres até que apareça um carácter não global. Por exemplo:

*404 corresponde a quaisquer caracteres404
404: * corresponde a quaisquer caracteres 404:
http://*.mysite.com/* corresponde a
http://qualquer sistema central.mysite.com/qualquer ficheiro
* corresponde a quaisquer caracteres

Efeito no desempenho

Quando configurar as opções para processar páginas de erro esporádico, aumenta a duração do tempo de processamento do pesquisador, porque todas as páginas pesquisadas de forma bem-sucedida terão de ser verificadas. É necessário mais tempo de processamento para procurar correspondências de padrões e determinar se deverá ser devolvida uma página ou um código de estado de substituição.

Suporte para pesquisa de sítios da Web protegidos

Ao especificar credenciais na consola de administração, pode activar o pesquisador da Web para aceder a conteúdo restrito, como documentos que necessitam de uma palavra-passe para o acesso.

Se um servidor da Web utilizar uma autenticação base de HTTP ou uma autenticação baseada em formulário HTTP para restringir o acesso ao sítios da Web, pode especificar as credenciais na configuração do pesquisador da Web que permite a pesquisa de páginas nos sítios da Web protegidos por palavra-passe. Também pode especificar as opções para configurar manualmente os ficheiros dos cookies.

Conceitos relacionados:

"Sítios da Web protegidos por autenticação base HTTP" "Sítios da Web protegidos por autenticação baseada em formulário" na página 148

Sítios da Web protegidos por autenticação base HTTP

Se um servidor da Web utilizar uma autenticação base HTTP para restringir o acesso ao sítios da Web, pode especificar as credenciais da autenticação que permitem ao pesquisador da Web aceder às páginas protegidas por palavra-passe.

Para determinar se um utilizador (ou aplicação cliente) tem permissão para aceder às páginas num sítio da Web, muitos servidores da Web utilizam um esquema de autenticação cliente chamado autenticação base de HTTP para estabelecer a identidade do utilizador. Normalmente, essa interacção é interactiva:

- Quando um agente de utilizador de HTTP (tal como um navegador) solicita uma página que está protegida por autenticação base de HTTP, o servidor da Web responde com um código de estado 401, que indica que o solicitador não está autorizado a aceder à página solicitada.
- O servidor da Web também desafia o solicitador a apresentar as credenciais que podem ser utilizadas para verificar se o utilizador está autorizado a aceder ao conteúdo restrito.
- O navegador apresenta o utilizador com um diálogo que solicita um nome do utilizador, uma palavra-passe e qualquer outra informação que é solicitada para constituir as credenciais do utilizador.
- O navegador codifica as credenciais, inclui-as quando repete o pedido da página protegida.
- Se as credenciais forem válidas, o servidor da Web responde com um código de retorno 200 e com os conteúdos da página solicitada.
- Os pedidos seguintes de páginas a partir do mesmo servidor da Web incluem, normalmente, as mesmas credenciais, permitindo ao utilizador autorizado aceder ao conteúdo restrito adicional sem especificar credenciais adicionais.

Após ser estabelecida a identidade de um utilizador, o servidor da Web e o agente de utilizador de HTTP, normalmente, trocam tokens chamados *cookies*, que dão a conhecer o estado de início de sessão do utilizador e que deve ser mantido entre pedidos de HTTP.

Como o pesquisador não é executado de forma interactiva, as credenciais que lhe permitem pesquisar páginas protegidas por palavra-passe têm de ser especificadas antes de o pesquisador iniciar a pesquisa. Ao criar um pesquisador da Web ou ao editar o espaço de pesquisa, especifique a informação acerca de cada sítio da Web protegido que tem de ser pesquisado.

Para especificar esta informação, tem de trabalhar mais perto dos administradores dos sítios da Web ou dos servidores da Web que estão protegidos pela autenticação base HTTP. Esses têm de lhe fornecer os requisitos de segurança dos sítios da Web a pesquisar, incluindo toda a informação utilizada para autenticar a identidade do pesquisador da Web e para determinar que o pesquisador tem permissão para pesquisar as páginas restritas.

Se a segurança tiver sido activada para a colecção quando esta foi criada, pode especificar tokens de segurança, tais como IDs de utilizador, IDs de grupo ou funções de utilizador para controlar o acesso a documentos quando configura o pesquisador. O pesquisador da Web associa estes tokens de segurança a todos os documentos pesquisados na árvore do sistema de ficheiros do URL raiz especificado. Os tokens são utilizados juntamente com os tokens de segurança ao nível de documentos que configura para a totalidade do espaço de pesquisa da Web.

A ordem dos URLs é importante. Depois de adicionar informação acerca de um sítio da Web protegido por palavra-passe, tem de o posicionar na ordem pela qual pretende que o pesquisador o processe. Liste primeiro os URLs mais específicos, e coloque os URLs mais genéricos mais abaixo na lista. Quando o pesquisador da Web avalia um possível URL, utiliza os dados de autenticação especificados para o primeiro URL da lista que corresponde ao possível URL.

Conceitos relacionados:

"Suporte para pesquisa de sítios da Web protegidos" na página 146

Sítios da Web protegidos por autenticação baseada em formulário

Se um servidor da Web utilizar formulários HTML para restringir o acesso aos sítios da Web, pode especificar as credenciais de autenticação que permitem ao pesquisador aceder às páginas protegidas por palavra-passe.

Para determinar se um utilizador (ou aplicação cliente) tem permissão para aceder às páginas num sítio da Web, muitos servidores da Web utilizam formulários HTML para estabelecer a identidade do utilizador. Normalmente, essa interacção é interactiva:

- Quando um agente do utilizador HTTP (tal como um navegador) solicitar uma página que esteja protegida por autenticação baseada em formulário, o servidor da Web verifica se o pedido inclui um cookie que estabeleça a identidade do utilizador.
- Se a cookie não estiver presente, o servidor da Web pede ao utilizador para introduzir os dados de segurança num formulário. Se o utilizador submeter o formulário, o servidor da Web devolve os cookies solicitados e o pedido da página protegida por palavra-passe pode continuar.
- Os pedidos futuros que incluem os cookies requeridos também podem continuar. O utilizador autorizado pode aceder a conteúdo restrito adicional sem que lhe seja pedido que preencha um formulário e especifique credenciais com cada pedido.

Como o pesquisador não é executado de forma interactiva, as credenciais que lhe permitem pesquisar páginas protegidas por palavra-passe têm de ser especificadas antes de o pesquisador iniciar a pesquisa. Ao criar um pesquisador da Web ou ao editar o espaço de pesquisa, especifique a informação acerca de cada sítio da Web protegido que tem de ser pesquisado.

Os campos que especifica correspondem aos campos que um utilizador interactivo preenche quando solicitado pelo navegador, e a quaisquer campos ocultos ou estáticos requeridos para um início de sessão bem sucedido.

Para especificar esta informação, tem de trabalhar mais perto dos administradores dos sítios da Web ou dos servidores da Web que estão protegidos pela autenticação baseada em formulário. Esses têm de lhe fornecer os requisitos de segurança dos sítios da Web a pesquisar, incluindo toda a informação utilizada para autenticar a identidade do pesquisador da Web e para determinar que o pesquisador tem permissão para pesquisar as páginas restritas.

A ordem dos padrões URL é importante. Depois de adicionar informação acerca de um sítio da Web protegido por palavra-passe, tem de o posicionar na ordem pela qual pretende que o pesquisador o processe. Liste primeiro os padrões de URLs mais específicos, e coloque os padrões de URLs mais genéricos mais abaixo na lista. Quando o pesquisador da Web avalia um possível URL, utiliza os dados do formulário especificados para o primeiro padrão de URL da lista que corresponde ao possível URL.

Utilizar um suplemento para pesquisar sítios WebSphere Portal seguros

Se a segurança global estiver activada em WebSphere Application Server, e se pretender pesquisar sítios WebSphere Portal seguros com o pesquisador da Web, tem de criar um suplemento de pesquisador para processar os pedidos de autenticação baseada em formulário. Para uma discussão acerca da autenticação baseada em formulário e um programa de amostra que pode adaptar ao suplemento de pesquisador da Web personalizado, consulte http://www.ibm.com/ developerworks/db2/library/techarticle/dm-0707nishitani.

O suplemento é necessário se utilizar o pesquisador da Web para pesquisar quaisquer sítios através de WebSphere Portal, incluindo os sítios IBM Web Content Manager e Quickr para sítios WebSphere Portal.

Conceitos relacionados:

"Suporte para pesquisa de sítios da Web protegidos" na página 146

Sítios da Web que são servidos por servidores proxy

Se o pesquisador não permitir o acesso directo a uma rede, pode configurar o pesquisador para utilizar um servidor proxy HTTP para aceder ao conteúdo que pretende pesquisar.

Se o acesso a uma rede TCP/IP não estiver disponível no computador em que o pesquisador da Web será executado, ou se o acesso for restrito a processos privilegiados, pode configurar o pesquisador da Web para utilizar um servidor proxy HTTP. Um proxy HTTP é um processo de recepção de pedidos HTTP numa porta especificada de um determinado sistema central. O servidor proxy retransmite os pedidos ao servidor da Web e retransmite as respostas do servidor da Web ao cliente que pede (pesquisador da Web). Um servidor proxy pode ser executado no mesmo computador que o pesquisador da Web, ou noutro computador.

Numa pesquisa não proxy, um pedido por um URL é enviado directamente para o sistema central. Por meio da pesquisa proxy, o pedido é enviado para o servidor proxy.

Quando cria um pesquisador da Web ou edita o espaço do pesquisador, especifique informações acerca dos servidores proxy que o pesquisador da Web utiliza quando pesquisa páginas no domínio do servidor proxy. Antes de adicionar um servidor proxy ao espaço do pesquisador, obtenha os nomes dos domínios que são assistidos pelo servidor proxy, o nome do sistema central ou endereço de IP do servidor proxy e o número da porta utilizada pelo servidor proxy.

Se o servidor proxy requerer autenticação, obtenha também um nome de utilizador e palavra-passe que o pesquisador possa utilizar para aceder às páginas assistidas pelo servidor proxy. O pesquisador da Web suporta apenas autenticação de proxy básica HTTP, como está descrito no protocolo de transferência de hipertexto RFC 2616, e a autenticação básica de Windows NT LAN Manager (NTLM). Outros tipos de autenticação, incluindo NTLM Versão 2, não são suportados.

Depois de adicionar um servidor proxy, tem de o seleccionar e posicionar na ordem pela qual pretende que o pesquisador o processe. Liste primeiro os nomes de domínio mais específicos, e coloque os nomes de domínio mais genéricos mais abaixo na lista. Quando o pesquisador da Web avalia um possível URL, utiliza os dados do servidor proxy especificados para o primeiro domínio da lista que corresponde ao possível URL. (Os URLs que não correspondem a nenhuma regra de proxy são assumidos como directamente acessíveis para o pesquisador.)

Administração de cookies

Normalmente, a administração de cookies ocorre automaticamente, não sendo requerida nenhuma acção de um administrador. Se necessário, pode especificar manualmente os cookies para uma sessão de pesquisa da Web.

Os cookies são tokens opacos que um servidor da Web devolve a um agente do utilizador como parte de um cabeçalho de resposta de HTTP. Apenas são significativos para o servidor da Web que os emite, e são utilizados para manter estado entre pedidos HTTP. Durante a autenticação cliente, o servidor da Web pode devolver um cookie que permite ao servidor determinar que um utilizador autenticado já iniciou a sessão. A presença de um cookie permite ao utilizador emitir pedidos adicionais para páginas no servidor da Web sem lhe ser pedido para reiniciar a sessão.

O pesquisador da Web guarda os cookies recebidos dos servidores da Web e utiliza-os durante a instância do pesquisador. Armazena os cookies num ficheiro cookies.ini, que é escrito de novo pelo pesquisador no fim de cada sessão do pesquisador. Quando o pesquisador pára, guarda todos os cookies não expirados e, em seguida, recarrega-os no início da sessão seguinte.

Se especificar os cookies manualmente, armazene-os num ficheiro separado e, em seguida, intercale-os com os cookies no ficheiro cookies.ini quando for necessário. O pesquisador não elimina os cookies não expirados, mas se um problema impedir a gravação de toda a colecção de cookies, não pretende perder os cookies manualmente especificados. Tem de intercalar os respectivos cookies que o pesquisador mantém automaticamente antes de iniciar a sessão de pesquisa.

Formato de cookie

Os cookies que pretende intercalar com o ficheiro cookies.ini tem de estar num determinado formato.

- Cada cookie tem de estar numa única linha. As linhas em branco e os comentários são permitidos, mas não serão mantidos no ficheiro cookies.ini.
- Cada cookie tem de ter o seguinte formato: CookieN(cookie length,URL length)cookie text,validation URL

Cookie

Uma palavra-chave requerida que indica o início duma entrada de cookie.

A palavra-chave de cookie contém espaços em branco e tem de ter um único dígito anexado ou 0, 1 ou 2. O dígito indica o tipo de cookie: versão-0 (Netscape), versão-1 (RFC2109), ou versão-2 (RFC2965). As listas da porta não são suportadas em cookies RFC2965.

comprimento_cookie

O comprimento em caracteres do texto do cookie associado.

comprimento_URL

O comprimento em caracteres do URL de validação associado.

texto_cookie

O conteúdo do cookie que vai ser enviado para o servidor da Web de origem. Esta cadeia (que representa o lado direito da directiva Conjunto-Cookie no cabeçalho de resposta HTTP) especifica o nome do cookie e o par de valor e qualquer outro conteúdo (tal como um caminho, uma definição de segurança, etc) a ser enviado com o cookie. Esta cadeia é seguida por um separador de vírgula (,).

URL_validação

O URL no qual o cookie é identificado. O URL é utilizado para determinar para onde o cookie deve ser enviado (por exemplo, fornecendo um nome do domínio e um nome do caminho). O URL de validação tem de satisfazer a segurança do servidor da Web de origem e as restrições de privacidade.

O exemplo seguinte é mostrado em duas linhas para leitura; os cookies que especificar têm de estar numa única linha:

Cookie0(53,40)ASPSESSIONIDQSQTACSD=SLNSIDFNLISDNFLSIDNFLSNL;path=/, https://www.ibm.com:443/help/solutions/

Configurar cookies para o pesquisador Web

Pode especificar manualmente os cookies numa sessão de pesquisa da Web, e intercalá-las com cookies que o pesquisador mantém.

Para configurar manualmente cookies num pesquisador Web:

- 1. Na consola administrativa, para o pesquisador Web no qual pretende especificar cookies.
- 2. Inicie sessão como administrador predefinido do IBM Content Analytics with Enterprise Search no servidor do pesquisador.
- **3**. Mude para o directório de dados do pesquisador que pretende configurar, em que *ID_sessão_pesquisador* é um ID que foi atribuído à sessão do pesquisador pelo sistema. Por exemplo:

```
ES_NODE_ROOT/data/col_56092.WEB_88534
```

- Edite o ficheiro cookies.ini, anexe as entradas de cookies manualmente especificadas às já listadas e, em seguida, guarde e saia do ficheiro. Certifique-se de que os respectivos cookies não substituem as já existentes.
- 5. Na consola administrativa, reinicie o pesquisador da Web que parou.

Configuração do espaço global da pesquisa Web

Pode configurar o espaço global de pesquisa global para pesquisadores da Web, o que lhe permite controlar melhor a remoção de URLs do índice.

Cada pesquisador da Web é configurado com um espaço de pesquisa que define os URLs que devem ser ou não pesquisados. Os URLs identificados que se encontram no espaço de pesquisa estão guardados (numa base de dados) para serem pesquisados posteriormente. Os URLs que não se encontram no espaço de pesquisa são eliminados. Se o pesquisador for iniciado com uma base de dados vazia, a definição do espaço de pesquisa e a base de dados permanecem consistentes enquanto o pesquisador é executado.

Às vezes, o pesquisador está parado, e o respectivo espaço de pesquisa é reduzido (por exemplo, por novas regras que proíbem as páginas de serem pesquisadas). Quando o pesquisador é reiniciado, a respectiva definição de espaço de pesquisa e a base de dados tornam-se inconsistentes. A base de dados contém URLs (alguns pesquisados, outros não) que não se encontram no novo e mais reduzido espaço de pesquisa. Se uma colecção tiver apenas um pesquisador da Web, este poderá restaurar a consistência, alterando os códigos de retorno de HTTP desses URLs para 760 (o que especifica que devem ser excluídos) e solicitando a remoção de páginas recém-excluídas do índice.

Se dividir o espaço de pesquisa entre dois ou mais pesquisadores da Web (por exemplo, para garantir que algumas páginas sejam pesquisadas com mais frequência que as restantes), cada pesquisador da Web manterá tabelas de bases de dados independentes (inicialmente vazias), e cada uma poderá pesquisar uma parte diferente do espaço de pesquisa da Web. O espaço de pesquisa do pesquisador original é então reduzido para o que quer que esteja à esquerda, depois de as partes a serem pesquisadas por outros pesquisadores serem removidas. Os problemas surgem quando o pesquisador original tenta restaurar a consistência, removendo as páginas movidas do índice. Como as páginas movidas estão a ser pesquisadas por outros pesquisadores, as páginas deviam permanecer no índice.

Ao configurar um espaço global do pesquisador de nível superior, pode identificar URLs que não devem ser pesquisados pelo pesquisador original, mas que também não devem ser removidos do índice. Os URLs que já não estão em qualquer espaço de pesquisa do pesquisador continuam a ser marcados para exclusão pelos processos de identificação, e são removidos do índice quando são pesquisados de novo.

O espaço de pesquisa global é definido por um ficheiro de configuração chamado global.rules, o que tem de existir no directório de configuração do pesquisador (a presença de um ficheiro global.rules activa a função do espaço global do pesquisador). Se este ficheiro existir, será lido durante a inicialização do pesquisador. Se este ficheiro não existir, o pesquisador opera com um espaço de pesquisa de nível único, e remove documentos do índice conforme seja necessário para manter a consistência entre a respectiva definição do espaço de pesquisa e a base de dados.

Se existir um espaço global de pesquisa, o pesquisador dirige os URLs como antes, mas pedirá a remoção de um URL do índice apenas se o URL não se encontrar em nenhum espaço de pesquisa da Web.

O ficheiro global.rules tem a mesma sintaxe que o ficheiro crawl.rules local, excepto que esse pode conter apenas regras do nome do domínio. Essa restrição permite ao espaço de pesquisa ser particionado entre pesquisadores apenas com base de nomes do sistema central DSN, não endereços de IP ou padrões de prefixo HTTP. Os URLs que são excluídos pelo prefixo URL ou por regras do endereço de IP no espaço de pesquisa local (como definido no ficheiro crawl.rules) não são afectados pelo espaço global do pesquisador; tais URLs ainda estão excluídos.

O espaço global de pesquisa é utilizado apenas para impedir a remoção de URLs, que são excluídos de um espaço de pesquisa de um pesquisador por uma regra do domínio local, do índice. As regras seguintes são aplicadas na seguinte ordem:

- Se um URL da base de dados do pesquisador for excluído por uma regra de prefixo local ou regra de endereço, o código de retorno 760 é atribuído ao URL e este é removido do índice. O URL não será pesquisado de novo.
- 2. Se um URL da base de dados do pesquisador for excluído por uma regra de domínio local, e não houver um espaço global de pesquisa, o código de retorno 760 é atribuído ao URL e este é removido do índice. O URL não será pesquisado de novo.

- 3. Se um URL da base de dados do pesquisador for excluído por uma regra do domínio local, mas explicitamente permitido por uma regra no espaço global do pesquisador, será atribuído o código de retorno 761 ao URL. O pesquisador não pesquisará o URL de novo, mas não será removida do índice (assume-se que esteja em algum espaço de pesquisa local do pesquisador).
- 4. Se um URL da base de dados do pesquisador for excluído por uma regra de domínio local e não explicitamente permitido por uma regra no espaço global do pesquisador, será atribuído o código de retorno 760 ao URL e este será removido do índice.

Como o espaço de pesquisa global é consultado apenas para impedir a eliminação de URLs já excluídos pelo espaço de pesquisa local, o resultado predefinido do espaço de pesquisa global, se não for aplicada uma regra a um possível URL, está proibido de ser pesquisado.

O ficheiro global.rules tem de existir no directório master_config de cada pesquisador que partilha o espaço global do pesquisador. Tem de editar com cuidado todas as cópias do ficheiro global.rules e dos ficheiros crawl.rules locais individuais para garantir que permaneçam mutuamente consistentes.

Directivas para não seguir ou não indexar

Pode melhorar a qualidade da procura, especificando directivas para o pesquisador da Web que controlam, se as ligações em páginas são seguidas ou se as páginas são indexadas.

Algumas páginas da Web têm directivas para não seguir ou para não indexar, que informam os robots (tais como o pesquisador da Web) para não seguir as ligações encontradas nessas páginas, para não incluir os conteúdos dessas páginas no índice, ou para não executar nenhuma dessas acções.

O controlo dessas definições pode melhorar a qualidade da pesquisa. Por exemplo, algumas páginas do directório podem conter milhares de ligações mas nenhum conteúdo útil; essas páginas deviam ser pesquisadas e as respectivas ligações seguidas, mas não há qualquer benefício em indexar as próprias páginas do directório.

Pode haver vezes em que quando pretende que o pesquisador não aceda a uma hierarquia inferior, mas as páginas de folha desejadas contêm ligações e não contêm directivas para não seguir. Como algumas dessas páginas são geradas automaticamente, não têm proprietários que possam inserir as directivas requeridas.

Para especificar as regras para pesquisar essas páginas, crie ou edite um ficheiro de configuração chamado followindex.rules. Utilize as seguintes directrizes quando especificar as regras neste ficheiro:

- As regras que configura têm de especificar prefixos URL (não pode identificar sítios da Web por meio de um endereço IP ou de um nome do sistema central DSN).
- Os prefixos URL podem incluir asteriscos (*) como um carácter global para permitir ou proibir sítios múltiplos com URLs semelhantes.
- A ordem é significativa (o pesquisador aplica a primeira regra que corresponde a um possível URL).
- As regras que de forma explícita permitem ou proíbem seguir ou indexar, substituem as outras definições, incluindo aquelas no documento de destino.

Substituir directivas para não seguir ou não indexar em páginas da Web

Pode especificar regras num ficheiro de configuração para controlar se o pesquisador segue ligações a páginas ou indexa páginas que contêm directivas para não seguir ou não indexar.

A menos que a página inclua uma etiqueta META robots, as directivas que especificar substituem as directivas existentes na página que vai ser pesquisada. O pesquisador da Web tenta observar o Protocolo de Exclusão de Robots e não utilizará o ficheiro followindex.rules se a página da Web incluir uma etiqueta de META robots. O ficheiro followindex.rules é utilizado para definir regras de robots adicionais para as páginas que não têm META robots. Caso a página inclua uma etiqueta META robots, como por exemplo <meta name="robots" content="noindex", nofollow">, a página não será pesquisada.</u>

Para substituir directivas para não seguir e para não indexar:

- 1. Na consola de administração, pare o pesquisador da Web no qual pretende configurar algumas regras.
- 2. Inicie sessão como administrador predefinido do IBM Content Analytics with Enterprise Search no servidor do pesquisador.
- Mude para o directório de configuração do pesquisador que pretende configurar, em que *ID_sessão_pesquisador* é um ID que foi atribuído à sessão do pesquisador pelo sistema. Por exemplo: ES NODE ROOT/master config/*col 56092.WEB 88534*
- 4. Crie ou edite um ficheiro chamado followindex.rules.
- 5. Escreva regras para o pesquisador no formato seguinte, em que *URLprefix* são os caracteres de início dos sítios da Web que pretende permitir ou proibir que sejam seguidos ou indexados:

forbid follow URLprefix allow follow URLprefix forbid index URLprefix allow index URLprefix

- 6. Guarde e feche o ficheiro.
- 7. Na consola administrativa, reinicie o pesquisador da Web que parou.

Pesquisadores do Web Content Management

Para incluir documentos IBM Web Content Manager numa colecção, crie um pesquisador Lista de valores geradores. O pesquisador Web Content Management apenas é facultado para suportar migrações de versões anteriores de IBM Content Analytics with Enterprise Search.

Credenciais de ligação do pesquisador

Quando cria o pesquisador, pode especificar as credenciais que permitem que o pesquisador se ligue às fontes a pesquisar. Também pode configurar as credenciais de ligação quando especifica as definições gerais de segurança do sistema. Se utilizar a última abordagem, vários pesquisadores e outros componentes do sistema podem utilizar as mesmas credenciais. Por exemplo, os servidores de procura podem utilizar as credenciais para determinar se um utilizador possui autorização para aceder ao conteúdo.

Formato de URL

Quando especificar os URLs a pesquisar, tem de utilizar o seguinte formato:

protocolo_http://nomesistemacentral_portal:número_porta/prefixo_portal /WCM_search_seed_servlet_path/searchseed?site=WCM_site_name&lib=WCM_library_name

O exemplo seguinte mostra o URL de um sítio no caminho predefinido de instalação de IBM Web Content Manager em WebSphere Portal:

http://portal.server.ibm.com:80/wps/wcmsearchseed/
searchseed?site=SiteTest01&lib=Web+Content

Se o nome do sítio ou da biblioteca contiver espaços, tem de substituir o espaço por um sinal de mais (+). Por exemplo, substitua Web Content por Web+Content.

Descrição geral da configuração

Para pesquisar sítios IBM Web Content Manager, tem de integrar WebSphere Portal e IBM Content Analytics with Enterprise Search. Para obter instruções sobre como configurar a integração, consulte a integração de *WebSphere Portal com IBM Content Analytics with Enterprise Search*.

É possível utilizar o pesquisador Web Content Management para pesquisar qualquer número de sítios IBM Web Content Manager. Quando configura o pesquisador, especifica os URLs para os sítios a pesquisar. Em seguida, o pesquisador transfere as páginas que estão ligadas a partir dos sítios especificados.

Os sítios a serem pesquisados têm de ser acessíveis pelo mesmo ID e palavra-passe de administrador de WebSphere Portal. Para pesquisar sítios que utilizem credenciais diferentes, tem de configurar um pesquisador do Web Content Management separada.

Tarefas relacionadas:

"Configurar credenciais de ligação" na página 380

Referências relacionadas:

"Formatos de URI no índice" na página 164

"Nomes de campos de índice reservados" na página 199

Integração do WebSphere Portal com IBM Content Analytics with Enterprise Search

Pesquisadores do WebSphere Portal

Para incluir páginas de sítios IBM WebSphere Portal numa colecção, crie um pesquisador Lista de valores geradores. O pesquisador WebSphere Portal apenas é facultado para suportar migrações de versões anteriores do IBM Content Analytics with Enterprise Search.

Credenciais de ligação do pesquisador

Quando cria o pesquisador, pode especificar as credenciais que permitem que o pesquisador se ligue às fontes a pesquisar. Também pode configurar as credenciais de ligação quando especifica as definições gerais de segurança do sistema. Se utilizar a última abordagem, vários pesquisadores e outros componentes do sistema podem utilizar as mesmas credenciais. Por exemplo, os servidores de

procura podem utilizar as credenciais para determinar se um utilizador possui autorização para aceder ao conteúdo.

Descrição geral da configuração

Para pesquisar sítios WebSphere Portal, tem de integrar WebSphere Portal e IBM Content Analytics with Enterprise Search. Para obter instruções sobre como configurar a integração, consulte a integração de *WebSphere Portal com IBM Content Analytics with Enterprise Search*.

É possível utilizar o pesquisador WebSphere Portal para pesquisar um único local WebSphere Portal. Ao configurar o pesquisador, especifique o URL para o sítio do portal a pesquisar. Em seguida, o pesquisador descarrega as páginas que estão ligadas a partir do URL de portal especificado. Para pesquisar outro sítio de portal, crie outro pesquisador.

Referências relacionadas:

Integração do WebSphere Portal com IBM Content Analytics with Enterprise Search

Copiar o URL para pesquisar no WebSphere Portal

Para reduzir a possibilidade de escrever um URL incorrecto, pode copiar e colar o URL do sítio que pretende pesquisar no campo apropriado quando configurar um pesquisador do WebSphere Portal ou Web Content Management.

Quando cria um pesquisador WebSphere Portal ou Web Content Management, especifica o URL do sítio no servidor WebSphere Portal que pretende pesquisar. Como os URLs são longos e normalmente contêm caracteres não ASCII codificados, poderá querer utilizar este procedimento para copiar o URL do servidor do WebSphere Portal e colá-lo na consola administrativa.

Para especificar o URL que pretende que o pesquisador pesquise:

- 1. Quando estiver pronto a especificar o URL ou URLs a pesquisar na consola de administração, inicie sessão no WebSphere Portal como administrador.
- **2**. Se utilizar o WebSphere Portal versão 6.1 ou 6.1.5, complete os seguintes passos no servidor WebSphere Portal:
 - a. Faça clique em Administração, no canto inferior esquerdo.
 - b. Faça clique em Gerir Procura na área de navegação à esquerda.
 - c. Na página Gerir Procura, faça clique na ligação Colecções de Procura.
 - d. Na página Gerir Procura, faça clique em **Serviço de Procura do Portal Predefinido**.
 - e. Na tabela Colecções de Procura, faça clique na colecção **Conteúdo do Portal**.
 - f. Na tabela Origens de Conteúdo, faça clique no ícone Editar na extremidade direita (o ícone do lápis), junto a **Origem do Conteúdo do Portal**.
 - g. Para o **Tipo de origem do conteúdo**, faça clique em **Sítio do Portal**. O URL do sítio é apresentado no campo **Reunir documentos ligados por esse URL**.
 - h. Copiar o URL para a área de transferência. Por exemplo, destaque o URL e retenha a tecla Ctrl enquanto prime a tecla Insert.
- **3**. Regresse à consola de administração e cole o URL que copiou no campo do URL do sítio.

Pesquisadores do Sistema de ficheiros Windows

Para incluir documentos que estão armazenados nos sistemas de ficheiros Microsoft Windows numa colecção, pode configurar um pesquisador do Sistema de ficheiros Windows.

Se instalar o servidor do pesquisador em AIX ou Linux, não pode pesquisar origens do sistema de ficheiros Windows com o pesquisador do Sistema de ficheiros Windows (o pesquisador não surge na lista de tipos de pesquisadores disponíveis). No entanto, pode instalar um servidor de agente que lhe permite pesquisar sistemas de ficheiros Windows remotos com o pesquisador do Agente para sistemas de ficheiros Windows.

Pode utilizar o pesquisador do Sistema de ficheiros Windows para pesquisar um número de sistemas de ficheiros Windows. Ao configurar o pesquisador, seleccione os directórios locais e remotos e os subdirectórios que pretenda pesquisar.

O ID de utilizador que o pesquisador utiliza para aceder a documentos a serem pesquisados deve ter os seguintes direitos de administrador Windows:

- Conteúdos da Pasta da Lista. Este direito permite ao pesquisador listar os documentos na pasta.
- Permissões de Leitura. Este direito permite ao pesquisador aceder a informações da lista de controlo de acesso (ACL) para cada documento.

Pesquisa de directórios de rede partilhados

O pesquisador do Sistema de ficheiros Windows pesquisa documentos de acordo com permissões de leitura que são especificadas para o administrador. O administrador é a conta de serviços do IBM Content Analytics with Enterprise Search.

Pode especificar um ID de utilizador e palavra-passe para os directórios a pesquisar. No entanto, o ID de utilizador e palavra-passe só são utilizados para estabelecer ligação com directórios de rede partilhados. O pesquisador pesquisa ficheiros de acordo com as permissões de leitura que estão definidas para este utilizador para directórios de rede partilhados e não para unidades locais.

As ligações a directórios de rede não são canceladas até que reinicie o serviço do IBM Content Analytics with Enterprise Search. Uma vez estabelecida uma ligação, é possível aceder ao directório com um ID de utilizador e palavra-passe incorrectos. No entanto, esta ligação só é permitida para sessões de descoberta e de pesquisador do Sistema de ficheiros Windows que estão sob o controlo do sistema. Para impedir possíveis riscos de segurança, certifique-se de que as autorizações para a conta de administrador (em que é executado o serviço do IBM Content Analytics with Enterprise Search) são definidas correctamente.

Para evitar problemas com a ligação futura a um directório de rede, especifique o mesmo ID de utilizador e palavra-passe para o mesmo directório de rede. Se especificar o ID de utilizar e palavra-passe incorrectos e reiniciar o serviço do IBM Content Analytics with Enterprise Search, o pesquisador de Sistema de ficheiros Windows poderá não conseguir executar a pesquisa, porque tenta estabelecer ligação ao directório com as credenciais incorrectas. A pesquisa pode ter êxito se a ligação à rede for estabelecida por outro pesquisador do Sistema de ficheiros Windows que esteja a utilizar o ID de utilizador e palavra-passe correctos.

Credenciais de ligação do pesquisador

Quando cria o pesquisador, pode especificar as credenciais que permitem que o pesquisador se ligue às fontes a pesquisar. Também pode configurar as credenciais de ligação quando especifica as definições gerais de segurança do sistema. Se utilizar a última abordagem, vários pesquisadores e outros componentes do sistema podem utilizar as mesmas credenciais. Por exemplo, os servidores de procura podem utilizar as credenciais para determinar se um utilizador possui autorização para aceder ao conteúdo.

Descrição geral da configuração

Ao criar o pesquisador, um assistente ajudá-lo-á nessas tarefas:

- Especifique as propriedades que controlam o modo como o pesquisador opera e utiliza os recursos do sistema. As propriedades do pesquisador controlam o procedimento do pesquisador para pesquisar todos os subdirectórios no espaço de pesquisa.
- Defina um agendamento para a pesquisa de sistemas de ficheiros.
- Seleccione os subdirectórios a pesquisar.

Pode especificar o número de níveis de subdirectórios que pretende que o pesquisador pesquise. Para pesquisar sistemas de ficheiros remotos, especifique também um ID de utilizador e uma palavra-passe que permita ao pesquisador aceder aos dados.

- Especifique as opções para fazer com que os documentos possam ser procurados em subdirectórios. Por exemplo, pode excluir certos tipos de documentos do espaço de pesquisa ou especificar um ID de utilizador e uma palavra-passe que permita ao pesquisador aceder aos ficheiros num determinado subdirectório.
- Configure as opções de segurança ao nível do documento. Se a segurança tiver sido activada quando a colecção foi criada, o pesquisador pode associar dados de segurança aos documentos existentes no índice. Esses dados permitem às aplicações fazer cumprir os controlos de acesso baseados nas listas de controlo de acesso armazenadas ou nos tokens de segurança.

Pode seleccionar uma opção para validar as credenciais do utilizador no momento em que um utilizador permitir uma consulta. Nesse caso, em vez de comparar as credenciais do utilizador com os dados de segurança indexados, o sistema compara as credenciais com as listas de controlo de acesso actuais que são actualizadas pela origem de dados original.

Para reforçar a segurança ao nível do documento, tem de garantir que o utilizador a informação de conta do domínio seja configurada correctamente no servidor do pesquisador.

Para maximizar o rendimento, o pesquisador do Sistema de ficheiros Windows pode detectar actualizações a ficheiros e determinar se um ficheiro necessita de ser pesquisado novamente sem o abrir.

Pesquisadores do sistema de ficheiros: configurar vários módulos

Para melhorar o desempenho do pesquisador, pode configurar os pesquisadores do sistema de ficheiros Sistema de ficheiros Windows e Sistema de ficheiros UNIX para utilizar vários módulos para pesquisar uma única árvore de directórios. Não existe suporte para esta tarefa na consola de administração.

Na configuração predefinida, os pesquisadores de sistema de ficheiros podem utilizar vários módulos, mas cada módulo é dedicado a uma única árvore de

directórios. Por exemplo, um módulo pesquisa o directório A e respectivos subdirectórios, enquanto outro módulo pesquisa o directório B e respectivos subdirectórios.

Para pesquisar de forma mais eficiente e reduzir a possibilidade de módulos inactivos, pode configurar os pesquisadores para utilizar vários módulos para pesquisar ficheiros na mesma árvore de directórios. Por exemplo, os módulos de árvore poderão partilhar o volume de trabalho da pesquisa de ficheiros no directório A.

Para configurar esta função, tem de criar uma extensão do ficheiro de configuração do pesquisador. A codificação de caracteres no ficheiro XML tem de ser UTF-8.

Para configurar os pesquisadores do sistema de ficheiros para utilizar vários módulos para uma única árvore de directórios:

- 1. Inicie sessão como administrador predefinido do IBM Content Analytics with Enterprise Search no servidor do pesquisador.
- Aceda ao directório ES_NODE_ROOT/master_config/nome-colecção.nomepesquisador do pesquisador que pretende configurar.
- 3. Crie um ficheiro XML de extensão de configuração denominado winfscrawler_ext.xml para um pesquisador do Sistema de ficheiros Windows ou um ficheiro unixfscrawler_ext.xml para um pesquisador do Sistema de ficheiros UNIX, e inclua o seguinte conteúdo:

<?xml version="1.0" encoding="UTF-8"?>

```
<ExtendedProperties>
<RemoveChild XPath="/Crawler/DataSources/Server/Target[@Name='target_name']
/NumberOfMaxTargetConnections"/>
<AppendChild XPath="/Crawler/DataSources/Server/Target[@Name='target_name']"
Name="NumberOfMaxTargetConnections">10</AppendChild>
</ExtendedProperties>
```

Onde *target_name* identifica o destino da origem de dados especificado no ficheiro winfscrawler.xml ou no ficheiro unixfscrawler.xml. No exemplo seguinte, o nome de destino é \\file_server1\share1:

<Crawler><DataSources><Server><Target Name="\\file_server1\share1" ID="FS1225436622812" hasTitle="True" Domain="MyDomain" isDomain="false">

Sugestão: Se pretender aplicar esta definição a todos os destinos do pesquisador, não especifique um destino. Por exemplo:

```
<?xml version="1.0" encoding="UTF-8"?>
```

```
<ExtendedProperties>
```

```
<RemoveChild XPath="/Crawler/DataSources/Server/Target/
NumberOfMaxTargetConnections"/>
```

```
<AppendChild XPath="/Crawler/DataSources/Server/Target"</pre>
```

Name="NumberOfMaxTargetConnections">10</AppendChild>

Em ambos os exemplos, o número de módulos do pesquisador está definido como 10 (<AppendChild XPath="/Crawler/DataSources/Server/Target" Name="NumberOfMaxTargetConnections">10</AppendChild>). Em vez de 10, especifique o número de módulos que pretende reservar para a pesquisa da árvore de directórios.

4. Reinicie o pesquisador que actualizou.

</ExtendedProperties>

Pesquisadores do sistema de ficheiros: aumentar o tamanho máximo do documento

É possível editar um ficheiro de configuração para permitir aos pesquisadoresSistema de ficheiros Windows e Sistema de ficheiros UNIX adicionarem documentos até 2 GB de tamanho no espaço de pesquisa. Não existe suporte para esta tarefa na consola de administração.

Para configurar um pesquisador do sistema de ficheiros para adicionar documentos até 2 GB ao espaço de pesquisa:

- 1. Inicie sessão como administrador predefinido do IBM Content Analytics with Enterprise Search no servidor do pesquisador.
- Aceda ao directório ES_NODE_ROOT/master_config/nome-colecção.nomepesquisador do pesquisador que pretende configurar.
- 3. Neste directório, crie um ficheiro XML de extensão de configuração denominado winfscrawler_ext.xml para um pesquisador de Sistema de ficheiros Windows ou unixfscrawler_ext.xml para um pesquisador de Sistema de ficheiros UNIX e inclua o seguinte conteúdo:

```
<?xml version="1.0" encoding="UTF-8"?>
<ExtendedProperties>
<RemoveChild XPath="/Crawler/MaxPageLength"/>
<AppendChild XPath="/Crawler" Name="MaxPageLength">2097152</AppendChild>
</ExtendedProperties>
```

4. Reinicie o pesquisador que actualizou.

Configurar suporte para aplicações de Data Listener

Pode expandir pesquisadores utilizando a API Data Listener para criar um pesquisador externo. As aplicações personalizadas da escuta de dados podem adicionar dados a uma colecção, remover dados de uma colecção ou instruir um pesquisador Web a visitar e revisitar URLs.

Importante: O Data Listener não será suportado em edições futuras. Utilize as APIs de procura e indexação (SIAPI - search and index) em vez das APIs Data Listener para desenvolver aplicações cliente. As informações seguintes são fornecidas para utilizadores que criaram previamente aplicações de Data Listener.

Uma aplicação cliente de escuta de dados permite a pesquisa de tipos de origens de dados que não podem ser pesquisados com os pesquisadores predefinidos. Antes de poder utilizar uma aplicação de Data Listener, terá de configurar as credenciais que permitem à aplicação aceder e a actualizar as colecções.

Quando a respectiva aplicação cliente de Data Listener for ligada ao Data Listener, terá de passar no ID da aplicação cliente, na palavra-passe e no ID da colecção a actualizar. Esta informação tem de corresponder com a informação que configura para a aplicação na consola de administração.

O Data Listener é iniciado automaticamente quando o sistema é iniciado. Se alterar o número de porta depois de configurar a aplicação na consola de administração, terá de reiniciar o Data Listener.

Configurar aplicações de Data Listener:

1. Faça clique em Sistema para abrir a vista Sistema.

- 2. Na página Data Listener, faça clique em **Configurar aplicações de Data Listener**.
- 3. Na página Aplicações de Data Listener, especifique o número de módulos que o Data Listener pode criar para processar pedidos a partir de aplicações cliente e o número de porta em que o Data Listener escuta os pedidos. Especifique também o número máximo de documentos, por colecção, que podem ser mantidos na memória temporária até que o analisador comece a analisá-los.
- 4. Faça clique em **Adicionar Aplicação de Data Listener** para adicionar informação acerca duma aplicação cliente.
- 5. Na página Adicionar Aplicação Data Listener, especifique as informações de autenticação que permitem às aplicações cliente Data Listener aceder a colecções. Os IDs cliente do Data Listener têm de ser únicos no sistema.
- 6. Seleccione as colecções que a aplicação de Data Listener conseguem actualizar:
 - Faça clique em **Todas as colecções** se pretender que a aplicação actualize todas as colecções.
 - Faça clique em Colecções específicas se pretender que a aplicação actualize apenas as colecções que especifica.

Quando selecciona esta opção, é apresentada uma lista de nomes de colecções. Seleccione a caixa de verificação **Seleccionar** para cada colecção que a aplicação pode actualizar.

- 7. Se tiver alterado o número de porta do Data Listener ou o número de documentos que podem ser mantidos na memória temporária, reinicie o Data Listener:
 - a. Faça clique em Sistema para abrir a vista Sistema.
 - b. Na página Data Listener, faça clique em 🔊 Reiniciar.

Tarefas relacionadas:

"Supervisionar Data Listener" na página 442

Suplementos do pesquisador personalizados

Quando configura propriedades para pesquisadores, pode especificar uma classe de Java a utilizar para fazer cumprir os controlos de acesso de nível de documento. Também pode utilizar a classe de Java para actualizar o índice, adicionando, modificando ou removendo metadados e conteúdo de documentos. Ao escrever um suplemento, também pode expandir a capacidade do pesquisador para pesquisar ficheiros de arquivo.

Um suplemento contém uma classe de Java que é chamada para cada documento que o pesquisador pesquisa. À classe de Java é transmitido o identificador de documentos (URI) do índice, tokens de segurança, metadados e conteúdo de documentos. A classe pode devolver tokens de segurança novos ou modificados, metadados e conteúdo, podendo também remover tokens de segurança, metadados e conteúdo.

Depois de os documentos no espaço de pesquisa serem pesquisados uma vez, o suplemento é chamado apenas para os documentos novos ou modificados. Para alterar os tokens de segurança, metadados ou conteúdo dos documentos que estão no índice, mas que não foram actualizados na origem de dados original, inicie uma pesquisa completa de todos os documentos no espaço de pesquisa e depois reconstrua o índice principal.

Suplementos para reforçar a segurança

A segurança ao nível do documento é reforçada para utilizar um ou mais tokens de segurança (uma cadeia delimitada por vírgulas) com cada documento que o pesquisador pesquisa. Os identificadores do grupo são normalmente utilizados como tokens de segurança.

Por predefinição, cada documento é atribuído a um token público que faz com que o documento fique disponível para todos. O token público pode ser substituído por um valor fornecido pelo administrador ou por um valor extraído de um campo no documento pesquisado.

O suplemento permite-lhe aplicar as respectivas regras de negócios para determinar o valor dos tokens de segurança nos documentos pesquisados. Os tokens de segurança utilizados com cada documento são armazenados no índice. São utilizados para filtrar documentos que correspondem aos tokens de segurança e garante que apenas os documentos que um utilizador tem permissão para consultar sejam devolvidos nos resultados da procura.

Suplementos para adicionar, modificar ou remover metadados

Os metadados do documento, tais como a data em que um documento foi modificado pela última vez, são criados para todos os documentos pesquisados. O suplemento do pesquisador permite-lhe aplicar as respectivas regras de negócios para determinar o valor dos metadados que devem ser indexados para cada documento.

Por exemplo, os metadados não podem ser extraídos de documentos binários através dos analisadores incorporados. Contudo, pode criar um suplemento pós análise e utilizar o atributo MetadataPreferred para adicionar metadados pesquisáveis para documentos após serem pesquisados.

Os metadados são criados como um par nome-valor. Os utilizadores podem procurar os metadados com uma consulta de texto livre ou com uma consulta que especifique o nome do campo dos metadados.

Suplementos para adicionar, modificar ou remover o conteúdo de documentos

O conteúdo de documentos engloba as partes de um documento que contêm conteúdo pesquisável e conteúdo que se pode tornar parte do resumo de documento dinâmico nos resultados da procura. O suplemento do pesquisador permite-lhe aplicar as suas próprias regras de negócios para determinar o conteúdo que deve ser indexado para cada documento.

Se tenciona pesquisar várias colecções ao mesmo tempo, poderá querer criar um suplemento para alterar a forma como os campos ordenáveis são apresentados nos resultados da pesquisa. Numa pesquisa federada, os valores do campo são considerados valores de cadeia, não numéricos. Se ordenar os resultados por tamanho de ficheiro, por exemplo, os valores do tamanho do ficheiro de cada colecção são colocados por ordem alfabética, não numérica. Um suplemento do pesquisador pode alterar a forma como os valores do tamanho do ficheiro são avaliados. Por exemplo:

• O suplemento pode modificar directamente o campo filesize adicionando zeros para criar um campo de cadeia de comprimento fixo.

 O suplemento pode definir um novo campo, como filesizesort, como uma cadeia de comprimento fixo que utiliza zeros adicionados para reforçar o tamanho fixo. Certifique-se de que define este campo para ser ordenado e devolvido na consola de administração. A vantagem desta abordagem é que permite ao campo filesize continuar a ser pesquisado como um campo paramétrico.

Com ambas as abordagens, pode modificar o ficheiro searchResultsTable.jsp para remover os zeros adicionados antes de mostrar os resultados de pesquisa, por isso os valores como 000012345 não são apresentados.

Suplementos do pesquisador da Web

Através das interfaces de programação de aplicações no pesquisador da Web, pode controlar o modo como os documentos são pesquisados e como são preparados para a análise. Por exemplo, pode adicionar campos ao cabeçalho do pedido HTTP que será utilizado quando o pesquisador pedir um documento. Depois de um documento ser pesquisado, e antes de ser analisado ou testemunhado, pode alterar o conteúdo, os tokens de segurança e os metadados. Também pode parar o envio do documento ao analisador.

Suplementos de ficheiros de arquivo

Ao escrever um suplemento, pode expandir os pesquisadores e activar o suporte para a pesquisa de formatos de ficheiros de arquivo diferentes de ZIP e TAR. Por exemplo, pode escrever um suplemento para suportar a pesquisa de documentos no formato LZH.

Modo não delimitado

Quando configurar alguns pesquisadores, pode seleccionar uma opção para executar o suplemento no modo não delimitado. Neste modo, o processo de suplemento é executado no âmbito do processo do pesquisador, o que pode melhorar o desempenho do suplemento.

Importante: Se o suplemento tiver um problema e não for possível efectuar a recuperação, o processo do pesquisador poderá ser terminado.

Conceitos relacionados:

E Suplementos do pesquisador

🕩 Criar um suplemento de pós-análise para o pesquisador da Web

Suporte para pesquisa de ficheiros de arquivo

Os pesquisadores podem extrair ficheiros de um ficheiro de arquivo (como um ficheiro ZIP ou TAR) para que os ficheiros individuais no arquivo possam ser indexados e pesquisados.

Formatos de ficheiros de arquivo suportados

São suportados os seguintes formatos de ficheiros de arquivo:

Extensão de ficheiro	Tipo MIME	Tipo de dados	Notas
.zip, .ZIP	aplicação/zip	zip	 Depende das capacidades do pacote java.utl.zip Suporta a compactação deflaccionada (método 8): Sem suporte para ficheiros codificados Sem suporte para zip64
.tar	aplicação/tar	tar	Formatos tar suportados: • GNU tar 1.13 • POSIX 1003.1-1998 (ustar) • POSIX 1003.1-2001 (pax)
.tar.gz, .tgz	não aplicável	tgz	Depende das capacidades do pacote java.utl.zip

Tabela 17. Formatos de ficheiro de arquivo suportados por pesquisadores

Restrições e directrizes

A detecção automática de páginas de códigos não está disponível para ficheiros que sejam extraídos de um ficheiro de arquivo. Ao extrair os ficheiros, o pesquisador utiliza a definição da página de códigos que está configurada para utilizar texto normal e tipos de documentos desconhecidos. Quando utiliza a consola de administração de para configurar definições de idioma e página de códigos para um pesquisador, especifica a página de códigos que o pesquisador deverá utilizar para documentos de texto normal e para documentos cuja página de códigos não possa ser detectada automaticamente.

Para determinar quando é necessário pesquisar de novo os ficheiros de um ficheiro de arquivo, o pesquisador utiliza a data de modificação existente nos dados do cabeçalho de entrada do arquivo para cada ficheiro. Quando supervisiona um pesquisador, as estatísticas apresentadas para documentos pesquisados, incluindo as estatísticas para documentos inseridos, actualizados e eliminados, incluem informações sobre ficheiros que foram extraídos de ficheiros de arquivo.

Para permitir que os pesquisadores pesquisem ficheiros de arquivo noutros formatos de ficheiro de arquivo, tais como ficheiros LZH, tem de gravar um suplemento do pesquisador e, em seguida, configurá-lo para utilizar o suplemento.

Formatos de URI no índice

O URI (Uniform Resource Identifier) de cada documento no índice indica o tipo de pesquisador que adicionou o documento à colecção.

Pode especificar URIs ou padrões de URI ao configurar categorias, âmbitos e ligações rápidas para uma colecção. Também pode especificar o URI quando é necessário remover documentos do índice, ou visualizar informações de estado detalhadas sobre um URI específico.

Efectue procuras na colecção para determinar os URIs ou padrões de URI de um determinado documento. Pode fazer clique nos URIs nos resultados da procura para obter os documentos pretendidos. Pode copiar o URI a partir dos resultados

da procura para utilizar o URI na consola de administração. Por exemplo, pode especificar um padrão de URI para associar automaticamente documentos que correspondam a esse padrão de URI a uma ligação rápida.

Quando especifica um URI ou um padrão URI, tem de especificar o formato codificado de URL para o URI e de se certificar que o URI não contém caracteres que não estão incluídos no conjunto de caracteres codificados US-ASCII. Para obter detalhes, consulte a RFC1738, a norma da Internet para URLs.

No exemplo seguinte, não é possível especificar o primeiro URI, que contém os caracteres hebraicos. No entanto, pode especificar o segundo URI, que é o formato codificado de URL do primeiro URI.

URI incorrecto

file:///c:/shared/hebrew/^^^^^

URI correcto

file:///c:/shared/hebrew/%D7%A2%D7%91%D7%A8%D7%99%D7%AA

Ficheiros de arquivo

O formato URI para documentos que são extraídos de um ficheiro de arquivo (como um ficheiro .zip ou .tar) e, em seguida pesquisados são: Original URI(? [&)ArchiveEntry=Nome Entrada (&ArchiveEntry=Nome Entrada)

Parâmetros

URI_Original

A localização do ficheiro de arquivo na origem de dados.

Nome_Entrada

O nome codificado por URL da entrada de arquivo no ficheiro de arquivo.

Exemplos

file:///d:/Arquivol.zip
file:///d:/Arquivol.zip?ArchiveEntry=Pasta1/PowerPoint.ppt
file:///d:/Arquivol.zip?ArchiveEntry=Pasta2/Text.txt

Pesquisadores do Agente para sistemas de ficheiros Windows

O formato de URI para documentos que são pesquisados por um pesquisador do Agente para sistemas de ficheiros Windows é:

file:///Host_Name/Directory_Name/.../File_Name

Parâmetros

A codificação de URL aplica-se a todos os campos.

Nome Sistema Central

O nome do sistema central ou endereço de IP do servidor onde o documento está localizado.

Nome Directório

O nome de um directório partilhado no domínio Windows.

Nome_Ficheiro

O nome do ficheiro.

Exemplo

file:////9.187.186.83/temp/test1/test12/Copy+%284%29+of+dumpstore_1.txt

Pesquisadores do Case Manager

O formato URI de documentos pesquisados por um pesquisador do Case Manager é o seguinte:

p8ce://nome_sistema_central:porta/armazenamento_objectos/id_série_versão/código_hash
[/número_elemento]?protocolo=http

Parâmetros

A codificação de URL aplica-se a todos os campos.

nome_sistema_central

Um nome do sistema central de um servidor no qual é executado do IBM FileNet Content Engine,

porta Um número de porta no qual é executado o Content Engine Web Service.

arquivo objectos

Um nome de um armazenamento de objectos no qual é armazenado um documento.

id_série_versão

Um identificador de documento único. O ID de série de versão é utilizado porque o ID de documento muda à medida que o documento é alterado consoante a versão, enquanto o ID de série de versão não muda.

código_hash

Para distinguir entre pastas, é adicionado um código hash ao caminho para o URI do objecto. No exemplo seguinte, 7584373 corresponde ao código hash do caminho de pasta /0bjectStore/CaseSolution/../CaseFolder/SubFolders:

p8ce://9.39.44.204:9080/wsi/FNCEWS40MT0M/ATOSAIX2/

{2D09F43F-3392-485E-B338-E67D68F04FA6}.7584373?protocol=http

número_elemento

Um índice de elementos de conteúdos. Esta variável é anexada apenas quando um URI indica um documento que contém elementos de conteúdo múltiplo.

protocol

Protocolo para aceder ao Serviço Web. Os valores válidos são http ou https.

Pesquisadores do Content Integrator

O formato URI de documentos pesquisados por um pesquisador do Content Integrator no modo de acesso de servidor é o seguinte:

vbr://Nome_Servidor/ID_Sistema_Repositório/ID_Persistente_RepositórioID
 /ID_Item/ID_Versão

/Tipo_Item/?[Page=Número_Página&] Propriedades_JNDI

O formato URI de documentos pesquisados por um pesquisador do Content Integrator no modo de acesso directo é o seguinte:

vbr:///ID_Sistema_Repositório/ID_Persistente_Repositório /ID_Item/ID_Versão /Tipo_Item/[?Page=Número_Página]

Parâmetros

A codificação de URL aplica-se a todos os campos.

Nome_Servidor

O nome do servidor do IBM Content Integrator.

ID_Sistema_Repositório

O ID do sistema do repositório.

ID_Persistente_Repositório

O ID persistente do repositório.

ID_Item

O ID do item.

ID_Versão

O ID da versão. Se o ID da versão estiver em branco, este valor indica a versão mais recente do documento.

Tipo_Item

O tipo do item (CONTENT ou FOLDER).

Número Página

O número da página.

Propriedades JNDI

As propriedades JNDI para o cliente da aplicação empresarial J2EE. Existem dois tipos de propriedades:

java.naming.factory.initial

O nome da classe do servidor da aplicação utilizado para criar o parâmetro identificador EJB.

java.naming.provider.url

O URL do serviço de nomenclatura do servidor da aplicação utilizado para pedir o parâmetro identificador EJB.

Exemplos

Documentum:

vbr://vbrsrv.ibm.com/Documentum/c06b/094e827780000302//CONTENT/? java.naming.provider.url=iiop%3A%2F%2Fmyvbr.ibm.com%3A2809& java.naming.factory.initial=com.ibm.websphere.naming.WsnInitContextFactory

FileNet PanagonCS:

vbr://vbrsrv.ibm.com/PanagonCS/4a4c/003671066//CONTENT/?Page=1& java.naming.provider.url=iiop%3A%2F%2Fmyvbr.ibm.com%3A2809& java.naming.factory.initial=com.ibm.websphere.naming.WsnInitContextFactory

Pesquisadores do Content Manager

O formato URI de documentos pesquisados por um pesquisador do Content Manager é o seguinte:

cm://Nome_Servidor/Nome_Tipo_Item/PID

Parâmetros

A codificação de URL aplica-se ao parâmetro PID.

Nome_Servidor

O nome do servidor de bibliotecas do IBM Content Manager Enterprise Edition.

Nome_Tipo_Item

O nome do tipo de item destino.

PID O identificador persistente do Content Manager EE.

Exemplo

```
cm://cmsrvctg/ITEMTYPE1/92+3+ICM8+icmnlsdb12+ITEMTYPE159+26+A1001001A
03F27B94411D1831718+A03F27B+94411D183171+14+1018
```

Pesquisadores do DB2

O formato URI de documentos pesquisados por um pesquisador do DB2 é o seguinte:

db2://Nome_Base_de_Dados/Nome_Tabela

/Nome1_Coluna_Identificador_Exclusivo/Valor1_Identificador_Exclusivo
[/Nome2_Coluna_Identificador_Exclusivo//Valor2_Identificador_Exclusivo/...
/NomeN_Coluna_Identificador_Exclusivo/ValorN_Identificador_Exclusivo]

Parâmetros:

A codificação de URL aplica-se a todos os campos.

Nome Base de Dados

O nome interno da base de dados ou o nome alternativo da base de dados.

Nome Tabela

O nome da tabela destino, incluindo o nome do esquema.

Nome1_Coluna_Identificador_Exclusivo O nome da primeira coluna Identificador Exclusivo na tabela.

- *Valor1_Identificador_Exclusivo* O valor da primeira coluna Identificador Exclusivo.
- NomeN_Coluna_Identificador_Exclusivo O nome da coluna Identificador Único *n*-ésimo na tabela.
- *ValorN_Identificador_Exclusivo* O valor da coluna Identificador Único *n*-ésimo.

Exemplos

Base de dados catalogada, local: db2://LOCALDB/SCHEMA1.TABLE1/MODEL/ThinkPadA20

Base de dados não catalogada, remota: db2://meuservidor.minhaempresa.com:50001/BDREMOTA/ESQUEMA2.TABELA2/NOME/DAVID

Pesquisadores do Domino Document Manager

O formato URI de documentos pesquisados por um pesquisador do Domino Document Manager é o seguinte:

dominodoc://Server_Name:Port_Number/Database_Replica_ID/Database_Path_and_Name
 /View_Universal_ID/Document_Universal_ID

/?AttNo=Attachment_Number&AttName=Attachment_File_Name

Parâmetros

A codificação de URL aplica-se a todos os campos.

Nome Servidor

O nome do servidor do Domino Document Manager.

Número_Porta

Opcional: O número da porta do servidor do Domino Document Manager.

ID_Réplica_Base_de_Dados O identificador da réplica da base de dados.
Nome_e_Caminho_da_Base_de_Dados

O nome de ficheiro e caminho do documento de base de dados NSF no servidor do Domino Document Manager destino.

ID_Universal_Vista

O ID Universal de Vista utilizado para pesquisar documentos do Domino Document Manager.

ID_Universal_Documento

O ID Universal do Documento definido no documento pesquisado.

Número_Anexo

Opcional: Um número consecutivo, a começar em zero, para cada anexo.

Nome_Ficheiro_Anexo

Opcional: O nome original do ficheiro anexo.

Exemplos

Um documento do Domino Document Manager:

dominodoc://dominodocsvr.ibm.com/49256D3A000A20DE/domdoc%2FADMN-6FAJXL.nsf /8178B1C14B1E9B6B8525624F0062FE9F/0205F44FA3F45A9049256DB20042D226

Um anexo de documento:

dominodoc://dominodocsvr.ibm.com/49256D3A000A20DE/domdoc%2FADMN-6FAJXL.nsf /8178B1C14B1E9B6B8525624F0062FE9F/0205F44FA3F45A9049256DB20042D226 ?AttNo=0&AttName=FicheiroAnexo.doc

Pesquisadores do Exchange Server

Uma vez que o IBM Content Analytics with Enterprise Search não consegue obter o URL dos anexos através do Outlook Web App (OWA), apresenta URLs alternativos para os itens anexados. Uma vez que o Exchange Server 2007 apenas suporta o navegador Internet Explorer, os utilizadores só podem aceder ao OWA do Exchange Server 2007 nesse navegador.

Quando os utilizadores fazem clique em títulos na página de resultados da aplicação de procura empresarial ou na prospecção de análise de conteúdo, é apresentado o item do Exchange Server correspondente através do OWA. Se o utilizador tiver o MailboxPermission para a caixa de correio que contém os resultados da procura, pode também abrir o item através do OWA. No entanto, se o utilizador tiver o MailboxFolderPermission ou o Delegation para a caixa de correio que contém os resultados da procura, terá de aceder ao seguinte URL antes de fazer clique no título para aceder ao item, em que *user's_primarySmtpAddress* é o endereço ao qual os resultados da pesquisa pertencem originalmente https://hostname/OWA/user's primarySmtpAddress/?cmd=contents

O pesquisadorExchange Server gera URIs originais para documentos pesquisados. O pesquisador utiliza IDs para o URI que são valores únicos entre os itens e anexos. Se um documento for um item, o URI é formatado da seguinte forma: exchadp://hostname/mailbox name/itemId=itemId&owa=owaURL

Se o documento não for um anexo, o URI é formatado da seguinte forma: exchadp://hostname/mailbox name/attachmentId=attachmentId&owa=owaURL

Pesquisadores do Exchange Server 2000 e 2003

O formato URI para documentos pesquisados por um Exchange Server pesquisador do 2000 e do 2003 é: exchange://Caminho_OWA[?useSSL=true]

Parâmetros

Caminho_OWA

O caminho OWA (Outlook Web Access) sem o protocolo.

useSSL=true

Adicionado quando o protocolo do caminho OWA original corresponde a HTTPS.

Exemplos

Corpo do documento:

exchange://exchangesvr.ibm.com/public/PastaRaiz1/Pasta1/Documento.EML

Anexo de documento:

exchange://exchangesvr.ibm.com/public/PastaRaiz1/Pasta1/Documento.EML/
FicheiroAnexo.doc

Activado para SSL:

exchange://exchangesvr.ibm.com/public/SalaEquipa/Pasta1/Documento.EML ?useSSL=true

Pesquisadores do FileNet P8

O formato URI de documentos pesquisados por um pesquisador do FileNet P8 é o seguinte:

p8ce://nome_sistema_central:porta/arquivo_objectos/id_objecto[/número_elemento]?protocol=http

Parâmetros

A codificação de URL aplica-se a todos os campos.

nome_sistema_central

Um nome do sistema central de um servidor no qual é executado do IBM FileNet Content Engine,

porta Um número de porta no qual é executado o Content Engine Web Service.

arquivo_objectos

Um nome de um armazenamento de objectos no qual é armazenado um documento.

id_objecto

Um Globally Unique Identifier (GUID) atribuído pelo Content Engine a um objecto armazenado. Uma cadeia de caracteres que contém 38 caracteres, o GUID consiste num parêntesis curvo esquerdo, 8 caracteres hexadecimais, um traço, 4 caracteres hexadecimais, um traço, 4 caracteres hexadecimais, um traço, 4 caracteres hexadecimais, um traço, 12 caracteres hexadecimais e um parêntesis curvo direito. Os parêntesis são codificados por regras de codificação de URLs. Por exemplo:

%7B1234abcd-56ef-7a89-9fe8-7d65cd43ba21%7D

número_elemento

Um índice de elementos de conteúdos. Esta variável é anexada apenas quando um URI indica um documento que contém elementos de conteúdo múltiplo.

protocol

Protocolo para aceder ao Serviço Web. Os valores válidos são http ou https.

Exemplo

p8ce://host.filenet.com:9080/STORE1/{1234abcd-56ef-7a89-9fe8-7d65cd43ba21}/2

Pesquisadores de Base de dados JDBC

O formato URI de documentos pesquisados por um pesquisador de Base de dados JDBC é o seguinte:

jdbc://URL DB/Nome Tabela

/Nome1_Coluna_Identificador_Exclusivo/Valor1_Identificador_Exclusivo /[Nome2_Coluna_Identificador_Exclusivo/Valor2_Identificador_Exclusivo /.../NomeN_Coluna_Identificador_Exclusivo/ValorN_Identificador_Exclusivo]

Parâmetros

A codificação de URL aplica-se a todos os campos.

URL_DB O URL para a base de dados.

Nome_Tabela

O nome da tabela destino, incluindo o nome do esquema.

- Nome1_Coluna_Identificador_Exclusivo O nome da primeira coluna Identificador Exclusivo na tabela.
- Valor1_Identificador_Exclusivo O valor da primeira coluna Identificador Exclusivo.
- *NomeN_Coluna_Identificador_Exclusivo* O nome da coluna Identificador Único *n*-ésimo na tabela.

ValorN_Identificador_Exclusivo O valor da coluna Identificador Único *n*-ésimo.

Exemplos:

Base de dados de DB2:

jdbc:db2://host01.svl.ibm.com:50000/SAMPLE/DB2INST1.ORG/DEPTNUMB/51

Base de dados da Oracle:

jdbc:oracle:thin:@/host01.svl.ibm.com:1521:ora/SCOTT.EMP/EMPN0/7934

Base de dados MS SQL Server 2000:

jdbc:microsoft:sqlserver://host01.svl.ibm.com:1433; DatabaseName=Northwind/dbo.Region/RegionID/100

Base de dados MS SQL Server 2005:

jdbc:sqlserver://host01.svl.ibm.com:1433; DatabaseName=Northwind/dbo.Region/RegionID/100

Pesquisadores do Notes

O formato URI de documentos pesquisados por um pesquisador do Notes é o seguinte:

[?AttNo=Attachment_Number&AttName=Attachment_File_Name]

Parâmetros

A codificação de URL aplica-se a todos os campos.

Nome_Servidor

O nome do servidor do Lotus Notes.

Número_Porta

O número da porta do servidor do Lotus Notes. O número da porta é opcional.

ID_Réplica_Base_de_Dados

O identificador da réplica da base de dados.

Nome_e_Caminho_da_Base_de_Dados

O nome de ficheiro e caminho da base de dados NSF no servidor do Lotus Notes destino.

ID_Universal_Vista

O ID Universal de Vista definido na base de dados destino. Este ID é especificado apenas quando o documento é seleccionado a partir de uma vista ou pasta. Se não designar uma vista nem uma pasta a pesquisar (por exemplo, se especificar que pretende pesquisar todos os documentos numa base de dados), o ID Universal de Vista não é especificado.

ID_Universal_Documento

O ID Universal do Documento definido no documento pesquisado pelo pesquisador.

Número_Anexo

Um número consecutivo, a começar em zero, para cada anexo. O número de anexo é opcional.

Nome_Ficheiro_Anexo

O nome original do ficheiro anexo. O nome de ficheiro anexo é opcional.

Exemplos

Um documento seleccionado para pesquisa por vista ou pasta: domino://dominosvr.ibm.com/49256D3A000A20DE/Database.nsf/ 8178B1C14B1E9B6B8525624F0062FE9F/0205F44FA3F45A9049256DB20042D226

Um documento não seleccionado para pesquisa por vista nem pasta:

domino://dominosvr.ibm.com/49256D3A000A20DE/Database.nsf// 0205F44FA3F45A9049256DB20042D226

Um anexo de documento:

domino://dominosvr.ibm.com/49256D3A000A20DE/Database.nsf// 0205F44FA3F45A9049256DB20042D226?AttNo=0&AttName=FicheiroAnexado.doc

Pesquisadores do Quickr for Domino

O formato URI de documentos pesquisados por um pesquisador do Quickr for Domino é o seguinte:

quickplace://Server_Name:Port_Number/Database_Replica_ID/Database_Path_and_Name
/View_Universal_ID/Document_Universal_ID
/?AttNo=Attachment Number&AttName=Attachment File Name

Parâmetros

A codificação de URL aplica-se a todos os campos.

Nome_Servidor

O nome do sistema central do servidor Quickr for Domino.

Número Porta

Opcional: O número da porta do servidor Quickr for Domino.

ID_Réplica_Base_de_Dados

O identificador da réplica da base de dados.

```
Nome_e_Caminho_da_Base_de_Dados
```

O nome do caminho e do ficheiro para a base de dados do documento NSF no destino de servidor Quickr for Domino.

ID_Universal_Vista

O View Universal ID utilizado para pesquisar documentos.

ID Universal Documento

O ID Universal do Documento definido no documento pesquisado.

Número Anexo

Opcional: Um número consecutivo, a começar em zero, para cada anexo.

Nome_Ficheiro_Anexo

Opcional: O nome original do ficheiro anexo.

Exemplos

Um documento:

```
quickplace://ltwsvr.ibm.com/49257043000214B3/QuickPlace%5Csampleplace
%5CPageLibrary4925704300021490.nsf
/A7986FD2A9CD47090525670800167225
/2B02B1DE3A82B2CE49257043001C2498
```

Um anexo de página:

```
quickplace://ltwsvr.ibm.com/49257043000214B3/QuickPlace%5Csampleplace
%5CPageLibrary4925704300021490.nsf
/A7986FD2A9CD47090525670800167225
/2B02B1DE3A82B2CE49257043001C2498
?AttNo=0&AttName==QPCons3.ppt
```

Pesquisadores Lista de valores geradores

O formato URI de documentos pesquisados por um pesquisador do Lista de valores geradores é o seguinte:

seedlist://Page_URL?pageID=Page_ID[&useSSL;=true]

Parâmetros

A codificação de URL aplica-se a todos os campos.

URL Página

O URL para o documento (exclusivo para cada documento).

ID Página

O identificador de objecto para o documento.

useSSL Quando o protocolo é HTTPS, &useSSL;=true é adicionado ao URI. Caso contrário, useSSL é omitido.

Exemplo

Protocolo HTTPS:

seedlist://quickrserver.ibm.com:10035/lotus/mypoc?uri=dm:bec6090046f1cd5
2bc5cfcb06e9f4550&verb;=view&pageID;=N1FSZUR1MkJQNjZSMDZQMUMwM1FPNjZCQzY
2SUw2SUhPNk1RQ0M2Uk80Nk9PNjVCRUM2UUs2TDFDMA==&useSSL;=true

Pesquisadores do SharePoint

O pesquisador do SharePoint não gera o seu próprio formato de URI do documento. Em contrapartida, cria um URL acessível para o URI do documento. O URL acessível pode ser alterado segundo a configuração Site and Form do servidor SharePoint. O pesquisador tenta obter o URL em forma de apresentação e anexar-lhe o ID do documento. Se o pesquisador estiver configurado para obter um URL junto de um campo específico, tentará utilizar o valor do campo enquanto URI. Este formato é útil para pesquisar listas que não gerem URLs baseadas no valor da chave principal. O formato predefinido é:

http://servidor/caminho_forma_apresentação?nome_campo_chave_principal=valor_chave_principal

Parâmetros

A codificação de URL aplica-se a todos os campos.

servidor

caminho_forma_apresentação

nome_campo_chave_principal

valor_chave_principal

Exemplo

https://sharepoint.example.ibm.com:9999/rootDir/Shared%20Documents/ Forms/DispForm.aspx?ID=5

Pesquisadores do Sistema de ficheiros UNIX

O formato URI de documentos pesquisados por um pesquisador do Sistema de ficheiros UNIX é o seguinte:

file:///Nome_Directório/Nome_Ficheiro

Parâmetros

A codificação de URL aplica-se a todos os campos.

Nome_Directório

O nome do caminho absoluto do directório.

Nome Ficheiro

O nome do ficheiro.

Exemplo

file:///home/utilizador/teste.doc

Pesquisadores do Web Content Management

O formato URI de documentos pesquisados por um pesquisador do Web Content Management é o seguinte:

wcm://URL_Página?pageID=ID_Página[&useSSL=true]

Parâmetros

A codificação de URL aplica-se a todos os campos.

URL_Página

O URL para o documento (exclusivo para cada documento).

ID_Página

O identificador da página.

useSSL Quando o protocolo é HTTPS, useSSL=true é adicionado ao URI. Caso contrário, useSSL é omitido.

Exemplos

Protocolo HTTP:

wcm://wp6server.ibm.com:9081/wps/wcm/minhaligação/Web+Content /Site01/SiteArea01/ContentTest01?pageID= 6QReDeJ9DI3R0663E03Q06L1E2MR47MH0C3Q862RD6J0863B0GJS86J9E0

Protocolo HTTPS:

wcm://wp6server.ibm.com:9444/wps/wcm/minhaligação/Web+Content/Site01 /SiteArea01/ContentTest01?pageID= 6QReDeJ9DI3R0663E03Q06L1E2MR47MH0C3Q862RD6J0863B0GJS86J9E0&useSSL=true

Pesquisadores do WebSphere Portal

O formato URI de documentos pesquisados por um pesquisador do WebSphere Portal é o seguinte:

wp6://URL_Página?portletURL=URL_Portlet?portletDefID=ID_Def_Portlet &pageID=ID_Página[&useSSL=true]

Parâmetros

A codificação de URL aplica-se a todos os campos.

URL_Página

O URL para o documento (exclusivo para cada documento).

URL_Portlet

O URL exclusivo para o documento.

ID_Def_Portlet

O identificador de definição da portlet.

ID_Página

O identificador da página.

useSSL Quando o protocolo é HTTPS, useSSL=true é adicionado ao URI. Caso contrário, useSSL é omitido.

Exemplos

Protocolo HTTP:

wp6://wp6server.ibm.com:9081/wps/meuportal/!ut/p/c1/04_SB8K8xLLM9MSSzPy 8xBz9CP0os3gjE59gQwMLQ0P_IDMnAyNHA3f3UEsTD1NjA6B8pFm8AQ7gaEBAdzjIPrz6_ Tzyc1P1C3IjDHQdFRUBTu-saA!!/d12/d0/Y2BkbGBgY1rDwMDEJ1XAwMggYxZvZ0ITbGh gYWjobuhmaGDka0Bu5uHqFRpkaAAAEisaBQ!! ?portletUr1=/wps/meuporta1/!ut/p/c1/04_SB8K8xLLM9MSSzPy8xBz9CP0os3gjE59 gQwMLQ0P_IDMnAyNHA3f3UEsTD1NjA6B8pFm8AQ7gaEBAdzjIPrz6_Tzyc1P1C3IjDHQdF RUBTu-saA!!/d12/d0/Y2BiUZnBwMqsyaBykYGBmS2tcDoDE40MebyRiU-woYGFoaG7oZu hgZGjgbuZh6uXY4ihAY0MG07Z0CBDAwDcXPkM

&portletDefID=6 24LS10811G1F102A0G6HEJUR10

&pageID=3 24LS108110R6B02A0GGU94LN00

Protocolo HTTPS:

wp6://wp6server.ibm.com:9444/wps/meuportal/!ut/p/c1/04_SB8K8xLLM9MSSzPy 8xBz9CP0os3gjE59gQwMLQ0P_IDMnAyNHA3f3UEsTD1NjA6B8pFm8AQ7gaEBAdzjIPrz6_ Tzyc1P1C3IjDHQdFRUBTu-saA!!/d12/d0/Y2BkbGBgY1rDwMDEJ1XAwMggYxZvZ0ITbGh gYWjobuhmaGDka0Bu5uHq5RtqaAAA50L41Q!! ?portletUr1=/wps/meuportal/!ut/p/c1/04_SB8K8xLLM9MSSzPy8xBz9CP0os3gjE59 gQwMLQ0P_IDMnAyNHA3f3UEsTD1NjA6B8pFm8AQ7gaEBAdzjIPrz6_Tzyc1P1C3IjDHQdF RUBTu-saA!!/d12/d0/Y2BiUZnBwMqsyaBykYGBmS2tcD0DE40MebyRiU-woYGFoaG7oZu hgZGjgbuZh6uXWZihAY0MGQ5Z31BDAwAk73P2 &portletDefID=6_24LS10811G1F102A0G6HEJMU10 &pageID=3_24LS108110R6B02A0GGU94T410&useSSL=true

Pesquisadores do Sistema de ficheiros Windows

Os formatos URI de documentos pesquisados por um pesquisador do Sistema de ficheiros Windows são os seguintes:

file:///Nome_Directório/Nome_Ficheiro
file:///Nome_Pasta_Rede/Nome_Directório/Nome_Ficheiro

Parâmetros

A codificação de URL aplica-se a todos os campos.

Nome_Directório

O nome do caminho absoluto do directório.

Nome_Ficheiro

O nome do ficheiro.

Nome_Pasta_Rede

Apenas para documentos em servidores remotos, o nome da pasta partilhada numa rede do Windows.

Exemplos

Sistema de ficheiros local: file:///d:/directório/teste.doc

Sistema de ficheiros de rede:

file:///filesvr.ibm.com/directório/ficheiro.doc

Conceitos relacionados:

"Categorias baseadas em regras" na página 235

"Classificação de documento baseada em padrões URI" na página 318

Tarefas relacionadas:

"Remover URIs do índice" na página 225

"Configurar categorias baseadas em regras" na página 237

"Configurar ligações rápidas" na página 295

"Configurar âmbitos" na página 298

"Visualizar detalhes sobre um URI" na página 436

Administração de análise e indexação

Para ampliar a capacidade de encontrar documentos relevantes ou explorar desvios e tendências nos dados, pode especificar as opções para como os documentos e os metadados devem ser analisados e categorizados antes de serem adicionados ao índice.Para assegurar que os utilizadores têm sempre acesso à informação mais recente, ocorrem automaticamente actualizações incrementais do índice quando novos documentos estão prontos para ser indexados.

Os serviços de indexação fornecem processamento de documentos e funções de indexação. Primeiro, o serviço de índice lê os documentos pesquisados na fila de ficheiros do armazenamento de dados. Para preparar um documento para indexação, as tarefas do processador de documentos extraem e analisam o texto e os metadados para cada documento. O processamento de documentos inclui:

- Utilizar analisadores, como os filtros de extracção de textos e analisadores de HTML, XML, e texto, para analisar conteúdo.
- Utilizar anotadores de Apache Unstructured Information Management Architecture (UIMA) para segmentar no conteúdo e extrair entidades.

Depois dos dados e metadados de texto serem analisados e segmentados, o serviço de indexação constrói um índice de texto principal. O serviço pode também construir outros índices, como um índice de categorias, uma cache de documento, etc. Com a indexação incremental, os documentos podem estar disponíveis para procura num período de tempo curto, sem indexar a colecção inteira.

Partições de índice para colecções de procura empresarial

Quando criar uma colecção de procura empresarial, pode especificar se quer criar múltiplas partições de índice. As partições activam o sistema para dimensionar vários milhões de documentos e documentos de índice em paralelo. As partições de índice são acedidas como se fossem um só índice.

Para evitar a degradação do desempenho, mantenha cada partição para menos de 20 milhões de documentos. Por exemplo, um índice com 40 milhões de documentos pode requerer 10 para que sejam devolvidos resultados de procura. Caso pretenda que tal colecção cresça, pode obter um melhor desempenho ao criar três partições de índice, cada uma com, no máximo, 15 milhões de documentos.

Conceitos relacionados:

- E Suporte linguístico para procura semântica
- Suporte Linguístico no IBM Content Analytics with Enterprise Search
- Conceitos básicos utilizados no processamento da análise de texto
- 🕩 Aplicações de procura semântica
- Fremos da consulta de procura semântica

Configuração de serviços de análise e indexação

é possível especificar as opções, referentes a toda a colecção, para como o conteúdo deve ser analisado e indexado.

Por exemplo, é possível configurar o analisador para detectar, automaticamente, documentos de entrada e analisá-los imediatamente após serem adicionados à colecção.

Configurar módulos para serviço de índice

Se tiver um grande número de colecções e recursos de memória suficientes, pode aumentar o número de módulos que estão disponíveis para indexar os documentos.

Para configurar o número de módulos de índice:

- 1. Na vista Colecções, expanda a colecção que pretende editar. No painel Analisar e Indexar, faça clique em **Configurar** > **Opções de análise**.
- Especifique o número máximo de módulos que serão iniciados quando o serviço de índice for iniciado e faça clique em OK. Tem de especificar pelo menos um módulo.
- 3. No painel Analisar e Indexar, reinicie os serviços de análise e indexação.

Tarefas relacionadas:

"Configurar sessões múltiplas para o extractor de texto" na página 275

Activar a análise avançada para termos compostos

Com a análise avançada para termos compostos, tanto a palavra inteira como cada parte de uma palavra composta são tratados como termo único. Sem a análise avançada para termos compostos, cada parte da palavra composta é tratada como termo único, mas a palavra inteira não.

Alguns idiomas acumulam termos de várias palavras em palavras únicas sem espaços (termos *compostos*). A análise avançada e a decomposição dos termos compostos são úteis para a procura de idiomas como o alemão e são críticas para a procura de idiomas como o coreano.

Se permitir a análise avançada de termos compostos, os utilizadores poderão procurar termos sem ter de utilizar caracteres globais para encontrar formas compostas dos termos da consulta. Por exemplo, uma procura por Organ (órgão) pode devolver documentos que contenham Organspender (dador de órgãos) mas não devolverá documentos que contenham Organsp*. Com a análise avançada para termos compostos activada, tanto Organ como Organsp* poderão devolver documentos que contenham Organspender.

Por exemplo, o composto Mustermann é dividido em duas categorias (muster e mann), as quais são armazenadas separadamente no índice. Quando a consulta de carácter global Musterma* é introduzida, os processos de procura não conseguem identificar Musterma como prefixo de uma palavra decomposta. Por conseguinte, não são encontrados documentos com o termo Mustermann. Se pretender que os utilizadores tenham a capacidade de introduzir consultas de carácter global para termos compostos, active a análise de termos compostos.

Os termos do vocabulário definidos pelo utilizador, como sinónimos e palavras hierárquicas, também se aplicam a partes compostas utilizadas como palavras únicas na consulta.

Para permitir uma análise avançada de termos compostos:

1. Na vista Colecções, expanda a colecção que pretende editar. No painel Analisar e Indexar, faça clique em **Configurar** > **Opções de análise**.

- 2. No campo **Termos compostos**, seleccione **Permitir análise avançada de termos compostos** e faça clique em **OK**.
- 3. No painel Analisar e Indexar, reinicie os serviços de análise e indexação.

Conceitos relacionados:

"Caracteres globais em consultas" na página 296

- E Suporte linguístico para procura semântica
- Suporte Linguístico no IBM Content Analytics with Enterprise Search

Activar suporte para procura de XML nativa

Se uma colecção incluir documentos XML, poderá permitir aos utilizadores utilizar a marcação XML quando procurarem documentos, permitindo a procura XML nativa da colecção.

Uma procura XML nativa, tal com os fragmentos XPath ou XML, pode fornecer vários resultados precisos da procura, explorando a marcação XML dos documentos na consulta. Os utilizadores podem especificar que um termo da consulta tem de ocorrer num determinado elemento ou atributo de XML.

Por exemplo, as facturas de um retalhista de computador em que re no formato XML pode conter no formato pode conter introdução de <encomendas> que incluem os elementos <empresa> e <tipocomputador>. Para obter as facturas que contêm encomendas de blocos de notas IBM, uma procura por palavra-passe de IBM e bloco de notas podem obter documentos que incluem os portáteis Dell e modelos de secretária da IBM.

Ao utilizar a procura XML, pode especificar que IBM tem de aparecer dentro do elemento <empresa>, esse bloco de notas tem de aparecer no elemento <tipocomputador>, e ambos os elementos têm de se encontrar sob o mesmo elemento , a<encomenda>. Desta forma, obtém facturas que examinam especialmente os blocos de notas de IBM.

Para permitir que os utilizadores consultem uma colecção com a procura XML nativa:

- 1. Na vista Colecções, expanda a colecção que pretende editar. No painel Analisar e Indexar, faça clique em **Configurar** > **Opções de análise**.
- 2. No campo **Procura XML nativa**, seleccione **Permitir que os utilizadores procurem documentos XML com a procura XML nativa** e faça clique em **OK**.
- 3. No painel Analisar e Indexar, reinicie os serviços de análise e indexação.

Conceitos relacionados:

"Correlacionamento de campo de índice para elementos XML" na página 206

- Suporte linguístico para procura semântica
- 🖙 Suporte Linguístico no IBM Content Analytics with Enterprise Search
- Aplicações de procura semântica
- 🖙 Termo da consulta de procura semântica

Facetas para colecções de procura empresarial

Numa colecção de procura empresarial, pode seleccionar facetas para filtrar e limitar os resultados da pesquisa para encontrar determinados documentos de interesse. Uma *faceta* faculta uma forma de organizar ou classificar o conteúdo para que os utilizadores possam refinar e navegar numa colecção por vários caminhos, em que cada caminho apresenta uma vista diferente ou perspectiva do conteúdo. Pode configurar tipos diferentes de facetas para ver o conteúdo de uma colecção de aspectos diferentes:

- Com uma *faceta plana*, todos os caminhos possíveis estão ao mesmo nível, como uma faceta denominada Estado que inclui valores para Nova Iorque, Califórnia, Virgínia, e assim por diante. Ao seleccionar facetas, pode limitar rapidamente os resultados a documentos que remetem para um determinado estado ou estados.
- Uma faceta hierárquica permite-lhe explorar níveis de classificação imbricados, como uma categoria de música que lhe permite pesquisar por género, nome do artista, nome da canção e assim por diante. É possível configurar facetas hierárquicas apenas para colecções de procura empresarial. Este tipo de faceta não é suportado em colecções de análise de conteúdo.
- Uma *faceta da data* é uma faceta de quatro níveis construída a partir de um campo de data paramétrico. A faceta consiste num valor de ano, mês, dia do mês e hora.
- Uma *faceta de intervalo decimal* permite-lhe categorizar o conteúdo segundo um intervalo de valores numéricos decimais, tais como uma faceta que ajuda a limitar uma pesquisa para modelos que estejam disponíveis num determinado intervalo de preços. Uma faceta assim pode ser representada da seguinte forma nos resultados da pesquisa:

```
Preço
Menos de 15000
15000 - 20000
20001 - 25000
25001 - 30000
Mais de 30000
```

• Uma *faceta do intervalo de data* permite-lhe categorizar o conteúdo segundo um intervalo de datas, como uma faceta que o ajuda a limitar uma pesquisa de registos de garantia de produtos para pedidos que foram submetidos esta semana, este mês, entre Janeiro e Março deste anos e assim por diante. Uma faceta assim pode ser representada da seguinte forma nos resultados da pesquisa:

```
Pedidos
Hoje
Esta semana
Este mês
01/01/2010 - 03/31/2010
```

Pesquisa facetada em colecções de procura empresarial

As facetas permitem-lhe ver aspectos diferentes dos resultados da pesquisa actual. O conteúdo estruturado pode ser associado a facetas pelos anotadores que adicionam valores de facetas a conteúdo como metadados, ou pode explicitamente criar facetas para uma colecção.

Para criar, explicitamente, uma faceta, cria, primeiro, campos de índice e, depois, correlaciona os campos de índice com a faceta.

Para activar a procura facetada de elementos XML e de elementos HTML, correlaciona os elementos com campos de índice que são correlacionados com facetas.

Para activar documentos para serem associados a facetas quando forem pesquisados e preparados para indexação, correlacione campos da origem de

dados ou campos de metadados para a faceta quando configurar opções de pesquisa. É possível correlacionar o campo de origem de dados ou o campo de metadados com um campo de índice que esteja correlacionado com uma faceta, ou pode especificar que o campo da origem de dados actual ou que o campo de metadados deve ser utilizado como uma faceta.

Para restringir um conjunto extenso de resultados, os utilizadores podem seleccionar facetas que serão apresentadas no formato de árvore ou num gráfico, no formato de gráfico de barras, dependendo da forma como a aplicação de procura empresarial está configurada. As facetas são devolvidas de forma dinâmica com base nos documentos que se encontram nos resultados da procura. Para uma faceta plana ou faceta hierárquica ser apresentada, um valor da faceta deve ocorrer em pelo menos um documento que corresponda às condições de pesquisa. Quando configurar uma faceta de intervalo decimal ou uma faceta de intervalo de dados, escolha ser os intervalos serão apresentados mesmo quando não corresponderem documentos na correspondência dos resultados da pesquisa.

Se um utilizador selecciona vários valores de facetas como, por exemplo, FacetA/dog e FacetB/cat, e os adicionar à pesquisa, a consulta é processada como um pedido Boolean AND: "FacetA/dog FacetB/cat". Se o utilizador seleccionar vários valores de facetas da mesma faceta como, por exemplo, FacetA/dog e FacetA/cat, a consulta será processada como um pedido Boolean OR: (FacetA/dog OR FacetA/cat).

Facetas numa procura de colecção federada

Quando os utilizadores procuram várias colecções de procura empresarial de uma só vez, são apresentados resultados agregados. Se várias colecções incluírem facetas, as facetas e as contagens de facetas também são agregadas. Quando configura o suporte para a procura facetada federada, é possível seleccionar o modo como os resultados federados são devolvidos:

- Uma federação de iteração. Esta opção federa os melhores resultados K de cada colecção.
- Várias iterações através de um algoritmo limite. Esta opção assegura os melhores resultados K federados. Esta abordagem necessita de obtenção adicional de aspectos e pode necessitar de mais tempo, consoante a distribuição de facetas nas colecções federadas.

O exemplo seguinte mostra o modo como os valores da faceta de duas colecções podem ser agregados nos resultado:

```
Colecção 1
Autor
- Roberts (12)
- Smith (11)
Colecção 2
Autor
- Roberts (3)
Tipo de ficheiro
- Word (5)
Resultados
Autor
- Roberts (15)
- Smith (11)
Tipo de ficheiro
- Word (5)
```

Restrição: As seguintes limitações aplicam-se a uma procura de colecção federada:

- Se as funções de análise estiverem activadas para a aplicação, o gráfico de série cronológica mostra o ano como a escala de tempo predefinida. Os utilizadores podem, então, pesquisar mais detalhadamente unidades de tempo mais específicas.
- A sinalização de documentos e a capacidade de criar categorias a partir de uma consulta não são suportadas. Para sinalizar documentos ou criar categorias, tem de procurar uma única colecção de cada vez.

Conceitos relacionados:

"Campos de índice" na página 191

"Facetas em colecções protegidas" na página 385

Tarefas relacionadas:

"Configurar facetas hierárquicas e simples para uma colecção de procura empresarial"

"Configurar facetas de intervalo decimal para uma colecção de procura empresarial" na página 184

"Configurar facetas de intervalo de data para uma colecção de procura empresarial" na página 186

"Configurar a árvore da faceta para colecções de análise de conteúdo" na página 251

"Desactivar a segurança de nível de documento" na página 414

Configurar facetas hierárquicas e simples para uma colecção de procura empresarial

Quando especificar as opções de análise e de índice para uma colecção de procura empresarial, pode criar facetas. Ao correlacionar campos de índice com facetas, permite que os utilizadores restrinjam os resultados da procura e visualizem documentos específicos do seu interesse.

Se alterar uma definição de faceta depois de pesquisar e indexar dados, o índice torna-se inconsistente. Para tornar as alterações disponíveis para os utilizadores que procurem a colecção, tem de reiniciar os pesquisadores que utilizam a faceta alterada de modo a que os documentos possam ser pesquisados e indexados novamente. Se a colecção estiver configurada para utilizar a memória cache de documentos, o índice pode ser reconstruído a partir da memória cache, sem voltar a pesquisar os documentos.

Facetas hierárquicas: Todos os campos numa configuração de facetas hierárquicas devem ter um valor para o documento coincidir. Por exemplo, dadas as seguintes condições, o documento não terá nenhum valor para a faceta de localização:

- A faceta de localização consiste nos campos país, estado e cidade.
- O documento tem país: US e estado: CA, mas não tem valor para o campo cidade.

Os valores das facetas hierárquicas ficam incoerentes, se o campo especificado tiver vários valores de campo. Por exemplo, dadas as seguintes condições, o documento poderá ter o valor de faceta de localização-JP-CA-San Jose (o valor US poderá ser substituído pelo valor JP).

- A faceta de localização consiste nos campos país, estado e cidade.
- O documento tem país: US, país: JP, estado: CA, e cidade: San Jose.

A faceta Data é um tipo particular de faceta hierárquica que é construída a partir de um campo de data paramétrico. Por exemplo, se o valor de um campo de data for 2010/04/01/24, o valor é analisado na seguinte faceta hierárquica, com valores distintos para o ano, mês, dia do mês e hora:

Para configurar facetas simples e hierárquicas:

- 1. Na vista Colecções, expanda a colecção que pretende configurar. No painel Análise e Indexação, faça clique em **Configurar** > **Facetas**.
- 2. Na página Definições de faceta, faça clique em Criar faceta.
- 3. Na página Criar uma Faceta, escreva um nome descritivo para a faceta e especifique se a faceta é uma faceta simples, hierárquica ou se é uma faceta de dados. O tipo de faceta que configurar determina a forma como ela é apresentada nos resultados da procura:
 - Os campos de índice que correlacionar com uma faceta simples irão aparecer todos no mesmo nível da árvore da faceta.
 - Os campos de índice que correlacionar com uma faceta hierárquica, inclusive a faceta de Data, estão imbricados hierarquicamente na árvore da faceta.
- 4. Seleccione o campo de índice que quer correlacionar com esta faceta. Depois de fazer clique em **OK**, a nova faceta aparece listada na página Definições de facetas com outras facetas que foram definidas para esta colecção.
- 5. Se quiser correlacionar mais campos de índice com uma faceta simples ou com uma faceta hierárquica, edite a faceta, seleccione a posição onde quer adicionar o campo de índice na árvore da faceta (para uma faceta simples, deve seleccionar a faceta raiz), faça clique em **Adicionar um campo** e, de seguida, seleccione o campo de índice que quer adicionar.

A árvore da faceta é actualizada de modo dinâmico para mostrar como os campos de índice serão listados ou imbricados quando os utilizadores procuram facetas nos resultados da procura. Reveja a árvore de facetas para se assegurar de que esta organização é a que pretende. Para modificar uma árvore da faceta, pode editar e eliminar campos de índice individuais.

Sugestão: Quando editar uma faceta de Data, a única alteração que pode fazer é especificar um nome diferente para a faceta. Se quiser alterar o campo de índice que está correlacionado com a faceta, deve eliminá-la e criar uma nova faceta.

- 6. Execute todas ou uma das seguintes acções para configurar a sua colecção para utilizar as facetas:
 - a. Configure os pesquisadores para utilizar facetas. Ao configurar o espaço de pesquisa para um pesquisador, pode especificar opções para pesquisar campos da origem de dados específicos e campos de metadados. Ao configurar estes campos, pode correlacionar campos individuais com campos de índice que havia correlacionado com as facetas previamente, ou pode seleccionar uma caixa de verificação para especificar que o campo pode ser usada para procura facetada. Depois, os utilizadores podem restringir os resultados para documentos que contêm estes campos ou campos de metadados, seleccionando facetas quando procurarem uma colecção.
 - b. Correlacionar elementos XML e elementos HTML com campos de índice. Ao configurar opções de análise para a colecção, pode correlacionar elementos XML ou HTML com os campos de índice que havia previamente correlacionado com facetas. Os utilizadores podem depois restringir os

resultados para partes específicas dos documentos XML e HTML seleccionando facetas quando pesquisarem uma colecção.

Conceitos relacionados:

"Facetas para colecções de procura empresarial" na página 179

Tarefas relacionadas:

"Correlacionar elementos XML com campos de índice" na página 207

"Correlação de elementos metadados HTML para os campos do índice" na página 210

Configurar facetas de intervalo decimal para uma colecção de procura empresarial

Ao criar facetas que se baseiam num intervalo numérico decimal, permite aos utilizadores restringir os resultados seleccionando um intervalo de valores, como registos de hospital para pacientes de determinada idade ou produtos para venda a vários preços.

Se alterar uma faceta de intervalo decimal, essa alteração reflecte-se na aplicação de procura empresarial após um curto espaço de tempo. Para disponibilizar de imediato essas alterações aos utilizadores, deve reiniciar os servidores da procura para a colecção.

Para configurar uma faceta de intervalo decimal:

- 1. Na vista Colecções, expanda a colecção que pretende configurar. No painel Análise e Indexação, faça clique em **Configurar** > **Facetas**.
- 2. Na página Definições de faceta, faça clique em Criar faceta.
- **3**. Na página Criar uma Faceta, escreva um nome descritivo para a faceta e especifique que a faceta é uma faceta de intervalo decimal.
- 4. Seleccione o campo de índice que quer correlacionar com esta faceta. A lista de campos de índice disponíveis inclui apenas campos decimais paramétricos. Depois de fazer clique em OK, a nova faceta aparece listada na página Definições de facetas com outras facetas que foram definidas para esta colecção.
- 5. Faça clique no ícone **Editar** para a faceta configurar os intervalos dos valores que pretende que esta faceta inclua.
- 6. Na página Editar uma Faceta, especifique como pretende que as facetas sejam organizadas na árvore da faceta quando os utilizadores procurarem os resultados da procura. Pode escolher se os intervalos que definir para esta faceta serão apresentados na árvore da faceta mesmo que nenhum dos documento nos resultados corresponda a um intervalo. Também pode especificar uma ordem de apresentação para os intervalos. Especifique se pretende listar os intervalos com um número mais elevado de correspondências a um nível superior na árvore da faceta ou se pretende listar os intervalos na ordem em que os organizou nesta página.
- Decida onde, na estrutura da faceta, pretende configurar um intervalo. Seleccione Raiz para adicionar um intervalo no nível superior, ou seleccione um intervalo previamente definido para adicionar um intervalo a um subnível, e depois faça clique em Adicionar Intervalo.
- Especifique um nome de apresentação significativo para o intervalo, como 0 a 100 ou Entre 50 e 60. Os intervalos que estão definidos ao mesmo nível têm de ter nomes únicos. O campo do nome do intervalo não pode estar vazio.
- 9. Faça clique no ícone **Editar** para expandir os campos de modo a definir o intervalo decimal.

10. Seleccione se pretende configurar um intervalo definindo valores ou se pretende criar um intervalo que englobe todos os valores restantes ao mesmo nível que este intervalo e não estejam explicitamente definidos. Por exemplo, se tiver a seguinte estrutura de intervalo:

Intervalo	Sub-intervalo	Definição de intervalo explícita
0 - 1000		0 < Preço ≤ 1000
1000 - 2000		1000 ≤ Preço ≤ 2000
	1500 - 1800	1500 ≤ Preço ≤ 1800
	Valores não explícito	S

O intervalo criado para todos os valores que não foram explicitamente definidos inclui valores superiores ou iguais a 1,000, inferiores a 1,500 ou superiores a 1,800, e inferiores ou iguais a 2,000.

11. Se criar um intervalo para todos os outros valores que não estejam explicitamente definidos, especifique se pretende que este intervalo inclua documentos que não têm o campo especificado ou documentos nos quais o campo está vazio. Por exemplo, partindo do princípio que a colecção inclui estes documentos:

```
Documento A: Preço=200
Documento B: Preço=100
Documento C:
```

O intervalo decimal inclui as seguintes definições:

```
Raiz
preço > 150
Criar um intervalo para todos os outros valores
```

Caso a caixa de verificação **Incluir documentos com valores nulos ou vazios** esteja limpa, o intervalo para todos os outros valores inclui apenas o Documento B porque o valor do Documento A é superior a 150 e o Documento C não possui um campo Preço. Caso esta caixa de verificação esteja seleccionada, o intervalo criado para todos os valores inclui tanto o Documento B como o Documento C.

- **12**. Execute os passos seguintes para definir um intervalo para uma faceta de intervalo decimal:
 - a. Seleccione ou escreva o valor mais baixo para o intervalo, como 0 ou 50. Caso pretenda definir um limite mais baixo para o intervalo, seleccione a caixa de verificação do limite mais baixo. Se a caixa de verificação estiver desmarcado, significa que não existe um limite mais baixo para o intervalo. Por exemplo, se a caixa de verificação do limite mais baixo estiver desmarcado para o intervalo Entre 50 e 60, o intervalo também inclui valores inferiores a 50. Se não especificar um limite mais baixo para o intervalo, tem de especificar um limite mais alto.
 - b. Seleccione o operador que define como os valores decimais na colecção serão comparados com o valor de intervalo mais baixo. Por exemplo, caso tenha seleccionado o operador inferior ou igual a (≤) para o limite mais baixo 50, os valores correspondem quando são superiores ou iguais a 50.
 - c. Especifique o valor mais alto para o intervalo, como 100 ou 60. Caso pretenda definir um limite mais alto para o intervalo, seleccione a caixa de verificação de limite mais alto. Se a caixa de verificação estiver desmarcado, significa que não existe limite superior para o intervalo. Por exemplo, se a caixa de verificação do limite mais alto for desmarcado para o intervalo Entre 50 e 60, o intervalo inclui valores que também são mais elevados do que 60. Se não especificar um limite superior para o intervalo, tem de especificar um limite mais baixo.

- d. Seleccione o operador que define a forma como os valores decimais na colecção serão comparados ao valor de intervalo mais elevado. Por exemplo, se seleccionar o operador menor que ou igual (≤) para o limite mais alto 60, os valores correspondem quando são inferiores ou iguais a 60.
- 13. Repita os passos anteriores para definir intervalos adicionais para esta faceta. Por exemplo, poderá criar uma faceta que inclua os intervalos Abaixo de 20, 20 a 50, 50 a 80 e Acima de 80 para representar todos os intervalos possíveis que os utilizadores queiram procurar.
- 14. Opcional: Elimine um intervalo que já não pretende usar com esta faceta seleccionando o intervalo e fazendo clique em **Remover Intervalo**.
- **15**. Execute todas ou uma das seguintes acções para configurar a sua colecção para utilizar as facetas:
 - a. Configure os pesquisadores para utilizar facetas. Ao configurar o espaço de pesquisa para um pesquisador, pode especificar opções para pesquisar campos da origem de dados específicos e campos de metadados. Ao configurar estes campos, pode correlacionar campos individuais com campos de índice que havia correlacionado com as facetas previamente, ou pode seleccionar uma caixa de verificação para especificar que o campo pode ser usada para procura facetada. Depois, os utilizadores podem restringir os resultados para documentos que contêm estes campos ou campos de metadados, seleccionando facetas quando procurarem uma colecção.
 - b. Correlacionar elementos XML e elementos HTML com campos de índice. Ao configurar opções de análise para a colecção, pode correlacionar elementos XML ou HTML com os campos de índice que havia previamente correlacionado com facetas. Os utilizadores podem depois restringir os resultados para partes específicas dos documentos XML e HTML seleccionando facetas quando pesquisarem uma colecção.

Conceitos relacionados:

"Facetas para colecções de procura empresarial" na página 179

Configurar facetas de intervalo de data para uma colecção de procura empresarial

Ao criar facetas que se baseiam num intervalo de data, vai permitir aos utilizadores restringir resultados seleccionando um intervalo de valores, como apólices de seguros que foram submetidas hoje, este mês ou durante um período de tempo específico.

Se alterar uma faceta do intervalo de data, a alteração reflecte-se na aplicação de procura empresarial após um curto período de tempo. Para disponibilizar de imediato essas alterações aos utilizadores, deve reiniciar os servidores da procura para a colecção.

Para configurar uma faceta do intervalo de data:

- 1. Na vista Colecções, expanda a colecção que pretende configurar. No painel Análise e Indexação, faça clique em **Configurar** > **Facetas**.
- 2. Na página Criar uma Faceta, escreva um nome descritivo para a faceta e especifique que a faceta é uma faceta de intervalo de data.
- **3**. Seleccione o campo de índice que quer correlacionar com esta faceta. A lista de campos de índice disponíveis inclui apenas campos de data paramétricos.

Depois de fazer clique em **OK**, a nova faceta aparece listada na página Definições de facetas com outras facetas que foram definidas para esta colecção.

- 4. Faça clique no ícone **Editar** para a faceta configurar os intervalos dos valores que pretende que esta faceta inclua.
- 5. Na página Editar uma Faceta, especifique como pretende que as facetas sejam organizadas na árvore da faceta quando os utilizadores procurarem os resultados da procura. Pode escolher se os intervalos que definir para esta faceta serão apresentados na árvore da faceta mesmo que nenhum dos documento nos resultados corresponda a um intervalo. Também pode especificar uma ordem de apresentação para os intervalos. Especifique se pretende listar os intervalos com o número mais elevado de correspondências num nível superior na árvore, ou se pretende listar os intervalos pela ordem em que os organizou nesta página.
- 6. Decida onde, na estrutura da faceta, pretende configurar um intervalo. Seleccione **Raiz** para adicionar um intervalo no nível superior, ou seleccione um intervalo previamente definido para adicionar um intervalo a um subnível, e depois faça clique em **Adicionar Intervalo**.
- Especifique um nome de apresentação significativo para o intervalo, como Anterior a 2010 ou Este Mês. Os intervalos que estão definidos ao mesmo nível têm de ter nomes únicos. O campo do nome do intervalo não pode estar vazio.
- 8. Faça clique no ícone **Editar** para expandir os campos para definir o intervalo de data.
- 9. Seleccione o tipo de intervalo que pretende criar:
 - Configure uma data absoluta para restringir o intervalo para uma data que ocorra numa data específica ou entre duas datas específicas, como de 1/4/2010 a 30/4/2010. Todos os subníveis que definir para um intervalo de data absoluto devem ser igualmente intervalos de datas absolutos.
 - Configure uma data relativa para representar intervalos de tempo relativos à data actual. Por exemplo, poderá querer definir um intervalo de data relativo para localizar documentos que foram publicados há um ano ou documentos para um evento que ocorreu entre dois a quatro dias atrás. Todos os subníveis que define para um intervalo de data relativo têm também de ser intervalos de data relativos.
 - Seleccione intervalos predefinidos para hoje, para esta semana, este mês e para este de modo a facilitar a visualização de resultados relativos para a data actual. Pode colocar estes intervalos predefinidos no nível superior de uma definição da faceta do intervalo de data.
 - Adicione um intervalo que abranja todos os valores restantes ao mesmo nível que este intervalo que não estejam definidos explicitamente.
- 10. Se criar um intervalo para todos os valores restantes ao mesmo nível deste intervalo que não estejam explicitamente definidos, especifique se pretende que este intervalo inclua documentos que não têm o campo especificado ou documentos nos quais o campo está vazio. Caso a caixa de verificação Incluir documentos com valores nulos ou vazios esteja limpa, serão excluídos da faceta aqueles documentos que não contenham campo de data ou que tenham valores vazios para o campo de data.
- 11. Execute os passos seguintes para definir valores absolutos para o intervalo:
 - a. Seleccione ou escreva o valor mais baixo para o intervalo, como 1/4/2010. Caso pretenda definir um limite mais baixo para o intervalo, seleccione a caixa de verificação do limite mais baixo. Se a caixa de verificação estiver desmarcado, significa que não existe um limite mais baixo para o intervalo.

Por exemplo, se a caixa de verificação do limite mais baixo estiver desmarcado para o intervalo Entre 1/4/2010 e 30/4/2010, o intervalo inclui valores que também são anteriores a 1/4/2010. Se não especificar um limite mais baixo para o intervalo, tem de especificar um limite mais alto.

- b. Seleccione o operador que define como os valores de data na colecção serão comparados com o valor de intervalo mais baixo. Por exemplo, se seleccionar o operador inferior a ou igual a (≤) para o limite mais baixo 1/4/2010, os valores correspondem quando são superiores ou iguais a 1/4/2010.
- c. Especifique o valor mais alto para o intervalo, como por exemplo 30/4/2010. Caso pretenda definir um limite mais alto para o intervalo, seleccione a caixa de verificação de limite mais alto. Se a caixa de verificação estiver desmarcado, significa que não existe limite superior para o intervalo. Por exemplo, se a caixa de verificação do limite mais alto estiver desmarcado para o intervalo Entre 1/4/2010 e 30/4/2010, o intervalo inclui também os valores que são posteriores a 30/4/2010. Se não especificar um limite superior para o intervalo, tem de especificar um limite mais baixo.
- d. Seleccione o operador que define como os valores de data na colecção serão comparados ao valor de intervalo mais alto. Por exemplo, se seleccionar o operador menos que ou igual a (≤) para o limite mais alto 30/4/2010, os valores correspondem quando são inferiores ou iguais a 30/4/2010.
- 12. Execute os passos seguintes para definir valores relativos para o intervalo:
 - a. Seleccione ou escreva as condições que definem o valor mais baixo no intervalo, como por exemplo dois dias antes da data actual ou seis meses após a data actual. Da mesma forma que configura valores de data absoluta, seleccione a caixa de verificação para indicar se pretende definir este valor como o limite mais baixo para o intervalo. Se não especificar um limite mais baixo para o intervalo, tem de especificar um limite mais alto.
 - b. Seleccione o operador que define como os valores de data na colecção serão comparados ao valor do intervalo mais baixo, como inferior a (<), ou inferior ou igual a (≤).
 - c. Seleccione ou escreva as condições que definem o valor mais alto no intervalo, como por exemplo um ano antes da data actual. Da mesma forma que configura valores de data absolutos, seleccione a caixa de verificação para indicar se pretende definir este valor como o limite mais alto para o intervalo. Se não especificar um limite superior para o intervalo, tem de especificar um limite mais baixo.
 - d. Seleccione o operador que define como os valores de data na colecção serão comparados ao valor de intervalo mais alto.
- 13. Repita os passos anteriores para definir intervalos de data adicionais para esta faceta. Por exemplo, pode criar uma faceta que inclua os intervalos Hoje, Este Mês, Este Ano e Há Um Ano para representar todos os intervalos possíveis que os utilizadores possam querer procurar.
- 14. Opcional: Elimine um intervalo que já não pretende mostrar para esta faceta seleccionando o intervalo e fazendo clique em **Remover Intervalo**.
- **15**. Execute todas ou uma das seguintes acções para configurar a sua colecção para utilizar as facetas:
 - a. Configure os pesquisadores para utilizar facetas. Ao configurar o espaço de pesquisa para um pesquisador, pode especificar opções para pesquisar campos da origem de dados específicos e campos de metadados. Ao

configurar estes campos, pode correlacionar campos individuais com campos de índice que havia correlacionado com as facetas previamente, ou pode seleccionar uma caixa de verificação para especificar que o campo pode ser usada para procura facetada. Depois, os utilizadores podem restringir os resultados para documentos que contêm estes campos ou campos de metadados, seleccionando facetas quando procurarem uma colecção.

b. Correlacionar elementos XML e elementos HTML com campos de índice. Ao configurar opções de análise para a colecção, pode correlacionar elementos XML ou HTML com os campos de índice que havia previamente correlacionado com facetas. Os utilizadores podem depois restringir os resultados para partes específicas dos documentos XML e HTML seleccionando facetas quando pesquisarem uma colecção.

Conceitos relacionados:

"Facetas para colecções de procura empresarial" na página 179

Dicionários lexicais para colecções de procura empresarial

Pode melhorar a qualidade e precisão dos resultados da procura integrando dicionários lexicais personalizados em colecções.

O IBM Content Analytics with Enterprise Search fornece dicionários incorporados para analisar e segmentar conteúdo. Pode criar um dicionário personalizado para assegurar que o analisador utiliza o vocabulário específico da empresa e as construções gramaticais ao preparar o conteúdo para o índice. Uma vez que as colecções de análise de conteúdo suportam dicionários de utilizador, pode configurar dicionários lexicais apenas para colecções de procura empresarial.

Os exemplos de como um dicionário lexical pode melhorar a qualidade da pesquisa, incluem:

- Se os dados incluírem pronomes possessivos, como da Rita, o analisador cria dois tokens: [Rita] e [da]. Uma procura de Rita devolve documentos com Rita e documentos com apóstrofe, o que provavelmente não é a intenção da consulta. Adicionar o termo da Rita a um dicionário lexical personalizado pode ajudar a assegurar que apenas documentos que contenham a cadeia da Rita são devolvidos nos resultados da pesquisa.
- Se os dados incluíram referências à Declaração Chinesa dos Direitos Humanos (世界人权宣言), o analisador predefinido cria dois tokens: [世界] e [

人权宣言], resultando em que muitos mais documentos são considerados correspondentes e devolvidos nos resultados da pesquisa. Adicionar o termo

completo, 世界人权宣言, a um dicionário lexical personalizado pode melhorar a precisão dos resultados certificando-se que apenas os documentos que referem essa Declaração Chinesa dos Direitos Humanos são devolvidos.

Em certos casos, uma palavra registada como sendo de um certo idioma não tem qualquer efeito em documentos escritos noutros idiomas. Por exemplo, uma palavra registada como inglês não tem qualquer efeito em documentos escritos em japonês. Noutros casos, uma palavra registada pode não ter qualquer efeito na pesquisa porque uma regra que foi definida para o sistema tem prioridade em relação ao dicionário específico da colecção. Para criar um dicionário lexical personalizado, os peritos no domínio ou na matéria têm de definir a terminologia específica da empresa que pretende utilizar para analisar o conteúdo num ficheiro XML. De seguida, utilizam a ferramenta ES_INSTALL_ROOT/bin/eslexdictbuilder para criar um ficheiro de dicionário (ficheiro .dic).

Para utilizar o dicionário personalizado, tem de carregar o dicionário para o sistema e associá-lo a uma ou mais colecções.

Depois de carregar um dicionário no sistema, não pode utilizar a consola de administração para modificá-lo. Para modificar o dicionário, tem de actualizar o ficheiro XML, reconstruir o dicionário e carregar o novo dicionário no sistema.

Tarefas relacionadas:

"Adicionar dicionários lexicais ao sistema"

"Associar um dicionário lexical personalizado a uma colecção de procura empresarial" na página 191

🖙 Criar um dicionário lexical para colecções de pesquisa

Referências relacionadas:

Ficheiros XML para dicionários lexicais para colecções de pesquisa

Adicionar dicionários lexicais ao sistema

Se criar um dicionário lexical personalizado para uma colecção de procura empresarial, deve adicioná-lo ao sistema antes de o conseguir utilizar para analisar o conteúdo.

Para criar um dicionário lexical, tem de utilizar a ferramenta ES_INSTALL_ROOT/bin/eslexdictbuilder.

Para adicionar um dicionário lexical personalizado ao sistema:

- 1. Faça clique em Sistema para abrir a vista Sistema.
- 2. Na página Análise, faça clique em **Configurar um dicionário lexical personalizado**.
- 3. Na página Dicionários Lexicais, faça clique em Adicionar Dicionário Lexical.
- Na página Adicionar um Dicionário Lexical, escreva um nome descritivo para o novo dicionário. O sistema utiliza este nome de apresentação para identificar o dicionário através da consola de administração.
- 5. Especifique a localização do ficheiro .dic. Se o ficheiro for menor que 8 MB, o ficheiro pode estar no computador local e o utilizador pode utilizar a ferramenta de procura para o localizar. Se o ficheiro for maior do que 8 MB, este deve estar no servidor do controlador. Deve escrever o caminho totalmente qualificado para o ficheiro.
- 6. Faça clique em **OK**. O seu dicionário lexical personalizado está listado na página Dicionários Lexicais. Se pretender alterar o nome de apresentação, edite o dicionário.

Para utilizar o dicionário personalizado, tem de o associar a pelo menos uma colecção de procura empresarial. Não pode remover um dicionário do sistema, excepto se o remover primeiro de quaisquer colecções que o estiverem a utilizar.

Conceitos relacionados:

"Dicionários lexicais para colecções de procura empresarial" na página 189

Tarefas relacionadas:

"Associar um dicionário lexical personalizado a uma colecção de procura empresarial"

🖙 Criar um dicionário lexical para colecções de pesquisa

Referências relacionadas:

Ficheiros XML para dicionários lexicais para colecções de pesquisa

Associar um dicionário lexical personalizado a uma colecção de procura empresarial

Se criou dicionários lexicais personalizados anteriormente e os adicionou ao sistema, pode seleccionar um deles e utilizá-lo numa colecção de procura empresarial.

O analisador pode aplicar as construções de vocabulário e gramática personalizadas ao analisar conteúdos, para além de aplicar dicionários integrados. A utilização de um dicionário lexical personalizado que contenha vocabulário específico da empresa pode melhorar a precisão dos resultados da procura.

Quando alterar a definição do dicionário lexical personalizado para uma colecção, essas alterações entram em vigor depois de reiniciar o analisador. As definições alteradas aplicam-se a novos dados e não afectam dados já existentes no índice. Para actualizar documentos que já se encontram no índice, tem de pesquisar e indexar de novo os documentos. Se a colecção for configurada para utilizar a memória cache do documento, pode reconstruir o índice a partir da memória cache sem voltar a pesquisar os documentos.

Para associar um dicionário lexical personalizado a uma colecção de procura empresarial:

- 1. Na vista Colecções, expanda a colecção que pretende configurar.No painel Analisar e Indexar, faça clique em **Configurar** > **Dicionários lexicais**.
- Seleccione o dicionário que pretende utilizar com esta colecção. Se não existirem dicionários disponíveis, ou se seleccionar Nenhum seleccionado, o analisador aplica os dicionários integrados ao analisar os conteúdos adicionados por si à colecção.
- 3. No painel Analisar e Indexar, reinicie o serviço de análise e indexação.

Conceitos relacionados:

"Dicionários lexicais para colecções de procura empresarial" na página 189 Tarefas relacionadas:

"Adicionar dicionários lexicais ao sistema" na página 190

🖙 Criar um dicionário lexical para colecções de pesquisa

Referências relacionadas:

Ficheiros XML para dicionários lexicais para colecções de pesquisa

Campos de índice

Os campos de índice podem ajudar os utilizadores a obter documentos de várias origens sem precisarem de conhecer os nomes dos campos actuais.

Por exemplo, diferentes origens de dados podem ter diferentes formas de armazenar informação sobre o autor como, por exemplo, proprietário, criador, originador e assim em diante. Pode criar um campo de índice que representa esta informação comum, tal como Autor (Author). Quando configurar opções de pesquisa para as várias origens de dados, pode correlacionar campos de origens de dados individuais e campos de metadados a este campo de índice partilhado. Quando os utilizadores consultam a colecção, os dados do autor são devolvidos da mesma forma de todas as origens de dados.

Criar e configurar campos de índice

Pode criar e configurar campos de índice para uma colecção de várias formas:

Definições de campo de índice

Quando configurar opções de análise e índice para uma colecção, pode seleccionar uma opção para configurar campos de índice. Quando criar um campo de índice, pode seleccionar atributos para como o campo pode ser consultado e apresentado nos resultados. Pode também importar definições de campo de índice a partir de ficheiros XML.

Facetas

Quando cria um campo de índice, é criada uma faceta com o mesmo nome do campo de índice se seleccionar o atributo **Procura facetada**. Quando configura facetas, pode seleccionar os campos de índice que pretende correlacionar na faceta.

Elementos XML e elementos de metadados HTML

Quando especifica os elementos XML ou os elementos HTML que pretende activar para fins de procura, pode correlacionar os elementos a campos de índice. Os campos de índice devem existir antes de correlacionar elementos XML e HTML aos mesmos.

Extractor de texto em campos de metadados

Depois de o extractor de texto extrair campos de metadados do conteúdo, pode correlacionar os campos de metadados nos campos de índice. Os campos de índice devem existir antes de correlacionar campos de metadados aos mesmos.

Campos de origens de dados e campos de metadados pesquisados

Quando edita um espaço de pesquisa e especifica opções para pesquisar origens de dados individuais ou campos de metadados, pode correlacionar os campos de origens de dados e campos de metadados em campos de índice. Também pode criar campos de índice a partir da origem de dados e campos de metadados pesquisados e seleccionar atributos para como os novos campos de índice podem ser pesquisados e apresentados nos resultados da pesquisa.

Colunas de ficheiro CSV

Quando importa ficheiros CSV para uma colecção, pode correlacionar colunas de cabeçalhos para campos de índice. Também pode criar campos de índice a partir de colunas importadas e seleccionar atributos para como os novos campos de índice podem ser pesquisados e apresentados nos resultados de procura.

CAS para ficheiros de correlação de índice

No ficheiro XML que define quais as funcionalidadesUIMA que são correlacionadas como campos no índice, pode correlacionar uma funcionalidade no campo de índice. Se o campo de índice não existir, pode configurar a anotação de correlação para criá-lo. Se a anotação de correlação não criar o campo de índice, crie-o quando configurar as definições do campo de índice para a colecção.

REST API

Quando adiciona conteúdo a uma colecção utilizando REST APIs, pode correlacionar campos e metadados em campos de índice.

Ligar a facetas que utilizam campos de índice

A página de Definições de Campos de Índice mostra informação sobre as facetas que são correlacionadas para cada campo de índice. Se fizer clique na etiqueta da faceta, pode editar como a faceta utiliza o campo de índice. Para uma colecção de procura empresarial, a ligação abre a página de definições da faceta. Para uma colecção de análise de conteúdo, a ligação abre a página de configuração de uma árvore de facetas.

Ligar a origens que utilizam campos de índice

A página Definições de Campos de Índice mostra informação sobre as várias origens que utilizam cada campo de índice. Se precisar de alterar como uma origem de entrada correlaciona os seus campos no campo de índice, pode fazer clique na ligação de origem de entrada e ir directamente para a página de configuração de origem.Informação sobre algumas origens não pode ser apresentada porque os dados correlacionados não estão guardados. Por exemplo, a correlação de dados não está disponível para campos de índice que são criados através de REST APIs.

Ligações podem ser fornecidas para os seguintes tipos de origens de entrada:

- Campos de origens de dados e campos de metadados pesquisados. O nome do
 pesquisador é apresentado na coluna Origens de entrada. Se fizer clique no
 nome do pesquisador, pode editar o espaço de pesquisa e alterar a definição de
 campo de índice. Por exemplo, pode alterar os atributos para como o campo
 pode ser consultado ou apresentado nos resultados da consulta.
- Extractor de texto em campos de metadados. O nome do campo de metadados é apresentado na coluna Origens de entrada. Se fizer clique na ligação, pode editar como o extractor de texto correlaciona o campo de metadados no campo de índice.
- Elementos XML e elementos de metadados HTML. O nome do elemento é apresentado na coluna Origens de entrada. Se fizer clique no nome do elemento, pode editar como o elemento é correlacionado no campo de índice, assim como seleccionar um campo de índice diferente para o elemento.

Restrição: Se seleccionar **Todos os elementos metadados HTML** ou **Todos os elementos Dublin Core Metadata** quando configura correlações de metadados HTML e campos de índice, as ligações de origem de entrada não podem ser apresentadas. As ligações de origem de entrada são fornecidas apenas para elementos metadados que tenha explicitamente correlacionado com um campo do índice.

• CAS para correlações de índice. Se fizer clique em **CAS para índice** na coluna **Origens de entrada**, pode editar as opções de processamento de texto para a colecção.

Conceitos relacionados:

"Facetas para colecções de procura empresarial" na página 179

"Correlacionamento de campo de índice para elementos XML" na página 206

"Correlacionamentos de campo de índice para metadados de HTML" na página 209

"Correlacionamentos com campo de índice para metadados extraídos de conteúdo binário" na página 212

"Filtros de campos" na página 221

Tarefas relacionadas:

"Configurar filtros de campos" na página 222

"Exportar documentos de um colecção de procura empresarial para uma colecção de análise de conteúdos" na página 304

Criar campos de índice

Ao especificar opções de análise e indexação para uma colecção, pode criar campos de índice para representar dados que apareçam frequentemente em várias origens. Ao correlacionar campos de origem, elementos e facetas com os campos de índice, pode ajudar a assegurar que os dados são obtidos e devolvidos de forma consistente a partir de várias origens na colecção.

Caso altere a definição de um campo de índice depois de os dados terem sido adicionados ao índice, tem de reconstruir o índice. Se a cache de documento não estiver activada para a colecção de procura empresarial, tem de:

- Reinicie os pesquisadores que utilizam o campo de índice alterado de forma a que os documentos possam ser pesquisados e indexados novamente.
- Importe novamente documentos que utilizam o campo de índice alterado para a colecção de forma a que os documentos possam ser indexados novamente.

Restrição: Quando configura um pesquisador, o sistema cria automaticamente campos de índice específicos para o tipo de pesquisador. Por exemplo, é criado automaticamente um campo de índice para o nome da base de dados quando configura um Base de dados JDBC. Estes nomes de campos são reservados por IBM Content Analytics with Enterprise Search e não pode especificá-los como nomes de campos de índice que adiciona ao sistema.

Para adicionar um campo de índice e utilizá-lo numa colecção:

- 1. Na vista Colecções, expanda a colecção que pretende editar. No painel Analisar e Indexar, faça clique em **Configurar** > **Campos de índice**.
- Na página Definições do Campo de Índice, faça clique em Criar Campo de Índice.
- 3. Na página Criar um Campo de Índice, seleccione os atributos que pretende utilizar para procurar este campo e para visualizar o valor do campo nos resultados da procura. Depois de fazer clique em **OK**, o novo campo de índice aparece listado juntamente com os outros campos de índice que foram definidos para esta colecção.

Sugestão: Caso seleccione o atributo **Procura facetada** para um novo campo de índice, será criada uma faceta com o mesmo nome do campo de índice. Uma vez criado o campo de índice, não pode criar uma faceta editando a definição do campo de índice e seleccionando o atributo **procura facetada**. Para correlacionar um campo de índice existente com uma faceta, crie ou edite facetas na árvore da faceta.

4. Configure facetas e origens para serem utilizadas no campo de índice. Pode, por exemplo, correlacionar o campo de índice com uma faceta, ou editar um espaço de pesquisa e correlacionar um campo de origem de dados com um campo de índice. Como exemplo, pode correlacionar o campo Proprietário numa origem de dados e o campo Criador noutra origem de dados com um campo de índice denominado Autor.

Depois de correlacionar campos de origem ou facetas com o campo de índice, a página Definições de Campo de Índice mostra as origens e facetas que estão a utilizar o campo de índice. Se as colunas **Origens de entrada** e **Facetas** não forem mostradas, pode editar o campo de índice para ver que pesquisadores e facetas estão a utilizar o campo de índice. Não pode apagar um campo de índice se as origens de entrada e as facetas o estiverem a utilizar.

Caso pretenda alterar o modo como uma origem ou faceta utiliza um campo de índice, faça clique na ligação da origem de entrada ou na ligação da faceta. Por exemplo, se fizer clique na ligação da etiqueta da faceta, irá abrir a página de configuração da faceta. As ligações não estão disponíveis para alguns tipos de origens de entrada, tal como para campos que sejam adicionados por REST APIs ou para certos tipos de correlacionamentos de elementos de metadados.

Conceitos relacionados:

"Atributos do campo de índice"

Tarefas relacionadas:

"Remover informações sobre as origens de entrada e facetas do campo de índice" na página 206

Referências relacionadas:

"Nomes de campos de índice reservados" na página 199

Atributos do campo de índice

Pode correlacionar os campos de conteúdos aos campos de índice, ao configurar as opções para o conteúdo de procura que adicionar à colecção, como quando configura os pesquisadores, importa os ficheiros CSV, cria facetas ou configura elementos XML. Os atributos que especificar para o conteúdo de procura são armazenados no índice. Os atributos permitem restringir o que os utilizadores podem consultar e o que os utilizadores podem ver nos resultados de consulta.

Devolvido

Pode pretender procurar alguns campos, mas não os mostrar nos resultados da procura, ou pode pretender visualizar um campo nos resultados da procura mesmo se não o consultar. Por exemplo, poderá ter a necessidade de consultar dados financeiros para obter um relatório significativo, mas poderá não querer mostrar salários dos funcionários nos resultados que também mostram nomes dos funcionários.

Para activar um campo a ser mostrado nos resultados da procura, seleccione a caixa de verificação **Retornável**

Procura facetada

Quando um campo é utilizado como uma faceta, os utilizadores podem facilmente ver quantas ocorrências do campo e valores do campo existem na pesquisa ou nos resultados da análise. Numa aplicação de procura empresarial, os utilizadores podem seleccionar facetas para restringir a sua vista dos resultados e identificar melhor os documentos de interesse. Num prospecção de análise de conteúdo, os utilizadores seleccionam facetas para tornar visível uma dimensão da colecção, para uma análise mais aprofundada.

Para permitir que um campo seja utilizado como faceta, seleccione a caixa de verificação **Pesquisa facetada**. Se seleccionar este atributo, é criada uma faceta com o mesmo nome que o campo do índice. Se desejar que a faceta não tenha o mesmo nome, não seleccione este atributo. Pode criar uma faceta com um nome diferente ao configurar a sua árvore da mesma, e correlacionar um ou mais campos de índice à faceta.

Procura de texto livre

O índice é um índice de texto completo com conteúdo de várias origens de dados. Pode procurar o conteúdo, especificando uma consulta simples em linguagem normal. Para permitir que um campo seja procurado com uma consulta de texto livre, seleccione a caixa de verificação **Procura de texto livre**.

Se seleccionar a caixa de verificação **Em resumo**, o valor do campo torna-se parte do resumo do documento dinâmico nos resultados e pode ser utilizado para detectar documentos duplicados.

Exemplo 1:

Uma pesquisa de texto livre pode ser tão simples como a seguinte consulta:

corrente da bicicleta

Para indicar as palavras que têm ou não de aparecer num documento, pode incluir notações especiais. Por exemplo, pode preceder um palavra por um sinal de adição (+) para especificar que um documento tem de conter aquela palavra para que a correspondência ocorra. Preceda uma palavra por um sinal de subtracção (-) para excluir documentos que contêm aquela palavra a partir dos resultados da procura. Inclua duas ou mais palavras entre aspas (") para procurar uma frase exacta.

Exemplo 2:

Na consulta de texto livre seguinte, uma correspondência ocorre apenas se um documento contiver a frase exacta ficção científica e não contiver a palavra robô:

+"ficção científica" -robô

Procura por campo

Uma procura por campo permite-lhe restringir o objecto da consulta a campos de dados específicos e a campos de metadados num documento. Por exemplo, pode especificar que certas palavras têm de existir no título de um documento.

Para permitir que um campo seja procurado por nome de campo, seleccione a caixa de verificação **Procura por campo**.

Exemplo

Para especificar uma pesquisa com campos, inclua o nome do campo ou a palavra ou frase que tem de existir nesse campo na sua consulta. A consulta seguinte pesquisa documentos que têm de conter a palavra ibm e a frase análise de Texto no campo do título:

title: ibm title: "análise de texto"

Correspondência exacta e dependência de maiúsculas e minúsculas

Se permitir que um campo suporte pesquisa com campos, pode especificar preferências reforçando correspondência exacta e dependência de maiúsculas e minúsculas. Uma pesquisa de correspondência exacta pode melhorar a qualidade dos resultados da pesquisa permitindo-lhe especificar consultas precisas. Com a pesquisa por correspondência exacta, pode consultar campos e obter apenas os documentos em que todo o valor do campo corresponde aos termos da consulta. Se o valor do campo ou elemento contiver menos conteúdo ou conteúdo adicional, não ocorre uma correspondência.

Sugestão: Ao determinar se existe uma correspondência exacta, o sistema converte os termos da consulta para minúsculas, remove os espaços extra na cadeia da consulta e executa a correspondência do padrão de caracteres globais. No entanto, não ocorre a formação de lemas nem a procura de sinónimos e as palavras de paragem não são removidas. Uma correspondência completa com um elemento XML requer que o nome do elemento, sem elementos imbricados, e o valor completo do elemento correspondam exactamente aos termos da consulta.

Pode controlar se pretende respeitar ou ignorar diferenças entre letras maiúsculas e minúsculas em consultas. Por exemplo, se reforçar a dependência de maiúsculas e minúsculas, uma pesquisa para uima não corresponde ao valor UIMA. Se especificar que se pode ignorar as maiúsculas e minúsculas, os utilizadores podem inserir uima e localizar documentos que contenham Uima, UIMA, e assim por diante.

É possível consultar campos para correspondências exactas ao utilizar as sintaxes de consulta suportadas SIAPI, XMLFrag2 ou XPath. Um sinal de igual (=) a preceder os termos da consulta indica que será efectuada uma pesquisa de correspondência exacta.

Para permitir que um campo seja procurado para correspondências exactas, seleccione a caixa de verificação **Correspondência exacta**. Se permitir que os utilizadores pesquisem documentos de XML com sintaxe de XML nativa ao configurar opções de análise para uma colecção, todos os elementos de XML podem ser pesquisados para correspondências exactas dos termos de pesquisa.

Exemplo 1:

Um campo chamado cor contém o valor azul-escuro.

- A consulta de correspondência exacta seguinte corresponde porque a consulta não contém mais termos: cor="azul-escuro"
- A consulta de correspondência exacta seguinte não corresponde porque a consulta inclui a palavra saia além de azul-escuro: cor="saia azul-escuro"
- A consulta de correspondência exacta seguinte não corresponde porque o campo cor também contém a palavra escuro: cor=azul

Exemplo 2:

Sem uma correspondência exacta, a seguinte consulta XMLFrag2 pode devolver documentos que especificam <diagnosis>intraductal carcinoma comedo type</diagnosis> ou outros termos que não correspondem exactamente aos termos da consulta:

@xmlf2::'<diagnosis>intraductal carcinoma</diagnosis>'

Com correspondência exacta, a seguinte consulta XMLFrag2 assegura que apenas os documentos devolvidos são os que em todo o conteúdo do valor de elemento XML corresponde aos termos da consulta:

@xmlf2::'=<diagnosis>intraductal carcinoma</diagnosis>'

Outros atributos de procura por campos

Se permitir que um campo suporte a procura por campos, pode especificar outras preferências para o modo como o campo é procurado e classificado. Poderá proceder do seguinte modo:

- Especifique um valor de factor hierárquico, de 0 a 100, para aumentar, automaticamente, a importância de documentos que correspondem a consultas deste campo e classificam estes documentos em posições mais elevadas nos resultados de procura.
- Expanda, automaticamente, consultas, para que um termo de consulta de texto livre seja tratado como uma procura por campos deste campo.
- Apresente, automaticamente, sugestões de ortografia para este campo.

Exemplo

Se um utilizador submeter uma consulta de texto livre para encontrar documentos que contenham IBM e se a expansão estiver activada para um campo de índice denominado empresa, a consulta é automaticamente expandida para (IBM OU empresa:IBM). Se especificar, também, um factor hierárquico de 5.0 para o campo empresa, a consulta que é processada é a de (IBM OU empresa:IBM^5.0). Se IBM ocorrer no conteúdo ou no campo da empresa, estes documentos são colocados em níveis hierárquicos superiores nos resultados da procura.

Procura paramétrica

Uma procura paramétrica é um tipo de procura por campo que lhe permite efectuar consultas de comparação ou avaliação em campos numéricos e de data e em metadados. Por exemplo, pode procurar documentos de um determinado tamanho ou que estão escritos após uma determinada data. Também pode procurar documentos com atributos maiores, menores ou iguais a um valor especificado.

Para ordenar os resultados numericamente, de acordo com um valor do campo, tem de activar o campo para procura paramétrica.

Para procurar um campo com uma consulta paramétrica, ou para ser capaz de organizar os números numericamente, seleccione a caixa de verificação **Procura paramétrica**. Também vai especificar se o campo contém uma data ou dados numéricos.

Exemplo 1:

A consulta seguinte procura itens que custam exactamente 50 dólares (ou qualquer moeda indexada no campo preço):

#preço::=50

Exemplo 2:

A consulta seguinte procura documentos com um tamanho do ficheiro maior do que 1024 mas menor ou igual a 2048:

#tamanhoficheiro::>1024<=2048</pre>

Ordenável por texto

Se uma origem de dados incluir campos, ou se os utilizadores estiverem a pesquisar documentos de XML ou HTML, pode achar conveniente permitir que os resultados sejam ordenados pelos valores de um campo em particular. Na aplicação de procura empresarial, são mostrados os nomes de todos os campos que foram configurados para serem campos ordenáveis. Os utilizadores podem optar por ordenar resultados alfabeticamente (de acordo com uma ordenação de cadeias), através da selecção de um dos campos listados em vez de ordenarem os resultados por relevância ou data do documento. Os utilizadores também podem decidir se os documentos devem ser dispostos por ordem ascendente ou descendente.

Os documentos de resultados que não contenham o campo de ordenação são apresentados no fim dos resultados da procura. Os documentos de resultados que contenham o campo de ordenação, mas que tenham sido indexados antes de o campo ter sido configurado para ser ordenável, também são apresentados no fim dos resultados da procura.

Para permitir que os utilizadores ordenem os resultados da procura por ordem alfabética num campo, seleccione a caixa de verificação **Texto ordenável**. Se o campo contiver valores numéricos, seleccione a caixa de verificação **Procura paramétrica** para especificar que os valores do campo podem ser utilizados para ordenar os resultados da procura numericamente.

Analisável

Se a colecção for configurada para utilizar um anotador de Unstructured Information Management Architecture (UIMA), pode configurar um campo para ser analisável. Quando os documentos são processados no canal, os anotadores adicionam anotações, ou metadados, aos documentos. Esta informação adicional, que é armazenada com os documentos no índice, pode ajudar a melhorar a relevância dos resultados de procura e permitir a análise estatística para documentos numa colecção de análises de conteúdo. Os utilizadores podem seleccionar facetas para explorar os resultados de análises quando consultarem colecções.

Para permitir que os utilizadores consultem as anotações que foram adicionadas aos documentos, seleccione a caixa de verificação **Analisável**. Se a opção **Analisável** não estiver activada para um campo, o anotador não vê o campo nem cria metadados para o mesmo.

Conceitos relacionados:

Sintaxe de consultas

Nomes de campos de índice reservados

Ao configurar um pesquisador, o sistema automaticamente cria campos de índice que são específico ao tipo de pesquisador. Estes nomes de campo de índice são reservados pelo IBM Content Analytics with Enterprise Search.

Por exemplo, se configurar um pesquisador de DB2 ou de Base de dados JDBC, o sistema cria automaticamente um campo de índice para o nome da base de dados.

Para além dos nomes de campos reservados por pesquisadores, os seguintes nomes de campos são reservados porque são utilizados na sintaxe de consulta de IBM Content Analytics with Enterprise Search:

```
date
docid
link
samegroupas
site
title
url
```

Ao configurar os campos de índice para uma colecção, não pode especificar nenhum dos nomes de campo reservado como nomes de campo de índice. A seguinte tabela mostra nomes de campos adicionais reservados pelos vários pesquisadores.

Tipo de pesquisador	Nomes de campos reservados
Agente para sistemas de ficheiros Windows	Directory_Name File_Name File_Size Last_Modified_Date Title
Case Manager	caseid casestate casestype classname crawled_date date filename filesize lastmodified_date mimetype objecttype path title
Content Integrator	\$ServerName\$ \$RepositorySystemID\$ \$RepositoryPersistentID\$ \$RepositoryType\$ \$ItemName\$ \$ItemClassName\$ \$CreatedDate\$ \$CreatedDate\$ \$ModifiedDate\$ \$RevisionUser\$ \$RevisionUser\$ \$RevisionUser\$ \$Filename\$ \$Filename\$ \$FolderID\$ \$Folderf\$ \$Folder\$ \$ArchiveCharset\$ \$EntryName\$ \$EntryDate\$ \$EntrySize\$ \$Date\$

Tabela 18. Nomes de campos reservados por pesquisadores

Tipo de pesquisador	Nomes de campos reservados
Content Manager	\$ServerName\$ \$ItemTypeName\$ \$CreatedDate\$ \$ModifiedDate\$ \$Mimetype\$ \$VersionID\$ \$ObjectType\$ \$ObjectType\$ \$SemanticType\$ \$SemanticType\$ \$SemanticType\$ \$OrgFileName\$ \$PartNumber\$ \$PartNumber\$ \$ArchiveType\$ \$ArchiveCharset\$ \$EntryDate\$ \$EntrySize\$ \$Date\$
DB2	\$DatabaseName\$ \$TableName\$ \$ContentColumnName\$ \$ArchiveType\$ \$ArchiveCharset\$ \$EntryName\$ \$EntryDate\$ \$EntrySize\$ \$Date\$
Domino Document Manager	\$ServerName\$ \$LibraryDBName\$ \$BinderDBName\$ \$BinderDBName\$ \$DocumentDBName\$ \$DocumentDBTitle\$ \$DocumentDBTitle\$ \$VersionID\$ \$VersionID\$ \$FileName\$ \$VersionS\$ \$CreatedDate\$ \$RodifiedDate\$ \$NRPCPort\$ \$NRPCPort\$ \$DocumentUNID\$ \$ArchiveType\$ \$ArchiveCharset\$ \$EntryName\$ \$EntryDate\$ \$EntryDate\$ \$EntryDate\$ \$EntrySize\$ \$EntrySize\$ \$EntrySize\$ \$EntrySize\$ \$Date\$ \$NoteID\$ \$SearchType\$ \$SearchType\$ \$IUSeSSL\$

Tabela 18. Nomes de campos reservados por pesquisadores (continuação)

Tipo de pesquisador	Nomes de campos reservados
Exchange Server e Exchange Server 2000 e 2003	\$ServerName\$ \$PublicFolderURL\$ \$Extension\$ \$MessageClass\$ \$CreatedDate\$ \$ModifiedDate\$ \$ArchiveType\$ \$ArchiveCharset\$ \$ArchiveCharset\$ \$EntryName\$ \$EntryDate\$ \$EntrySize\$ \$Date\$
FileNet P8	content_size crawled_date docname filename lastmodified_date
Base de dados JDBC	\$DatabaseName\$ \$TableName\$ \$ContentColumnName\$ \$ArchiveType\$ \$ArchiveCharset\$ \$EntryName\$ \$EntryDate\$ \$EntrySize\$ \$Date\$
NNTP	Subject Message-ID From Newsgroups References Date Organization
Notes	\$ServerName\$ \$DatabaseName\$ \$DatabaseTitle\$ \$FileName\$ \$Extension\$ \$CreatedDate\$ \$ModifiedDate\$ \$Protocol\$ \$NRPCPort\$ \$DIIOPPort\$ \$DocumentUNID\$ \$View\$ \$ArchiveType\$ \$ArchiveCharset\$ \$EntryDate\$ \$EntryDate\$ \$EntrySize\$ \$EntrySize\$ \$EntrySize\$ \$Date\$ \$NoteID\$ \$SearchType\$ \$SearchType\$ \$IsUseSSL\$

Tabela 18. Nomes de campos reservados por pesquisadores (continuação)

Tipo de pesquisador	Nomes de campos reservados
Quickr for Domino	\$ServerName\$ \$PlaceName\$ \$PlaceTitle\$ \$RoomName\$ \$RoomTitle\$_ \$FileName\$ \$FileName\$ \$Extension\$ \$CreatedDate\$ \$CreatedDate\$ \$CreatedDate\$ \$CreatedDate\$ \$CreatedDate\$ \$Protocol\$_ \$Protocol\$_ \$Protocol\$_ \$Protocol\$_ \$Protocol\$_ \$Protocol\$_ \$NRPCPort\$_ \$DIIOPPort\$_ \$DIIOPPort\$_ \$DocumentUNID\$ \$View\$ \$ArchiveCharset\$ \$ArchiveCharset\$ \$ArchiveCharset\$ \$EntryName\$ \$EntryDate\$ \$EntryDate\$ \$EntrySize\$ \$Date\$ \$NoteID\$ \$SearchType\$ \$IsUseSSL\$
Lista de valores geradores	
SharePoint	crawled_date listname listtitle sitetitle

Tabela 18. Nomes de campos reservados por pesquisadores (continuação)

Tipo de pesquisador	Nomes de campos reservados
Sistema de ficheiros UNIX	\$Directory\$ \$FileName\$ \$Extension\$ \$ModifiedDate\$ \$FileSize\$
	\$Title\$ \$ArchiveType\$ \$ArchiveCharset\$ \$EntryName\$ \$EntryDate\$
	_\$Entry5ize\$ _\$Date\$
Web Content Management	\$PageID\$ _\$PortletID\$ _\$PortletURL\$ _\$Title\$ _\$Description\$ _\$FileSize\$ \$Date\$
	\$PageTitle\$ \$PortletTitle\$ \$ServerVersion\$ \$Author\$ \$Owner\$
	\$EffectiveDate\$ \$LastModifiedDate\$ \$EffectiveDateStr\$ \$LastModifiedDateStr\$ \$Modifier\$ \$AuthoringTemplate\$ %ContemtD%
WebSphere Portal	\$ContentID\$ \$PortletID\$ \$PortletURL\$ \$Title\$ \$Description\$ \$FileSize\$ \$Date\$ \$PageTitle\$ \$PortletTitle\$ \$ServerVersion\$
Sistema de ficheiros Windows	\$Directory\$ \$FileName\$ \$Extension\$ \$ModifiedDate\$ \$FileSize\$ \$Title\$ \$ArchiveType\$ \$ArchiveCharset\$ \$EntryName\$ \$EntryDate\$ \$EntryDize\$ \$EntrySize\$ \$Date\$

Tabela 18. Nomes de campos reservados por pesquisadores (continuação)
Importar e exportar campos de índice

Ao configurar campos de índice para uma colecção, pode importar definições de campo de índice que tenha definido anteriormente num ficheiro XML. Pode também exportar campos de índice da consola administrativa para um ficheiro XML.

Se o ficheiro XML que pretende importar estiver no sistema local, o tamanho máximo do ficheiro é de 8 MB. Se o ficheiro XML se encontrar no servidor do controlador, o ficheiro poderá ser maior do que 8 MB.

O ficheiro XML que importar terá de ter os seguintes elementos, em que cada elemento de <campo> especifica o nome do campo de índice e atributos para a utilização desse campo de procura:

```
<?xml version="1.0" encoding="UTF-8"?>
<fields>
<field name="fieldname">
        <field name="fieldname">
        <fieldsearchable enabled="true"/>
        <fieldsearchable enabled="false" tokenized="true" casesensitive="false"/>
        <parametricsearchable enabled="false"/>
        <returnable enabled="true"/>
        <analyzable enabled="true"/>
        </field>
</fields>
```

Para importar e exportar definições de campo de índice:

- Na vista Colecções, expanda a colecção que pretende editar. No painel Analisar e Indexar, faça clique em Configurar > Campos de índice.
- 2. Importar definições de campo de índice a partir de um ficheiro XML:
 - a. Na página Definições de Campo de Índice, faça clique em Importar Campos de Índice.
 - b. Na página Importar Definições de Campos de Índice, especifique a localização do ficheiro que contém as suas definições de campos de índice. Se o ficheiro XML se encontrar no sistema local, pode navegar até encontrar o caminho. Se o ficheiro XML se encontrar no servidor do controlador, tem de escrever o caminho totalmente qualificado.
 - c. Faça clique em **Seguinte** para carregar o ficheiro XML. São apresentadas as definições dos campos de índice.
 - d. Faça clique na caixa de verificação **Importar** para cada campo de índice que pretende adicionar à colecção e faça clique em **Terminar**. Os campos de índice importados são listados juntamente com os outros campos de índice que foram definidos para esta colecção.

Se pretender alterar uma definição de campo de índice depois de a importar, seleccione o campo de índice na página Definições de Campo de Índice e faça clique no ícone **Editar**.

- 3. Exporte as definições dos campos de índice para um ficheiro XML:
 - a. Na página Definições de Campos de Índice, faça clique na caixa de verificação Exportar para cada campo de índice que pretende exportar.
 - b. Faça clique em Exportar de Campos de Índice.
 - **c.** Quando lhe for pedido, especifique a localização onde pretende guardar o ficheiro XML.

"Atributos do campo de índice" na página 195

Referências relacionadas:

"Nomes de campos de índice reservados" na página 199

Remover informações sobre as origens de entrada e facetas do campo de índice

Caso pretenda prevenir que as informações sobre origens de entrada e facetas correlacionadas com campos de índice sejam mostradas na página Definições do Campo de Índice, pode editar um ficheiro de configuração. Não existe suporte para esta tarefa na consola de administração.

A página Definições de Campo de Índice mostra todos os campos de índice definidos para uma colecção. A informação mostrada para cada campo de índice inclui as origens de entrada e facetas que estejam a utilizar o campo de índice, bem como as definições de atributo para o modo como o campo de índice pode ser consultado e apresentado nos resultados da consulta.

Obter informações sobre as origens de entrada e facetas associadas demora mais tempo. Caso não queira passar por este atraso, pode remover estas colunas de dados da página. Se remover esta informação, pode ver quais os pesquisadores e facetas que estão a utilizar um campo de índice editando a definição do campo de índice.

Para remover as colunas **Origens de entrada** e **Facetas** da tabela de campos de índice:

- Inicie sessão como administrador predefinido do IBM Content Analytics with Enterprise Search no servidor do controlador e edite o ficheiroES_INSTALL_ROOT/ jetty/adminapp/ESAdmin/WEB-INF/config.properties.
- Defina os valores das seguintes propriedades para true: disable.fieldDefinitions.showInputSources=true disable.fieldDefinitions.showFacets=true

Para mostrar estas colunas novamente, defina as propriedades para false.

3. Para que estas alterações se tornem efectivas, reinicie a sessão de administração: esadmin admin restart

Tarefas relacionadas:

"Criar campos de índice" na página 194

Correlacionamento de campo de índice para elementos XML

Ao correlacionar elementos XML com campos de índice, permite que os utilizadores consultem partes específicas de documentos XML.

Os documentos XML estão a tornar-se mais comuns, uma vez que contêm texto semi-estruturado e não estruturado. A estrutura de XML é encapsulada e utiliza um contexto explicitamente definido por elementos XML em redor do texto. Por exemplo, o nome de um autor poderá ser apresentado da seguinte forma:

<autor>João Silva</autor>

Neste contexto, o texto João Silva identifica o autor de um documento XML.

Normalmente, todos os dados de texto de um documento de XML são indexados. Ao correlacionar elementos de XML com campos de índice, pode utilizar as informações de estrutura de documentos XML para suportar consultas mais específicas.Por exemplo, pode tornar todos os dados de um elemento XML pesquisáveis por nome de campo e permitir que sejam devolvidos nos resultados de procura, correlacionando o nome do elemento XML com um campo de índice que tenha esses atributos. Também pode tornar o correlacionamento mais preciso correlacionando atributos do elemento XML (e os valores desses atributos) com campos de índice.

Ao configurar opções de análise para uma colecção, especifique quais os elementos XML que serão correlacionados com campos de índice. Ao correlacionar elementos XML com campos de índice, permite que os utilizadores procurem os valores desses elementos especificando os nomes dos campos em consulta. As consultas que procuram campos designados podem fornecer resultados da procura mais precisos do que consultas de texto livre que pesquisam todo o conteúdo de documento.

Por exemplo, se a sua colecção incluir documentos XML e se o utilizador especificar que os elementos <title> e <author> serão marcados como campos de índice no índice, os utilizadores podem consultar directamente estes elementos. Uma procura de autor:Silva pesquisa os documentos de XML que tenham Silva nos elementos que estão correlacionados com um campo designado autor.

Noutro exemplo, um elemento de XML designado <resumo> pode conter informações que são úteis para mostrar nos resultados da procura. Se correlacionar o elemento <summary> com um campo de índice e especificar que o valor deste elemento deve ser mostrado nos resultados de procura, o conteúdo do elemento faz parte do documento dos resultados.

Quando correlaciona um elemento de XML com um nome de campo, o texto dentro do elemento é pesquisável sob o nome de campo que especificar. No entanto, se o elemento de XML incluir atributos, os valores de atributo não são indexados, nem directamente pesquisáveis. Para consultar os valores de atributos, tem de configurar opções de análise na consola de administração e activar procura de XML nativa. Após activar procura de XML nativa, a estrutura do documento XML fica disponível no índice e pode consultar o documento especificando restrições de consulta XPath. Por exemplo:

@xmlxp::'//elementName[@attributeName="attributeValue"]'

Conceitos relacionados:

"Campos de índice" na página 191

- France da consulta de procura semântica
- 🕩 Marcação XML na análise e procura

Tarefas relacionadas:

"Correlacionar elementos XML com campos de índice"

- "Configurar pesquisadores para pesquisar ficheiros XML" na página 209
- "Activar suporte para procura de XML nativa" na página 179

Correlacionar elementos XML com campos de índice

Quando correlacionar um elemento XML com um campo de índice, especifique quais os utilizadores de elementos XML que pode procurar especificando um nome de campo numa consulta. É necessário definir campos de índice antes de poder correlacionar com eles elementos XML.

Podem existir várias correlações de campos de XML por colecção, mas apenas uma correlação de elementos raiz de XML. O elemento raiz de um documento de XML aplica as correlações de forma conveniente.

Quando criar uma correlação de campos XML, ou adicionar, alterar ou eliminar campos numa correlação de campos XML existente, as correlações novas e alteradas aplicam-se somente a novos dados, e não afectam dados já existentes no índice. Para actualizar documentos que já se encontram no índice, tem de pesquisar e indexar de novo os documentos. Se a colecção estiver configurada para utilizar a cache de documento, pode voltar a indexar os documentos sem os voltar a pesquisar.

Para correlacionar elementos XML com campos de índice:

- Na vista Colecções, expanda a colecção que pretende configurar. No painel Analisar e indexar, faça clique em Configurar > Correlacionamento de campo de índice para elementos XML.
- Na página Correlacionamentos de Campos de XML, faça clique em Criar Correlações de XML. A página Criar uma Correlação de Campos de XML é aberta.
- No campo Nome do elemento raiz de XML, escreva o nome do elemento raiz: pessoal.

Certifique-se de que o nome que especifica aqui corresponde exactamente ao elemento raiz nos documentos XML que pretende procurar. Quando analisa e indexar os documentos XML, o sistema selecciona qual a correlação que será utilizada de acordo com o nome do elemento raiz.

4. No campo **Nome da correlação de XML**, escreva um nome para este conjunto de regras de correlação de campos de XML.

Depois de criar um conjunto de regras de correlação de XML, este nome será apresentado na página Correlações de Campos de XML, e seleccione este nome para adicionar, eliminar ou alterar as regras de correlação.

- 5. Especifique quais os elementos XML, atributos e valores de atributo que pretende correlacionar com campos de índice:
 - a. Faça clique em **Adicionar campo** para adicionar uma linha em branco à lista de regras de correlação de campos.
 - b. Escreva o nome do elemento que pretende correlacionar com um campo de índice.
 - c. Opcional: Escreva o nome de um atributo deste elemento ou do valor de atributo que pretende correlacionar com um campo de índice.
 - d. Seleccione o campo de índice com o qual pretende correlacionar este elemento. Os atributos definidos para o campo de índice seleccionado são aplicados automaticamente e não os poderá alterar.

"Correlacionamento de campo de índice para elementos XML" na página 206

"Campos de índice" na página 191

Tarefas relacionadas:

"Configurar facetas hierárquicas e simples para uma colecção de procura empresarial" na página 182

Configurar pesquisadores para pesquisar ficheiros XML

Para procurar campos XML, é necessário especificar as opções para o modo como os documentos são pesquisados. Por exemplo, para procurar documentos XML que estão armazenados numa base de dados do IBM DB2, é necessário configurar a base de dados de origem e configurar as opções para pesquisar e procurar as tabelas de base de dados.

Este procedimento mostra como configurar um pesquisador do DB2 para pesquisar documentos XML. É possível adaptar este procedimento para configurar outros tipos de pesquisadores.

Para permitir aos utilizadores procurar colecções que contêm documentos XML:

- 1. Configure a tabela da base de dados de origem. Para o IBM DB2, o tipo de dados da coluna da base de dados que contém a cadeia XML que pretende poder pesquisar pode ser CLOB, DBCLOB ou XML.
- 2. Configure o pesquisador do DB2.
 - **a**. Na página Seleccionar tabelas DB2 Individuais para configurar, faça clique no ícone **Editar opções** para a tabela que contém o conteúdo XML.
 - b. Na área Opções adicionais para conteúdo do documento, seleccione a coluna XML no campo Coluna com conteúdo grande binário ou de texto e insira text/xml no campo Tipo MIME.
- **3**. Guarde a configuração do pesquisador e, em seguida, inicie o pesquisador para pesquisar a tabela da base de dados.
- 4. Para permitir aos utilizadores procurar o conteúdo XML, deve também seguir os procedimentos para correlacionar elementos XML a campos de índice.

Conceitos relacionados:

"Correlacionamento de campo de índice para elementos XML" na página 206

Correlacionamentos de campo de índice para metadados de HTML

Ao correlacionar elementos de metadados de HTML com campos de índice, é possível permitir aos utilizadores consultar atributos dos elementos de HTML.

Os elementos de metadados em documentos HTML são semelhantes aos atributos do documento nos quais fornecem informação acerca do documento, como está formatado e como pode ser acedido na Web. Por exemplo:

```
<meta http-equiv="Content-Type" content="text/html; charset=utf-8" />
<meta name="copyright" content="(C) Copyright IBM Corporation 2005" />
<meta name="content.owner" content="(C) Copyright IBM Corporation 2005" />
<meta name="security" content="public" />
<meta name="abstract" content="Este tópico descreve um produto IBM." />
<meta name="format" content="XHTML" />
```

É possível associar, ou correlacionar, elementos de metadados de HTML com campos de índice. Quando configura opções de análise para uma colecção, especifica os elementos de metadados de HTML que devem ser correlacionados com os respectivos campos de índice. Ao correlacionar elementos de metadados de HTML com campos de índice, permite aos utilizadores encontrar documentos com aqueles elementos, especificando os nomes do campo de índice em consultas. As consultas que efectuam procuras em campos específicos podem fornecer resultados da procura mais exactos do que as consultas de texto livre que efectuam procuras em todo o conteúdo dos documentos.

Por exemplo, se a respectiva colecção incluir documentos HTML, e especificar que os direitos de autor e os elementos dos metadados abstractos devem ser indexados como campos de índice, os utilizadores poderão consultar esses elementos específicos. Uma procura por direitos de autor:IBM encontra os documentos HTML que têm IBM nos metadados direitos de autor.

Quando correlaciona elementos de metadados de HTML com campos de índice, especifica se pretende correlacionar todos os elementos de metadados de HTML, apenas os elementos que pertencem ao conjunto Dublin Core de elementos de metadados, ou apenas os elementos de metadados de HTML que especifica. Para obter uma descrição dos elementos no conjunto de elementos de metadados Dublin Core, consulte o sítio da web Dublin Core Metadata Initiative.

Conceitos relacionados:

"Campos de índice" na página 191

Tarefas relacionadas:

"Correlação de elementos metadados HTML para os campos do índice"

"Indexação de texto alternativo de imagem em documentos HTML" na página 211

Correlação de elementos metadados HTML para os campos do índice

Ao correlacionar um elemento de metadados HTML para o campo de um índice, está a especificar quais são os utilizadores de elementos metadados HTML que podem efectuar procuras especificando o nome de um campo numa consulta.

É necessário definir os campos do índice antes que possa correlacionar os elementos metadados HTML com eles.

Quando criar uma correlação de campos HTML, ou adicionar, alterar ou eliminar campos numa correlação de campos HTML existente, as correlações novas e alteradas aplicam-se somente a novos dados, e não afectam dados já existentes no índice. Para actualizar documentos que já se encontram no índice, tem de pesquisar ou importar novamente os documentos e depois reconstruir o índice. Se a colecção estiver configurada para utilizar a cache de documentos, poderá voltar a indexar os documentos sem ter de os voltar a pesquisar ou importar.

Depois de correlacionar elementos para os campos do índice, será apresentada uma ligação para o elemento na página Definições do campo do índice. Ao clicar na ligação, poderá editar o modo como o elemento está correlacionado com o campo do índice.

Restrição: Se seleccionar **Todos os elementos metadados HTML** ou **Todos os elementos metadados Dublin Core**, as ligações de origem de entrada não serão exibidas. As ligações de origem de entrada são fornecidas apenas para elementos metadados que tenha explicitamente correlacionado com um campo do índice.

Para correlacionar elementos metadados HTML para os campos do índice:

- Na vista Colecções, expanda a colecção que pretende configurar. No painel Analisar e indexar, faça clique em Configurar > Correlacionamento de campo de índice para elementos de metadados HTML.
- 2. Na página Correlacionamentos para campos HTML, especifique quais os elementos metadados HTML que quer correlacionar com os campos do índice:
 - a. Faça clique em **Adicionar campo** para adicionar uma linha em branco à lista de regras de correlação de campos.
 - b. Escreva o nome do elemento de metadados que pretende correlacionar.
 - c. Seleccione o campo de índice com o qual pretende correlacionar este elemento. Os atributos definidos para o campo de índice seleccionado são aplicados automaticamente e não os poderá alterar.

Os utilizadores podem consultar os nomes dos campos correlacionados para localizar documentos HTML com metadados específicos. Por exemplo, se correlacionou um elemento de metadados HTML designado descrição com um campo de índice designado abstracto, os utilizadores podem introduzir uma consulta semelhante à seguinte para localizar documentos HTML sobre computadores Thinkpad:

abstracto:thinkpad

Conceitos relacionados:

"Correlacionamentos de campo de índice para metadados de HTML" na página 209

"Campos de índice" na página 191

Tarefas relacionadas:

"Configurar facetas hierárquicas e simples para uma colecção de procura empresarial" na página 182

Indexação de texto alternativo de imagem em documentos HTML

As regras predefinidas de análise não indexam o conteúdo de atributos alternativos em elementos HTML, como, por exemplo, . Para assegurar a indexação do texto alternativo, tem de especificar regras num ficheiro de configuração do analisador.

Para configurar o analisador para adicionar conteúdo de texto alternativo ao índice:

- Inicie sessão como o administrador predefinido de IBM Content Analytics with Enterprise Search no servidor do controlador e edite o ficheiro ES_NODE_ROOT/master_config/ID_Colecção.indexservice/parser_config.xml, onde ID_Colecção identifica a colecção que pretende personalizar.
- 2. Adicione a seguinte regra ao ficheiro e guarde as alterações:

```
<Rule>
<Conditions>
<Element Name="img" />
</Conditions>
<Actions>
<AttributeAction FieldName="alttext" FieldNameType="fixed" ValueFrom="alt" />
</Actions>
</Rule>
```

 Na consola administrativa, crie um campo de índice denominado texto alternativo. Active o atributo pesquisável de texto completo para o campo de índice.

- 4. Reinicie os serviços de análise e indexação para a colecção.
- 5. Volte a pesquisar ou volte a importar documentos para que possam ser novamente analisados e indexados.

"Correlacionamentos de campo de índice para metadados de HTML" na página 209

Tarefas relacionadas:

"Criar campos de índice" na página 194

Correlacionamentos com campo de índice para metadados extraídos de conteúdo binário

Para controlar o modo como os utilizadores consultam os documentos binários que adiciona a uma colecção, pode correlacionar metadados extraídos pelo extractor de texto com campos de índice. Ao correlacionar os metadados com campos de índice, permite aos utilizadores consultar partes específicas de documentos binários.

O IBM Content Analytics with Enterprise Search pode extrair texto e metadados de vários tipos de formatos binários, como, por exemplo, documentos Microsoft Word ou Adobe portable document format (PDF). O extractor de texto utiliza o Oracle Outside In Content Access para proceder à extracção. Alguns dos metadados extraídos são automaticamente correlacionados com os campos de índice. Por exemplo, um título extraído de um documento Microsoft PowerPoint é automaticamente correlacionado com um título mencionado de campo de índice para que os utilizadores possam ver o título extraído nos resultados da procura.

Para controlar o modo como os utilizadores consultam e vêem os metadados extraídos, pode especificar as suas preferências para correlacionamento de metadados extraídos com campos de índice. Por exemplo, pode preferir ver o nome de um ficheiro nos resultados em vez do texto do título extraído. É necessário definir os campos de índice antes de poder correlacionar os campos de metadados extraídos com os mesmos.

Nomes de campos de metadados predefinidos

Quando selecciona os metadados que pretende correlacionar, pode manter o nome que o extractor de texto atribuiu ao campo de metadados, seleccionar um nome de uma lista de nomes de campos de metadados candidatos predefinidos ou introduzir um nome de campo diferente.

O tipo de metadados extraídos de documentos binários dependem do Oracle Outside In Content Access. Se introduzir um nome de campo de metadados em vez de seleccionar um a partir da lista de nomes predefinidos, tem de saber o nome do Oracle Outside In Content Access para os metadados. Os nomes dos metadados não são sensíveis a maiúsculas.

Tabela 19. Nomes de campos de metadados predefinidos

abstracto	countmmclips	headingpairs	recordedby
conta	countnotes	palavra-chave	recordeddate
endereço	countpages	idioma	referência
anexos	countparas	lastprintdate	revisiondate
autorização	countslides	lastsavedate	revisionnotes
backupdate	countslideshidden	lastsavedby	revisionnumber
basefilelocation	countwords	linksdirty	scalecrop
billto	creationdate	mailstop	secondaryauthor
blindcopy	departamento	gestor	secção
carboncopy	destino	assunto	segurançaorigem
categoria	disposição	office	estado
checkedby	divisão	operadorproprietário	assunto
cliente	doccomment	presentationformat	título
empresa	docnumber	primaryauthor	titleofparts
completeddate	doctype	projecto	typist
countbytes	editminutes	editor	versiondate
countchars	editor	propósito	versionnotes
countcharswithspaces	forwardto	receivedfrom	versionnumber
countlines	grupo		

"Campos de índice" na página 191

Tarefas relacionadas:

"Correlacionar metadados extraídos para campos de índice"

"Associar tipos de documentos ao extractor de texto" na página 269

Correlacionar metadados extraídos para campos de índice

Após o extractor de texto ter extraído metadados dos documentos binários, pode correlacionar os metadados para campos de índice. Quando correlaciona metadados extraídos para um campo de índice, permite aos utilizadores consultar partes específicas de documentos binários.

É necessário definir os campos de índice antes de poder correlacionar os campos de metadados extraídos com os mesmos.

Quando adiciona, altera ou elimina correlacionamentos de campos de índice, os correlacionamentos novos e alterados são aplicados aos novos dados e não têm qualquer efeito nos dados que já estão no índice. Para actualizar documentos que já estão no índice, tem de voltar a pesquisar ou a importar os documentos e, em seguida, tem de reconstruir o índice. Se a colecção estiver configurada para utilizar a memória cache de documentos, poderá voltar a indexar os documentos sem ter de os voltar a pesquisar ou importar.

Para correlacionar metadados extraídos de documentos binários para campos de índice:

- Na vista Colecções, expanda a colecção que pretende configurar. No painel Analisar e indexar, faça clique em Configurar > Correlacionamento de campo de índice para metadados extraídos. O Correlacionar metadados extraídos para campos de índice apresenta os metadados que foram extraídos pelo extractor de texto e os correlacionamentos de campo de índice predefinidos.
- Opcional: No campo Nome do campo de metadados, seleccione ou insira um nome diferente para os metadados:

- Seleccione um nome de campo diferente a partir da lista de campos de metadados predefinidos. Se seleccionar um campo, a lista de campos disponíveis incluí apenas campos para os quais os correlacionamentos ainda não estão definidos.
- Introduza um novo nome de campo. Se introduzir um novo nome de campo, são apresentados termos candidatos que comecem pelo texto que introduziu. Os campos de metadados não são sensíveis a maiúsculas/minúsculas e têm de ser únicos.
- 3. Opcional: Seleccione a caixa de verificação Correlacionar para conteúdo do documento se pretende correlacionar os metadados para conteúdo do documento em vez de um campo de índice. Quando seleccionado, o valor dos metadados é processado como se estivesse no corpo do documento em vez de ser activado para procura por campo.
- 4. No campo **Nome do campo**, seleccione o campo de índice para o qual pretende correlacionar os metadados. Os atributos definidos para o campo de índice seleccionado são aplicados automaticamente. Para alterar os atributos, tem de editar a definição de campo de índice.
- 5. Opcional: Faça clique em Adicionar Correlacionamento para adicionar uma linha em branco no topo dos campos da lista. Seleccione ou introduza o nome do campo de metadados e, em seguida, correlacione os metadados para o conteúdo do documento ou um campo de índice.
- 6. Se a colecção utiliza uma cache de documento, reconstrua o índice a partir da cache. Se a colecção não utilizar uma cache, volte a pesquisar ou a importar os documentos para a colecção e, em seguida, reconstrua o índice.

"Correlacionamentos com campo de índice para metadados extraídos de conteúdo binário" na página 212

"Campos de índice" na página 191

Processamento de texto personalizado

Pode melhorar a qualidade e precisão dos resultados da procura integrando algoritmos de processamento de texto personalizados em colecções.

O IBM Content Analytics with Enterprise Search suporta a Apache Unstructured Information Management Architecture (UIMA), a qual consiste numa estrutura que permite criar, identificar, compor e implementar funções de análise de texto. Os programadores de aplicações criam e testam algoritmos de análise para que o conteúdo seja procurado e, em seguida, criam um arquivo de motor de processamento (ficheiro .pear) que inclui todos os recursos necessários para utilizar o arquivo. Para poder consultar colecções utilizando algoritmos de análise personalizados, é necessário adicionar o arquivo (que contém o motor de análise de texto) ao sistema.

O componente lógico de análise num motor de análise de texto denomina-se *anotador*. Cada anotador efectua tarefas de análise linguística específicas. Um motor de processamento de texto pode conter qualquer número de anotadores ou pode ser composto por vários motores de análise de texto, cada um contendo anotadores próprios personalizados.

As informações produzidas pelos anotadores são referidas como *resultados da análise*. Os resultados da análise, que correspondem às informações que pretende procurar, são escritos numa estrutura de dados denominada *estrutura de análise comum*.

Ao configurar opções de processamento de texto para uma colecção, efectue as seguintes tarefas:

- Seleccione o motor de análise de texto que pretende utilizar para anotar documentos na colecção.
- Se a colecção contiver documentos XML com marcações significativas e pretender utilizar esta marcação na análise de texto personalizada, pode associar ficheiros de correlação à colecção e correlacionar a saída dos ficheiros de correlacionamento XML com a estrutura de análise comum.

Por exemplo, pode correlacionar o conteúdo dos elementos <destinatário> e <cliente> com anotações de Pessoa na estrutura de análise comum. Estas anotações podem, em seguida, ser acedidas por anotadores personalizados, que poderão detectar informações adicionais (por exemplo, poderão detectar o género). Pode também correlacionar anotações de Pessoa com o índice, permitindo aos utilizadores procurar Pessoas sem ser necessário conhecer os nomes originais dos elementos XML.

Se pretender autorizar os utilizadores a especificar elementos de XML originais nas consultas, não é necessário definir quaisquer correlações de XML. Como alternativa, pode configurar opções de análise e activar a procura de XML nativa para a colecção.

• Correlacione a estrutura de análise comum com o índice, o que permite que os documentos anotados sejam pesquisados com a procura semântica.

Por exemplo, dependendo das entidades e relações detectadas pelos anotadores, os utilizadores podem procurar conceitos que ocorram na mesma frase (por exemplo, uma pessoa específica e qualquer nome de competidor) ou uma palavra-chave e um conceito (por exemplo, o nome Alexandre e um número de telefone).

• Correlacione a estrutura de análise comum com uma base de dados relacional. Pode correlacionar dados com tabelas da IBM DB2 ou da Oracle. Este tipo de correlação permite que os resultados da análise sejam utilizados em aplicações de bases de dados como, por exemplo, prospecção de dados. Também permite que sejam utilizadas consultas SQL para procurar os dados fora do IBM Content Analytics with Enterprise Search.

- 🕒 Integração de análise de texto personalizada
- Conceitos básicos utilizados no processamento da análise de texto
- Fluxo de trabalho para integração de análise personalizada
- Algoritmos de análise de texto
- 🖙 Aplicações de procura semântica
- 🖙 Termo da consulta de procura semântica

Tarefas relacionadas:

"Adicionar motores de análise de texto ao sistema"

"Associar um motor de análise de texto a uma colecção" na página 217

"Correlacionar elementos de XML com a estrutura de análise comum" na página 218

"Correlacionar a estrutura de análise comum com o índice" na página 219

"Correlacionar a estrutura de análise comum com uma base de dados relacional" na página 220

Adicionar motores de análise de texto ao sistema

Se criar um motor de análise de texto personalizado, terá de adicioná-lo ao sistema antes de o conseguir utilizar para procura ou análise de texto. As colecções podem utilizar o motor para analisarem e anotarem documentos e melhorarem a precisão do resultados da procura.

Os programadores de aplicações conseguem criar um arquivo do motor de processamento (.pear) que adere ao contexto de UIMA para a análise de texto. O arquivo inclui todos os recursos necessários para consultar colecções. Para poder consultar colecções utilizando algoritmos de análise personalizados, é necessário adicionar o arquivo (que contém o motor de análise de texto) ao sistema.

Depois de adicionar um motor de análise de texto ao sistema, pode alterar o seu nome de apresentação e seleccionar uma opção para visualizar a origem XML. A origem de XML mostra-lhe a informação que é produzida por este motor.

Se um motor de análise de texto for associado a uma colecção, não conseguirá remover o motor de análise de texto do sistema.

Para adicionar um motor de análise de texto personalizado ao sistema:

- 1. Faça clique em Sistema para abrir a vista Sistema.
- 2. Na página Análise, faça clique em Configurar motores de análise de texto.
- 3. Na página Motores de Análise de Texto, faça clique em Adicionar Motor de Análise de Texto.
- 4. Na página Adicionar um Motor de Análise de Texto, escreva um nome descritivo para o novo motor. O sistema utiliza este nome de apresentação para identificar o motor de análise de texto ao longo da consola de administração.
- 5. Especifique se pretende utilizar o descritor no seu arquivo de motores de processamento personalizados, de modo a isolar os respectivos componentes dos anotadores linguísticos que são facultados com o IBM Content Analytics with Enterprise Search. Se tiver exportado o ficheiro .pear do Content Analytics Studio, poderá evitar conflitos permitindo que o sistema configure um gestor de caminho de classe localizado e de recursos internos para o motor

de análise de texto personalizado. Para activar esta opção numa configuração de vários servidores, terá de assegurar que todos os servidores utilizam o mesmo caminho de directórios de dados (ES_NODE_ROOT).

- 6. Especifique a localização do ficheiro .pear. Se o ficheiro for menor do que 64 MB, o ficheiro pode estar no computador local e o utilizador pode procurar para localizar o ficheiro. Se o ficheiro tiver mais de 64 MB, tem de estar no servidor do controlador e é necessário escrever o caminho totalmente qualificado para o ficheiro.
- 7. Faça clique em **OK**. O motor de análise de texto é apresentado na página Motores de Análise de Texto.

Conceitos relacionados:

"Processamento de texto personalizado" na página 214

- Fluxo de trabalho para integração de análise personalizada
- 🕩 Integração de análise de texto personalizada
- Conceitos básicos utilizados no processamento da análise de texto

🕒 Marcação XML na análise e procura

Tarefas relacionadas:

"Associar um motor de análise de texto a uma colecção"

🖙 Criar elementos XML para o ficheiro de mapeamento de estrutura comum

Associar um motor de análise de texto a uma colecção

Se os motores de análise de texto personalizada estiverem associados ao sistema, pode seleccionar um para utilizá-lo com uma colecção. Os utilizadores podem especificar consultas semânticas quando procuram a colecção e melhoram a qualidade e a precisão dos resultados da procura.

Se um motor de análise de texto já estiver associado a esta colecção, poderão ocorrer as acções seguintes quando associa um motor diferente:

- Se seleccionar **Nenhuma análise personalizada**, todas as correlações de análise de texto definidas anteriormente para a colecção são repostas. A colecção começa a utilizar os valores predefinidos do sistema.
- Se seleccionar o nome de um motor de análise de texto personalizada diferente, todas as correlações de análise de texto definidas anteriormente para a colecção são retidas. Por exemplo, se mudar do motor_1 para o motor_2, o motor_2 herda os ficheiros de correlação XML que configurou para o motor_1.

Para utilizar um motor de análise de texto com uma colecção:

- Na vista Colecções, expanda a colecção que pretende configurar. No painel Analisar e Indexar, faça clique em Configurar > Opções de processamento de texto.
- Faça clique em Seleccionar um motor de análise de texto. Se não tiverem sido adicionados ao sistema nenhuns motores de análise de texto, ou se a colecção utilizar os algoritmos de análise predefinidos, o nome do motor será Predefinição.
- 3. Na página Seleccionar um Motor de Análise de Texto para esta Colecção, seleccione o nome do motor que pretende utilizar com esta colecção. Se não existirem motores de análise de texto disponíveis, ou se seleccionar Nenhuma análise personalizada, o analisador aplica regras de análise de texto predefinidas ao anotar e preparar documentos para o índice.

- 4. Se o arquivo de processamento personalizado tiver sido exportado a partir do Content Analytics Studio, especifique a vista CAS (Common analysis structure estrutura de análise comum) que pretende utilizar para análise de texto personalizada. Ao utilizar uma viste CAS separada, poderá evitar potenciais conflitos entre os componentes linguísticos Content Analytics Studio no seu anotador personalizado e os componentes linguísticos que estão incorporados nos anotadores fornecidos.
- 5. Para motores de processamento personalizados tais como anotadores de multimédia que precisem de analisar o documento original pesquisado, especifique o nome da vista CAS a utilizar para o conteúdo do documento original. Certifique-se de que este nome de vista corresponde ao nome de vista especificado no anotador de multimédia ou no ficheiro XML descritor de UIMA.

"Processamento de texto personalizado" na página 214

- Fluxo de trabalho para integração de análise personalizada
- Integração de análise de texto personalizada
- Conceitos básicos utilizados no processamento da análise de texto

Tarefas relacionadas:

"Adicionar motores de análise de texto ao sistema" na página 216

Criar anotadores de multimédia

Correlacionar elementos de XML com a estrutura de análise comum

Se a colecção contiver documentos XML com marcações significativas, e pretender utilizar esta marcação para permitir que os utilizadores consultem o índice ou tabelas de bases de dados relacionais através da procura semântica, pode correlacionar os elementos XML com a estrutura de análise comum.

O tamanho máximo de um ficheiro de correlação é de 8 MB.

Para permitir que os processos de análise de texto personalizada acedam a elementos específicos em documentos de XML, ou para correlacionar vários elementos de XML com um Tipo comum para utilização na procura semântica, pode criar ficheiros de correlação personalizados. Os ficheiros de correlação têm de aderir à estrutura UIMA para análise de texto.

Quando adiciona ficheiros de correlação a uma colecção que utilize um motor de análise de texto personalizada, irá permitir que os elementos XML dos documentos de origem sejam correlacionados com anotações da estrutura de análise comum. Estas anotações podem depois ser utilizadas pelos seus motores de análise de texto personalizada. Pode correlacionar a estrutura de análise comum com o índice e permitir aos utilizadores consultar as anotações quando pesquisarem a colecção com a procura semântica.

Por exemplo, pode correlacionar o conteúdo do destinatário e os elementos do cliente com as anotações da Pessoa na estrutura de análise comum. Estas anotações podem, em seguida, ser acedidas por anotadores personalizados, que poderão detectar informações adicionais (por exemplo, poderão detectar o género). Pode também correlacionar anotações de Pessoa com o índice, permitindo aos utilizadores procurar Pessoas sem ser necessário conhecer os nomes originais dos elementos XML.

Se pretende permitir que os utilizadores especifiquem os elementos de XML originais em consultas, não necessita de configurar ficheiros de correlação. Como alternativa, pode configurar opções de análise e activar a procura de XML nativa para a colecção.

Correlacionar elementos de XML com a estrutura de análise comum:

- Na vista Colecções, expanda a colecção que pretende configurar. No painel Analisar e Indexar, faça clique em Configurar > Opções de processamento de texto.
- 2. Na área Correlacionar elementos de XML com a estrutura de análise comum, faça clique em Adicionar Correlação.
- **3**. Na página Correlacionar Elementos de XML com a Estrutura de Análise Comum, escreva um nome de apresentação descritivo para o ficheiro de correlação.
- 4. Especifique a localização do ficheiro. Se o ficheiro de correlação se encontrar no sistema local, pode navegar até o localizar. Se o ficheiro de correlacionamento se encontrar no servidor do controlador, tem de escrever o caminho totalmente qualificado.
- 5. Faça clique em **OK**. O novo ficheiro de correlação é adicionado à página Opções de Processamento de Texto.

Conceitos relacionados:

"Processamento de texto personalizado" na página 214

- Fluxo de trabalho para integração de análise personalizada
- 🕩 Integração de análise de texto personalizada
- Conceitos básicos utilizados no processamento da análise de texto
- Marcação XML na análise e procura

Tarefas relacionadas:

Criar elementos XML para o ficheiro de mapeamento de estrutura comum

Correlacionar a estrutura de análise comum com o índice

Pode especificar quais são os resultados da análise de texto de uma estrutura de análise comum que devem ser correlacionados com o índice, e que devem ficar disponíveis para os utilizadores que consultam uma colecção, utilizando procura semântica.

O tamanho máximo de um ficheiro de correlação é de 8 MB.

Ao correlacionar a estrutura de análise comum com o índice, permitirá que os utilizadores especifiquem consultas semanticamente exactas e melhorem a qualidade dos resultados da procura.

Por exemplo, dependendo das entidades e relações detectadas pelos anotadores, os utilizadores podem procurar conceitos que ocorram na mesma frase (como uma pessoa específica e qualquer nome de competidor), ou uma palavra-chave e um conceito (como o nome Alexandre e um número de telefone).

Para correlacionar a estrutura de análise comum com o índice:

 Na vista Colecções, expanda a colecção que pretende configurar. No painel Analisar e Indexar, faça clique em Configurar > Opções de processamento de texto.

- 2. Na área **Correlacionar a estrutura de análise comum com o índice**, faça clique em **Seleccionar um ficheiro de correlação**.
- **3**. Na página Seleccionar um Ficheiro de Correlação para esta Colecção, seleccione o ficheiro de correlação que pretende utilizar com a colecção:
 - Para utilizar as regras de correlação predefinidas, seleccione Predefinição.
 - Para correlacionar uma estrutura de análise comum personalizada com o índice, especifique a localização do ficheiro de correlação. Se o ficheiro se encontrar no sistema local, pode navegar até o localizar. Se o ficheiro se encontrar no servidor de controlador, escreva o caminho totalmente qualificado.
- 4. Faça clique em **OK**. O ficheiro de correlação que especificou é apresentado na página Opções de Processamento de Texto.

"Processamento de texto personalizado" na página 214

- Fluxo de trabalho para integração de análise personalizada
- Integração de análise de texto personalizada
- Conceitos básicos utilizados no processamento da análise de texto
- Ficheiros de mapeamento de índice

Tarefas relacionadas:

🕩 Criar a estrutura de análise comum para o ficheiro de mapeamento de índices

Correlacionar a estrutura de análise comum com uma base de dados relacional

Pode especificar quais são os resultados da análise de texto de uma estrutura de análise comum que pretenda correlacionar com uma base de dados relacional, para utilização em aplicações de base de dados.

O tamanho máximo de um ficheiro de correlação é de 8 MB.

Ao correlacionar a estrutura de análise comum com tabelas de base de dados relacionais, permite que os dados sejam utilizados por aplicações de base de dados. Por exemplo, os utilizadores podem especificar consultas de SQL fora do IBM Content Analytics with Enterprise Search para procurar nas anotações que foram adicionadas pela análise de texto. Também pode utilizar a informação sobre a continuação do processamento de dados, tal como a informação em aplicações na prospecção de dados.

Pode utilizar um ficheiro de correlação de cada vez para correlacionar uma estrutura de análise comum com uma base de dados relacional. Se carregar um novo ficheiro de correlação, o ficheiro de correlação actual será substituído.

Para correlacionar a estrutura de análise comum com uma base de dados relacional:

- Na vista Colecções, expanda a colecção que pretende configurar. No painel Analisar e Indexar, faça clique em Configurar > Opções de processamento de texto.
- 2. Na área Correlacionar a estrutura de análise comum com uma base de dados relacional, faça clique em Adicionar Correlação.

- 3. Na página Correlacionar a Estrutura de Análise Comum com uma Base de Dados Relacional, escreva um nome de apresentação descritivo para o ficheiro de correlação que pretende utilizar para correlacionar informações com uma base de dados relacional.
- 4. Especifique a localização do ficheiro de correlação. Se o ficheiro se encontrar no sistema local, pode navegar até o localizar. Se o ficheiro se encontrar no servidor de controlador, escreva o caminho totalmente qualificado.
- 5. Faça clique em **OK**. O nome de apresentação do ficheiro de correlação é mostrado na página Opções de Processamento de Texto.

"Processamento de texto personalizado" na página 214

- Fluxo de trabalho para integração de análise personalizada
- 🕩 Integração de análise de texto personalizada
- Conceitos básicos utilizados no processamento da análise de texto
- Ficheiros de mapeamento de bases de dados

Tarefas relacionadas:

Criar a estrutura de análise comum para o ficheiro de mapeamento de bases de dados

Filtros de campos

Para que o utilizador compreenda melhor os dados e possa comparar segmentos de informação aparentemente distintos, pode especificar opções para combinar, separar e alterar valores de campo de diferentes formas.

Por exemplo, o utilizador poderá pretender:

- Concatenar vários valores de campo num único campo para que os dados possam ser analisados como uma única faceta
- Dividir um valor de campo único em vários valores de campo inserindo um delimitador para separar os valores
- Realize operações de texto básicas num valor de campo, como, por exemplo, ajustar texto, correlacionar texto (como, por exemplo, correlacionar o valor N/A com nulo) e alterar a utilização de maiúsculas/minúsculas (como, por exemplo, alterar as minúsculas para maiúsculas)
- Substituir um valor de campo por um outro valor
- Transformar vários campos booleanos num único campo de texto de vários valores e utilizar os nomes de campo originais como novos valores de campo

Quando cria um filtro de campo, selecciona o campo de índice a que o filtro deve ser aplicado. Os valores de todos os campos e elementos HTML que estejam correlacionados com o campo de índice serão filtrados segundo as regras que especificar ao definir o filtro. Após correlacionar o campo de índice com uma faceta, pode seleccionar a faceta ao explorar a colecção.

Por exemplo, parta do princípio que está a examinar dados de uma sala de urgências que contenham vários campos acerca de testes de diagnóstico realizados e cada campo contenha o valor Sim ou Não. Em vez de explorar cada um dos campos independentemente, poderá ter uma maior compreensão se conjugar os campos e os tratar como uma faceta única. Neste caso, os nomes originais destes campos tornam-se o valor do campo de índice que, por sua vez, é correlacionado com uma faceta única. Explorar as estatísticas de teste de diagnóstico combinadas pode ser ainda mais significativo quando explora esta faceta com outra faceta como, por exemplo, a fonte principal de pagamento dos testes.

Outro exemplo: parta do princípio que a colecção do utilizador inclui várias representações de medicamentos como, por exemplo, uma lista separada por delimitadores num campo ou vários campos com valores diferentes:

MEDICAMENTOS: Tylenol, Claritin, Flonase

MEDICAMENTO1: Tylenol MEDICAMENTO2: Claritin MEDICAMENTO3: Flonase MEDICAMENTO4: N/A

Em ambos os casos, poderá ter uma maior compreensão conjugando todos os valores de campo para que as estatísticas possam ser representadas e exploradas como uma faceta única.

Conceitos relacionados:

"Campos de índice" na página 191

Tarefas relacionadas:

"Configurar filtros de campos"

Configurar filtros de campos

Pode configurar filtros para analisar valores de campo de diferentes modos como, por exemplo, conjugar valores de campo num único campo de vários valores para que os valores possam ser analisados como uma faceta única.

Caso adicione ou modifique filtros de campo após analisar e indexar documentos, a qualidade da procura deteriorar-se-á até os documentos serem novamente pesquisados e o índice ser reconstruído. Se a colecção for configurada para utilizar a memória cache dos documentos, pode reconstruir o índice sem voltar a pesquisar os documentos.

Para configurar um filtro para um campo de índice:

- 1. Na vista Colecções, expanda a colecção que pretende configurar. No painel Análise e Indexação, faça clique em **Configurar** > **Filtros de campos**.
- 2. Na página Filtros de campos, faça clique em Criar novo filtro.
- 3. Especifique um nome descritivo para o filtro.
- Seleccione um campo de índice previamente definido com qual quer correlacionar o filtro. Todos os campos correlacionados com o campo de índice serão analisados de acordo com a regra deste filtro.
- 5. Seleccione a condição que causa a filtragem do campo. Por exemplo, caso o valor de campo tenha de conter um valor específico para o filtro ser aplicado, seleccione Valor de campo contém o seguinte valor e, de seguida, especifique o valor devido.
- 6. Seleccione a acção a efectuar pelo analisador, caso se verifique a condição seleccionada. Por exemplo, poderá querer substituir o actual valor de campo por outro valor especificado pelo utilizador.
- 7. Faça clique em **OK**. O novo filtro aparece no final da lista de filtros na página Filtros de campo.
- 8. Opcional: Caso existam outros filtros, seleccione o filtro e faça clique no botão para o mover para outra posição na lista. A ordem segundo a qual especifica os filtros corresponde à ordem utilizada pelo analisador para os aplicar.

"Campos de índice" na página 191

"Filtros de campos" na página 221

Campos de data e formatos de data personalizados

A data do documento é crítica para a análise de texto, especialmente para explorar como os dados são alterados ao longo do tempo e para observar desvios e tendências.

Para preservar a forma como as datas são calculadas, o IBM Content Analytics with Enterprise Search inclui um campo de índice paramétrico denominado Data, que não pode editar ou remover. No entanto, quando configura o analisador para uma colecção, é possível especificar formatos de data personalizados para garantir que os dados de data adicionados à colecção são correlacionados com este campo de índice e correctamente indexados.

Os dados de data podem ser configurados para uma colecção de vários modos:

- Se configurar um pesquisador, é possível correlacionar os campos de origem de dados e campos de metadados com o campo de índice Data.
- Se importar ficheiros CSV para uma colecção, é possível especificar o formato dos valores de data.
- Se adicionar documentos a uma colecção utilizando a API de administração REST, a API pode identificar os campos de data.
- Se correlacionar elementos HTML e XML com campos de índice, é possível correlacionar elementos com o campo de índice Data.
- Quando configura facetas para uma colecção, é possível correlacionar o campo de índice Data com facetas.
- Se associar um anotador UIMA com uma colecção, o anotador pode produzir valores de data para a faceta da data.

Faceta de data

O campo de índice Data é utilizado como a data do documento nos resultados da consulta. O valor deste campo é convertido numa faceta da data, em colecções de análise de conteúdo, de modo a que possa ser utilizado para comparar cronologias, desvios e tendências. Em colecções de pesquisa, os utilizadores podem utilizar a faceta de data para limitar os resultados.

A faceta de data consiste nos seguintes componentes de caminho: data, *ano*, *mês*, *dia*, *hora*. Os níveis dos componentes de caminho não podem ser alterados.

Quando o analisador detectar um valor de dados, converte o valor em hora epoch (o número de milissegundos desde dia 1 de Janeiro, 1970, 00:00:00 GMT), como 1235487600000. Os caracteres na cadeia são processados como o número de milissegundos desde a data epoch.

Além do campo de índice Data predefinido, é possível configurar outros campos a serem utilizados como campos de data. Neste caso, tem de especificar que o campo de origem de dados ou o campo de metadados é um campo de índice paramétrico e tem de especificar que o campo contém dados de data. Quando o analisador detectar campos de data paramétricos, o valor de campo é convertido em hora epoch.

Formatos de data detectados por predefinição

O analisador pode, automaticamente, detectar os seguintes formatos de data e hora, pela ordem especificada nesta tabela. Além de estes formatos, é possível configurar o analisador para reconhecer formatos de data personalizados para o conteúdo que deseja incluir numa colecção.

Tahela 20	Detectar	automaticamente	formatos de data
	Delectar,	automaticamente,	ionnalos de dala

Formato de data	Valor de amostra
RFC 1123	Domingo, 06 Nov 1994 08:49:37 GMT
RFC 850	Domingo, 06-Nov-94 08:49:37 GMT
asctime	Domingo Nov 6 08:49:37 1994
 ISO8601. Apenas é suportada a data do calendário. As representações não suportadas são: Representações com precisão reduzida Representações truncadas Representações expandidas Representação de fracções decimais Representação com designador de zona 	2004-02-05
RFC 1123 sem fuso horário	Domingo, 06 Nov 1994 08:49:37
RFC 850 sem fuso horário	Domingo, 06-Nov-94 08:49:37
<pre>Formato de data e hora para o local predefinido da colecção. Obtido através da classe de Java DateFormat.getDateInstance(): • DateFormat.getDateTimeInstance() • DateFormat.getDateInstance(DateFormat.FULL) • DateFormat.getDateInstance(DateFormat.LONG) • DateFormat.getDateInstance(DateFormat.MEDIUM) • DateFormat.getDateInstance(DateFormat.SHORT)</pre>	

Formatos de data personalizados

Quando configura as opções de análise e de índice para uma colecção, é possível especificar os formatos de data personalizados para garantir que os dados de data que incluir na colecção são correctamente indexados. O analisador testa os seus formatos de data personalizados (no seguimento da ordem que especificar) para analisar valores de data e, depois, testa os formatos de data predefinidos. O primeiro valor analisado com sucesso é utilizado como data.

Quando configura formatos de data personalizados, especifica:

- A cadeia do formato, como EEE, d MMM aaaa HH:mm:ss Z (por exemplo, Qua, 4 Jul 2001 12:08:56 -0700). A cadeia pode estar em qualquer formato suportado pela classe Java SimpleDateFormat.
- O fuso horário e locale para a data. O locale e fuso horário da colecção são seleccionados por predefinição.
- A ordem em que os seus formatos de data personalizados devem ser aplicados. Após adicionar um novo formato personalizado, é possível deslocá-lo para a primeira, última ou qualquer posição na lista.

Os seus formatos de data personalizados aplicam-se a todo o conteúdo de data que seja adicionado ou configurado para uma colecção. Para que as suas alterações se tornem efectivas, tem de reiniciar o analisador. Para aplicar as alterações a

documentos no índice, reconstrua o índice ou, se a colecção utilizar memória cache de documento, reconstrua o índice a partir da memória cache.

Apresentar datas nos resultados da pesquisa

É possível utilizar qualquer um dos métodos seguintes para controlar a forma como os campos de data são apresentados nos resultados da consulta:

- Edite o ficheiro de propriedades para a aplicação, como o ficheiro config.properties para a aplicação da procura empresarial. Na propriedade date.fields, especifique uma lista separada por espaços dos campos que devem ser formatados como dados de data nos resultados da consulta. O formato da data apresentada corresponde às definições do locale no navegador Web.
- Execute o personalizador da aplicação, expanda o separador Resultados e inclua os nomes dos campos que serão formatados como datas no campo Campos de data. O formato da data apresentada corresponde às definições do locale no navegador Web.

Tarefas relacionadas:

"Personalizar aplicações através da utilização de uma interface de personalizador" na página 347

Referências relacionadas:

"Ficheiros de configuração da aplicação" na página 325

Remover URIs do índice

Para impedir os utilizadores de procurar documentos numa colecção, poderá remover os URIs desses documentos do índice.

Quando remover um URI do índice, não o remova do espaço de pesquisa. Da próxima vez que o pesquisador pesquisar o documento, ou da próxima vez que o índice for reconstruído a partir da memória cache de documentos, o URI será incorporado no índice e será novamente disponibilizado para procura. Para remover um URI do espaço de pesquisa, terá de actualizar as regras de pesquisa para excluir o documento, e então parar e reiniciar o pesquisador.

Para remover URIs de documentos específicos do índice:

- Na vista Colecções, expanda a colecção que pretende configurar.No painel Analisar e Indexar, faça clique em Configurar > URIs para os remover do índice.
- 2. Na página Remover URIs do Índice, escreva os URIs (ou os padrões URI) que pretende remover do índice. Por exemplo:

http://domain.org/hr/*
db2://knowledgeManagement/ROI*
cm://enterprise/finance*

Referências relacionadas:

"Formatos de URI no índice" na página 164

Processos globais

Os serviços de indexação efectuam a análise global dos documentos numa colecção para determinar os resultados de correlação, detectar documentos duplicados, gerar imagens de miniaturas e atribuir documentos a categorias.

Quando configura as opções de análise e de índice para uma colecção, é possível especificar opções que influenciam o modo como os utilizadores exploram o índice e vêem os resultados. Poderá proceder do seguinte modo:

- Agende processos globais. Para conservar os recursos do sistema, a análise global ocorre quando o índice está oculto e não estão a ser adicionados activamente documentos ao índice. Devido aos processos globais apenas serem executados quando a construção do índice está parada, poderá pretender configurar agendamentos para garantir que apenas são executados quando os pesquisadores não estão a adicionar, de modo activo, conteúdo ao índice:
 - Criar um agendamento para executar a análise da ligação. Se seleccionar a opção de classificação estatística para classificar documentos por número de ligações a cada documento quando cria uma colecção ou edita as definições da colecção, a análise da ligação é automaticamente activada.
 - Criar um agendamento para gerar imagens de miniatura. Se seleccionar a opção para gerar miniaturas ao criar uma colecção, as imagens de miniaturas são criadas para determinados tipos de documentos. Os utilizadores podem ver as imagens de miniatura quando vêem os resultados da consulta.
 - Criar um agendamento para detectar documentos duplicados. Se seleccionar a opção para detectar documentos duplicados quando cria uma colecção ou edita as definições da colecção, são identificados documentos idênticos ou quase idênticos na colecção. Pode melhorar a qualidade da pesquisa assegurando que os resultados da pesquisa não contêm vários documentos com o mesmo ou quase o mesmo conteúdo.
- Configure tarefas de análise global personalizadas. Utilize a análise global personalizada para implementar lógica de personalização que analise todo o conjunto de documentos.
- Configure categorias baseadas em regras. As categorias permitem-lhe agrupar documentos que partilham características comuns para que os utilizadores possam explorar apenas documentos que correspondam aos critérios para se tornarem membros desse grupo.
- Configure e implemente categorias que são compiladas através da análise de agrupamentos. Na análise do agrupamento, o sistema extrai os termos propostos dos metadados de subconjuntos (agrupamentos) de documentos no índice.

Análise de ligações

Ao criar uma colecção e seleccionar a opção para classificar documentos pelo número de ligações aos documentos, ocorre automaticamente a análise das ligações.

Na análise de ligações, os documentos ligados são automaticamente recuperados e indexados com o texto âncora na ligação. Em consequência, os documentos ligados são pesquisáveis por palavras-chave que ocorrem no texto âncora, mesmo que as palavras-chave não apareçam no documento de origem.

Para assegurar que a quantidade máxima de recursos do sistema será dedicada a construir o índice, o processo de análise de ligações apenas é executado quando o processo de construção do índice está em pausa, quando, por exemplo, a construção do índice aguarda novos documentos para indexar.

Ao configurar as opções de análise e de indexação, poderá especificar um agendamento que controle quando o processo de análise de ligações é executado, por exemplo, todos os dias ou a cada seis horas. Ao supervisionar o analisador, poderá registar como muitos documentos foram analisados, consultar quando o processo está agendado para ser executado novamente e activar ou desactivar o agendamento. Se não especificar um agendamento de análise de ligações, o processo é activado para ser executado quando se inicia a construção do índice, embora o processo não seja executado até que a construção do índice entre no estado de espera.

Se alterar o modelo de classificação estática para a opção de classificação de documentos pelo número de ligações depois de criar uma colecção, tem de reconstruir todo o índice dos documentos em cache.

Detecção de documentos duplicados

A detecção de documentos duplicados é uma técnica que é utilizada para impedir que os resultados da procura contenham documentos múltiplos com o mesmo ou quase com o mesmo conteúdo.

A qualidade da procura poderá ser degradada se cópias múltiplas dos mesmos documentos (ou aproximadamente os mesmos) forem listadas nos resultados da procura. A análise de documentos duplicados ocorre apenas quando ambas as seguintes condições forem verdadeiras:

- A colecção utiliza o modelo de classificação baseado em ligações. Este modelo aplica-se a pesquisadores que pesquisam sítios da Web, tais como o pesquisador da Web ou o pesquisador do WebSphere Portal.
- A segurança das colecções está desactivada.

Durante a análise global, os processos de indexação detectam duplicados através do varrimento do conteúdo dos documentos para cada documento. Se dois documentos tiverem o mesmo conteúdo, são tratados como duplicados.

Quando especificar se um campo ou um campo de metadados constitui conteúdo do documento, o conteúdo desses campos é adicionado ao resumo dinâmico do documento nos resultados da procura, que pode ter um impacto se o documento for apresentado nos resultados da procura. Se a detecção duplicada aproximada estiver activada na aplicação (a propriedade NearDuplicateDetection no método setProperty é definida como Yes), os documentos com títulos e resumos semelhantes são ocultados quando um utilizador visualiza os resultados da procura.

Num grupo de documento duplicados, um documento é o principal e os outros são as duplicações. Todos os documentos no grupo de duplicações têm a mesma representação canónica do conteúdo. Durante a indexação, o conteúdo (tokens) do documento principal é indexado. Para os documentos duplicados, só os tokens de metadados são indexados. Quando o documento principal é eliminado do índice, o duplicado seguinte torna-se o principal. Quando os utilizadores procuram a colecção, só o documento principal é devolvido.

Configurar a detecção de documentos duplicados

Quando criar uma colecção, pode especificar se pretende activar detecção de documentos duplicados para a colecção. Também pode activar ou desactivar detecção duplicada de documentos alterando opções gerais para a colecção.

Se activar esta função para uma colecção, pode configurar agendamentos para controlar quando o processo de detecção é executado. Porque a detecção de documentos duplicados só é executada quando a construção do índice faz uma pausa, poderá querer agendar uma hora para que execute de forma a assegurar de que apenas é executada quando os pesquisadores não estiverem a adicionar activamente o conteúdo do índice.

Ver documentos duplicados no prospecção de análise de conteúdo

No prospecção de análise de conteúdo, os utilizadores podem especificar preferências para ver informações acerca de documentos duplicados. Por exemplo, os utilizadores podem ver a percentagem dos documentos que correspondem às condições de procura actuais que são duplicados. Também podem seleccionar um documento e ver uma lista de documentos semelhantes ao documento seleccionado. Os utilizadores também podem definir uma preferência de escala móvel para controlar quão semelhantes ao documento seleccionado têm de ser os documentos para que possam ser incluídos na lista de documentos semelhantes.

Conceitos relacionados:

"Análise de documentos duplicados e segurança de colecção" na página 382

Criação de miniaturas

Ao criar uma colecção, pode especificar se pretende que as imagens miniatura sejam criadas para determinados tipos de documentos. Ao visualizar imagens miniatura nos resultados da procura, utilizadores podem avaliar melhor a importância dos documentos sem gastar tempo a abrir documentos para ver o conteúdo real.

Restrição: Se instalar o IBM Content Analytics with Enterprise Search no sistema operativo Linux com System *z*, a criação de miniaturas não é suportada.

Para activar a criação de miniaturas para uma colecção, primeiro tem de activar a memória cache de documentos. Para assegurar que a quantidade máxima de recursos do sistema será dedicada a construir o índice, as imagens miniaturas serão armazenadas nesta memória cache separada. O processo de criação de miniaturas é executado apenas quando processo de construção do índice está parado, como quando a construção do índice aguarda a indexação de novos documentos.

Ao configurar as opções de análise e de indexação, poderá especificar uma agenda que controle quando o processo de geração de miniaturas é executado, como, por exemplo, todos os dias ou a cada seis horas. Ao supervisionar o analisador, poderá registar quantas imagens miniatura foram criadas, consultar quando o processo está agendado para ser executado novamente e activar ou desactivar a agenda. Se não especificar um agendamento para a criação de miniaturas, o processo fica em fila de espera para ser executado quando a construção do índice for iniciada, ainda que o processo não seja executado até que a construção do índice entre em estado de espera.

Quando os utilizadores procuram uma colecção que inclua imagens de miniatura, podem seleccionar se pretendem mostrar ou ocultar as imagens quando visualizam

os resultados da procura. Também pode especificar se pretendem que as imagens apareçam no lado esquerdo ou direito da página. Se um utilizador fizer clique numa imagem miniatura, a imagem é temporariamente apresentada no centro da página.

Como administrador, pode alterar o tamanho de visualização das imagens de miniatura.

Supervisionar a construção do índice

A criação de miniaturas é executada num único servidor e num único módulo, ainda que estejam configurados vários servidores de processamento de documentos. O mesmo conversor de sessão é partilhado por diversas colecções. Isto significa que os documentos em diversas colecções são processados através de um único módulo no sistema.

Quando a construção do índice de texto está em execução para processar novos documentos ou para processar um pedido de reconstrução do índice, a criação de miniaturas fica suspensa até que os serviços de indexação terminem de processar os documentos.

Quando supervisionar a construção do índice, o progresso do processo da criação de miniaturas é apresentado. Este progresso é um rácio entre o número de documentos processados e o número total de documentos para os quais a criação de miniaturas é suportada, tais como documentos em formato PDF. Por exemplo, se o progresso é de 50% e o número total de documentos que suportam a criação de miniaturas é 200, a criação de miniaturas processou 100 documentos.

Ao fazer a reconstrução do índice, a criação de miniaturas processa novamente todos os documentos para verificar se as imagens em miniatura precisam de ser actualizadas. Neste caso, o progresso começa nos 0%. Se já existir uma miniatura para o documento e o documento não foi submetido a qualquer processamento adicional, será utilizada a miniatura existente. Deste modo, a miniatura anterior continua a ser passível de ser procurada a não ser que o documento esteja determinado como tendo sido processado novamente.

Tipos MIME suportados

As imagens miniatura podem ser criadas para documentos nos formatos seguintes:

```
application/vnd.lotus-freelance
application/x-lotus-123
application/vnd.lotus-wordpro
application/rtf
application/x-msexcel
application/msword
application/x-mspowerpoint
application/vnd.openxmlformats-officedocument.spreadsheetml.sheet
application/vnd.openxmlformats-officedocument.presentationml.presentation
application/vnd.openxmlformats-officedocument.wordprocessingml.document
application/pdf
application/x-js-taro
application/vnd.visio
application/vnd.stardivision.writer
application/vnd.stardivision.calc
application/vnd.stardivision.impress
application/vnd.sun.xml.writer
application/vnd.sun.xml.calc
application/vnd.sun.xml.impress
```

```
application/vnd.oasis.opendocument.text
application/vnd.oasis.opendocument.spreadsheet
application/vnd.oasis.opendocument.presentation
```

Gerar miniaturas em AIX

Para gerar imagens em miniatura num sistema AIX, o X Server tem de estar em execução. Além disso, a variável de ambiente DISPLAY para o administrador deve remeter para o X Server.

Para activar a geração de miniaturas num sistema AIX:

- 1. No servidor de processamento do documento, inicie o servidor X. Para obter um desempenho melhor, active a criação de miniaturas no servidor do controlador e todos os servidores de processamento do documento.
- 2. Inicie sessão como administrador predefinido do IBM Content Analytics with Enterprise Search em modo GUI.
- Confirme se a variável de ambiente DISPLAY está definida para o administrador. Por exemplo, execute o comando export e verifique se a saída de dados é semelhante à seguinte linha: DISPLAY=:0.0

Se a variável de ambiente DISPLAY não estiver definida, defina-a.

- Confirme se o administrador pode abrir o ecrã. Por exemplo, execute o comando xclock e verifique se é apresentado um relógio. Se o relógio não for apresentado, contacte o administrador do AIX e peça assistência.
- 5. Se o sistema IBM Content Analytics with Enterprise Search estiver em execução, pare-o e reinicie-o:

```
esadmin system stopall
esadmin system startall
```

- 6. Continue com sessão iniciada como administrador de análise de conteúdos enquanto o processo de criação de miniatura estiver em execução. Poderá bloquear o ecrã para garantir que a sessão não é terminada.
- 7. Na consola de administração de IBM Content Analytics with Enterprise Search, supervisione uma colecção activada para gerar miniaturas e reconstruir o índice da memória cache de documentos ou iniciar os pesquisadores para executar uma pesquisa completa para que o índice possa ser reconstruído.

Alterar a predefinição DISPLAY:

A variável do ambiente DISPLAY para o administrador é definida quando o IBM Content Analytics with Enterprise Search está instalado e guardado no ficheiro \$ES_INSTALL_ROOT/nodeinfo/es.cfg no servidor do controlador. A definição corresponde à variável de ambiente DISPLAY que está definida para utilizador raiz. Siga estes passos se quiser utilizar um valor diferente como, por exemplo, **DISPLAY=:1.0** em vez de **DISPLAY=:0.0**:

- 1. Inicie sessão como o administrador predefinido do IBM Content Analytics with Enterprise Search no servidor do controlador e edite o ficheiro es.cfg para especificar o valor de DISPLAY que quer usar.
- Pare o sistema IBM Content Analytics with Enterprise Search: esadmin stop stopccl.sh
- 3. Termine a sessão.
- 4. Inicie sessão como administrador outra vez.
- 5. Reinicie o sistema IBM Content Analytics with Enterprise Search:

```
startccl.sh
esadmin start
```

Resolução de problemas:

Utilize as seguintes sugestões de detecção e correcção de problemas caso as imagens miniatura não apareçam nos resultados de procura após seguir o procedimento neste tópico:

- Certifique-se de que memória cache de documentos e o suporte de miniaturas estão activados para a colecção.
- Certifique-se de que o sistema X Windows está em execução e de que as fontes estão correctamente configuradas.
- Certifique-se de que reiniciou o sistema IBM Content Analytics with Enterprise Search.
- Certifique-se de que reconstruiu o índice da memória cache de documentos ou de que voltou a pesquisar origens de dados que contêm documentos para os quais pretende gerar imagens miniatura.

Alterar o tamanho das imagens de miniatura

Caso pretenda alterar a forma como as imagens são apresentadas nos resultados da procura, pode editar um ficheiro de configuração para alterar as propriedades de altura e largura.

Poderá querer aumentar o tamanho das imagens de miniatura, por exemplo, para ser capaz de identificar melhor os documentos que correspondem aos termos da consulta antes de abrir algum documento.

Para alterar o tamanho de visualização das imagens de miniatura:

- Na consola de administração, crie uma colecção e active a memória de cache do documento, bem como a criação de miniaturas. Certifique-se de que a colecção inclui documentos com tipos MIME que são suportados para criação de miniaturas.
- 2. Caso estejam em execução, pare os serviços de índice e de análise para a colecção.
- 3. Edite o ficheiro ES_NODE_ROOT/master_config/collection_id.indexservice/
 converter_config.xml e mude os valores para as propriedades HeightLimit e
 WidthLimit. Altere, por exemplo, o valor predefinido, 360, para 640:
 cproperty name ="HeightLimit" value="640">
 cproperty name ="WidthLimit" value="640">
- 4. Reinicie os serviços de análise e de indexação.
- Pesquise documentos que são do tipo MIME suportado. Quando os utilizadores pesquisam a colecção e definem as respectivas preferências para ver imagens de miniatura, as imagens de miniatura são apresentadas no tamanho que especificou.

Análise global personalizada

Em acréscimo às tarefas de análise global predefinidas que ocorrem durante o processo de indexação, é possível configurar o IBM Content Analytics with Enterprise Search para executar tarefas de análise global personalizadas.

Restrição: A análise global personalizada apenas está disponível para colecções que utilizem IBM InfoSphere BigInsights. Jaql tem de estar instalado no servidor BigInsights.

É possível utilizar a análise global personalizada para obter informações ao examinar todo o conjunto de documentos, em vez de examinar cada documento individualmente. Por exemplo, tenha em consideração os seguintes casos de exemplo de utilização:

Para cada documento, conte o número de vezes que é mencionado por outro documento

Suponha que a sua colecção consiste em documentos, como, por exemplo, artigos de publicações e patentes que mencionam outros documentos, e que pretende calcular o número de vezes que cada documento é mencionado noutro documento. É utilizado um anotador personalizado durante o processamento do documento para extrair informações sobre os documentos mencionados em cada documento. É configurada, então, uma tarefa de análise global personalizada para contar o número de menções de cada documento em todo o conjunto de documentos. Este valor é, então, guardado num novo campo de cada documento mencionado. No prospecção de análise de conteúdo, um utilizador pode ordenar os documentos de acordo com este valor para determinar quais os documentos mais frequentemente mencionados por outros documentos na colecção.

Conte o número de entidades designadas que foram extraídas de documentos numa colecção

Durante o processamento do documento, o IBM Content Analytics with Enterprise Search extrai entidades designadas de cada documento. É possível configurar uma tarefa de análise global personalizada para contar quantas vezes cada entidade designada ocorre em todos os documentos da colecção. Este valor é, então, guardado num ficheiro para que outra aplicação possa utilizar os dados.

A lógica da análise global personalizada é implementada ao criar um script Jaql (linguagem de consulta para JSON). As entradas de dados para o script são os campos, facetas e texto extraídos do conteúdo durante a etapa de processamento do documento. A produção de dados do script pode ser armazenada como campos ou facetas do documento no índice do IBM Content Analytics with Enterprise Search. Também é possível especificar o seu script Jaql para guardar a produção de dados num ficheiro ou noutro formato para que outra aplicação possa utilizar os dados.

O script Jaql tem de ser incluído num ficheiro de arquivo que tenha a extensão de ficheiro .zip. Em acréscimo ao script Jaql, o ficheiro de arquivo tem de conter o ficheiro de configuração de análise global personalizada (install.jaql) e quaisquer outros ficheiros que sejam necessários para o script Jaql personalizado.

Após desenvolver um script de análise personalizada, configure uma tarefa de análise global personalizada para uma colecção na consola administrativa para especificar quais os campos e facetas a transferir para o script para análise. No painel Analisar e Indexar na consola administrativa, faça clique em **Configurar** > **Processamento global** > **Análise global personalizada** e faça clique no ícone **Adicionar**.

Por predefinição, a sua tarefa de análise global personalizada é automaticamente executada após conclusão da indexação. Como alternativa, é possível especificar

um agendamento para configurar a frequência de execução da tarefa, como, por exemplo, todos os dias ou de seis em seis horas.

Por exemplo, os seus pesquisadores são agendados para serem executados várias vezes ao dia. Após o pesquisador terminar, o processo de indexação é automaticamente iniciado. Após o processo de indexação terminar, a tarefa de análise global é automaticamente iniciada por predefinição. Por serem operações sequenciais, a indexação e a análise global não podem ser executadas em simultâneo. Durante a análise global, nenhum documento pode ser indexado. Se a sua tarefa de análise global personalizada necessitar de um longo período de tempo para ser executada, os documentos são pesquisados regularmente mas não são indexados enquanto a tarefa de análise global estiver a ser executada. Nesse caso, se pretender indexar documentos regularmente, mesmo que os documentos recém-indexados não tenham os resultados da análise global, é possível desactivar a execução automática. Em vez disso, agende a execução de uma análise global, uma vez por dia, de modo a que os documentos disponham, eventualmente, do resultado da análise global.

Devido a ser necessário armazenar os dados de entrada e de saída para a tarefa de análise global personalizada, é necessário espaço de disco adicional. A quantidade de espaço de disco adicional que é necessária depende de quais os dados que são gerados e utilizados pela tarefa de análise global personalizada. Os dados de entrada são armazenados em cada nó do BigInsights e os dados de saída são armazenados no servidor principal do IBM Content Analytics with Enterprise Search.

Tarefas relacionadas:

🖙 Criar e implementar um suplemento de análise global personalizada

Referências relacionadas:

🕩 Suplementos de exemplo para análise global personalizada

Categorias

As categorias permitem agrupar documentos que possuem características comuns. Os utilizadores só poderão explorar os documentos que cumprem os critérios para pertencer a esse grupo.

Se associar documentos com categorias e as suas aplicações suportarem essa capacidade, os utilizadores podem reduzir os resultados ao explorar apenas os documentos que pertencem a categorias específicas.

Quando configura uma categoria baseada em regras, especifique as regras para associar documentos à categoria. Pode agrupar documentos que possuem o mesmo padrão URI. Também pode agrupar documentos que contêm conteúdos específicos como, por exemplo, documentos que contêm ou excluem determinadas palavras ou expressões.

Ao analisar agrupamentos de documentos, o sistema pode extrair termos de metadados propostos a partir de subconjuntos (*agrupamentos*) de documentos no índice. Ao configurar uma categoria com base na análise de agrupamentos de documentos, seleccione a proposta de agrupamento que contenha as palavras que pretende utilizar para classificar documentos.

Se alterar uma categoria após pesquisar e indexar documentos, a qualidade de procura deteriora-se até os documentos serem pesquisados e indexados novamente. Se a colecção for configurada para utilizar a memória cache de documentos, pode reconstruir o índice sem voltar a pesquisar documentos.

Categorias baseadas em regras

Pode configurar regras para controlar os documentos que são associados a categorias numa colecção.

Para configurar regras para categorizar documentos, especifique que pretende utilizar categorias baseadas em regras quando criar a colecção. Também pode activar a categorização baseada em regras ao editar as definições da colecção. Para aplicar categorias a documentos no índice, todos os documentos da colecção devem ser pesquisados e o índice feito novamente. Se a colecção for configurada para utilizar a memória cache de documentos, pode reconstruir o índice sem voltar a pesquisar documentos.

O analisador utiliza as regras que especifica para utilizar documentos com uma ou mais categorias:

- Se um documento passar de pelo menos uma regra para uma categoria, o analisador utiliza o documento com a categoria.
- Se um documento passar de pelo menos uma regra para várias categorias, o analisador utiliza o documento com todas as categoria.
- Se um documento não passar de nenhuma regra para uma categoria, o analisador não utilizará o documento com uma categoria. Os utilizadores podem pesquisar este documento e obtê-lo quando pesquisarem a colecção, mas não podem procurar uma categoria e esperar recuperar o documento.

Ao administrar a árvore de categorias para uma colecção, decida onde é que pretende adicionar uma nova categoria na hierarquia de categorias. Também utiliza a árvore de categorias para seleccionar a categoria que pretende editar. Depois pode adicionar regras para categorizar documentos, eliminar regras ou alterar o conteúdo de determinadas regras.

Quando configura uma regra para categorizar documentos, escolhe se pretende utilizar a URI ou o conteúdo do documento para determinar se o documento pertence à categoria.

Regras de padrões de URI

Uma regra URI aplica-se ao documento URI. Especifique um URI parcial (um padrão), e os documentos que têm o padrão especificado nos respectivos URIs passam a regra.

Por exemplo, se especificar que o texto da regra é /hr/, então o primeiro URI do exemplo que se segue passa a regra, mas o segundo URI não:

file:///corporate/hr/medicalform.doc
http://company.com/hr2/medicalform.htm

Como todos os URIs são considerados como padrões, o sistema ignora quaisquer asteriscos que especifique como carácter global no início ou fim do padrão. Por exemplo, */hr/* e /hr/ correspondem ao mesmo conjunto de URIs.

As regras de padrão URI não são sensíveis a maiúsculas. Se um URI contiver espaços, o padrão de URI deve aderir às regras para codificar URIs. O exemplo seguinte mostra formas correctas e incorrectas de especificar um URI para um caminho de sistema de ficheiros do Windows: URI incorrecto: file:///c:/program files/ URI correcto: file:///c:/program+files/

Regras de conteúdo do documento

As regras de conteúdo do documento são expressas no mesmo formato que uma consulta. Se o documento for válido para uma consulta, passa a regra. Quando configurar a regra, especifique as palavras e expressões que os documentos devem conter ou excluir. Também pode seleccionar o idioma para aplicar regras de stemming de palavras.

Por exemplo, a regra seguinte especifica que se um documento contiver ou a palavra hr ou a frase recursos humanos, o documento passa a regra:

rh "recursos humanos"

Noutro exemplo, a regra que se segue especifica que, se o título de um documento contiver a palavra "saúde", mas não a frase "benefícios dos empregados", o documento passa a regra:

+title:saúde -title:"benefícios dos empregados"

As regras de conteúdo estão sujeitas às mesmas normalizações linguísticas que as consultas de API de Procura e de Índice (SIAPI). No entanto, a sintaxe para regra de conteúdo suporta um subconjunto das operações disponíveis na sintaxe da consulta SIAPI. Só são permitidos os seguintes operadores de consulta em regras de conteúdo:

- + Precede um termo com um sinal de adição para indicar que o termo tem de existir no documento.
- Precede um termo com um sinal de subtracção para indicar que o termo não tem de existir no documento.
- "" Incluir dois ou mais termos entre aspas para indicar que a frase exacta tem de existir no documento.

nome_campo:

Preceda um termo ou frase de um nome de campo para indicar que o termo ou frase tem (ou não tem) de ocorrer no campo de documento especificado.

São suportados todos os campos de conteúdo e metadados que estejam configurados para serem campos pesquisáveis na colecção. Não são suportados as seguintes palavras-chave de campo e tipos de campo SIAPI:

- site:
- url:
- link:
- docID:
- samegroupas:
- campos paramétricos
- tokens de segurança
- atributos (como \$source, \$language, \$doctype, etc.)

Sintaxe de consultas

Referências relacionadas:

"Formatos de URI no índice" na página 164

Árvores de categorias

Uma categoria permite-lhe visualizar todas as categorias baseadas em regras numa colecção. Utilize a árvore de categorias para criar categorias, eliminar categorias e editar as regras que utilizam documentos com categorias.

Uma árvore de categorias, também chamada de taxonomia está disposta numa hierarquia. A árvore começa com a categoria raiz e todas as outras categorias provêm da categoria raiz. Pode imbricar qualquer número de categorias e subcategorias para pôr à disposição dos utilizadores diferentes opções para procurar e obter documentos.

Por exemplo, se um documento passar as regras em várias categorias, está associado a todas essas categorias. Quando os utilizadores exploram categorias nos resultados da procura, o facto de um documento pertencer a várias categorias aumenta a probabilidade de os utilizadores o encontrarem.

Quando administra a árvore de categorias, pode controlar os documentos pertencentes a uma ou mais categorias, imbricando novas categorias sob as categorias existentes. Quando cria uma categoria, especifique se será criada ao nível da raiz ou como uma subcategoria de outra categoria. Utilize também a árvore de categorias para eliminar categorias da colecção e para alterar as regras para utilizar documentos com categorias. Ao editar uma categoria, pode mudar o nome da categoria, adicionar ou eliminar as regras de categorização, ou alterar o conteúdo de regras individuais.

Se um utilizador procurar uma categoria de alto nível, serão procurados documentos correspondentes aos critérios da procura nessa categoria e em todas as subcategorias pertencentes. Se um utilizador procurar uma categoria que não tem subcategorias adicionais, será procurada apenas essa categoria.

Configurar categorias baseadas em regras

Pode criar várias categorias para uma colecção e cada categoria pode conter várias regras. As regras determinam quais os documentos associados automaticamente à categoria.

Se as aplicações permitirem o suporte para categorias, os utilizadores podem limitar resultados ao seleccionar categorias no painel de navegação de facetas. Quando um utilizador seleccionar uma categoria, apenas os documentos que pertencem à categoria seleccionada são apresentados nos resultados da procura.

Importante: Se alterar regras de categorias após pesquisar e indexar documentos, o índice fica incoerente. Para assegurar a exactidão dos resultados de procura, todos os documentos na colecção devem ser pesquisados ou novamente importados e, de seguida, novamente indexados. Se a colecção for configurada para utilizar a memória cache dos documentos, pode reconstruir o índice sem voltar a pesquisar ou importar os documentos.

Se um utilizador tiver autoridade para criar regras de categorias, as regras que esse utilizador adicionar são apresentadas na árvore de categorias. Pode consultar os IDs de utilizador dos utilizadores da aplicação que adicionaram as regras e editar ou remover estas regras. A menos que o utilizador da aplicação também tenha a autoridade para reconstruir o índice de categorias, as categorias não se reflectem na aplicação do utilizador até que reconstrua o índice de categorias.

Para configurar uma categoria baseada em regras:

- Na vista Colecções, expanda a colecção que pretende configurar. No painel Analisar e Indexar, faça clique em Configurar > Processos globais > Categorias baseadas em regras.
- 2. Na página categorias Baseadas em Regras, seleccione a categoria que pretende configurar.
 - Seleccione a categoria raiz predefinida caso pretenda adicionar uma categoria ao nível de topo na árvore de categorias. Não pode editar a definição da categoria raiz predefinida ou remover a categoria raiz predefinida da árvore.
 - Seleccione uma categoria individual na árvore se quiser adicionar uma subcategoria a uma categoria. Pode editar as regras de categoria, mudar a ordem das subcategorias ou remover a categoria ou subcategoria da árvore. Pode imbricar subcategorias para definir a hierarquia em que as categorias são apresentadas quando os utilizadores consultam uma colecção.
- **3**. Adicione uma categoria. Tem de especificar o caminho da categoria, que é um ID que permite que a categoria seja utilizada programaticamente. Os seguintes caracteres especiais e caracteres alfanuméricos são suportados: hífen (-), sublinhado (_), 0-9, a-z e A-Z.

Também pode especificar um nome de apresentação para a categoria. Se não especificar um nome de apresentação, o caminho de categoria será utilizado como o nome de apresentação.

Sugestão: O caminho de categoria assegura a singularidade de cada categoria. Se especificar o nome da categoria, deve querer garantir que cada nome também é único. Os utilizadores podem ficar confusos ao ver mais do que uma categoria com o mesmo nome ao consultarem uma colecção.

- 4. Na árvore de categorias, seleccione a categoria que adicionou e faça clique no botão para editar regras de categorias. É apresentada uma lista de regras actualmente definidas. Pode ver o nome da regra, o tipo, o conteúdo, o ID do utilizador que adicionou a regra através da aplicação (caso ela exista), o idioma de consulta a utilizar com documentos que pertençam à categoria e a data e a hora a que a regra foi modificada pela última vez. Pode alterar a forma como as regras são apresentadas:
 - Faça clique no cabeçalho da coluna pela qual pretende ordenar as regras. Para inverter a ordem, faça novamente clique no cabeçalho da coluna.
 - Limite a lista a regras que correspondam a um padrão especificado pelo utilizador. No campo **Filtrar por**, seleccione como pretende filtrar as regras (pelo nome da regra, pelo texto que define a regra ou pelo ID de utilizador) e, de seguida, escreva os caracteres que as regras têm de conter. À medida que escreve, a lista de regras é dinamicamente actualizada para mostrar apenas as regras correspondentes.
- 5. Adicione uma regra:
 - a. Na página Regras, faça clique em Adicionar.
 - b. Na página Adicionar Regra, escreva um nome para a regra, caso esteja a configurar regras para uma colecção de procura. O nome da regra não é necessário para as colecções de análise de conteúdo.

A regra pode ser baseada em padrões de URI ou conteúdo do documento:

Regras de padrões de URI

Para utilizar o URI de um documento para determinar se este pertence à categoria, faça clique em **padrão de URI** e, em seguida, especifique o padrão de URI. Se o texto especificado existir no URI, o documento é associado à categoria. Por exemplo:

file:///c:/program+files/finance

Regras de conteúdo do documento

Para determinar se um documento pertence à categoria através da consulta de conteúdo pesquisável, faça clique em **Conteúdo do documento**, seleccione o idioma dos documentos e, em seguida, especifique as palavras e frases que têm ou não de aparecer no conteúdo do documento. Se um documento incluir ou excluir as palavras especificadas, o documento é associado à categoria.

Exprima a regra no mesmo formato que uma consulta. Por exemplo:

+finance -accounting +title:"ano fiscal"

O operador da consulta rulebased::category-id não é suportado.

Depois de fazer clique em **OK**, a página Regras mostra a regra que criou. Pode adicionar outra regra, editar uma regra ou remover uma regra que adicionou anteriormente.

- c. Depois de terminar a configuração das regras para esta categoria, faça clique em **OK** para voltar à página Categorias Baseadas em Regras.
- 6. Opcional: Altere a ordem das categorias na árvore. Pode alterar a estrutura hierárquica da árvore de categorias de dois modos:

Categorias que podem ser arrastadas

Pode seleccionar e arrastar subcategorias para diferentes locais na árvore. Não pode arrastar a categoria raiz e não pode arrastar uma subcategoria para uma posição de nível superior abaixo da categoria raiz. O sistema faculta indicadores que lhe indicam onde pode largar uma categoria ou subcategoria que seleccionou para arrastar. Pode utilizar a tecla Ctrl para seleccionar várias subcategorias da mesma categoria e arrastá-las ao mesmo tempo para uma nova posição na árvore. Não pode seleccionar e arrastar subcategorias que pertençam a categorias diferentes.

Mover categorias

Seleccione a categoria raiz para organizar categorias de nível superior ou seleccione uma categoria para organizar as subcategorias dessa categoria. Ao fazer clique no ícone **Editar** na área **Ordem de categorias**, é apresentada uma lista de todas as subcategorias da categoria seleccionada. Seleccione uma subcategoria e faça clique nos botões devidos para mover a subcategoria para cima, para baixo ou para a primeira ou última posição da categoria.

Referências relacionadas:

"Formatos de URI no índice" na página 164

Configurar e implementar conjuntos de documentos

Para melhorar a relevância dos resultados de procura, é possível configurar o IBM Content Analytics with Enterprise Search para testar agrupamentos de documentos no índice e, de seguida, configurar uma colecção para categorizar documentos com base na análise de palavras nos agrupamentos. A categorização de documentos que é baseada na análise de conjuntos envolve:

- Configurar o sistema para criar agrupamentos testando um subconjunto de documentos e extraindo palavras.O resultado desta tarefa de agrupamento de documentos é uma *proposta de agrupamento*, a qual consiste em agrupamentos que contêm palavras candidatas para a classificação de documentos.
- Categorizar documentos ao implementar a proposta de agrupamento. Este processo adiciona metadados a documentos com base na análise de agrupamento e cria uma base de dados de conhecimento interno. A base de dados de conhecimento é utilizada para classificar todos os documentos no índice em categorias baseadas em regras.

Quando os utilizadores consultam a colecção, limitam as resultados a documentos que foram categorizados quando a análise de agrupamento foi implementada. Em acréscimo, se a procura conceptual estiver activada para a colecção, os utilizadores podem procurar documentos que correspondam conceptualmente aos termos da consulta.

Se activar a criação de agrupamentos de documentos depois da construção de um índice, é necessária uma reconstrução total do índice, para implementar a criação de agrupamentos de documentos. Se a colecção for configurada para utilizar a memória cache dos documentos, pode reconstruir o índice sem voltar a pesquisar ou importar os documentos.

Para categorizar documentos com base na análise de agrupamentos:

- Expanda a colecção que pretende configurar. Se o suporte para conjuntos de documentos não foi activado quando a colecção foi criada, clique em Acções para editar as definições da colecção e activar a criação de conjuntos de documentos.
- No painel Analisar e Indexar, clique em Configurar > Processamento global > Conjuntos de documentos.
- 3. Crie uma proposta de agrupamento:
 - a. Na página Tarefas de Criação de Agrupamentos de Documentos, introduza um nome descritivo para a tarefa de criação de agrupamentos de documentos.
 - b. Introduza o número de agrupamentos que pretende que o sistema crie ao testar documentos. O valor predefinido é 100.
 - c. Introduza o número de documentos que pretende que o motor de criação de conjuntos de unidades teste quando extrair palavras e criar agrupamentos. Os documentos são extraídos do índice através de amostragem aleatória. O valor predefinido é 5000.
 - d. Introduza o número de documentos a incluir a partir de conjuntos expandidos de amostras. Se este valor não for definido, todos os documentos no índice de texto para a colecção são incluídos. Este parâmetro não está disponível caso utilize algoritmos Latent Dirichlet Allocation (LDA) ou K-means.
 - e. Seleccione o algoritmo de análise de agrupamento que pretende aplicar. Um debate detalhado sobre estes algoritmos e sobre as respectivas diferenças não é abrangido por este documento. Em resumo:
 - Se seleccionar Latent Dirichlet Allocation detectar agrupamentos por amostras, aprender por todas, o algoritmo LDA é aplicado a um conjunto de documentos de amostra (o número de amostras especificado no campo Número de amostras). De seguida, é efectuada a classificação de um conjunto de documentos maior (o número especificado de documentos no campo Número de documentos agrupados) ao detectar agrupamentos e
treinando uma base de dados de conhecimento. Esta base de dados de conhecimento é utilizada para classificar uma colecção inteira de documentos, com base nos documentos de amostra, e é utilizada para suportar a procura conceptual. Ao aplicar este algoritmo, uma base de dados de conhecimento aprende mais palavras do que quando o algoritmo LDA é aplicado individualmente.

 Se seleccionar Latent Dirichlet Allocation - detectar agrupamentos em partições, aprender por todas, o algoritmo LDA é aplicado a um conjunto de documentos de amostra (o número de amostras especificado no campo Número de amostras). De seguida, este conjunto base de agrupamentos detectados é aperfeiçoado por mais documentos (o número de documentos especificado no campo Número de documentos agrupados). Ao aplicar este algoritmo, os agrupamentos são aperfeiçoados por um número superior de documentos e uma base de dados de conhecimento aprende mais palavras do que quando o algoritmo LDA é aplicado individualmente.

Restrição: Se configurar a criação de conjuntos de documentos para uma colecção que seja executada num servidor IBM InfoSphere BigInsights, este é o único algoritmo que pode aplicar.

- f. Faça clique em **Iniciar** para iniciar a tarefa de criação de conjuntos de documentos e crie a proposta de agrupamento.
- 4. Opcional: Adicione os agrupamentos e edite o conteúdo dos agrupamentos:
 - a. Na página Tarefas de Criação de Agrupamentos de Documentos, faça clique em **Iniciar** para a tarefa que pretende aperfeiçoar e volte a executar.
 - b. Na página Editar uma Proposta de Agrupamento:
 - Faça clique em **Adicionar um Agrupamento** para adicionar uma fila à lista de agrupamentos na proposta. É possível especificar palavras adicionais que pretende utilizar para categorizar documentos.
 - Adicione e remova palavras candidatas. Para remover uma palavra do agrupamento, seleccione a palavra da lista e faça clique em **Remover**. Para adicionar uma palavra ao agrupamento, introduza a palavra no campo facultado e faça clique em **Adicionar uma Palavra**.
 - c. Faça clique em **OK** para aplicar as suas alterações à proposta de agrupamento.
- 5. Opcional: Alterar o nome de agrupamentos e remover agrupamentos:
 - a. Na página Tarefas de Criação de Agrupamentos de Documentos, faça clique em **Editar** para a tarefa que pretende modificar.
 - b. Na página Editar uma Proposta de Agrupamento:
 - Mude o nome de quaisquer agrupamentos a que pretenda alterar o nome.
 - Faça clique em **Eliminar** para quaisquer agrupamentos que pretenda remover da proposta de agrupamento.
 - **c.** Faça clique em **OK** para aplicar as suas alterações à proposta de agrupamento.
- 6. Configure a implementação do agrupamento:
 - a. No painel Analisar e indexar, clique em **Configurar** > **Processamento global** > **Implementação de Conjunto**.
 - b. Especifique como pretende aplicar a proposta de agrupamento para categorizar documentos no índice:

- Insira uma etiqueta para a categoria que deve ser apresentada no painel de navegação da faceta do prospecção de análise de conteúdo ou da aplicação de procura empresarial.
- Seleccione a proposta de agrupamento que pretende implementar.
- Seleccione a política que pretende utilizar para categorizar documentos em conjuntos. Pode aplicar as palavras no agrupamento mais relevante como metadados a documentos no índice, aplicar palavras de agrupamentos com classificações de relevância acima de um limiar especificado pelo utilizador ou aplicar palavras do agrupamento mais relevante acima de um limiar especificado pelo utilizador. Caso um documento não corresponda aos critérios especificados, este não é categorizado.
- 7. Opcional: Adicione as classificações do agrupamento ao índice para que possam ser utilizadas para classificação com base em categorias ou procura conceptual. No painel Procura ou no painel Análise para a sua colecção, faça clique em Configurar > Opções do servidor de procura e certifique-se de que as opções para as classificações de categoria que influenciam os resultados de procura estão activadas.
- 8. Inicie a tarefa de implementação de agrupamento para categorizar documentos no índice. Expanda **Processos Globais** no painel Analisar e Indexar, adicione e faça clique em **Iniciar** para iniciar a tarefa de **Implementação de agrupamento**. É apresentado o progresso da tarefa.

Conceitos relacionados:

"Planos de decisão e pontuações por categoria do Content Classification" na página 504

Tarefas relacionadas:

"Configurar os campos de índice e pontuações do Content Classification" na página 506

Anotadores

Os anotadores podem extrair conceitos, palavras, frases, classificações e entidades designadas de conteúdo não estruturado e marcar essas extracções como anotações. As anotações são adicionadas ao índice como tokens ou facetas e são utilizadas como a origem da análise de conteúdo.

Para que o conteúdo seja analisado por anotadores, deve activar os anotadores que desejar utilizar. Alguns dos anotadores avançados para os análise de texto não estão disponíveis para colecções de procura empresarial. Além disso, nem todos os anotadores suportam dicionários definidos pelo utilizador, regras definidas pelo utilizador ou configurações personalizadas para análise de texto.

Adicionalmente aos anotadores fornecidos com IBM Content Analytics with Enterprise Search, pode configurar cada colecção para utilizar um anotador personalizado.Pode utilizar o Content Analytics Studio como alternativa a desenvolver, manualmente, anotadores com o kit de desenvolvimento de software (SDK) UIMA. O Content Analytics Studio é um componente instalado separadamente do IBM Content Analytics with Enterprise Search.

Os anotadores para colecções de procura empresarial e colecções de análises de conteúdos.

Pode permitir que os seguintes anotadores analisem e anotem o conteúdo não estruturado nas colecções de procura empresarial e de análise de texto:

Anotador de Identificação de Idioma

Determina automaticamente o idioma de um documento de entrada com base no conteúdo do documento e nas opções de idioma que foram configuradas para a colecção e para o pesquisador. Este anotador está sempre activado.

Anotador de Análise Linguística

Analisa os atributos sintácticos básicos do texto de entrada e determina o intervalo de frases e tokens e a parte do discurso que cada token representa. Para colecções de análise de conteúdo, este anotador também analisa a categoria gramatical que cada token representa. Este anotador está sempre activado.

Anotador do Content Classification

Adiciona campos de metadados a documentos baseados em campos e categorias definidos num servidor do IBM Content Classification.

Anotador de reconhecimento da entidade designada

Utiliza as regras predefinidas para extrair nomes de pessoas, localizações e organizações do texto de entrada e correlaciona esta informação com facetas. Os nomes de faceta predefinidos são Pessoa, Localização e Organização, mas pode especificar nomes de faceta diferentes ou correlacionar a informação com facetas planas existentes.

Anotador personalizado

Activa um anotador personalizado que foi desenvolvido para a empresa. Para utilizar um anotador personalizado, tem de adicionar o motor de análise de texto personalizado ao sistema e configurar uma ou mais colecções para que o utilizem.

Anotadores para colecções de análise de conteúdo

Numa colecção de análise de conteúdo, também pode activar os seguintes anotadores:

Anotador de procura de dicionários

Procura palavras definidas pelo utilizador no texto de entrada e marca as palavras com o caminho de faceta associado. O caminho de faceta é utilizado pelo anotador de Correspondência de Padrões.

Anotador de correspondência de padrões

Capta padrões que estão construídos com uma ou mais palavras no texto de entrada. O texto é correlacionado com facetas predefinidas para as partes da categoria gramatical como substantivos e verbos, e padrões de frases como uma sequência de substantivos.

Conceitos relacionados:

"Anotador do Content Classification" na página 503

- Anotadores
- Content Analytics Studio
- 🕒 Anotador de identificação de idioma
- Anotador de análise linguística
- 🖙 Anotador de reconhecimento da entidade designada
- Anotador de procura de dicionários
- Anotador de correspondência de padrões

Tarefas relacionadas:

"Activar anotadores para uma colecção"

"Activar o anotador Voz do cliente" na página 245

Activar anotadores para uma colecção

Quando especificar opções de análise e indexação para uma colecção, pode especificar quais os anotadores que pretende incluir para melhorar a obtenção de informações importantes.

Os anotadores analisam conteúdos e adicionam anotações (categorias) que são indexados com os documentos de origem. As anotações ajudam as aplicações de procura a devolver resultados altamente relevantes. Numa prospecção de análise de conteúdo, as anotações fornecem a base para a análise de conteúdos que permite aos utilizadores explorar o conteúdo em contexto. Por exemplo, um anotador pode reconhecer strings de texto como nomes de empresas.

Para activar os anotadores para uma colecção:

- 1. Na vista Colecções, expanda a colecção que pretende configurar.
- 2. No painel Analisar e Indexar, faça clique em **Configurar** > **Anotadores** para ver quais os anotadores que podem ser activados para a colecção actual.
- **3**. Faça clique na caixa de verificação para cada anotador que pretende utilizar para processar documentos.
- 4. Opcional: Edite quaisquer anotadores que pretende alterar. Por exemplo, pode configurar as facetas que prefere utilizar para o reconhecimento de entidades nomeadas e edite o plano ou servidor de decisão do Content Classification.

Numa colecção de análise de conteúdos, também pode editar os dicionários do utilizador ou editar os ficheiros de regras que são utilizados para correspondência de padrões.

- 5. Se seleccionar a caixa de verificação Anotador Personalizado:
 - a. Adicione o motor de análise de texto personalizada ao sistema. Abra a vista Sistema, faça clique na página Analisar, e seleccione de seguida a opção de configuração de motores de análise de texto.
 - b. Configure a colecção para utilizar o motor de análise de texto personalizada. No painel Analisar e Indexar, faça clique em Configurar > Opções de processamento de texto.
- 6. Para que as alterações se tornem efectivas, reinicie o analisador.

Conceitos relacionados:

"Anotadores" na página 243

"Anotador do Content Classification" na página 503

- Anotadores
- Content Analytics Studio
- 🕒 Anotador de identificação de idioma
- Anotador de análise linguística
- 🖙 Anotador de reconhecimento da entidade designada
- 🕒 Anotador de procura de dicionários
- Anotador de correspondência de padrões

Tarefas relacionadas:

"Activar o anotador Voz do cliente"

Activar o anotador Voz do cliente

O IBM Content Analytics with Enterprise Search fornece uma amostra de um anotador para extrair e indexar os padrões de voz do cliente do texto. O anotador de amostra de Voz do cliente suporta Inglês e Japonês.

Poderá pretender capturar os padrões de voz do cliente se, por exemplo, precisar de avaliar os comentários dos clientes. Ao activar este anotador, poderá explorar as facetas de Voz do cliente no prospecção de análise de conteúdo. Ao identificar rapidamente as preocupações dos clientes, poderá transformar os comentários em informações accionáveis.

Para activar este anotador para uma colecção:

- 1. Inicie sessão como o administrador predefinido do IBM Content Analytics with Enterprise Search no servidor de controlador.
- Edite o ficheiro ES_NODE_ROOT/master_config/collection_ID.indexservice/ category_tree.xml, no qual collection_ID identifica as colecções de análise de conteúdos nas quais pretende activar a análise do padrão de voz do cliente. Altere o atributo visible da categoria _voc para true.
- Copie os ficheiros no directório adic para o directório ES_NODE_ROOT/ master_config/collection_ID.indexservice/resource/adic.
- 4. Copie os ficheiros no directório pattern para o directório ES_NODE_ROOT/ master_config/collection_ID.indexservice/resource/pattern.

5. Na consola administrativa, inicie ou reinicie os serviços de análise e de indexação para a colecção e, em seguida, implemente ou volte a implementar os recursos de análise. Para aplicar a análise do padrão de voz dos clientes em documentos indexados anteriormente, volte a construir o índice a partir da cache do documento.

Conceitos relacionados:

"Anotadores" na página 243

Tarefas relacionadas:

"Activar anotadores para uma colecção" na página 244

Recursos de análise e análise de conteúdo

Quando configurar uma colecção de análise de conteúdos, pode especificar as opções de análise sintáctica, de análise e de indexação de documentos que não é possível especificar numa colecção de procura empresarial.

Recursos de análise

Poderá personalizar a análise do conteúdo de modo a assegurar que sejam extraídos do texto conceitos e relações significativas para a sua empresa. Quando os documentos são processados no canal de análise, são adicionadas as anotações e outros metadados de análise ao índice. Os utilizadores podem utilizar a prospecção de análise de conteúdo para explorar os resultados desta análise aprofundada. Por exemplo, os utilizadores podem comparar frequências, desvios, tendências, e correlações entre diferentes facetas de conteúdo ao longo do tempo.

Quando alterar a árvore da faceta, os dicionários do utilizador ou as regras de análise textual definidas, terá que implementar novamente os recursos da análise textual. Se quiser aplicar os novos recursos de análise (dicionários e regras) em documentos que já estão presentes no índice, terá que reconstruí-lo.

Árvore de Facetas

Pode especificar quais são os campos de índice que devem ser correlacionados para facetas que os utilizadores podem seleccionar quando exploram a colecção com a prospecção de análise de conteúdo. Especifique também se uma faceta deve ser apresentada na prospecção de análise de conteúdo ou utilizada apenas para objectivos de análise.

Dicionários de utilizador personalizados

Pode criar e editar ficheiros de dicionário personalizado, bem como importar um ficheiro de valor separado por vírgulas (CSV Comma-separated value) que contenha as palavras e sinónimos que queira incluir num dicionário. Quando os processos de análise forem executados, as definições nos dicionários são usadas para extrair e anotar texto. Por exemplo, pode incluir terminologia específica da empresa, especificar os termos que devem estar disponíveis como facetas na prospecção de análise de conteúdo, quais os termos que são valores de faceta e aqueles que são sinónimos.

Regras de análise de texto personalizadas

É possível criar e editar ficheiros que especifiquem regras de análise de texto personalizadas para a colecção. Quando os processos de análise de conteúdo são executados, as regras são utilizadas para extrair padrões do texto como, por exemplo, palavras ligadas por relações gramaticais. As regras permitem que sejam extraídos e analisados conceitos, o que ajuda a proporcionar significado ao conteúdo, de uma forma que poderá não ser transmitida pela análise de palavras por si só.

Índice de faceta opcional

Se tiver um grande número de facetas, pode reduzir o consumo de memória armazenando os dados da faceta num índice de faceta em separado.

Definições para a colecção de análise de conteúdos

As opções para a análise sintáctica e para a indexação de conteúdos que se seguem estão disponíveis apenas para as colecções de análise de conteúdos. Algumas desta opções só estão disponíveis depois de criar um índice para a colecção.

Facetas de metadados de documentos automáticos

Ao criar uma colecção, pode seleccionar uma caixa de verificação para criar automaticamente facetas para determinados atributos de documento: Extensão do ficheiro, tamanho do ficheiro e data de última modificação. Pode editar estas facetas tal como edita outras facetas numa colecção e utilizá-las para as finalidades da sua empresa. Por exemplo, pode personalizar as facetas e permitir que os utilizadores da prospecção de análise de conteúdo explorem documentos por tipo de ficheiro, explorem documentos vazios, com menos de 1 KB ou mais do que 1 MB de tamanho e explorem documentos que foram modificados pela última vez há uma semana, no último ano ou há cinco anos.

Termos de facetas de interesse

Ao criar uma colecção, configurar uma análise e opções de índice, ou ao editar as definições de uma colecção, pode permitir que o sistema detecte automaticamente relações entre substantivos e verbos e advérbios próximos no texto. Para ajudar os analistas a identificar áreas que necessitem de investigação, podem ser mostrados como facetas de *Termos de interesse* os termos específicos de um domínio que sejam considerados extremamente significativos através deste tipo de análise na prospecção de análise de conteúdo.

Análise de postura

Quando cria uma colecção ou edita definições da colecção, pode activar o sistema para detectar expressões positivas e negativas no texto. Quando configura opções de análise e índice, especifique palavras e frases para ajudar o analisador a categorizar se a oração transmite postura e, caso transmita, qual é a postura geral. Na prospecção de análise de conteúdo, os analistas podem avaliar a postura enquanto exploram a colecção, consultar expressões positivas e negativas no contexto e consultar como a postura positiva e negativa é alterada ao longo do tempo.

Facetas da vista contextual

Ao configurar as opções de análise e de indexação, pode também configurar vistas que permitem aos utilizadores focar num conteúdo que ocorre num contexto específico. Ao combinar campos que representam um assunto comum para uma vista contextual, os utilizadores só podem explorar os documentos que pertencem a essa vista contextual.

Campos utilizados como facetas de data

Após a adição do conteúdo a uma colecção, pode especificar quais são os campos que devem ser utilizados como facetas da data na prospecção de análise de conteúdo. Ao editar as definições da colecção, é apresentada uma caixa de verificação para cada campo de data paramétrico. Os campos que activar para serem utilizados como facetas de data podem ajudar os utilizadores a explorar a colecção através de análise temporal. Por exemplo, os utilizadores podem analisar as mudanças na frequência e correlação ao longo do tempo com as vistas Série Cronológica, Tendências e Desvios. Nestas vistas os utilizadores podem seleccionar a faceta de data que pretendem utilizar para analisar conteúdo.

Escalas de tempo e escalas de tempo cíclicas a serem analisadas

Ao editar as definições da colecção, pode especificar quais as escalas de

tempo e as escalas de tempo cíclicas devem ser analisadas e indexadas. Os valores que seleccionar estão disponíveis para os analistas nas vistas temporais da prospecção de análise de conteúdo. Por predefinição, estão activadas as escalas de tempo e escalas de tempo cíclicas que se seguem: ano, mês, dia, meses do ano, dias do mês e dias da semana.

Pode especificar que sejam analisadas unidades temporais mais granulares, tais como horas, minutos ou os minutos de uma hora. No entanto, ao analisar o conteúdo por horas ou minutos, o tamanho do índice aumenta consideravelmente. Seleccione as opções de escala de tempo de horas e minutos, apenas se necessitar desse nível de pormenor para analisar a respectiva colecção.

O crescimento do índice depende da quantidade de datas únicas necessitam de ser indexadas para os documentos na colecção. Por exemplo, se milhões de documentos tiverem todos a mesma data, a quantidade de dados indexados é de 365 * 24. Pelo contrário, se 100 documentos tiverem todos datas diferentes, a quantidade de dados indexados é de 365 * 24 * 100.

Facetas para colecções de análise de conteúdo

Numa colecção de análise de conteúdo, selecciona facetas para explorar conteúdo analisado e identificar padrões, tendências e desvios nos dados ao longo do tempo. Determinar quais as facetas que são apresentadas aos analistas e quais os dados que contribuem para cada faceta é uma tarefa de concepção fulcral para uma prospecção de conteúdo bem sucedida.

Procura facetada em colecções de análise de conteúdo

Numa colecção de análise de conteúdo, as facetas derivam de documentos, de vários modos. As facetas podem ser associadas a conteúdo como parte do processamento de documentos e é possível configurar, explicitamente, facetas para a colecção:

- Processamento linguístico automático. Quando o conteúdo não estruturado é adicionado a uma colecção, os anotadores de idioma natural incorporados analisam o texto e associam diferentes partes do texto a facetas definidas pelo sistema. Os exemplos deste tipo de faceta incluem:
 - Parte da categoria gramatical, que tem subfacetas como adjectivos, substantivos e verbos.
 - O componente de expressões, com subfacetas para sequências de substantivos, expressões de adjectivo-substantivo e expressões de substantivo-verbo.
- Campos e metadados:
 - Se criar um campo de índice e activar a procura facetada como atributo de campo, é criada uma faceta com o mesmo nome que o do campo.
 - Quando configura pesquisadores, correlaciona campos de origem e metadados com campos de índice existentes que estão activados para procura facetada. Também é possível correlacionar campos de origem e metadados para criar novos campos de índice. Se activar a procura facetada como atributo de campo, é criada uma faceta com o mesmo nome do campo ou dos metadados.
 - Quando importar ficheiros CSV, é possível correlacionar colunas com campos de índice existentes que estão activados para a procura facetada. Também é

possível correlacionar colunas para criar novos campos de índice. Se activar a procura facetada como atributo de campo, é criada uma faceta com o mesmo nome do que o cabeçalho da coluna.

- Quando configura os elementos XML, é possível correlacionar elementos com campos de índice existentes que estão activados para a procura facetada.
- Quando configura os elementos de metadados HTML, é possível correlacionar elementos com campos de índice existentes que estão activados para a procura facetada.
- Quando adiciona conteúdo a uma colecção utilizando as APIs REST, é possível correlacionar campos e metadados com campos de índice que estão activados para a procura facetada.
- Definição de faceta explícita. Após criar campos de índice, é possível criar a sua árvore de faceta ao correlacionar os campos de índice a facetas específicas. Também é possível criar facetas sem correlacionar campos de índice com estas.
- APIs REST. É possível utilizar APIs para criar facetas para uma colecção.
- Regras e dicionários de análise de texto personalizado. É possível definir regras para associar texto com facetas. Também é possível configurar dicionários para associar palavras e sinónimos a facetas. As regras e os dicionários são aplicados quando é adicionado conteúdo não estruturado a uma colecção.
- Correlacionamento de anotação. Quando é aplicada a lógica do anotador, os valores da faceta são adicionados ao conteúdo como metadados. Quando cria um anotador que pretende exportar do Content Analytics Studio, é possível correlacionar uma função de uma anotação com um campo de índice existente ou criar um novo campo de índice. É possível activar o campo de índice para a procura facetada ou correlacionar o campo de índice com uma faceta.
- Correlacionamento de cas2index. É possível utilizar um ficheiro XML para definir que funções do UIMA correlacionam com os campos de índice. É possível criar este ficheiro XML, ou o mesmo pode ser criado como parte da fase de correlacionamento da anotação. É possível activar os campos de índice para a procura facetada ou correlacionar os campos de índice com facetas.
- Criação automática de facetas para metadados de data. Quando edita as definições da colecção após o conteúdo ser adicionado a uma colecção, é apresentada uma caixa de verificação para cada campo de data paramétrico. Os campos que activar para serem utilizados como facetas de data podem ajudar os utilizadores a explorar a colecção através de análise temporal.
- Criação automática de facetas para metadados de documentos. Quando cria uma colecção de análise de conteúdo, é possível especificar que as facetas devem ser criadas automaticamente para o tamanho do ficheiro, extensão do ficheiro e campos de metadados de data da última modificação. É possível personalizar estas facetas automáticas para que se adeqúem às finalidades da empresa do utilizador, como, por exemplo, permitir aos utilizadores de prospecção de análise de conteúdo explorar documentos que foram modificados pela última vez no mês passado, na semana passado ou no ano passado.

Devido à finalidade das facetas, numa colecção de análise de conteúdo, ser a de ajudar o utilizador a explorar conteúdo analisado, todas as facetas definidas para a colecção estão sempre disponíveis em prospecção de análise de conteúdo, ao contrário de serem dependentes de conteúdo que corresponda às condições de consulta actual.

Configurar a árvore da faceta para colecções de análise de conteúdo

Uma forma de criar facetas para uma colecção de análise de conteúdos é através da configuração da árvore de facetas.

Só um utilizador pode configurar a árvore de facetas de cada vez. Se tentar configurar facetas enquanto outro utilizador estiver a configurar dicionários, será apresentada uma mensagem. Faça clique em **OK** para interromper o outro utilizador e impedi-lo de fazer alterações até que tenha terminado. Faça clique em **Cancelar** para sair e aguardar que o outro utilizador termine, antes de proceder às suas alterações.

Para configurar uma árvore de facetas:

- Na vista Colecções, expanda a colecção que pretende configurar. No painel Análise e Indexação, faça clique em Configurar > Recursos de análise > Árvore de facetas.
- 2. Na página Árvore de facetas, seleccione a faceta que pretende editar ou a faceta onde pretende adicionar uma nova faceta:
 - Seleccione a faceta raiz predefinida, se quiser adicionar a faceta ao nível de topo na árvore de facetas. Tem de seleccionar a faceta raiz para adicionar uma faceta de intervalo decimal ou uma faceta de intervalo de data. Não pode correlacionar campos de índice com a faceta raiz predefinida, nem editar a definição da faceta raiz predefinida, nem tão pouco remover da árvore a faceta de raiz predefinida.
 - Seleccione uma faceta individual na árvore. Pode adicionar uma subfaceta à faceta, editar a definição da faceta, alterar a ordem das subfacetas, correlacionar campos de índice com a faceta ou remover a faceta da árvore. Pode imbricar subfacetas consoante necessário para definir a hierarquia na qual as facetas serão apresentadas na prospecção de análise de conteúdo.
- 3. Adicione uma faceta. Tem de especificar um caminho de faceta, o qual é um ID que permite à faceta ser utilizada programaticamente. Pode também especificar um nome de apresentação para a faceta, com o qual a faceta será apresentada na prospecção de análise de conteúdo. Se não especificar um nome de apresentação, será utilizado o caminho da faceta.

Seleccione o tipo de faceta que pretende adicionar: uma faceta padrão, uma faceta de intervalo decimal ou uma faceta de intervalo de data.

Pode também especificar se a faceta deve ser apresentada na prospecção de análise de conteúdo, bem como se pretende que a contagem da frequência processada para a faceta ascendente inclua a contagem de frequência desta faceta.

Depois de especificar informações sobre a nova faceta ou sub-faceta, clique em **Adicionar** para a adicionar à árvore de facetas.

- 4. Seleccione uma faceta ou subfaceta e edite-a:
 - Pode alterar a estrutura hierárquica da árvore de facetas de dois modos:

Arrastar e largar

Pode seleccionar e arrastar facetas para diferentes locais na árvore. Não pode arrastar a faceta raiz. O sistema faculta indicadores que lhe indicam onde pode largar uma faceta que seleccionou para arrastar. Pode utilizar a tecla Ctrl para seleccionar várias subfacetas da mesma faceta ascendente e arrastá-las ao mesmo tempo para uma nova posição na árvore. Não pode seleccionar e arrastar subfacetas que pertençam a facetas diferentes.

Janela Ordem de facetas

Seleccione a faceta raiz para organizar facetas de nível superior ou seleccione uma faceta para organizar as subfacetas dessa faceta. Ao fazer clique no ícone **Editar** na área **Ordem de facetas**, é apresentada uma lista de todas as subfacetas da faceta seleccionada. Seleccione uma subfaceta e faça clique nos botões devidos para mover a subfaceta para cima, para baixo ou para a primeira ou última posição da árvore.

- Pode adicionar, alterar ou remover campos de índice correlacionados com a faceta ou subfaceta. Ao editar o campo **Correlações de campos**, está a seleccionar os campos de índice de uma lista de campos de índice disponíveis para a colecção. Para uma faceta padrão, pode correlacionar vários campos de índice com a faceta. No caso de uma faceta de intervalo, pode correlacionar um campo de índice com a faceta e a lista de campos de índice disponíveis inclui apenas campos decimais paramétricos ou campos de data paramétricos.
- No caso de uma faceta de intervalo decimal ou uma faceta de intervalo de data, edite o campo **Intervalos** para especificar um ou mais valores numéricos ou de intervalo de data.
- 5. Opcional: Seleccione uma faceta ou subfaceta que já não quer utilizar com as colecções de análise de conteúdos e faça clique em **Remover**.
- 6. Quando terminar de configurar as facetas, faça clique em OK.
- 7. Faça clique em **Recursos de análise** e inicie a tarefa de implementação de recursos para garantir que as suas alterações são implementadas no índice.

A reimplementação de recursos actualiza a análise que é aplicada a novos documentos que sejam adicionados ao índice. É necessária uma reconstrução do índice, se quiser aplicar alterações a documentos existentes.

Conceitos relacionados:

"Facetas para colecções de procura empresarial" na página 179

Tarefas relacionadas:

"Supervisionar recursos de análise" na página 441

"Configurar facetas de intervalo decimal para uma colecção de análise de conteúdos"

"Configurar facetas de intervalo de data para uma colecção de análise de conteúdos" na página 255

Configurar facetas de intervalo decimal para uma colecção de análise de conteúdos

Ao criar facetas que se baseiam num intervalo numérico decimal, permite aos utilizadores explorar colecções limitando a vista a um intervalo de valores numéricos, como registos de hospital para pacientes de determinada idade ou produtos para venda a vários preços.

Se alterar uma faceta de intervalo decimal, a alteração reflecte-se na prospecção de análise de conteúdo após um curto período de tempo. Para disponibilizar de imediato essas alterações aos utilizadores, deve reiniciar os servidores da procura para a colecção.

Para configurar uma faceta de intervalo decimal:

 Na vista Colecções, expanda a colecção que pretende configurar. No painel Análise e Indexação, faça clique em Configurar > Recursos de análise > Árvore de facetas.

- 2. Na página Árvore de facetas, seleccione a faceta raiz.
- **3.** Adicione uma faceta. Tem de especificar um caminho de faceta, o qual é um ID que permite à faceta ser utilizada programaticamente. Pode também especificar um nome de apresentação para a faceta, com o qual a faceta será apresentada na prospecção de análise de conteúdo. Se não especificar um nome de apresentação, será utilizado o caminho da faceta.

Para o tipo de faceta, seleccione faceta de intervalo decimal.

Especifique se a faceta deve ser apresentada na prospecção de análise de conteúdo e se pretende que a contagem de frequência processada para a faceta ascendente inclua a contagem de frequência desta faceta.

Após especificar informações sobre a nova faceta de intervalo, faça clique em **Adicionar** para a adicionar à árvore de facetas.

- 4. Faça clique no ícone **Editar** junto ao campo **Correlações de campo** e seleccione o campo de índice que quer correlacionar com esta faceta. A lista de campos de índice disponíveis inclui apenas campos decimais paramétricos.
- 5. Faça clique no ícone **Editar** junto ao campo **Intervalos** para configurar os intervalos de valores que pretende que esta faceta inclua.
- 6. Na página Editar intervalos, especifique como pretende que as facetas sejam organizadas na área de navegação da faceta da prospecção de análise de conteúdo. Pode escolher se os intervalos que definir para esta faceta serão apresentados mesmo se nenhum documento dos resultados corresponder a um intervalo. Também pode especificar uma ordem de apresentação para os intervalos. Especifique se pretende listar os intervalos com um número mais elevado de correspondências a um nível superior na árvore da faceta ou se pretende listar os intervalos na ordem em que os organizou nesta página.
- Decida onde, na estrutura da faceta, pretende configurar um intervalo. Seleccione Raiz para adicionar um intervalo no nível superior, ou seleccione um intervalo previamente definido para adicionar um intervalo a um subnível, e depois faça clique em Adicionar Intervalo.
- Especifique um nome de apresentação significativo para o intervalo, como 0 a 100 ou Entre 50 e 60. Os intervalos que estão definidos ao mesmo nível têm de ter nomes únicos. O campo do nome do intervalo não pode estar vazio.
- **9**. Seleccione se pretende configurar um intervalo definindo valores ou se pretende criar um intervalo que englobe todos os valores restantes ao mesmo nível que este intervalo e não estejam explicitamente definidos. Por exemplo, se tiver a seguinte estrutura de intervalo:

Intervalo	Sub-intervalo	Definição de intervalo explícita
0 - 1000		0 < Preço ≤ 1000
1000 - 2000		1000 ≤ Preço ≤ 2000
	1500 - 1800	1500 ≤ Preço ≤ 1800
	Valores não explícitos	5

O intervalo criado para todos os valores que não foram explicitamente definidos inclui valores superiores ou iguais a 1,000, inferiores a 1,500 ou superiores a 1,800, e inferiores ou iguais a 2,000.

10. Se criar um intervalo para todos os outros valores que não estejam explicitamente definidos, especifique se pretende que este intervalo inclua documentos que não têm o campo especificado ou documentos nos quais o campo está vazio. Por exemplo, partindo do princípio que a colecção inclui estes documentos:

Documento A: Preço=200 Documento B: Preço=100 Documento C:

O intervalo decimal inclui as seguintes definições:

```
Raiz
preço > 150
Criar um intervalo para todos os outros valores
```

Caso a caixa de verificação **Incluir documentos com valores nulos ou vazios** esteja limpa, o intervalo para todos os outros valores inclui apenas o Documento B porque o valor do Documento A é superior a 150 e o Documento C não possui um campo Preço. Caso esta caixa de verificação esteja seleccionada, o intervalo criado para todos os valores inclui tanto o Documento B como o Documento C.

- 11. Execute os passos seguintes para definir um intervalo para uma faceta de intervalo decimal:
 - a. Seleccione ou escreva o valor mais baixo para o intervalo, como 0 ou 50. Caso pretenda definir um limite mais baixo para o intervalo, seleccione a caixa de verificação do limite mais baixo. Se a caixa de verificação estiver desmarcado, significa que não existe um limite mais baixo para o intervalo. Por exemplo, se a caixa de verificação do limite mais baixo estiver desmarcado para o intervalo Entre 50 e 60, o intervalo também inclui valores inferiores a 50. Se não especificar um limite mais baixo para o intervalo, tem de especificar um limite mais alto.
 - b. Seleccione o operador que define como os valores decimais na colecção serão comparados com o valor de intervalo mais baixo. Por exemplo, caso tenha seleccionado o operador inferior ou igual a (≤) para o limite mais baixo 50, os valores correspondem quando são superiores ou iguais a 50.
 - c. Especifique o valor mais alto para o intervalo, como 100 ou 60. Caso pretenda definir um limite mais alto para o intervalo, seleccione a caixa de verificação de limite mais alto. Se a caixa de verificação estiver desmarcado, significa que não existe limite superior para o intervalo. Por exemplo, se a caixa de verificação do limite mais alto for desmarcado para o intervalo Entre 50 e 60, o intervalo inclui valores que também são mais elevados do que 60. Se não especificar um limite superior para o intervalo, tem de especificar um limite mais baixo.
 - d. Seleccione o operador que define a forma como os valores decimais na colecção serão comparados ao valor de intervalo mais elevado. Por exemplo, se seleccionar o operador menor que ou igual (≤) para o limite mais alto 60, os valores correspondem quando são inferiores ou iguais a 60.
- Repita os passos anteriores para definir intervalos adicionais para esta faceta. Por exemplo, poderá criar uma faceta que inclua os intervalos Abaixo de 20, 20 a 50, 50 a 80 e Acima de 80 para representar todos os intervalos possíveis que os utilizadores queiram procurar.
- **13.** Opcional: Elimine um intervalo que já não pretende usar com esta faceta seleccionando o intervalo e fazendo clique em **Remover Intervalo**.
- 14. Faça clique em **OK** para guardar todos os intervalos que configurou para esta faceta de intervalo decimal.

Tarefas relacionadas:

"Supervisionar recursos de análise" na página 441

"Configurar a árvore da faceta para colecções de análise de conteúdo" na página 251

Configurar facetas de intervalo de data para uma colecção de análise de conteúdos

Ao criar facetas que se baseiam num intervalo de data, vai permitir aos utilizadores explorar colecções limitando a vista a um intervalo de valores de data como, por exemplo, apólices de seguros que foram submetidas hoje, este mês ou durante um período de tempo específico.

Se alterar uma faceta de intervalo de data, a mudança reflecte-se na prospecção de análise de conteúdo após um curto período de tempo. Para disponibilizar de imediato essas alterações aos utilizadores, deve reiniciar os servidores da procura para a colecção.

Para configurar uma faceta do intervalo de data:

- Na vista Colecções, expanda a colecção que pretende configurar. No painel Análise e Indexação, faça clique em Configurar > Recursos de análise > Árvore de facetas.
- 2. Na página Árvore de facetas, seleccione a faceta raiz.
- **3.** Adicione uma faceta. Tem de especificar um caminho de faceta, o qual é um ID que permite à faceta ser utilizada programaticamente. Pode também especificar um nome de apresentação para a faceta, com o qual a faceta será apresentada na prospecção de análise de conteúdo. Se não especificar um nome de apresentação, será utilizado o caminho da faceta.

Para o tipo de faceta, seleccione faceta de intervalo de data.

Especifique se a faceta deve ser apresentada na prospecção de análise de conteúdo e se pretende que a contagem de frequência processada para a faceta ascendente inclua a contagem de frequência desta faceta.

Após especificar informações sobre a nova faceta de intervalo, faça clique em **Adicionar** para a adicionar à árvore de facetas.

- 4. Faça clique no ícone **Editar** junto ao campo **Correlações de campo** e seleccione o campo de índice que quer correlacionar com esta faceta. A lista de campos de índice disponíveis inclui apenas campos de data paramétricos.
- 5. Faça clique no ícone **Editar** junto ao campo **Intervalos** para configurar os intervalos de valores que pretende que esta faceta inclua.
- 6. Na página Editar intervalos, especifique como pretende que as facetas sejam organizadas na área de navegação da faceta da prospecção de análise de conteúdo. Pode escolher se os intervalos que definir para esta faceta serão apresentados mesmo se nenhum documento dos resultados corresponder a um intervalo. Também pode especificar uma ordem de apresentação para os intervalos. Especifique se pretende listar os intervalos com um número mais elevado de correspondências a um nível superior na árvore da faceta ou se pretende listar os intervalos na ordem em que os organizou nesta página.
- Decida onde, na estrutura da faceta, pretende configurar um intervalo. Seleccione Raiz para adicionar um intervalo no nível superior desta faceta de intervalo ou seleccione um intervalo definido anteriormente para adicionar um intervalo num subnível e, em seguida, faça clique em Adicionar intervalo.

- 8. Especifique um nome de apresentação significativo para o intervalo, como Anterior a 2010 ou Este Mês. Os intervalos que estão definidos ao mesmo nível têm de ter nomes únicos. O campo do nome do intervalo não pode estar vazio.
- 9. Seleccione o tipo de intervalo que pretende criar:
 - Configure uma data absoluta para restringir o intervalo para uma data que ocorra numa data específica ou entre duas datas específicas, como de 1/4/2010 a 30/4/2010. Todos os subníveis que definir para um intervalo de data absoluto devem ser igualmente intervalos de datas absolutos.
 - Configure uma data relativa para representar intervalos de tempo relativos à data actual. Por exemplo, poderá querer definir um intervalo de data relativo para localizar documentos que foram publicados há um ano ou documentos para um evento que ocorreu entre dois a quatro dias atrás. Todos os subníveis que define para um intervalo de data relativo têm também de ser intervalos de data relativos.
 - Seleccione intervalos predefinidos para hoje, para esta semana, este mês e para este de modo a facilitar a visualização de resultados relativos para a data actual. Pode colocar estes intervalos predefinidos no nível superior de uma definição da faceta do intervalo de data.
 - Adicione um intervalo que abranja todos os valores restantes ao mesmo nível que este intervalo que não estejam definidos explicitamente.
- 10. Se criar um intervalo para todos os valores restantes ao mesmo nível deste intervalo que não estejam explicitamente definidos, especifique se pretende que este intervalo inclua documentos que não têm o campo especificado ou documentos nos quais o campo está vazio. Caso a caixa de verificação Incluir documentos com valores nulos ou vazios esteja limpa, serão excluídos da faceta aqueles documentos que não contenham campo de data ou que tenham valores vazios para o campo de data.
- 11. Execute os passos seguintes para definir valores absolutos para o intervalo:
 - a. Seleccione ou escreva o valor mais baixo para o intervalo, como 1/4/2010. Caso pretenda definir um limite mais baixo para o intervalo, seleccione a caixa de verificação do limite mais baixo. Se a caixa de verificação estiver desmarcado, significa que não existe um limite mais baixo para o intervalo. Por exemplo, se a caixa de verificação do limite mais baixo estiver desmarcado para o intervalo Entre 1/4/2010 e 30/4/2010, o intervalo inclui valores que também são anteriores a 1/4/2010. Se não especificar um limite mais baixo para o intervalo, tem de especificar um limite mais alto.
 - b. Seleccione o operador que define como os valores de data na colecção serão comparados com o valor de intervalo mais baixo. Por exemplo, se seleccionar o operador inferior a ou igual a (≤) para o limite mais baixo 1/4/2010, os valores correspondem quando são superiores ou iguais a 1/4/2010.
 - c. Especifique o valor mais alto para o intervalo, como por exemplo 30/4/2010. Caso pretenda definir um limite mais alto para o intervalo, seleccione a caixa de verificação de limite mais alto. Se a caixa de verificação estiver desmarcado, significa que não existe limite superior para o intervalo. Por exemplo, se a caixa de verificação do limite mais alto estiver desmarcado para o intervalo Entre 1/4/2010 e 30/4/2010, o intervalo inclui também os valores que são posteriores a 30/4/2010. Se não especificar um limite superior para o intervalo, tem de especificar um limite mais baixo.

- d. Seleccione o operador que define como os valores de data na colecção serão comparados ao valor de intervalo mais alto. Por exemplo, se seleccionar o operador menos que ou igual a (≤) para o limite mais alto 30/4/2010, os valores correspondem quando são inferiores ou iguais a 30/4/2010.
- 12. Execute os passos seguintes para definir valores relativos para o intervalo:
 - a. Seleccione ou escreva as condições que definem o valor mais baixo no intervalo, como por exemplo dois dias antes da data actual ou seis meses após a data actual. Da mesma forma que configura valores de data absoluta, seleccione a caixa de verificação para indicar se pretende definir este valor como o limite mais baixo para o intervalo. Se não especificar um limite mais baixo para o intervalo, tem de especificar um limite mais alto.
 - b. Seleccione o operador que define como os valores de data na colecção serão comparados ao valor do intervalo mais baixo, como inferior a (<), ou inferior ou igual a (≤).
 - c. Seleccione ou escreva as condições que definem o valor mais alto no intervalo, como por exemplo um ano antes da data actual. Da mesma forma que configura valores de data absolutos, seleccione a caixa de verificação para indicar se pretende definir este valor como o limite mais alto para o intervalo. Se não especificar um limite superior para o intervalo, tem de especificar um limite mais baixo.
 - d. Seleccione o operador que define como os valores de data na colecção serão comparados ao valor de intervalo mais alto.
- 13. Repita os passos anteriores para definir intervalos de data adicionais para esta faceta. Por exemplo, pode criar uma faceta que inclua os intervalos Hoje, Este Mês, Este Ano e Há Um Ano para representar todos os intervalos possíveis que os utilizadores possam querer procurar.
- 14. Opcional: Elimine um intervalo que já não pretende mostrar para esta faceta seleccionando o intervalo e fazendo clique em **Remover Intervalo**.
- **15**. Faça clique em **OK** para guardar todos os intervalos que configurou para esta faceta de intervalo de data.

Tarefas relacionadas:

"Supervisionar recursos de análise" na página 441

"Configurar a árvore da faceta para colecções de análise de conteúdo" na página 251

Configurar dicionários do utilizador para colecções de análise de conteúdo

Pode criar dicionários de utilizador para registar palavras e termos equivalentes com facetas. Quando os documentos são analisados, o processamento em linguagem natural extrais palavras e termos equivalentes a partir de dados não estruturados para criar índices estatísticos. Esses índices fornecem uma visão mais aprofundada do conteúdo da colecção.

Só um utilizador pode configurar dicionários de cada vez. Se tentar configurar um dicionário enquanto outro utilizador estiver a configurar dicionários, será apresentada uma mensagem. Faça clique em **OK** para interromper o outro utilizador e impedi-lo de fazer alterações até que tenha terminado. Faça clique em **Cancelar** para sair e aguardar que o outro utilizador termine, antes de proceder às suas alterações.

Para configurar uma colecção para utilizar um dicionário de utilizador personalizado:

- Na vista Colecções, expanda a colecção que pretende configurar. No painel Analisar e Indexar, faça clique em Configurar > Recursos de Análise > Dicionários de utilizador.
- 2. Na página Dicionários de Utilizador, introduza o nome de um ficheiro para o dicionário que pretende criar e faça clique em Criar Dicionário. Para importar palavras e termos equivalentes de um ficheiro de valor separado por vírgulas (CSV), faça clique em Importar Ficheiro CSV. Especifique se pretende actualizar um dicionário ou criar um novo.
- 3. Seleccione um dicionário e faça clique em Editar.
- 4. Seleccione um modo para trabalhar com termos do dicionário:

Árvore de Facetas

Seleccione este modo para correlacionar palavras aplicadas e termos equivalentes a uma faceta. A árvore de facetas que foi configurada para a colecção é apresentada. Seleccione uma faceta, seleccione as caixas de verificação de uma ou mais palavras na lista **Palavras Aplicadas** e faça clique no botão de seta para a direita [>>] para registar as palavras seleccionadas com a faceta seleccionada. Palavras aplicadas são as palavras e termos equivalentes podem ser extraídos e analisados estatisticamente.

Se não estiver configurada uma árvore de facetas para esta colecção, poderá correlacionar palavras com uma faceta predefinida chamada As Minhas Palavras.

Candidatas

Seleccione este modo para adicionar palavras candidatas ao dicionário ou para especificar que as palavras aplicadas não devem ser utilizadas na análise de texto. As palavras que adiciona são candidatas a tornarem-se palavras aplicadas, mas não são aplicadas quando os documentos são analisados.

Para começar a utilizar uma palavra candidata para análise de texto, seleccione a palavra e faça clique no botão seta para a direita [>>] para mover a palavra para a lista das palavras aplicadas e, depois, registe a palavra com uma faceta. Se já não quiser utilizar uma palavra aplicada na análise de texto, seleccione-a e clique no botão de seta para a esquerda [<<] para mover a palavra para a lista de palavras candidatas.

- 5. Configure a sua terminologia personalizada na área Palavras Aplicadas:
 - Faça clique em **Adicionar Palavras** para adicionar palavras e palavras compostas ao dicionário para que possa correlacioná-las com facetas. Introduza uma palavra ou palavra composta, tal como cruise control, por linha.
 - Sobre Termos Equivalentes, faça clique no ícone para adicionar e remover termos que são conceptualmente equivalentes ou sinónimos de uma palavra. Por exemplo, vidro reflector pode ser um termo equivalente para a palavra espelho.
 - Sobre Faceta, faça clique no ícone para eliminar o registo entre uma palavra e uma faceta, se já não quiser documentos que contenham a dita palavra, quando os utilizadores exploram a faceta.
 - Sobre **Opções de Palavra**, faça clique para definir formas gramaticais adicionais para uma palavra ou termos equivalentes, tais como acrescentar um plural ou forma do conjuntivo. Pode, também, especificar como quer que

a faceta seja apresentada aos utilizadores no prospecção de análise de conteúdo. Se seleccionar a caixa de verificação para mostrar a faceta, especifique uma etiqueta para a faceta.

Seleccione a caixa de verificação **Ignorar maiúsculas e minúsculas** se quiser permitir que ocorram correspondências de palavras independentemente de maiúsculas e minúsculas. Se esta caixa de verificação não for seleccionada, significa que o sistema deve aplicar uma lógica incorporada para determinar correspondências de palavras, não ocorrendo a correspondência sensível a maiúsculas e minúsculas. Esta caixa de verificação não está disponível para alguns idiomas, tais como Árabe, Chinês e Japonês.

- **6**. Depois de terminar de definir palavras de dicionário e de correlacionar as palavras com facetas, guarde as suas alterações.
- 7. Faça clique em **Recursos de análise** e inicie a tarefa de implementação de recursos para garantir que as suas alterações são implementadas no índice.

A reimplementação de recursos actualiza a análise que é aplicada a novos documentos que sejam adicionados ao índice. É necessária uma reconstrução do índice, se quiser aplicar alterações a documentos existentes.

Conceitos relacionados:

"Termos de facetas de interesse para colecções de análise de conteúdos" na página 261

Personalizar dicionários de utilizador para colecções de análise de texto

Personalizar dicionários de utilizador para colecções de análise de texto

Tarefas relacionadas:

"Supervisionar recursos de análise" na página 441

"Importar um ficheiro CSV para um dicionário de utilizador"

Importar um ficheiro CSV para um dicionário de utilizador

Pode actualizar um dicionário de utilizador ou criar um novo dicionário importando palavras de um ficheiro de valor separado por vírgula (CSV).

Só um utilizador pode configurar dicionários de cada vez. Se tentar configurar um dicionário enquanto outro utilizador estiver a configurar dicionários, será apresentada uma mensagem. Faça clique em **OK** para interromper o outro utilizador e impedi-lo de fazer alterações até que tenha terminado. Faça clique em **Cancelar** para sair e aguardar que o outro utilizador termine, antes de proceder às suas alterações.

Num cenário normal, um utilizador poderá criar um relatório em formato CSV após explorar uma colecção de análise de conteúdo. Este procedimento inclui passos para gerar um ficheiro CSV a partir do prospecção de análise de conteúdo que contém apenas valores de facetas, sem contagens de frequência, valores de correlação ou comentários.

Pode utilizar outros métodos para criar o ficheiro CSV. Caso utilize, tem de se certificar de que o conteúdo do ficheiro CSV corresponde ao seguinte exemplo, em que a primeira palavra de cada linha é registada como termo de base e todas as palavras subsequentes são registadas como termos equivalentes: termo,equivalente1,equivalente2,equivalente3

Caso importe o ficheiro CSV para um dicionário existente e o ficheiro de dicionário destino contenha uma palavra igual à do ficheiro CSV, só são actualizadas os termos equivalentes e correlação de facetas.

Para gerar e importar um ficheiro CSV para um dicionário do utilizador:

- 1. Abra o prospecção de análise de conteúdo, seleccione uma faceta na vista Facetas para explorar os valores que são candidatos para o dicionário do utilizador.
- 2. Faça clique no ícone para criar um relatório e seleccione **Ficheiro CSV** como formato de saída.
- 3. Seleccione **Tabela de valor de faceta** no campo **Conteúdo do ficheiro CSV** e certifique-se de que a caixa de verificação para mostrar informações adicionais como comentários de cabeçalho está limpa. Depois de fazer clique em **Submeter**, é criado um ficheiro CSV que contém apenas valores de facetas na localização que especificar.
- Abra a consola de administração. Na vista Colecções, expanda a colecção que pretende configurar. No painel Analisar e Indexar, faça clique em Configurar > Recursos de análise > Dicionários de utilizador.
- 5. Na página Dicionários de utilizador faça clique em Importar Ficheiro CSV.
- 6. Seleccione um dicionário que pretenda actualizar ou escreva um nome único para um dicionário que pretenda adicionar à colecção. Os dicionários são armazenados como ficheiros *nome_dicionário.*fdic.xml.
- 7. Faça clique em **Procurar** e seleccione o ficheiro CSV que criou ou qualquer ficheiro CSV que contenha as palavras e termos equivalentes que pretende adicionar ao dicionário.
- Seleccione a codificação de caracteres para o conteúdo do ficheiro CSV como, por exemplo, UTF-8.
- **9**. Para correlacionar todas as palavras do ficheiro CSV importado com uma ou mais facetas, seleccione as facetas na árvore de facetas apresentada.
- 10. Faça clique em **OK** para fechar a janela e, de seguida, guarde as alterações.
- 11. Na página Dicionários de utilizador, seleccione o dicionário que adicionou ou actualizou e faça clique em Editar. Edite o dicionário da mesma forma como editaria qualquer dicionário. Por exemplo, pode adicionar ou remover palavras, adicionar e remover termos equivalentes e configurar opções de palavras, tais como definir formas gramaticais adicionais das palavras.
- 12. Faça clique em Recursos de análise e inicie a tarefa de implementação de recursos para garantir que as suas alterações são implementadas no índice. A reimplementação de recursos actualiza a análise que é aplicada a novos documentos que sejam adicionados ao índice. É necessária uma reconstrução do índice, se quiser aplicar alterações a documentos existentes.

Tarefas relacionadas:

"Configurar dicionários do utilizador para colecções de análise de conteúdo" na página 257

Configurar ficheiros de regras para colecções de análise de conteúdo

É possível configurar regras personalizadas para analisar conteúdos. As regras são aplicadas durante o processamento do idioma natural e a anotação resultante é incorporada no índice. Os utilizadores podem explorar os efeitos das regras no prospecção de análise de conteúdo.

Só um utilizador pode configurar regras de cada vez. Se tentar configurar regras enquanto outro utilizador estiver a configurar regras, será apresentada uma mensagem. Faça clique em **OK** para interromper o outro utilizador e impedi-lo de fazer alterações até que tenha terminado. Faça clique em **Cancelar** para sair e aguardar que o outro utilizador termine, antes de proceder às suas alterações.

Para configurar uma colecção para utilizar um ficheiro de regras personalizado:

- Na vista Colecções, expanda a colecção que pretende configurar. No painel Analisar e Indexar, faça clique em Configurar > Recursos de análise > Regras de análise de texto personalizadas.
- 2. Na página Regras de Análise de Texto, seleccione o ficheiro que pretende editar ou introduza novo nome de ficheiro, para um que pretenda criar.
- **3**. Quando o ficheiro de regras for apresentado, escreva, copie, cole ou edite regras para os padrões que pretende extrair destes documentos. Faça clique em **OK** para guardar as alterações.
- 4. Faça clique em **Recursos de análise** e inicie a tarefa de implementação de recursos para garantir que as suas alterações são implementadas no índice.

A reimplementação de recursos actualiza a análise que é aplicada a novos documentos que sejam adicionados ao índice. É necessária uma reconstrução do índice, se quiser aplicar alterações a documentos existentes.

Conceitos relacionados:

Ficheiros de regras personalizados para colecções de análise de conteúdos

Ficheiros de regras personalizadas para colecções de análise de texto

Tarefas relacionadas:

"Supervisionar recursos de análise" na página 441

Indice da faceta opcional para colecções de análise de conteúdo

Para uma colecção de análise de conteúdo, pode permitir que o sistema utilize um índice à parte para armazenar dados de configuração de faceta.

Por exemplo, se tiver um grande número de facetas, pode reduzir o consumo de memória armazenando as facetas num índice à parte. Caso não utilize um índice de facetas, os dados de configuração de facetas são armazenados no índice de documentos.

Pode permitir que o sistema utilize um índice da faceta quando criar uma colecção de análise de conteúdo. Para uma colecção existente, também pode activar esta função ao editar as definições da colecção.

Para criar o índice da faceta opcional, deve iniciar uma construção de índice da faceta. Para fazê-lo, abra a vista Colecções, expanda o painel Analisar e Indexar para uma colecção de análise de conteúdo, expanda a secção **Recursos de Análise** e reconstrua o índice da faceta opcional.

Este índice não é actualizado por incrementos à semelhança de um índice de documentos. Tem de reconstruir periodicamente o índice de facetas para incorporar alterações.

Tarefas relacionadas:

"Supervisionar recursos de análise" na página 441

Termos de facetas de interesse para colecções de análise de conteúdos

Para uma colecção de análise de conteúdo, pode permitir que o sistema detecte automaticamente relações entre substantivos e verbos e advérbios próximos no texto.

Para ajudar os analistas a identificar áreas que necessitam de investigação, termos específicos de um domínio que sejam considerados extremamente significativos podem ser apresentados como facetas de *Termos de interesse* no prospecção de análise de conteúdo. Por exemplo, ao explorar documentos acerca de acidentes de viação, são apresentados termos que normalmente são mencionados em participações como, por exemplo, travão, volante, limpa-vidros, porta, etc.

A identificação de termos de interesse deve-se à proximidade de substantivos, verbos e advérbios no texto. A análise pode ocorrer independentemente do domínio ou do idioma do documento de origem. Por exemplo, determinados advérbios transmitem potenciais problemas independentemente do assunto.

Não é necessário modificar regras de dicionários ou regras de padrão para que estas relações sejam identificadas no texto. Contudo, pode melhorar a análise de conteúdo registando os substantivos que são identificados como termos interessantes num dicionário de utilizador.

Caso não permita que o sistema detecte automaticamente termos de interesse específicos de um domínio, um analista poderá ter de investigar todos os substantivos na faceta Substantivo para identificar os termos significativos. Caso esta função esteja activada, o analista vê instantânea e automaticamente candidatos prováveis para investigar e pode rapidamente identificar substantivos interessantes para adicionar a um dicionário de utilizador.

Os termos de interesse podem ser detectados em qualquer idioma suportado para colecções de análise de conteúdo.

Pode permitir que o sistema identifique automaticamente os termos de interesse quando criar uma colecção de análise de conteúdo. Para uma colecção de análise de conteúdo existente, pode também activar esta função ao editar as definições da colecção.

Tarefas relacionadas:

"Configurar dicionários do utilizador para colecções de análise de conteúdo" na página 257

Referências relacionadas:

Idiomas suportados

Configurar análise de postura para colecções de análise de conteúdo

E possível configurar o analisador para atribuir postura positiva, negativa ou nenhuma postura a expressões extraídas do texto.

Para configurar o analisador para reconhecer a postura, tem, primeiro, de activar a análise de postura. Caso não tenha activado a análise de postura quando a colecção foi criada, é possível activá-la ao editar as definições da colecção.

Quando configura opções de análise e de indexação, especifica palavras e expressões para ajudar o analisador a categorizar se uma frase transmite postura e, se assim for, qual a postura.

No prospecção de análise de conteúdo, os analistas podem avaliar a postura ao explorar facetas e documentos na colecção, ao ver expressões positivas e negativas contextualizadas e ao ver a mudança da postura positiva e negativa ao longo do tempo. Caso um documento não contenha expressões suficientes para transmitir uma postura positiva ou negativa, o documento é classificado como ambivalente.

Os analistas podem limitar os resultados ao adicionar valores da faceta seleccionada à consulta e ao seleccionar um botão para procurar apenas documentos com postura positiva ou apenas documentos com postura negativa. Este tipo de procura não significa que apenas sejam devolvidos documentos do tipo de postura especificada. Se um documento incluir expressões positivas e negativas, o documento é incluído na procura e ambos os tipos de expressões nesse documento podem ser devolvidas nos resultados.

Para configurar a análise de postura para uma colecção de análise de conteúdo:

- 1. Na vista Colecções, expanda a colecção que pretende configurar. No painel Análise e Indexação, clique em **Configurar** > **Análise de postura**.
- Na página Configuração de Análise de Postura, faça clique no ícone de editar para o idioma a ser aplicado quando analisar a postura do conteúdo. Os idiomas suportados são inglês e japonês.
- **3**. Especifique, no mínimo, uma expressão. Se especificar mais do que uma expressão de um determinado tipo, introduza cada expressão em linhas separadas:
 - Especifique as palavras e expressões que devem ser reconhecidas como expressões positivas quando o conteúdo é analisado.
 - Especifique as palavras e expressões que devem ser reconhecidas como expressões negativas quando o conteúdo é analisado.
 - Especifique as palavras e expressões que não devem ser reconhecidas como expressões de postura quando o conteúdo é analisado.
- 4. Opcional: Repita estes passos para especificar expressões para um idioma diferente. A página Configuração de Análise de Postura mostra o número de expressões positivas, expressões negativas e expressões bloqueadas que estão definidas para cada idioma.
- 5. Para aplicar as alterações, implemente novamente os recursos de análise e reconstrua o índice. Se a memória cache de um documento não estiver activada para a colecção, implemente os recursos de análise e volte a pesquisar ou volte a importar os documentos.

Tarefas relacionadas:

"Supervisionar recursos de análise" na página 441

Configurar vistas contextuais para colecções de análise de conteúdo

Para criar uma vista contextual, deve configurar o sistema para analisar campos específicos numa colecção de análise de conteúdo. Quando os utilizadores procuram na colecção, podem escolher explorar apenas as secções de documentos que pertencem à vista contextual.

Todos os campos que inclui numa vista contextual devem ser analisados e activados para serem mostrados nos resumos dos documentos. Quando configura campos de origem, campos de metadados, campos de índice e opções para importar ficheiros CSV, especifica atributos para como os campos podem ser procurados. Seleccione a caixa de verificação **Analisável** e a caixa de verificação **Em resumo** para cada campo que quer incluir numa vista contextual.

Vistas contextuais permitem aos utilizadores focarem-se no conteúdo que ocorre num contexto particular. Por exemplo, a expressão *componente de substituição* pode ter várias conotações dependendo se ocorrer num campo que contém queixas dos clientes ou num campo que contém resoluções. Se combinar campos que representam um assunto comum numa vista contextual, facilmente os utilizadores podem procurar os campos que pertencem a essa vista contextual. Também podem escolher ver apenas esses campos quando os documentos de procura resumem os resultados.

No prospecção de análise de conteúdo, os utilizadores podem especificar quais as vistas contextuais que pretendem procurar. Também podem especificar preferências sobre como os resumos dos documentos são apresentados, tal como se os resumos incluem a identificação da vista contextual ou se apresentam apenas o texto. Utilizadores autorizados podem também seleccionar as vistas contextuais que pretendem procurar quando utilizam o construtor da consulta para construir a consulta.

Importante: Se configurar as vistas contextuais, será necessário mais tempo para construir as consultas de índice e processo.

Para configurar vistas contextuais para que os utilizadores se possam focar em partes específicas dos documentos quando explorar a colecção de análise de conteúdo:

- 1. Na vista Colecções, expanda a colecção que pretende configurar. No painel Analisar e Indexar, faça clique em **Configurar** > **Vistas contextuais**.
- 2. Na página Configurar Vistas Contextuais, faça clique no botão para adicionar uma nova vista contextual.
- **3.** Especifique um identificador único para a vista. Pode especificar uma etiqueta de apresentação para a vista. Se existir uma etiqueta, os utilizadores vêem a etiqueta quando exploram a colecção no prospecção de análise de conteúdo. Caso contrário, os utilizadores vêem o ID da vista contextual.
- 4. Seleccione os campos que pretende incluir na vista contextual e fala clique em Adicionar. A lista de campos disponíveis incluem todos os campos de origem, campos de metadados e campos de índice que podem ser analisados e mostrados em resumos de documentos.
- 5. Se adicionar mais do que uma vista contextual, utilize os botões na página Configurar Vistas Contextuais para mover as vistas para a ordem em que serão apresentadas no prospecção de análise de conteúdo.
- 6. Quando terminar a configuração das vistas contextuais, reconstrua o índice para aplicar as alterações aos documentos existentes.

Detecção do formato do documento

É utilizada uma correlação predefinida da extensão de URL e o tipo MIME é utilizado para determinar tipos de documentos e o tipo de analisador a utilizar com cada documento.

Ao editar o ficheiro parser_config.xml e o ficheiro mimetypes.xml, pode sobrepor e expandir a correlação predefinida de extensões de URL e tipos MIME para tipos de analisador. Estes ficheiros definem regras para correlação de extensões de ficheiros ou tipos MIME para tipos de analisador. Por exemplo, pode correlacionar uma extensão de ficheiro como, por exemplo, .content, e especificar que os documentos desse tipo deverão ser analisados pelo analisador de HTML.

Formatos de documento diferentes têm representações internas diferentes. O sistema IBM Content Analytics with Enterprise Search utiliza filtros internos e de terceiros para analisar documentos e muitos documentos são analisados com serviços de analisador especializados para um determinado formato.

A detecção do formato do documento e a atribuição do analisador ocorrem do seguinte modo:

- O algoritmo para detecção do formato do documento verifica a extensão do URL do documento processado. A extensão do nome do ficheiro, que faz parte dos metadados definidos pelo pesquisador, também é considerada ao detectar o formato do documento.
- 2. O sistema verifica o tipo MIME do documento, que faz parte dos metadados definidos pelo pesquisador.
- **3**. O sistema tenta atribuir o tipo de analisador correcto a cada documento. Para documentos HTML, texto (TXT) e XML, o sistema atribui um tipo de analisador que é específico para cada formato de documento.

Para outros formatos de documentos, o sistema utiliza o extractor de texto. A tecnologia de filtragem de documentos do extractor de texto é baseada na tecnologia do Oracle Outside In Content Access. Esta tecnologia adquiriu, anteriormente, o nome de Stellent, e os nomes de alguns ficheiros de configuração incluem o termo Stellent.

O extractor de texto suporta várias centenas de formatos de documentos, mas apenas um subconjunto de filtros de documentos são activados no IBM Content Analytics with Enterprise Search. Pode, contudo, editar ficheiros de configuração para permitir outros tipos de documentos analisados pelo extractor de texto.

Importante: Os filtros de documentos que adiciona e que não pertencem ao subconjunto de filtros de documentos que estão activados na configuração do sistema predefinido não foram testados e não são suportados.

Se o sistema não conseguir identificar o formato de um documento, o documento é rejeitado. Pode ver uma mensagem de erro que indica que o tipo do documento não é suportado.

Para determinar o tipo de documento e o tipo de analisador, o sistema executa os passos seguintes:

a. Compara o nome do ficheiro para as regras no ficheiro mimetypes.xml. Se o nome do ficheiro não for especificado, o sistema compara a extensão URL (a extensão da ID do documento) às regras.

- b. Compara o tipo MIME com as regras no ficheiro mimetypes.xml para obter o tipo normalizado.
- c. Compara o tipo de documento (tipo MIME normalizado) às regras no ficheiro parser_config.xml.

Se o tipo de analisador for stellent, poderá ver uma mensagem de erro se o extractor de texto não reconhecer o formato do documento. O erro pode ocorrer se:

- O documento tiver erros.
- O documento não estiver num formato que o extractor de texto suporte. Para solucionar este problema, tem de adicionar os formatos de ficheiro rejeitados ao ficheiro stellentTypes.cfg. É também necessário actualizar o ficheiro mimetypes.xml ou o ficheiro parser_config.xml para especificar que o tipo ou extensão MIME dos formatos de documentos rejeitados devem ser associados com o extractor de texto.

Conceitos relacionados:

"Tipos de documentos associados com analisadores de colecção e extractor de texto" na página 267

Tarefas relacionadas:

"Associar tipos de documentos ao extractor de texto" na página 269

- "Configurar sessões múltiplas para o extractor de texto" na página 275
- "Analisar texto bidireccional em ficheiros PDF" na página 275

"Indexação de documentos .docm com URLs que incluam consultas" na página 276 **Referências relacionadas**:

"Tipos de documentos suportados por predefinição"

Tipos de documentos suportados por predefinição

Ao detectar o formato do documento, só são avaliados certos tipos de documentos.

Os formatos de documento seguintes são detectados e analisados automaticamente por serviços do analisador de colecção incorporados:

HTML Texto normal XML

Por predefinição, os formatos de documento que se seguem são analisados pelo extractor de texto:

Adobe Portable Document Format (PDF) Lotus 1-2-3 Lotus Freelance Graphics Lotus Word Pro Apenas Sistema Ichitaro Microsoft Excel (versões até 2010) Microsoft PowerPoint (versões até 2010) Microsoft Visio Microsoft Word (versões até 2010) Rich Text Format (RTF) StarOffice/OpenOffice Calc StarOffice/OpenOffice Impress StarOffice/OpenOffice Draw StarOffice/OpenOffice Draw Os formatos de ficheiro Office Open XML em Microsoft Office e formatos OpenDocument em OpenOffice.org são tratados sem ser necessário efectuar alterações aos ficheiros de configuração.

Para analisar outros tipos de documentos, tem de actualizar ficheiros de configuração (parser_config.xml, mimetypes.xml, e stellenttypes.cfg) para especificar regras para correlacionar tipos específicos de documentos para um serviço de analisador da colecção ou extractor de texto.

Conceitos relacionados:

"Detecção do formato do documento" na página 265

Tarefas relacionadas:

"Associar tipos de documentos ao extractor de texto" na página 269

Tipos de documentos associados com analisadores de colecção e extractor de texto

Para assegurar que os documentos num espaço de pesquisa são analisados com eficácia e precisão, pode criar ficheiros de configuração para especificar que tipos de documentos serão analisados pelo analisador da colecção ou quais serão analisados pelos filtros de documentos de extracção de texto.

A maioria dos formatos de documentos são processados por analisadores HTML ou XML incorporados. certos tipos de documentos são tipicamente não analisados (como os documentos Postscript), e outros tipos de documentos são processados por funções de extracção de texto (como os tipos de documento do Microsoft Word, Microsoft Excel, Microsoft PowerPoint, Lotus Freelance, Lotus 123, PDF, RT, e Ichitaro).

Porque os metadados podem induzir em erro, os documentos em texto simples e HTML poderão ser enviados erroneamente para o extractor de texto e depois devolvidos a um dos analisadores incorporados, uma situação que pode afectar o desempenho. Para outros documentos, pode não ser possível detectar o tipo de documento, de modo que os documentos são ignorados. Para evitar esta situação, pode criar ficheiros de configuração para controlar onde e quão diferente são analisados os tipos de documentos.

Associar os tipos de documentos com o analisador de colecção e o extractor de texto envolve as seguintes tarefas:

- Configurar tipos de documentos para o analisador da colecção. Este passo envolve a modificar os ficheiros de configuração que correlacionam os tipos de documentos com o analisador que é utilizado por uma colecção. O ficheiro parser_config.xml e o ficheiro mimetypes.xml são criados quando criar uma colecção com regras predefinidas.
- Configurar tipos de documentos para o extractor de texto. Este passo envolve criar um ficheiro de configuração que correlaciona os tipos de documentos com os filtros documentos de extracção de texto utilizados por uma colecção. Pode criar um desses ficheiros de configuração por colecção.
- **3**. Parar e reiniciar o analisador. Para que as alterações tenham efeito, utilize a consola administrativa para parar e reiniciar o analisador para a colecção.

Conceitos relacionados:

"Detecção do formato do documento" na página 265

Mapeamento do tipo MIME e do tipo de analisador

Para associar tipos de documentos em particular a um analisador de colecções, tem de modificar o ficheiro mimetypes.xml e o ficheiro parser_config.xml. Não existe suporte para esta tarefa na consola de administração.

Extensões de correlação para tipos MIME

O ficheiro ES_NODE_ROOT/master_config/collection_ID.indexservice/ mimetypes.xml define a correlação da extensão de URL ou a extensão do nome do ficheiro para tipos MIME. Este ficheiro tem uma ou mais regras para correlacionar extensões para um tipo MIME no elemento /Mimetypes/ExtensionMapping. No exemplo seguinte, a extensão .doc é correlacionada para o tipo MIME application/msword:

```
<MappedMimetype Name="application/msword">
<Extension>.doc</Extension>
</MappedMimetype>
```

O exemplo seguinte mostra várias extensões que estão todas correlacionadas para o tipo MIME application/vnd.lotus-wordpro:

```
<MappedMimetype Name="application/vnd.lotus-wordpro">
<Extension>.lwp</Extension>
<Extension>.mwp</Extension>
<Extension>.sty</Extension>
<Extension>.sam</Extension>
</MappedMimetype>
```

Correlacionar tipos MIME para tipos de analisador

O ficheiro ES_NODE_ROOT/master_config/collection_ID.indexservice/ parser_config.xml define a correlação dos tipos MIME para tipos de analisador. Este ficheiro tem uma ou mais regras para correlacionar tipos MIME para um tipo de analisador no elemento /ParserService. No seguinte exemplo, os tipos MIME application/msword e application/x-msexcel são correlacionados com o extractor de texto (tipo de analisador stellent):

```
<ParserMapping>
<ParserName>stellent</ParserName>
<Mimetype>application/x-msexcel</Mimetype>
<Mimetype>application/msword</Mimetype>
</ParserMapping>
```

Pode correlacionar tipos MIME para os seguintes tipos de analisador:

- html Os documentos são analisados pelo analisador de HTML.
- text Os documentos são analisados pelo analisador Text.
- xml Os documentos são analisados pelo analisador XML.

stellent

Os documentos são analisados pelo extractor de texto.

terminador

Os documentos são rejeitados pelo analisador.

Combinando os exemplos do tipo MIME e do tipo analisador, os documentos que têm a extensão .doc são analisados pelo extractor de texto porque esta extensão

está correlacionada com o tipo MIME application/msword, que, por sua vez, está correlacionado com o tipo de analisador stellent.

Analisar tipos de documentos desconhecidos

Se um tipo de documento é desconhecido (por exemplo, se um documento não tiver uma extensão do nome do ficheiro ou não houver um tipo MIME associado ao documento), pode configurar regras para impedir o analisador de deixar cair o documento.

Se o analisador não reconhecer um formato de ficheiro, o analisador tenta analisar o documento com o analisador de HTML predefinido. Se um conteúdo não estiver em formato HTML, o analisador rejeita o documento. O documento rejeitado é indexado com um código de razão.

Para evitar esta situação, configure o analisador para utilizar o analisador de ASCII facultado para tipos de documentos desconhecidos:

- 1. Inicie sessão como o administrador predefinido do IBM Content Analytics with Enterprise Search no servidor de controlador.
- Edite o ficheiro de configuração do analisador, em que collection_ID identifica a colecção que desejar configurar:

ES_NODE_ROOT/master_config/collection_ID.indexservice/parser_config.xml

3. Altere o analisador denominado como o analisador predefinido para o analisador de texto incorporado. Por exemplo, mude <ParserName>terminator</ ParserName> para <ParserName>text</ParserName>:

```
<DefaultParserMapping>
<ParserName>terminator</ParserName>
</DefaultParserMapping>
```

<DefaultParserMapping> <ParserName>text</ParserName> </DefaultParserMapping>

4. Para que as alterações tenham efeito, utilize a consola administrativa para parar e reiniciar o analisador para a colecção.

Associar tipos de documentos ao extractor de texto

Para especificar quais os tipos de documentos que serão analisados pelos filtros de documentos do extractor de texto, pode criar um ficheiro de configuração stellenttypes.cfg. Não existe suporte para esta tarefa na consola de administração.

O extractor de texto suporta filtros para os seguintes tipos de documentos: Adobe Portable Document Format (PDF) Lotus 1-2-3 Lotus Freelance Graphics Lotus Word Pro Ichitaro Microsoft Excel Microsoft PowerPoint Microsoft Visio Microsoft Word Formato de Rich Text (RTF, Rich Text Format) A tecnologia de filtragem de documentos do extractor de texto é baseada na tecnologia do Oracle Outside In Content Access. Para uma lista completa dos tipos de documentos que podem ser filtrados, consulte *Outside In Technology: Supported File Formats* no sítio da Oracle.

Caso pretenda incluir tipos de documentos adicionais num índice, e os tipos de documentos sejam suportados pelos filtros do Oracle Outside In Content Access, tem de configurar regras de análise no ficheiro de configuração stellenttypes.cfg.

Importante: Os filtros de documentos que adiciona e que não pertencem ao subconjunto de filtros de documentos que estão activados na configuração do sistema predefinido não foram testados e não são suportados.

O ficheiro de configuração stellenttypes.cfg especifica:

- Aceite as regras para os tipos de documentos que irão ser analisados pelo extractor de texto. Um tipo de documento corresponde a um dos tipos reconhecidos pela biblioteca do Oracle Outside In Content Access.
- Regras nativas para os tipos de documentos que irão ser enviados de volta para o analisador da colecção, para processamento com um dos analisadores incorporados. Esta acção é necessária, uma vez que o analisador de colecções pode enviar um documento para o extractor de texto com erros, devido a metadados incorrectos.
- Regras de rejeição para os tipos de documentos que irão ser rejeitados por não serem suportados.

Se o ficheiro de configuração stellenttypes.cfg não existir no directório ES_NODE_ROOT/master_config/*ID_colecção*.stellent/, serão utilizadas as regras de análise predefinidas para o extractor de texto.

O ficheiro de configuração lista tipos de documentos e o modo como são processados. O formato do ficheiro é uma sequência de linhas, em que cada linha é uma regra que corresponde a um dos seguintes formatos:

accept DEFAULT
accept ALL doc_type
accept outsideIn_type doc_type
native DEFAULT
native outsideIn_type doctype
reject outsideIn_type

doc_type

Trata-se do valor a ser utilizado para a categoria de consulta de tipo doc. É possível procurar documentos por tipo de documento. Por exemplo, um utilizador poderá especificar \$doctype::pdf para procurar documentos PDF.

outsideIn_type

É um dos valores do tipo de filtro na biblioteca do Oracle Outside In Content Access, tal como o FI_123R1.

DEFAULT

Significa que a lista de tipos aceites ou nativos, dependendo do tipo de regra, inclui todas as regras predefinidas. Esta opção permite expandir a configuração predefinida, em vez de substituí-la.

All Significa que todos os tipos que não se encontrem explicitamente listados são aceites com a categoria de tipo de documento especificado.

As regras no ficheiro de configuração são processadas do seguinte modo:

- 1. Se existir uma regra reject para *outsideIn_type*, o documento não é aceite.
- 2. Se existir uma regra native para *outsideIn_type* (incluindo a regra de análise predefinida se for especificado native DEFAULT), o documento é enviado novamente para o analisador incorporado para além do valor para o token *doc_type* que é especificado por esta regra. O valor de *doc_type* tem de ser txt, htm ou xml, indicando texto simples, HTML ou XML, respectivamente.
- **3**. Se existir uma regra accept para *outsideIn_type* (incluindo a lista predefinida se for especificado accept DEFAULT), o documento é aceite.
- 4. Caso contrário, se for especificado accept ALL, o documento é aceite.
- 5. Caso contrário, o documento é rejeitado e não será analisado.

Se o tipo de documento for aceite, é utilizado o valor *doc_type* especificado na regra aplicada. Este valor é devolvido ao analisador de colecções juntamente com o conteúdo analisado.

Para associar tipos de documentos ao extractor de texto:

- 1. Inicie sessão como o administrador predefinido do IBM Content Analytics with Enterprise Search no servidor de controlador.
- Utilize um editor de texto para criar o ficheiro ES_NODE_ROOT/master_config/ ID_colecção.stellent/stellenttypes.cfg, no qual ID_colecção identifica a colecção que pretende configurar. Especifique as regras do extractor de texto, em seguida, guarde e saia do ficheiro.
- Para que as alterações tenham efeito, utilize a consola administrativa para parar e reiniciar o analisador para a colecção.

Exemplos

No ficheiro de configuração seguinte, a sessão do extractor de texto aceita documentos no formato Microsoft Visio, para além da lista predefinida de tipos de documentos suportados.

accept DEFAULT accept FI_VISIO3 visio accept FI_VISIO4 visio accept FI_VISIO5 visio accept FI_VISIO6 visio

No ficheiro de configuração seguinte, os documentos Postscript são aceites e pesquisáveis com um tipo do documento de ps, os documentos em formato de mapa de píxeis X (XPM) são devolvidos ao analisador de texto incorporado, os documentos no formato de imagem PNG são rejeitados e todos os outros de ficheiros são aceites e tornados pesquisáveis com um tipo de documento de outro.

accept DEFAULT accept FI_POSTSCRIPT ps native FI_XPIXMAP txt reject FI_PNG accept ALL other

Conceitos relacionados:

"Correlacionamentos com campo de índice para metadados extraídos de conteúdo binário" na página 212

"Detecção do formato do documento" na página 265

Tarefas relacionadas:

"Configurar sessões múltiplas para o extractor de texto" na página 275

Referências relacionadas:

"Tipos de documentos suportados por predefinição" na página 266

"Regras de análise predefinidas para o extractor de texto"

Regras de análise predefinidas para o extractor de texto

Se não criar um ficheiro de configuração para mapear tipos de ficheiros para filtros de documentos do extractor de texto, o analisador utiliza regras predefinidas para analisar documentos.

O extractor de texto aceita e analisa os seguintes tipos de documentos:

ACCEPT	FI 123R1	123
ACCEPT	FT_123R2	123
ACCEPT	FI 123R3	123
ACCEPT	FT 123R4	123
ACCEPT	FT 123R6	123
ACCEPT	FT 123R9	123
ACCEPT		vls
ACCEPT		xlsx
ACCEPT		xls
ACCEPT		vls
ACCEPT		xls
ACCEPT		xls
ACCEPT	FI FXTPOWERPOINT4	nnt
ACCEPT		nnt
ACCEPT		ppt nrz
ACCEPT	FT_FREELANCE3	nrz
ACCEPT		ixw
ACCEPT		isw
ACCEPT		itd
ACCEPT		ndf
ACCEPT		ndf
ACCEPT	FI POWERPOINT2	nnt
ACCEPT		nnt
ACCEPT	FI POWERPOINT2007	nntx
ACCEPT		nnt
ACCEPT		nnt
ACCEPT	FI POWERPOINT7	nnt
ACCEPT		nnt
ACCEPT	FT_POWERPOINT97	nnt
ACCEPT	FI POWERPOINTMAC3	nnt
ACCEPT	FT_POWERPOINTMAC4	nnt
ACCEPT	FI POWERPOINTMACB3	nnt
ACCEPT		nnt
ACCEPT	FI RTF	rtf
ACCEPT	FT RTF.1	rtf
ACCEPT	FI STAROFFICEWRITER8	odt
ACCEPT	FI STAROFFICEDRAW8	oda
ACCEPT	FI STAROFFICEIMPRESS8	odp
ACCEPT	FI STAROFFICECALC8	ods
ACCEPT	FI_STAROFFICECALC6	SXC
ACCEPT	FI_STAROFFICEDRAW6	sxd
ACCEPT	FI STAROFFICFIMPRESS6	sxi
		<i></i>

3/11
sdc
sdd
sdw
vsd
doc
docx
doc
lwp
lwp

O extractor de texto devolve os seguintes tipos de documentos ao analisador da colecção para processamento com um dos analisadores incorporados:

NATIVE	FI 7BITTEXT	txt
NATIVE	FI ANSI	txt
NATIVE	FI ANSI8	txt
NATIVE	FI ARABIC 710	txt
NATIVE	FI ARABIC 720	txt
NATIVE	FI ARABIC WINDOWS	txt
NATIVE	FI [_] ASCII [_]	txt
NATIVE	FI ASCII8	txt
NATIVE	FI CENTRALEU 1250	txt
NATIVE	FI ⁻ CHINESEBIG5	txt
NATIVE	FI CHINESEGB	txt
NATIVE	FI_CYRILLIC1251	txt
NATIVE	FI_CYRILLICKOI8	txt
NATIVE	FI_EBCDIC_1026	txt
NATIVE	FI_EBCDIC_273	txt
NATIVE	FI_EBCDIC_277	txt
NATIVE	FI_EBCDIC_278	txt
NATIVE	FI_EBCDIC_280	txt
NATIVE	FI_EBCDIC_284	txt
NATIVE	FI_EBCDIC_285	txt
NATIVE	FI_EBCDIC_297	txt
NATIVE	FI_EBCDIC_37	txt
NATIVE	FI_EBCDIC_500	txt
NATIVE	FI_EBCDIC_870	txt
NATIVE	FI_EBCDIC_871	txt
NATIVE	FI_HANGEUL	txt
NATIVE	FI_HEBREW_E0	txt
NATIVE	FI_HEBREW_OLDCODE	txt
NATIVE	FI_HEBREW_PC8	txt
NATIVE	FI_HEBREW_WINDOWS	txt
NATIVE	FI_HTML	htm
NATIVE	FI_HTML_ARABIC_ASM0708	htm
NATIVE	FI_HTML_ARABIC_DOS	htm
NATIVE	FI_HTML_ARABIC_ISO	htm
NATIVE	FI_HTML_ARABIC_MAC	htm
NATIVE	FI_HTML_ARABIC_WINDOWS	htm
NATIVE	FI HTML BALTIC ISO	htm

NATIVE	ET HIML BALITIC WINDOWS	htm
		110111
NAIIVE	FI_HIML_CENTRALEUROPEAN_DOS	htm
NATIVE	ET HTML CENTRALEUROPEAN ISO	htm
NATIVE		h+m
NATIVE		II CIII
NAIIVE	FI_HIML_CENTRALEUROPEAN_WINDOWS	htm
NATIVE	FT_HTML_CHINESEBIG5	htm
NATIVE		h+m
NATIVE	FI_HIML_CHINESEEUC	n un
NATIVE	FI HTML CHINESEGB	htm
NATIVE	FT HTML CHINESESIMPLIFIED FUC	htm
NATIVE		h+m
NATIVE	FI_HIML_CHINESESIMPLIFIED_WINDOWS	num
NATIVE	FI HTML CHINESETRADITIONAL WINDOWS	htm
NATIVE	FT HTML CYRTLLIC DOS	htm
NATIVE		h+m
NATIVE	FI_HIML_CYRILLIC_ISU	ntm
NATIVE	FI HTML CYRILLIC KOI8R	htm
NATIVE	FT HTML CYRTLLIC MAC	htm
		1
NATIVE	FI_HIML_CIRILLIC_WINDOWS	num
NATIVE	FI HTML CYRILLIC1251	htm
NATIVE		htm
		110111
NAIIVE	FI_HIML_EBCDIC_1026	htm
NATIVE	FI HTML EBCDIC 273	htm
NATIVE		htm
		I CIII
NATIVE	FI_HIML_EBUDIU_278	ntm
NATIVE	FI HTML EBCDIC 280	htm
NATIVE		htm
		11011
NAIIVE	FI_HIML_EBUDIU_285	nτm
NATIVE	FI HTML EBCDIC 297	htm
NATIVE		htm
		110111
NAIIVE	FI_HIML_EBCDIC_500	nτm
NATIVE	FI HTML EBCDIC 870	htm
NATIVE	FTHTML FRCDIC 871	htm
		L +
NATIVE	FI_HIML_GREEK_ISU	ntm
NATIVE	FI HTML GREEK MAC	htm
NATIVE	FT_HTML_GREEK_WINDOWS	htm
		1
NATIVE	FI_HIML_HEBREW_DUS	ntm
NATIVE	FI HTML HEBREW ISO VISUAL	htm
NATIVE	FT HTML HERREW WINDOWS	htm
		h+m
NATIVE	FI_HIML_JAPANESE_MAC	num
NATIVE	FI HTML JAPANESE SHIFTJIS	htm
NATIVE	FT_HTML_JAPANESEEUC	htm
NATIVE		h+m
NATIVE	FI_HIML_JAPANESEJIS	n un
NATIVE	FI HTML JAPANESESJIS	htm
NATIVE	FT_HTMI_KORFAN JOHAB	htm
NATIVE		h+m
NATIVE	FI_HIML_KOREAN_WINDOWS	II UII
NATIVE	FI HTML KOREANHANGUL	htm
NATIVE	FT_HTML_LATIN2	htm
NATIVE		h+m
NATIVE		II CIII
NAIIVE	FI_HIML_IHAI_WINDOWS	htm
NATIVE	FI HTML TURKISH DOS	htm
NATIVE		htm
		11 UIII L +
NAIIVE	FI_HIML_IUKKISH_MAC	ntm
NATIVE	FI HTML TURKISH WINDOWS	htm
NATIVE	FT HTML VIETNAMESE WINDOWS	htm
		11 UIII 6 +
NAIIVE	FI_TIML_WESTERNEURUPEAN_ISU	ntm
NATIVE	FI HTML WESTERNEUROPEAN MAC	htm
NATIVE	FT HTMI WESTERNEUROPEAN WINDOWS	htm
NATIVE		h+
NAIIVE		ntm
NATIVE	FI_JAPANESE_EUC	txt
NATIVE	FIJAPANESEJIS	t.x+
NATTVE		+
NAIIVE		ιXΰ
NATIVE	FI_MAC	txt
NATIVE	FT MAC8	† x †
		640
NAIIVE	FI_PPZKHIML	nτm
NATIVE	FI SHIFTJIS	txt
NATIVE	FT_UNICODF	† x †
		++++
NAIIVE		ιXť
NATIVE	FI W2KHTML	htm
NATIVE	FT_WMI	ym]
	· *_***	7001

NATIVE	FI WML CHINESEBIG5	xml
NATIVE	FI_WML_CHINESEEUC	xm1
NATIVE	FI WML CHINESEGB	xml
NATIVE	FI_WML_CYRILLIC1251	xm1
NATIVE	FI WML CYRILLICKOI8	xml
NATIVE	FI WML JAPANESEEUC	xm1
NATIVE	FI_WML_JAPANESEJIS	xml
NATIVE	FI WML JAPANESESJIS	xml
NATIVE	FI_WML_KOREANHANGUL	xm1
NATIVE	FI WML LATIN2	xml
NATIVE	FIXHTML	htm
NATIVE	FI XL2KHTML	htm
NATIVE	FIXML	xm1
NATIVE	FI_XML_DOCTYPE_HTML	htm

Tarefas relacionadas:

"Associar tipos de documentos ao extractor de texto" na página 269

Configurar sessões múltiplas para o extractor de texto

Para colecções com vários módulos de analisador, é possível melhorar o desempenho do analisador configurando colecções para utilizar mais do que uma sessão de extractor de texto.

Para analisar documentos que são tratados pelo extractor de texto de forma mais eficaz, é possível configurar colecções para utilizar mais do que uma sessão de extractor de texto. Com sessões múltiplas de extractor de texto, podem ser processados mais documentos de forma simultânea.

Para configurar uma colecção para utilizar múltiplas sessões de extractor de texto:

- 1. Inicie sessão como o administrador predefinido do IBM Content Analytics with Enterprise Search no servidor de controlador.
- Edite o ficheiro ES_NODE_ROOT/master_config/ID_colecção.indexservice/ collection.properties, em que ID_colecção identifica a colecção que pretende configurar.
- **3**. Defina o valor do parâmetro **NumberOfStellentParsers** como sendo o número de sessões de extractor de texto que pretende utilizar.
- 4. Para que as alterações tenham efeito, utilize a consola administrativa para parar e reiniciar o analisador para a colecção.

Conceitos relacionados:

"Detecção do formato do documento" na página 265

Tarefas relacionadas:

"Configurar módulos para serviço de índice" na página 178

"Associar tipos de documentos ao extractor de texto" na página 269

Analisar texto bidireccional em ficheiros PDF

Processar texto bidireccional em ficheiros PDF para corresponder a ordem de leitura lógica do texto está para além do âmbito da tecnologia de visualização do extractor de texto. Se preferir, pode configurar o analisador para utilizar, em alternativa, a tecnologia PDFBox 1.0.0.

O extractor de texto não garante a ordem correcta do texto que extrai de ficheiros PDF. Com texto bidireccional em ficheiros PDF, a ordem em que o texto é analisado é para não corresponder provavelmente à ordem de leitura lógica do texto. Esta limitação causa um problema ao processar ficheiros PDF que estão

escritos em idiomas do Médio Oriente, tais como o Hebraico e o Árabe, que são predominantemente escritos da direita para a esquerda (bidireccional).

Se as colecções suportarem principalmente documentos escritos em Hebraico ou Árabe, poderá utilizar a tecnologia Apache PDFBox para analisar documentos PDF, em vez do analisador de texto predefinido. Se activar esta função:

- Apenas são suportados o Inglês e idiomas bidireccionais para analisar ficheiros PDF.
- O texto nos ficheiros PDF é extraído e ordenado por coordenadas. Se o ficheiro PDF contiver áreas de texto que partilham as mesmas coordenadas de localização, o texto poderá parecer alterado. Por exemplo, se um PDF contém a cadeia "description of company structure" e o texto foi alterado para "description of company business", algumas letras de uma palavra poderão ser movidas para a palavra seguinte e alguns poderão ser ignoradas porque as áreas de texto original e de texto alterado se sobrepõem.

Para activar o sistema para utilizar a tecnologia Apache PDFBox ao analisar texto bidireccional em ficheiros PDF:

- 1. Inicie sessão como o administrador predefinido do IBM Content Analytics with Enterprise Search no servidor de controlador.
- 2. Pare os serviços de análise e indexação.
- **3**. Edite o ficheiro de configuração do analisador, em que *collection_ID* identifica a colecção que desejar configurar:

ES_NODE_ROOT/master_config/collection_ID.indexservice/parser_config.xml

4. Elimine os caracteres de comentários para as linhas seguintes sob o elemento ParserService:

```
<ParserMapping>
<ParserName>pdf</ParserName>
<Mimetype>application/pdf</Mimetype>
</ParserMapping>
```

5. Remova a linha <Mimetype>application/pdf</Mimetype> do extractor de texto (<ParserName>stellent</ParserName>:

```
<ParserMapping>
<ParserName>stellent</ParserName>
...
<Mimetype>application/pdf</Mimetype>
...
</ParserMapping>
```

6. Para que as alterações se tornem eficazes, utilize a consola de administração para interromper e reiniciar o analisador para a colecção e depois indexe outra vez os documentos na colecção.

Conceitos relacionados:

"Detecção do formato do documento" na página 265

Indexação de documentos .docm com URLs que incluam consultas

Se configurar um pesquisador, como, por exemplo, o pesquisador da Web, para pesquisar documentos .docm, e o URL para um documento .docm incluir uma cadeia de consulta, o conteúdo não é indexado. Para assegurar que o conteúdo destes tipos de ficheiros é indexado, tem de especificar regras num ficheiro de configuração do analisador.

Para indexar documentos .docm com URLs que incluam consultas:
- Inicie sessão como o administrador predefinido de IBM Content Analytics with Enterprise Search no servidor do controlador e edite o ficheiro ES_NODE_ROOT/master_config/ID_Colecção.indexservice/mimetypes.xml, onde ID_Colecção identifica a colecção que pretende personalizar.
- 2. Adicione as seguintes linhas sob o elemento /Mimetypes/ MimetypeNormalization:

<NormalizedMimetype Name="application/vnd.ms-word.document.macroEnabled.12"> <Mimetype>application/vnd.ms-word.document.macroenabled.12</Mimetype> </NormalizedMimetype>

- 3. Reinicie o sistema IBM Content Analytics with Enterprise Search.
- 4. Reinicie o pesquisador e execute uma pesquisa completa para que os documentos possam ser pesquisados e indexados.

Conceitos relacionados:

"Detecção do formato do documento" na página 265

Suporte de idiomas e páginas de códigos

O processamento linguístico é processado de forma diferente pelo analisador e pelos servidores de procura.

Para fins de processamento linguístico, o analisador não distingue entre idiomas e locales. No entanto, se um utilizador procurar uma colecção que inclua documentos em múltiplos idiomas, os servidores de procura permitem que os resultados da procura se limitem a um idioma ou locale específico.

Por exemplo, se os metadados de um documento em inglês especificarem en_US para o locale do documento, este é indexado quer como um documento em inglês (en), quer como um documento que utiliza o locale dos E.U.A. para inglês(en_US). Este tipo de indexação permite que certas informações específicas do locale, como números, datas e horas, sejam correctamente representadas. Quando os utilizadores pesquisarem a colecção, o documento pode ser encontrado independentemente de o utilizador procurar documentos en ou en_US.

Se um documento for indexado apenas pelo código de linguagem, como en, o documento só será indexado pelo código de linguagem e não pelo locale. Se os utilizadores procurem documentos en_US na colecção, por exemplo, o documento não será encontrado.

O sistema fornece suporte linguístico para os seguintes idiomas e códigos de linguagem de dois caracteres, conforme documentado na norma ISO 639:

Idiomas de texto simples:

en=inglês sq=albanês az=azerbaijanês-latim bg=búlgaro be=bielorrusso ca=catalão hr=croata cs=checo da=dinamarquês nl=neerlandês et=estónio fi=finlandês fr=francês de=alemão el=grego hu=húngaro is=islandês id=indonésio in=indonésio it=italiano kk=cazague lv=letão lt=lituano lo=laosiano mk=macedónio ms=malaio mt=maltês no=norueguês nb=norueguês (Bokmal) pl=polaco pt=português ro=romeno

ru=russo sr=cirílico da Sérvia sh=latim da Sérvia sk=eslovaco sl=esloveno es=espanhol sv=sueco tr=turco uk=ucraniano cy=galês

Idiomas ideográficos:

Para chinês simplificado e tradicional, são utilizados códigos de linguagem expandidos em vez de códigos de dois caracteres.

zh-CN=chinês (simplificado) zh-TW=chinês (tradicional) ja=japonês ko=coreano

Idiomas de texto complexos:

ar=árabe as=assamês bn=bengali gu=guzerate iw=hebraico he=hebraico hi=hindi kn=canarês ml=malaiala mr=marata or=oriya pa=punjabi ta=tâmil te=telugu th=tailandês ur=urdu vi=vietnamita

O sistema pode detectar automaticamente muitos destes idiomas, além de poder detectar automaticamente a página de códigos que é utilizada em documentos de texto simples. Quando configura um pesquisador, pode desactivar a detecção automática de idiomas e páginas de códigos se pretender especificar qualquer idioma ou página de códigos explícito a utilizar.

Conceitos relacionados:

"Análise linguística de documentos em chinês, japonês e coreano" na página 282

Referências relacionadas:

"Detecção automática de idiomas"

"Detecção automática de páginas de códigos" na página 281

Detecção automática de idiomas

O sistema IBM Content Analytics with Enterprise Search pode processar documentos em praticamente qualquer idioma.

O sistema pode detectar automaticamente todos os idiomas suportados pelo anotador de Identificação de Idioma fornecido. Se souber qual o idioma dos seus documentos, pode especificar o idioma quando configurar um pesquisador em vez de permitir que o sistema detecta automaticamente o idioma. Ao criar uma colecção, pode especificar os idiomas dos documentos na colecção e especificar a ordem pela qual o processo de detecção do idioma irá tentar processar os documentos. Por exemplo, se uma colecção inclui muitos documentos em inglês e alguns documentos em francês, poderá configurar o processo de detecção automática de idioma para tentar primeiro o inglês. Se o processo determinar que o idioma do documento não é o inglês, irá tentar de seguida determinar se os documentos estão em francês.

Conceitos relacionados:

"Suporte de idiomas e páginas de códigos" na página 279

Anotador de identificação de idioma

Detecção automática de páginas de códigos

O sistema IBM Content Analytics with Enterprise Search suporta documentos numa variedade de páginas de códigos.

Para ficheiros de texto, o sistema pode detectar automaticamente as páginas de códigos que se seguem. Para outros formatos de documentos, o sistema utiliza metadados no documento, como elementos de metadados de HTML, para detectar a página de códigos. Se souber qual a página de códigos dos seus documentos, pode especificar a página de códigos a utilizar quando configurar um pesquisador em vez de permitir que o sistema detecte automaticamente a página de códigos.

Formatos de codificação Unicode:

UTF-8 UTF-16BE UTF-16LE

Formatos de codificação de byte múltiplo:

Shift-JIS ISO-2022-CN ISO-2022-JP ISO-2022-KR GB18030 EUC-JP EUC-KR

Formatos de codificação de byte único:

ISO-8859-1:	dinamarquês, neerlandês, alemão, inglês, francês, italiano, norueguês, português, espanhol, sueco
ISO-8859-2:	checo, húngaro, polaco, romeno
ISO-8859-5:	russo
ISO-8859-6:	árabe
ISO-8859-7:	grego
ISO-8859-8:	hebraico, hebraico em ordem visual
ISO-8859-9:	turco
Windows-1250:	checo, húngaro, polaco, romeno
Windows-1251:	russo
Windows-1252:	dinamarquês, neerlandês, alemão, inglês, francês, italiano,
	norueguês, português, espanhol, sueco
Windows-1253:	grego
Windows-1254:	turco
Windows-1255:	hebraico
Windows-1256:	árabe
KOI8-R:	russo

A detecção de caracteres é uma operação imprecisa. O processo de detecção da página de códigos tenta identificar o conjunto de caracteres (charset) que melhor corresponda às características dos dados de bytes, mas tem uma natureza parcialmente estatística e os resultados não podem ser garantidos com correctos.

Para uma maior exactidão, os dados de entrada devem estar num idioma simples. Também são necessárias, no mínimo, algumas centenas de bytes de texto normal no idioma.

Se ocorrer uma disparidade entre a codificação detectada e as codificações suportadas, o sistema utiliza a página de códigos predefinida para a colecção. **Conceitos relacionados**:

"Suporte de idiomas e páginas de códigos" na página 279

Análise linguística de documentos em chinês, japonês e coreano

Para melhorar a possibilidade de recuperação de documentos escritos nos idiomas chinês, japonês e coreano, pode especificar as opções de análise linguística.

Estão disponíveis três opções para as colecções de procura empresarial:

- Morfológica, baseada em unicode, segmentação, que utiliza espaço em branco como o delimitador entre palavras. Para utilizar esta abordagem, seleccione a opção para não usar segmentação de n-gram.
- A segmentação n-gram, que considera sequências sobrepostas de um número de caracteres como uma única palavra. Para idiomas como o chinês, o japonês e o coreano, a segmentação n-gram pode devolver resultados melhores da procura do que a segmentação de espaços em branco baseada em Unicode.

Se configurar dicionários de sinónimos para expandir termos de procura, os processos de procura aplicam a expansão de forma diferente a colecções que utilizam a segmentação n-gram.

 Esta abordagem de segmentação híbrida combina a segmentação específica do idioma baseada no dicionário com segmentação n-gram baseada em regras não utilizando o dicionário para analisar e efectuar a segmentação de palavras em documentos.

Para colecções de análise de conteúdo, pode especificar que o analisador vai usar a análise morfológica ou a abordagem híbrida.

Escolha o método de segmentação que pretende utilizar para analisar documentos quando cria uma colecção. Se quiser activar, desactivar ou alterar mais tarde o método de segmentação, edite a colecção e configure as opções de análise.

Conceitos relacionados:

"Suporte de idiomas e páginas de códigos" na página 279

"Expansão de sinónimos ao procurar colecções n-gram" na página 283

E Suporte linguístico para segmentação não baseada em dicionários

🕩 Suporte linguístico para segmentação híbrida

Tarefas relacionadas:

"Categorizar caracteres SBCS e numéricos como categorias n-gram"

Categorizar caracteres SBCS e numéricos como categorias n-gram

O processamento predefinido dos caracteres SBCS (single-byte character set - conjunto de caracteres de byte único) e numéricos no segmentador de espaços em branco e n-gram consiste em tratar esses caracteres como categorias segmentadas por espaços em branco.

Poderá alterar este comportamento para categorizar caracteres numéricos enquanto categorias n-gram, ou categorizar tanto os caracteres SBCS como os numéricos enquanto categorias n-gram.

Irá escolher a abordagem de categorizar caracteres SBCS e numéricos ao criar uma colecção. Se quiser activar, desactivar ou alterar o método de segmentação mais tarde, edite a colecção e configure opções de análise. As opções são:

- Tratar os caracteres SBCS e numéricos como categorias que são segmentadas por espaço em branco.
- Categorizar caracteres numéricos como categorias n-gram e tratar caracteres SBCS mediante segmentação por espaço em branco.
- Categorizar caracteres SBCS e numéricos como categorias n-gram.

Conceitos relacionados:

"Análise linguística de documentos em chinês, japonês e coreano" na página 282

- 🕩 Suporte linguístico para segmentação não baseada em dicionários
- 🖙 Suporte linguístico para segmentação híbrida

Expansão de sinónimos ao procurar colecções n-gram

Quando uma consulta é processada, os termos da consulta podem ser expandidos para incluir sinónimos de palavras na consulta. Se a consulta estiver especificada num idioma para o qual a análise morfológica é suportada como o inglês, as palavras são extraídas da consulta com base na análise morfológica. Os sinónimos são então expandidos das palavras extraídas.

Para colecções n-gram, uma consulta num idioma n-gram como o japonês é segmentada sem análise morfológica. Os processos de consulta aplicam a expansão de sinónimos a toda a sequência de caracteres consecutivos, que consiste em caracteres apenas segmentados em n-gram. Os processos de consulta também aplicam a expansão de sinónimos a toda a sequência de caracteres consecutivos que consiste nos caracteres que utilizam a segmentação de espaços.

Exemplos

Nestes exemplos, os caracteres em maiúsculas representam os caracteres japoneses que estão segmentados em n-gram. Os caracteres em minúsculas são os caracteres latinos. Numa colecção n-gram, os caracteres latinos são segmentados com base na segmentação de espaços em branco sem a análise morfológica.

- <AB> é definido num dicionário de sinónimos. Uma consulta que inclui a sequência <AB> expande os sinónimos. Uma consulta que inclua <ABC>, <XAB> ou <XABC> não expande os sinónimos.
- <computador> é definido num dicionário de sinónimos. Uma consulta que inclua <computador> expande os sinónimos. Uma consulta que inclua <computadores>, <meucomputador> ou <teucomputador> não expande os sinónimos.
- Quando um termo da consulta consiste numa combinação de caracteres japoneses e latinos, os limites da sequência segmentada de n-gram e a segmentação de espaços em branco é detectada nos termos da consulta. As duas regras precedentes são aplicadas a cada sequência. Por exemplo:
 - Se for especificado <ABcomputadorCD> como um termo de consulta, as sequências <AB>, <computador> e <CD> são extraídas. Estes termos encontram-se no dicionário de sinónimos, e a expansão de sinónimos é aplicada a todos os termos.

 Se for especificado <ABCmeucomputadorCD> como um termo de consulta, as sequências <ABC>, <meucomputador> e <CD> são extraídas. Estes termos não se encontram no dicionário de sinónimos, e os sinónimos não serão expandidos.

Conceitos relacionados:

"Análise linguística de documentos em chinês, japonês e coreano" na página 282

Administração do servidor de procura

As opções que pode especificar para os servidores da procura incluem a utilização do espaço da cache para resultados da procura, o controlo da duração máxima da apresentação dos resumos dos documentos nos resultados da procura, a associação de dicionários personalizados para melhorar a qualidade da procura, devolver URIs predefinidos nos resultados da procura quando determinados termos aparecerem na consulta e configurar suporte introdução adiantada

Quando um utilizador submete uma consulta, os servidores da procura utilizam o índice para localizar rapidamente os documentos relevantes. Os servidores de pesquisa utilizam o armazém de dados, que contém os dados analisados e segmentados, para obter metadados dos documentos relevantes. Os metadados podem incluir, mas não estar limitados ao documento URI, ao título, à descrição, à data, ao tipo de dados, etc.

Quando configura os servidores da procura para uma colecção, especifica opções que influenciam a forma como as consultas são processadas, incluindo opções que podem afectar o desempenho da consulta:

Configurar uma cache de procura

Para optimizar o desempenho da consulta, pode especificar que os resultados da procura (as respostas às consultas) devem ser armazenados numa cache, e pode configurar a quantidade de espaço a reservar para resultados da procura colocados na cache.

Configurar a duração máxima de apresentação para resumos de documentos

A maioria dos documentos dos resultados mostram um resumo do conteúdo do documento para ajudar os utilizadores a decidir se o documento é aquele que pretendem obter. Pode especificar o espaço que deve ser utilizado nos resultados da procura para visualizar essa informação do resumo.

Utilizar dicionários personalizados

Se os programadores de aplicações tiverem criado dicionários personalizados para palavras de paragem de sinónimos, ou para palavras hierárquicas, pode especificar os dicionários a serem utilizados quando os utilizadores procurarem a colecção.

Influenciar o processamento de consultas e correcção ortográfica

Pode especificar opções para influenciar o rendimento de consultas, sugestões de ortografia e ordenação de documentos, incluindo como as classificações de categoria IBM Content Classification podem influenciar a classificação total do documento.

Configurar ligações rápidas

Pode determinar previamente os URIs a serem retornados para certas palavras-chave e frases. Quando os utilizadores especificarem as palavras-chave ou as frases numa consulta, o URI predefinido será devolvido com os resultados da procura. Os URIs de ligação rápida são devolvidos em adição aos URIs que os servidores da procura devolvem ao procurar o índice.

Configurar suporte para caracteres globais

Pode permitir que os utilizadores incluam um carácter global nos termos de consulta e que procurem palavras que correspondam a um padrão especificado.

Configurar âmbitos

Pode limitar o âmbito de documentos que os utilizadores podem pesquisar. Quando os utilizadores procurarem a colecção, procuram apenas os documentos que pertencem à pontuação, e não todo o índice.

Resuma os documentos da mesma origem nos resultados da pesquisa

Pode agrupar documentos que correspondem a um URI ou padrão URI no índice, e mostrar apenas o início dos documentos dos resultados nos resultados da procura (os utilizadores podem especificar as opções para visualizar os documentos dos resultados resumidos).

Configurar suporte de consulta de introdução adiantada

Pode especificar opções para mostrar potenciais correspondências de termos de consulta à medida que um utilizador escreve uma consulta.

Influenciar o rendimento da consulta e a correcção ortográfica

Quando configura os servidores de procura, pode configurar opções para influenciar o rendimento de consultas, sugestões de ortografia e ordenação de documentos, incluindo como as classificações de categoriaIBM Content Classification podem influenciar a classificação total do documento.

Processamento e rendimento da consulta

As opções seguintes permitem-lhe influenciar o processamento e o desempenho da consulta:

- Pode permitir aos utilizadores inserirem sintaxe de consulta especial, *:*, para localizar todos os documentos na colecção. Esta consulta necessita de tempo para processar, por isso poderá não querer activar esta capacidade se o número de documentos na colecção for grande.
- Na pontuação baseada em texto, a proximidade dos termos da consulta pode influenciar a forma como os documentos são classificados nos resultados. Pode melhorar o rendimento da consulta para uma colecção grande especificando que não pretende que a proximidade do termo da consulta seja um factor aquando da informatização do documento.
- Se a aplicação necessitar de respostas de consulta rápidas e não necessitar de uma nova chamada total, pode especificar um limite para o tempo em que o índice vai demorar a ser digitalizado quando uma consulta é processada. Quando é atingido o limiar que especificar, a digitalização pára e são devolvidos resultados truncados.
- Pode impor um limite no número de documentos devolvidos nos resultados de pesquisa.
- Pode controlar a frequência com que os servidores de pesquisa verificam actualizações ao índice.

Sugestões para correcções ortográficas

As opções seguintes permitem-lhe controlar como as sugestões para as correcções ortográficas são facultadas:

• Pode activar e desactivar a apresentação das correcções ortográficas.

- Pode especificar um limiar para mostrar correcções ortográficas. Por exemplo, se especificar 100, as correcções ortográficas são informatizadas quando uma consulta devolve menos de 100 resultados.
- Pode especificar o número mínimo de caracteres iniciais numa palavra que vai ser comparada antes de ser apresentada uma correcção ortográfica. As correcções ortográficas devem cumprir o comprimento do prefixo que especificar antes de serem sugeridos na aplicação de procura empresarial.
- Pode controlar a correspondência semelhante entre termos da consulta e sugestões ortográficas. Se especificar um valor baixo, pode demorar algum tempo a informatizar sugestões para correcções ortográficas e são devolvidas correcções mais ambíguas.

"Classificação baseada em texto" na página 313

"Planos de decisão e pontuações por categoria do Content Classification" na página 504

Tarefas relacionadas:

"Configurar os campos de índice e pontuações do Content Classification" na página 506

Caches de procura

Quando a carga nos servidores de procura é relativamente elevada, pode melhorar o desempenho colocando os resultados da procura em cache.

Quando os servidores de procura processam pedidos de procura, verificam primeiro se já existem na cache resultados para a mesma consulta. Se os servidores de procura encontrarem a resposta apropriada à consulta, poderão devolver rapidamente resultados da procura ao utilizador. Se os servidores resposta de procura não encontrarem a resposta apropriada à consulta, procurarão no índice.

Quando a cache de procura estiver cheia, os resultados da procura mais antigos e os resultado para consultas pouco frequentes são retirados para criar espaço para novos resultados da procura.

Para a consola de administração, pode activar a colocação da pesquisa em memória cache e também especificar a capacidade da memória cache (o número de respostas da consulta que podem ser colocadas na cache em simultâneo).

Quando altera as opções da cache de procura, tem de reiniciar os servidores de procura para que as alterações se tornem efectivas.

Tarefas relacionadas:

"Configurar uma cache de procura"

Configurar uma cache de procura

Pode activar ou desactivar a memória cache de procura de uma colecção. Também pode especificar opções para controlar o tamanho da memória cache de procura.

Para configurar a memória cache de procura:

 Expanda a colecção que pretende configurar. No painel Procurar para uma colecção de procura empresarial ou no painel Análise para uma colecção de análise de conteúdo, faça clique em Acções > Configurar > Opções do servidor de procura.

- 2. Seleccione a caixa de verificação Utilizar a cache de procura.
- No campo Número máximo de entradas na memória cache, escreva o número máximo de respostas a consulta que a memória cache de procura pode conter.
- 4. Para tornar as alterações eficazes, supervisione os servidores da procura e reinicie os processos do servidor.

"Caches de procura" na página 287

Resumo dinâmico

O resumo dinâmico é uma técnica que determina as expressões de um documento que representam melhor o conceito de que o utilizador está à procura.

O resumo dinâmico tenta capturar frases em documentos que contêm um grande número e variedade de termos de procura. Algumas frases, ou partes de frases, são seleccionadas e combinadas para construir o resumo que é apresentado nos resultado. Os termos de procura são destacados no resumo.

Ao configurar as opções do servidor da procura para uma colecção, pode especificar o comprimento máximo da apresentação para resumir os documentos nos resultados da procura. Uma vez que o resumo inclui caracteres de destaque, a memória tampão devolvida à aplicação será maior do que o valor máximo especificado. No entanto, o comprimento de apresentação não excederá o valor especificado, apesar de o resumo poder ser mais curto (dependendo dos dados de resumo extraídos do documento origem).

Tarefas relacionadas:

"Personalizar resumos de documentos"

Personalizar resumos de documentos

Pode personalizar a quantidade de informação apresentada em resumos de documentos, especificando opções para o servidor da procura na consola de administração.

Para que o valor de um campo contribua para o resumo de documento dinâmico, tem de activar os atributos de **Procura de texto livre** e **Em resumo** para o campo. Os atributos são especificados quando configurar os campos de índice para a colecção ou quando correlacionar origens de entrada, como, por exemplo, campos de origem ou colunas CSV, com campos de índice.

O valor que especifica para o comprimento máximo de apresentação de resumos de documentos utiliza o valor que especifica como o número de instruções que cada resumo pode conter. O valor que resulta no resumo de documento mais curto tem precedência.

Por exemplo, se especificar um limite de quatro instruções, o resumo de documento contém apenas quatro instruções, mesmo se o comprimento de apresentação permitir mais caracteres que o número total de caracteres nessas instruções. Por outro exemplo, um limite de 10 frases combinado com um limite de 500 caracteres para o comprimento de apresentação, pode resultar num resumo de documento com menos de 10 frases.

Para configurar um comprimento de exibição para resumos de documentos:

- Expanda a colecção que pretende configurar. No painel Procurar para uma colecção de procura empresarial ou no painel Análise para uma colecção de análise de conteúdo, faça clique em Acções > Configurar > Opções do servidor de procura.
- Especifique o comprimento máximo de uma frase para os resumos do documento. Quando os utilizadores virem os resultados da procura, os resumos de documentos não excederão o valor que especifica.
- **3**. Especifique o número de frases que cada documento pode conter (os resumos podem conter um máximo de 10 frases).
- 4. Para que as alterações se tornem efectivas, reinicie os servidores de procura.

- "Resumo dinâmico" na página 288
- "Atributos do campo de índice" na página 195

Tarefas relacionadas:

"Criar campos de índice" na página 194

Dicionários de sinónimos personalizados

Para melhorar a qualidade dos resultados da procura, pode permitir que os utilizadores procurem sinónimos dos respectivos termos de consulta ao efectuarem procuras numa colecção.

Se criar um dicionário de sinónimos, adicioná-lo ao sistema e associá-lo a uma colecção, os utilizadores podem procurar documentos que contenham sinónimos dos respectivos termos de consulta quando efectuarem procuras na colecção. Ao expandir consultas desta forma, é mais provável que os utilizadores encontrem todos os documentos de interesse e não apenas documentos que tenham correspondência exacta com os respectivos termos de consulta. Uma vez que define quais as palavras que se tratam de sinónimos entre si, ao criar o dicionário de sinónimos, permite ajudar a garantir que os utilizadores encontram documentos relevantes sem ser necessário especificar todas as variantes do termo de consulta.

Por exemplo, a organização poderá utilizar acrónimos e abreviaturas para fazer referência a departamentos, equipamento e assim sucessivamente, ou os documentos nas colecções poderão conter vocabulário específico da indústria. Ao criar um dicionário de sinónimos, assegura que as consultas que incluam um acrónimo (por exemplo, ACL) devolvem documentos que abordem a expansão desse acrónimo (por exemplo, ACLs, listas de controlo de acessos, controlos de acesso, etc.).

A linguagem de consulta suporta sinónimos ao permitir que os utilizadores anexem um operador til como prefixo a um termo de consulta. Por exemplo, a consulta ~WAS poderá devolver documentos que abordem o WebSphere Application Server. Os programadores de aplicações também podem disponibilizar o suporte de sinónimos através das propriedades das consultas, que não requerem qualquer sintaxe em especial.

Os dicionários de sinónimos contêm variantes de palavras e apresentam as seguintes características:

 As palavras não são específicas de um idioma, mas pode ser utilizadas em diferentes idiomas. Existe apenas um dicionário de sinónimos por colecção. As palavras não são flexionadas. Todas as flexões possíveis têm de ser adicionadas à lista de sinónimos. Por exemplo, uma flexão poderá apresentar a forma singular e plural da palavra (por exemplo, ACL e ACLs).

Os termos adicionados a um dicionário de sinónimos tratam-se, na sua maioria, de equivalentes semânticos exactos, o que significa que se um termo A for um sinónimo do termo B, logo B é um sinónimo de A. Sempre que A for utilizado numa consulta, B pode ser utilizado e vice versa.

No entanto, também pode adicionar termos que correspondam a diferentes utilizações de um termo, incluindo variantes genéricas ou mais específicas do termo. Por exemplo, pode ter um grupo de sinónimos que inclua edifício e casa e outro grupo que inclua banco, margem e cooperativa de crédito.

Quando menos exacta for a relação entre os termos, maior o resultado da procura, embora alguns resultados da procura possam não ser relevantes para a consulta. A API de Procura e Índice fornece métodos que permitem aos utilizadores seleccionar os sinónimos apropriados ao submeterem um pedido de procura e métodos que mostram aos utilizadores quais os termos expandidos a quais os sinónimos.

Para criar um dicionário de sinónimos, um especialista em matéria de colecções tem de criar uma lista de sinónimos no formato XML ou trabalhar com um programador de aplicações para criar um ficheiro XML. Tem de ser utilizada uma ferramenta facultada, **essyndictbuilder**, para converter o ficheiro XML num ficheiro binário (.dic).

Um administrador principal carrega o ficheiro binário no sistema e atribui-lhe um nome de apresentação. Os administradores da colecção podem seleccionar um dicionário de sinónimos a utilizar para procurar documentos numa colecção quando configuram opções do servidor da procura para uma colecção.

Restrição: Após adicionar um dicionário de sinónimos personalizado ao sistema, não pode editá-lo. Para rever os sinónimos disponíveis para uma colecção, tem de:

- 1. Actualizar o ficheiro XML de origem.
- 2. Converter a origem de XML num novo ficheiro de dicionário.
- 3. Remover o dicionário de sinónimos antigo das colecções que o utilizam.
- 4. Eliminar o dicionário de sinónimos antigo do sistema.
- 5. Adicionar o dicionário de sinónimos novo ao sistema.
- 6. Associar o novo dicionário de sinónimos às colecções que o forem utilizar

É possível escrever um script que inclua estes passos e, em seguida, utilizar o script para voltar a implementar o dicionário no sistema.

🕩 Dicionários de sinónimos personalizados para colecções de pesquisa

Tarefas relacionadas:

"Adicionar dicionários de sinónimos ao sistema"

"Associar um dicionário de sinónimos a uma colecção"

🖙 Criar um ficheiro XML para sinónimos para colecções de pesquisa

F Criar um dicionário de sinónimos para colecções de pesquisa

Adicionar dicionários de sinónimos ao sistema

Se criar dicionários de sinónimos personalizados para procurar os documentos numa colecção, terá de associar os dicionários ao sistema. Mais tarde, pode seleccionar que dicionário de sinónimos pretende utilizar para procurar uma colecção.

O tamanho máximo de um dicionário de sinónimos é 8 MB.

Para adicionar um dicionário de sinónimos ao sistema:

- 1. Faça clique em Sistema para abrir a vista Sistema.
- 2. Na página Procura, faça clique em **Configurar dicionários de sinónimos**.
- **3**. Na página Configurar Dicionários de Sinónimos, faça clique em **Adicionar Dicionário de Sinónimos**.
- 4. Na página Adicionar um Dicionário de Sinónimos, escreva um nome de apresentação único para o dicionário de sinónimos
- 5. Especifique a localização do ficheiro .dic. Se o ficheiro se encontrar no sistema local, pode navegar até o localizar. Se o ficheiro se encontrar no servidor de controlador, escreva o caminho totalmente qualificado.
- 6. Faça clique em **OK**. É adicionado ao sistema o dicionário de sinónimos personalizado, ficando disponível para ser associado às colecções.

Conceitos relacionados:

"Dicionários de sinónimos personalizados" na página 289

Tarefas relacionadas:

"Associar um dicionário de sinónimos a uma colecção"

🕩 Criar um ficheiro XML para sinónimos para colecções de pesquisa

🖙 Criar um dicionário de sinónimos para colecções de pesquisa

Associar um dicionário de sinónimos a uma colecção

Se houver dicionários de sinónimos associados ao sistema, pode seleccionar um para utilizar quando procurar uma colecção. Se um termo de consulta corresponder a um termo no dicionário, os documentos dos resultados que contêm os sinónimos desse termo também são devolvidos nos resultados da procura.

Para associar um dicionário de sinónimos a uma colecção:

- Expanda a colecção que pretende configurar. No painel Procurar para uma colecção de procura empresarial ou no painel Análise para uma colecção de análise de conteúdo, faça clique em Acções > Configurar > Opções do servidor de procura.
- No campo Nome do dicionário de sinónimos, seleccione o dicionário de sinónimos que pretende utilizar quando os utilizadores consultam esta colecção.

A lista de dicionários de sinónimos disponíveis inclui todos os dicionários de sinónimos que foram adicionados ao sistema.

Conceitos relacionados:

"Dicionários de sinónimos personalizados" na página 289

Tarefas relacionadas:

"Adicionar dicionários de sinónimos ao sistema" na página 291

- 🖙 Criar um ficheiro XML para sinónimos para colecções de pesquisa
- Li Criar um dicionário de sinónimos para colecções de pesquisa

Dicionários de palavras de paragem personalizados

Para melhorar a qualidade dos resultados da procura, pode especificar que certas palavras devam ser automaticamente removidas dos termos da consulta durante o processamento da consulta.

Um dicionário de palavras de paragem contém termos específicos da empresa que são frequentemente utilizados, e assim, não são úteis como termos da consulta. Ao excluir essas palavras das consultas, pode assegurar que os utilizadores não obtenham dados com os documentos dos resultados , apenas marginalmente relevantes (somente os documentos que correspondem a outros termos na consulta serão devolvidos). Durante o processamento da consulta, os servidores da procura removem as palavras de paragem das consultas. As palavras removidas incluem palavras de paragem no respectivo dicionário personalizado e palavras de paragem que estão predeterminadas para pesquisa (tal como preposições comuns e artigos).

O reconhecimento de palavras de paragem específicas do idioma é executado por predefinição. Este processo remove frequentemente palavras comuns como um e o duma consulta. Tem a necessidade de definir um dicionário de palavras de paragem personalizado apenas para a empresa ou palavras de paragem específicas de domínio.

Quando uma consulta é processada, as palavras de paragem são removidas antes de serem dadas sugestões de ortografia. Se todas as palavras de uma consulta forem palavras de paragem, não será removida nenhuma palavra de paragem durante o processamento da consulta. Para assegurar que os resultados da procura são devolvidos, a remoção de palavras de paragem é desactivada quando todos os termos da consulta forem palavras de paragem. Por exemplo, se a palavra carro for uma palavra de paragem e o utilizador procurar a palavra carro, os resultados da procura conterão documentos que correspondem à palavra carro. Se procurar carro volvo, os resultados da procura só conterão documentos que correspondem à palavra volvo.

Para criar um dicionário de palavras de paragem, um especialista em matéria de colecções terá de criar uma lista de palavras de paragem em formato XML ou trabalhar com um programador de aplicações para criar o ficheiro XML. Tem de ser utilizada uma ferramenta facultada, **esstopworddictbuilder**, para converter o ficheiro XML num ficheiro binário (.dic).

Um administrador principal carrega o ficheiro binário no sistema e atribui-lhe um nome de apresentação. Os administradores de colecção podem seleccionar um dicionário de palavras de paragem para procurar documentos numa colecção quando configuram opções do servidor da procura para uma colecção.

Restrição: Depois de adicionar um dicionário de palavras de paragem personalizado ao sistema, não pode editá-lo. Para rever as palavras de paragem disponíveis para o processamento da consulta, tem de:

- 1. Actualizar o ficheiro XML de origem.
- 2. Converter a origem de XML num novo ficheiro de dicionário.
- Remover o dicionário de palavras de paragem antigo das colecções que o utilizam.
- 4. Eliminar o dicionário de palavras de paragem antigo do sistema.
- 5. Adicionar o dicionário de palavras de paragem novo ao sistema.
- 6. Utilizar o dicionário de palavras de paragem novo com as colecções que o utilizam.

É possível escrever um script que inclua estes passos e, em seguida, utilizar o script para voltar a implementar o dicionário no sistema.

Conceitos relacionados:

Dicionários de palavras de paragem personalizados

Tarefas relacionadas:

"Adicionar dicionários de palavras de paragem ao sistema"

- "Associar um dicionário de palavras de paragem a uma colecção" na página 294
- 🖙 Criar um ficheiro XML para palavras de paragem
- 🕩 Criar um dicionário de palavras de paragem

Adicionar dicionários de palavras de paragem ao sistema

Se criar dicionários de palavras de paragem personalizados para remover palavras de consultas, tem de adicionar os dicionários ao sistema. Mais tarde, pode seleccionar que dicionário de palavras de paragem pretende utilizar para procurar uma colecção.

O tamanho máximo de um dicionário de palavras de paragem é 8 MB.

Para adicionar um dicionário de palavras de paragem ao sistema:

- 1. Faça clique em Sistema para abrir a vista Sistema.
- 2. Na página Procura, faça clique em **Configurar dicionários de palavras de paragem**.
- **3**. Na página Configurar dicionários de palavras de paragem, faça clique em **Adicionar dicionário de palavras de paragem**.
- 4. Na página Adicionar dicionário de palavras de paragem, escreva o único nome de apresentação do dicionário.
- 5. Especifique a localização do ficheiro .dic. Se o ficheiro se encontrar no sistema local, pode navegar até o localizar. Se o ficheiro se encontrar no servidor de controlador, escreva o caminho totalmente qualificado.
- 6. Faça clique em **OK**. O dicionário de palavras de paragem é adicionado ao sistema e fica disponível para ser associado a colecções.

"Dicionários de palavras de paragem personalizados" na página 292

Tarefas relacionadas:

"Associar um dicionário de palavras de paragem a uma colecção"

- Criar um ficheiro XML para palavras de paragem
- 🖙 Criar um dicionário de palavras de paragem

Associar um dicionário de palavras de paragem a uma colecção

Se os dicionários de palavras de paragem forem utilizados com o sistema, pode seleccionar um para utilizar quando estiver à procura de uma colecção. Se um termo da consulta corresponder a um termo no dicionário, o termo será removido da consulta antes de ser processado.

Para utilizar um dicionário de palavras de paragem com uma colecção:

- Expanda a colecção que pretende configurar. No painel Procurar para uma colecção de procura empresarial ou no painel Análise para uma colecção de análise de conteúdo, faça clique em Acções > Configurar > Opções do servidor de procura.
- No campo Nome do dicionário de palavras de paragem, seleccione o dicionário de palavras de paragem que pretende utilizar quando os utilizadores consultam essa colecção.

A lista de dicionários disponíveis inclui todos os dicionários de palavras de paragem que foram adicionados ao sistema.

Conceitos relacionados:

"Dicionários de palavras de paragem personalizados" na página 292

Tarefas relacionadas:

"Adicionar dicionários de palavras de paragem ao sistema" na página 293

- 🖙 Criar um ficheiro XML para palavras de paragem
- Criar um dicionário de palavras de paragem

Trabalhar com ligações rápidas

As ligações rápidas são documentos devolvidos nos resultados da procura sempre que um utilizador submeter uma consulta que inclua palavras e frases específicas.

Utilize a consola de administração para configurar ligações rápidas a uma colecção.

Conceitos relacionados:

"Ligações rápidas"

Tarefas relacionadas:

"Configurar ligações rápidas" na página 295

Ligações rápidas

As ligações rápidas permitem-lhe fornecer aos utilizadores ligações aos documentos predeterminados a serem relevantes para os termos da consulta.

Uma ligação rápida é uma URI que o sistema inclui automaticamente nos resultados da pesquisa quando uma consulta incluir determinadas palavras ou frases. Normalmente, os URIs de ligação rápida aparecem no início da lista de resultados, que ajuda a assegurar que os utilizadores consultem os documentos que predeterminou a serem relevantes para a consulta.

As ligações rápidas são devolvidas adicionalmente a outros resultados da procura. Os processos de procura procuram o índice por documentos que correspondem aos termos da consulta, e devolvem URIs para os documentos em adição aos URIs de ligação rápida.

Quando configura uma ligação rápida, pode especificar um título descritivo e um resumo do URI para ajudar os utilizadores a reconhecer o documento e a determinar rapidamente se é um documento que pretendem recuperar.

Por exemplo, para o URI http://www.ibm.com/education/us/, pode utilizar um título, tal como Educação IBM nos Estados Unidos, e fornecer o resumo Soluções, produtos e recursos para profissionais, educadores e estudantes nos Estados Unidos.

Para utilizar ligações rápidas em colecções, a opção para mostrar ligações rápidas tem de estar disponível na aplicação de procura empresarial. Em algumas aplicações de procura, os utilizadores podem ter a capacidade de activar e desactivar a devolução de ligações rápidas ao procurarem a colecção.

Conceitos relacionados:

"Trabalhar com ligações rápidas" na página 294

Configurar ligações rápidas

Para criar uma ligação rápida para uma colecção, utilize um URI de um documento com as palavras-chave que activam a respectiva inclusão nos resultados da procura.

Não necessita de reiniciar os servidores da procura para que as alterações se tornem efectivas.

Para configurar uma ligação rápida:

- 1. Na vista Colecções, no painel Procura ou no painel Análise da respectiva colecção, faça clique em **Configurar** > **Ligações rápidas**.
- 2. Na página Ligações Rápidas, faça clique em Criar Ligação Rápida.
- 3. Especifique as palavras-chave e as frases que fazem com que esta ligação rápida seja devolvida nos resultados da procura, que o URI do documento que predeterminou é relevante para esta consulta, assim como outras opções para essa ligação rápida.

Pode especificar uma palavra-chave, várias palavras-chave ou uma frase (duas ou mais palavras incluídas entre aspas) por linha. Palavras-chave separadas por um espaço (não pode utilizar uma vírgula para delimitar as palavras-chave). Prima a tecla Enter para iniciar uma nova linha.

4. Faça clique em OK.

A nova ligação é listada na página Ligações Rápidas com as outras ligações rápidas que pertencem a esta colecção. Quando os utilizadores visualizam ligações rápidas nos resultados da pesquisa, as ligações rápidas são apresentadas por ordem descendente por classificação do documento.

"Trabalhar com ligações rápidas" na página 294

Referências relacionadas:

"Formatos de URI no índice" na página 164

Caracteres globais em consultas

Pode permitir que os utilizadores incluam um carácter global nos termos de consulta e que procurem palavras que correspondam a um padrão especificado.

Um termo de consulta de carácter global é um termo que contém um asterisco (*) enquanto símbolo de substituição para vários caracteres, ou um ponto de interrogação (?) enquanto símbolo de substituição para um único carácter. Quando um utilizador submete uma consulta que inclua um carácter global, os resultados da procura incluem todos os documentos no índice que correspondam ao termo de consulta, bem como todos os documentos no índice que correspondam ao padrão representado pelo carácter global.

Por exemplo, o carácter global de seguimento no termo de consulta mar* pode corresponder a marcador, marinho e martelo. Para outro exemplo, a consulta I?M pode corresponder a IBM, ICM, IRM, etc.

Ao configurar opções de carácter global para um índice, pode optar pela forma em que este suporte será fornecido:

- Pode restringir o carácter global ao último carácter de um termo de consulta (um carácter global à direita)
- Pode permitir que o carácter global ocorra em qualquer parte de um termo de consulta. O carácter global não pode ocorrer num nome de campo.

Quaisquer alterações efectuadas às definições de caracteres globais tornar-se-ão efectivas da próxima vez que ocorrer a criação do índice principal.

Expansão da consulta

Para expandir consultas e aplicar regras de correspondência de padrões quando os utilizadores submetem consultas que contenham caracteres globais, especifique a quantidade de variantes de um termo de consulta que constitui uma correspondência. Por exemplo, se especificar 50, podem qualificar-se até 50 variantes de um termo de consulta como correspondências do termo de consulta.

Para ilustrar este exemplo, o termo de consulta tecn* corresponde às palavras técnico, técnica, tecnologia e até 50 palavras diferentes que comecem com os caracteres tecn.

Embora a expansão da consulta tenha apenas um impacto mínimo no tamanho do índice, pode afectar negativamente o desempenho da consulta. Os processos de procura têm de interagir com todas as expansões possíveis do termo de consulta de carácter global, até ao limite especificado nas definições de carácter global.

Suporte para caracteres globais em consultas

O conjunto de expansões de um termo de consulta de carácter global contém todos os termos no índice que possam ser obtidos substituindo o carácter global por sequências arbitrárias de caracteres. O conjunto é determinado da seguinte forma:

- Se uma colecção suportar caracteres globais que ocorram em qualquer parte de um termo de consulta, qualquer termo de consulta que contenha um carácter global será interpretado como um termo de carácter global.
- O conjunto contém, no máximo, o número máximo de expansões configurado pelo administrador da colecção. Se o índice contiver um número de expansões superior a este valor, estas serão ignoradas. (Os resultados da procura indicam se foram ignoradas quaisquer expansões de caracteres globais.)
- Se o termo de consulta for um termo de campo, o carácter global tem de aparecer após o especificador de campo (por exemplo, fieldname:*sphere). O nome do campo não pode conter o sinal de dois pontos (:).
- Os caracteres globais são suportados apenas em termos de texto simples e não em nomes de elementos XML, nomes de atributos nem valores de atributos. Não é suportado um termo constituído apenas por um carácter global.

Tarefas relacionadas:

"Activar a análise avançada para termos compostos" na página 178

"Configurar opções para caracteres globais em consultas"

Configurar opções para caracteres globais em consultas

Quando configurar servidores de pesquisa para uma colecção, pode especificar se pretende activar utilizadores a incluir caracteres globais em termos de consulta.

Um termo de consulta de carácter global é um termo que contém um asterisco (*) enquanto símbolo de substituição para vários caracteres ou um símbolo de percentagem (%) como símbolo de substituição para caracteres únicos. Quando submeter uma consulta que inclua um carácter global, os resultados da pesquisa incluem todos os documentos no índice que correspondem ao termo da consulta e todos os documentos no índice que correspondam ao padrão representado pelo carácter global.

Os caracteres globais não têm restrições e podem ocorrer em qualquer parte num termo da consulta. Quando especifica as opções do carácter global, as suas alterações tornar-se-ão efectivas da próxima vez que ocorrer uma actualização do índice.

Para configurar o suporte de caracteres globais em consultas:

- 1. Expanda a colecção que pretende configurar. No painel Procura para uma colecção de procura empresarial ou no painel Análise para uma colecção de análise de conteúdos, faça clique em **Configurar** > **Caracteres globais**.
- 2. Especifique quantas variações do termo da consulta se qualificam como correspondência. Por exemplo, se especificar 50, até 50 variantes de um termo de consulta podem constituir correspondências do termo de consulta. O termo da consulta tec* corresponde a termos técnicos e tecnologia, e até 50 termos diferentes que começam pelos caracteres tec.

Conceitos relacionados:

"Caracteres globais em consultas" na página 296

Âmbitos

Configurar uma pontuação quando pretender apresentar os utilizadores com uma vista limitada duma colecção.

Uma pontuação é um grupo de URIs relacionados num índice. Quando configurar uma pontuação, limite os documentos que pode visualizar na colecção. Quando os utilizadores procuram a colecção, procuram apenas os documentos no âmbito, não todo o índice. Para utilizar esta função, as respectivas aplicações de procura têm de incluir suporte para procurar pontuações.

Quando criar uma pontuação, especifique um intervalo de URIs no índice que os utilizadores podem procurar. A limitação de documentos que os utilizadores podem procurar ajuda a garantir que os documentos nos resultados da procura sejam específicos para as informações que os utilizadores pesquisam.

Por exemplo, pode criar uma pontuação que inclui os URIs para o departamento de Suporte Técnico e outro âmbito que inclui os URIs para o departamento de Recursos Humanos. Se a aplicação de procura suportar pontuações, os utilizadores no departamento de Suporte Técnico obtêm documentos do âmbito Suporte Técnico e os utilizadores no departamento de Recursos Humanos obtêm documentos do âmbito Recursos Humanos.

Pode criar as pontuações que pretender, apesar de a criação de demasiadas pontuações poder afectar o desempenho. Configure pontuações para que a maior parte dos pedidos de procura tenha de filtrar apenas uma ou duas pontuações. Uma vez que as pontuações podem conter URIs ou padrões URI inteiros, o mesmo documento pode pertencer a mais de uma pontuação.

Tarefas relacionadas:

"Configurar âmbitos"

Configurar âmbitos

Quando configurar um âmbito para uma colecção, especifique os URIs, ou os padrões de URI, para um intervalo de documentos no índice. Quando os utilizadores consultam uma colecção, os documentos que procuram podem ser limitados aos documentos que correspondem ao âmbito.

Para configurar um âmbito:

- 1. Na página Colecções, expanda a colecção para a qual pretende definir um âmbito.
- No painel Procura para uma colecção de procura empresarial ou no painel Análise para uma colecção de análise de conteúdos, faça clique em Configurar > Âmbitos.
- 3. Na página Âmbitos, faça clique em Criar âmbito.
- 4. Especifique um nome para o âmbito e para os URIs e os padrões URI que definem os limites do âmbito. Também pode especificar os URIs e os padrões URI que pretende excluir do âmbito.

O novo âmbito é listado na página Âmbito com os outros âmbitos que pertencem a esta colecção.

Se quiser que as alterações se façam notar imediatamente, reinicie os servidores de procura.

Conceitos relacionados: "Âmbitos" na página 297 Referências relacionadas: "Formatos de URI no índice" na página 164

URIs contraídos

O IBM Content Analytics with Enterprise Search pode organizar os resultados da procura de modo a que os documentos de origens que tenham o mesmo prefixo de URI surjam contraídos nos resultados da procura.

Quando os resultados são contraídos, normalmente o melhor resultado aparece alinhado à esquerda. Um ou mais resultados da classificação inferior são agrupados e indentados abaixo do melhor resultado.

Para contrair documentos dos resultados que têm prefixos URI diferentes como um único grupo, pode utilizar os prefixos URI com um nome do grupo que cria. Por exemplo, se tiver três servidores para administrar dados financiais, pode agrupar documentos dos três servidores nos resultados da procura e contrair os resultados da classificação inferior abaixo dos documentos dos resultados superiores.

As aplicações de procura podem utilizar o prefixo URI ou o nome do grupo para contrair documentos nos resultados da procura como, por exemplo, mostrando apenas os dois principais documentos de resultados da procura. Se forem devolvidos mais do que dois documentos de resultados com o mesmo prefixo URI (ou documentos que pertençam ao mesmo grupo de URIs), pode seleccionar uma opção para visualizar os resultados contraídos.

Para pesquisar todos os documentos que estejam no mesmo grupo que o prefixo de URI especificado na consulta, utilize a sintaxe de consulta (samegroupas:prefixo URI).

Como organizar prefixos URI e nomes do grupo

Quando utilizar a consola de administração para configurar regras para contrair resultados da procura, especifique os prefixos URI dos documentos que pretende contrair e, opcionalmente, utilize os prefixos URI com um nome do grupo.

A ordem dos prefixos URI que configura é importante. O servidor de controlador utiliza a ordem dos prefixos URI quando calcula o valor de cada URI numa colecção. Para cada URI:

- 1. O servidor de controlador examina sequencialmente os prefixos URI nas regras para contrair os resultados da procura.
- 2. Quando o servidor de controlador encontra o primeiro prefixo URI que corresponde a um prefixo de um documento no índice, associa o nome do grupo (ou o prefixo URI, se a regra não especificar um nome do grupo) como um termo de procura extra para o documento.

Se não for possível corresponder um documento da Web a um prefixo URI, o servidor de controlador utiliza o nome do sistema central do URL como prefixo URI. Se não for possível corresponder um documento da NNTP a um prefixo URI, o servidor de controlador utiliza o primeiro ID de mensagem existente no valor do cabeçalho de referência como prefixo URI.

Depois de adicionar um prefixo URI à lista daqueles que devem ser contraídos nos resultados da procura, tem de posicionar o prefixo URI na ordem que pretende que o servidor de controlador o examine e, potencialmente, utilize como termo de procura extra com documentos no índice:

- Quando adiciona um prefixo URI e não o associa a um nome de grupo, pode seleccionar o prefixo URI individual e movê-lo para cima e para baixo na lista.
- Quando adiciona um prefixo URI e o associa a um nome de grupo, o grupo inteiro de prefixos URI que pertençam ao mesmo grupo é movido sempre que mover um prefixo URI para cima e para baixo na lista. A ordem dos prefixos URI num grupo não é relevante; a selecção de um prefixo URI individual selecciona automaticamente todo o grupo.

Tarefas relacionadas:

"Contrair URIs nos resultados da procura"

Contrair URIs nos resultados da procura

Pode especificar as opções para agrupar e contrair os documentos dos resultados das origens que tenham o mesmo prefixo URI. Também pode criar um nome do grupo que permita contrair em conjunto os documentos dos resultados com prefixos URI diferentes.

As alterações que efectuar para contrair os resultados da procura só entram em vigor da próxima vez que o índice for reconstruído.

Para especificar opções para contrair resultados da procura:

- 1. Na página Colecções, expanda a colecção que pretende configurar.
- No painel Procurar para uma colecção de procura empresarial ou no painel Análise para uma colecção de análise de conteúdos, faça clique em Configurar > Resultados da procura contraídos.
- 3. Na página Contrair Resultados da Procura, faça clique em Adicionar Prefixo URI.
- 4. Na página Adicionar um Prefixo URI para Contrair Resultados, escreva o prefixo URI para os documentos que pretende contrair nos resultados da procura. Por exemplo:

```
http://finance/ROI/
http://server1.com/finance/
db2://LOCALDB/SCHEMA1.TABLE1/
exchange://exchangesvr.ibm.com/public/TeamRoom/Pasta1/
```

- 5. Pode escrever um nome do grupo descritivo que pretende utilizar com este prefixo URI. Para contrair os documentos dos resultados de várias origens como um único grupo, escreva o mesmo nome do grupo quando adicionar cada prefixo URI.
- 6. Faça clique em OK.
- 7. Na página Contrair resultados da procura, posicione a nova regra na ordem que pretende que o serviço de índice o digitalize:
 - Se adicionou um prefixo URI e não o utilizou com o nome do grupo, o novo prefixo URI aparece no fim da lista. Utilize as teclas de seta para o mover para a posição correcta.
 - Se tiver utilizado o novo prefixo URI com um nome do grupo, o novo prefixo URI aparece no fim do conjunto de prefixos URI que pertencem ao mesmo grupo. Utilize as teclas de seta para mover todo o grupo de prefixos URI para a posição correcta.

- 8. Para alterar o prefixo URI ou o nome do grupo, seleccione o prefixo URI e faça clique em *∠* Editar.
- 9. Para remover um prefixo URI da lista, seleccione o prefixo URI e faça clique em **Remover**.

"URIs contraídos" na página 299

Suporte de introdução adiantada para consultas

Se activar o suporte de introdução adiantada para uma colecção, os utilizadores podem ver potenciais correspondências para termos da consulta à medida que escrevem uma consulta e seleccionam uma das correspondências sugeridas para executar a consulta.

Pode especificar se as sugestões serão feitas de consultas anteriores, de termos no índice ou de ambas as origens. Os utilizadores podem escolher as sugestões que pretendem ver quando configurarem preferências na aplicação.Se optarem por ver sugestões de ambas as origens, também podem especificar a ordem em que as sugestões serão apresentadas.

Cada correspondência sugerida inclui o número de resultados potenciais que poderão ser devolvidos. Este número é uma estimativa que difere do número actual de resultados que podem ser devolvidos.

Restrição: O suporte para introdução adiantada não suporta consultas em hebraico e tailandês.

Sugestões de consultas anteriores

Para activar os utilizadores a verem sugestões de introdução adiantada que sejam concebidas de consultas anteriores, tem de activar a colecção a utilizar o índice de registo da consulta quando criar a colecção.

Apenas as consultas que devolvem resultados são armazenadas no índice do registo da consulta. Nos resultados da pesquisa, as sugestões são armazenadas por frequência, com as consultas que são submetidas com mais frequência no topo da lista.

Quando configurar as opções de introdução adiantada para a colecção, especifique quantos dias as consultas serão retidas no índice de registo da consulta. O número que especificar depende do carregamento da consulta na sua empresa. Por exemplo, se tiver um número baixo de consultas, poderá especificar um grande número de dias para facultar um conjunto ainda maior para localizar potenciais correspondências. Se especificar um número de dias baixo, os utilizadores podem não ver tantas sugestões. Por exemplo, se especificar dois dias, os utilizadores vêem sugestões de consultas submetidas ontem e hoje, mas essas consultas não estão disponíveis como potenciais correspondências depois de amanhã.

Sugestões de termos no índice

Nos resultados da pesquisa, as sugestões são ordenadas por frequência. A frequência indica quantos documentos no índice contêm o termo sugerido. Sugestões com a frequência mais elevada surgem no topo da lista.

Se for configurada uma colecção para utilizar segmentação n-gram, a opção de introdução adiantada predefinida vai mostrar sugestões de consultas anteriores. Se as sugestões vão ou não surgir de termos no índice depende do método de segmentação n-gram:

- Se o método se baseia em idiomas que não usam espaços em branco como delimitadores de palavras, o processamento de documentos considera sequências sobrepostas de um número de caracteres como uma única palavra. Neste caso, as sugestões de introdução adiantada não podem surgir de termos no índice. As sugestões devem surgir de consultas anteriores.
- Se o método se baseia em segmentação morfológica e de n-gram (também conhecida como segmentação híbrida), o processamento de documentos inclui segmentação específica de idioma, segmentação baseada no dicionário e segmentação n-gram baseada em regras. Neste caso, os tokens morfológicos são indexados. As sugestões de introdução adiantada podem surgir de termos no índice e de consultas anteriores.

Procura segura

Se a segurança estiver activa para uma colecção, é desactivado por predefinição o suporte de introdução adiantada. Esta abordagem destina-se a impedir que os utilizadores vejam sugestões de documentos que não estejam autorizados a aceder. Se activar o suporte de introdução adiantada para colecções protegidas, poderá querer especificar filtros para evitar que sejam apresentadas sugestões que não pretende que os utilizadores vejam.

Devem ser especificados filtros como expressões regulares. Por exemplo, se quiser suprimir sugestões que comecem por pagar e impedir que sejam mostradas como potenciais correspondências em termos de consulta de utilizadores, insira a expressão regular pagar.*. Consulte a documentação Java sobre a construção de expressões regulares para obter mais informações e exemplos.

Suporte de colecção federada

Quando os utilizadores pesquisarem várias colecções, podem surgir sugestões de todas as colecções. Caso ocorram os mesmos termos em várias colecções, as frequências são adicionadas em conjunto para produzir a contagem da frequência final. Por exemplo:

A colecção A devolve os resultados seguinte:

cerca de	30	resultados
adicionais	20	resultados
contra	10	resultados

A colecção B devolve os seguintes resultados:

contabiliza	30	resultados
cerca de	20	resultados
após	10	resultados

Estes resultados são combinados quando as sugestões de introdução adiantada são ordenados, de forma a que os utilizadores vejam as seguintes sugestões de introdução adiantada:

50	resultados
30	resultados
20	resultados
10	resultados
10	resultados
	50 30 20 10 10

Configurar suporte de introdução adiantada

Quando configurar os servidores de procura para uma colecção, pode especificar opções para mostrar potenciais correspondências para termos da consulta à medida que o utilizador escreve uma consulta.

Para activar o suporte de introdução adiantada para consultas, tem de activar o índice do registo de consultas para a colecção quando criar a colecção.

Para activar o suporte de consulta de introdução adiantada para uma colecção:

- Expanda a colecção que pretende configurar. No painel Procura para uma colecção de procura empresarial ou no painel Análise para uma colecção de análise de conteúdos, faça clique em Acções > Configurar > Opções de introdução adiantada.
- 2. Especifique durante quanto tempo pretende manter as consultas submetidas pelos utilizadores no índice do registo de consultas. O valor que especificar influencia a quantidade de correspondências potenciais que se baseiam em consultas anteriores que estão disponíveis para serem mostradas a utilizadores à medida que escrevem consultas.
- 3. Seleccione os quadrados de confirmação para os tipos de sugestões que pretende facultar. Pode mostrar aos utilizadores sugestões que se obtêm de consultas anteriores que são armazenadas no registo da consulta e mostram aos utilizadores sugestões que se obtêm de termos no índice. Quando os utilizadores configuram preferências na aplicação, podem substituir as selecções que efectuar aqui. Por exemplo, os utilizadores podem escolher se pretendem ver sugestões de introdução adiantada e, se for o caso, se pretendem ver sugestões de consultas anteriores, de termos no índice ou de ambas as origens.
- 4. Opcional: Caso pretenda impedir que determinadas consultas ou termos do índice sejam sugeridos aos utilizadores, pode escrever padrões de expressões regulares para remover as sugestões. Separar vários filtros com um carácter de barra vertical (|). Por exemplo, para impedir cadeias que comecem por ibm sejam apresentadas como potenciais correspondências para os termos da consulta dos utilizadores, insira a expressão regular ibm.*.

Consulte a documentação Java sobre a construção de expressões regulares para obter mais informações e exemplos.

Configurar sinalizadores de documentos

Pode configurar sinalizadores que pode ser usados pelos utilizadores para classificar documentos quando consultam uma colecção. Por exemplo, pode criar um sinalizador denominado Confidencial para que os utilizadores possam sinalizar todos os documentos empresariais confidenciais quando revêem os resultados da consulta. Pode definir até 64 sinalizadores por colecção.

Por predefinição, nenhum utilizador tem privilégios de sinalização de documentos. O modo como autoriza utilizadores depende de os utilizadores terem iniciado sessão na aplicação:

- Se a aplicação de segurança de início de sessão for activada, abra a página Segurança e aceda à área de segurança ao nível do sistema. Faça clique em Acções > Adicionar um utilizador de aplicação ou grupo, especifique um ID de utilizador ou ID de grupo e seleccione a caixa de verificação para gerir sinalizadores de documentos.
- Se a aplicação de segurança do inicio da sessão não for activada, ou se quiser conceder privilégios a utilizadores da aplicação sem adicionar IDs de

utilizadores ou nomes de grupos ao sistema, atribua o privilégio de gerir sinalizadores de documentos seleccionando **Acções** > **Especificar privilégios do utilizador na aplicação predefinida**.

Para configurar sinalizadores de documentos:

- Na vista Colecções, expanda a colecção que pretende configurar. No painel Procurar para uma colecção de procura empresarial ou no painel Análise para uma colecção de análise de conteúdo, faça clique em Configurar > Sinalizadores de documentos.
- 2. No campo Etiqueta raiz na página Sinalizadores de documentos, especifique o nome que pretende que apareça no painel de navegação de faceta da aplicação como entrada de nível superior. Por exemplo, se especificar Sinalizadores, todos os sinalizadores que configurar são listados sob esta entrada no painel de navegação da faceta. Os utilizadores podem seleccionar o sinalizador de nível superior para explorar a colecção. Por exemplo, na vista Facetas, no prospecção de análise de conteúdo, os utilizadores podem ver uma lista de todos os sinalizadores e respectivas frequências.
- **3**. Seleccione um sinalizador e faça clique nos botões para o mover para cima ou para baixo na lista, ou para o início ou fim da lista. A ordem pela qual dispõe os sinalizadores aqui controla a ordem pela qual são listados quando os utilizadores exploram resultados da consulta e gerem sinalizadores.

Após concluir a configuração dos sinalizadores, as alterações são imediatamente aplicadas.

Tarefas relacionadas:

"Exportar documentos de um colecção de procura empresarial para uma colecção de análise de conteúdos"

Exportar documentos de um colecção de procura empresarial para uma colecção de análise de conteúdos

Pode exportar documentos de um colecção de procura empresarial para uma colecção de análise de conteúdos. Após ter aplicado recursos de análise para analisar os documento exportados, pode utilizar o prospecção de análise de conteúdo para explorar o conteúdo analisado.

Por exemplo, pode ter uma colecção de procura empresarial que contém informações sobre investigações de acidentes. Ao procurar a colecção pode encontrar registos da investigação, mas com a prospecção da colecção pode identificar correlações, desvios e tendências que o podem ajudar a obter conhecimento das principais causas dos acidentes.

Se criar uma colecção através da exportação de documentos sinalizados, os dados de configuração seguintes também serão exportados da colecção de procura empresarial:

- Definições de campo definidas pelo utilizador
- · Correlacionamentos de campos entre campos do pesquisador e campos de índice
- Definições da faceta. As facetas são copiadas como categorias para a colecção de análise de conteúdos. As facetas simples e as facetas de intervalo são copiadas sob a categoria Raiz pelos mesmos correlacionamentos de nome e campo. As facetas hierárquicas são copiadas como uma categoria segundo estas regras:
 - É criada uma categoria com o mesmo nome sob a categoria Raiz. O caminho da faceta é criado automaticamente, tais como faceta0, faceta1, etc.

 Sob esta categoria, é criada uma categoria para cada campo que está associado à faceta e a estrutura hierárquica é inalterável. O nome da categoria e o caminho da faceta correspondem ao nome do campo.

Se exportar documentos sinalizados numa colecção existente, não são exportados dados da configuração da colecção de procura empresarial. Caso necessário, tem de editar a colecção de análise de conteúdo, para configurar definições de campo e faceta para os documentos exportados.

Esta tarefa envolve vários passos executados pelos administradores e utilizadores da aplicação:

- Um administrador configura uma colecção de procura empresarial para suportar a exportação de documentos para uma colecção de análise de conteúdos e define sinalizadores para sinalizar os documentos a exportar. Um administrador também autoriza que utilizadores específicos ou todos os utilizadores exportem os documentos para a colecção de análise de conteúdos.
- 2. Os utilizadores autorizados sinalizam os documentos que pretendem exportar e submetem o pedido de exportação.
- 3. Um administrador correlaciona campos nos documentos exportados para campos de índice para que o conteúdo possa ser analisado pelos recursos de análise. O administrador pode supervisionar pedidos para verificar o estado e o histórico da actividade de exportação e para visualizar quais foram os sinalizadores seleccionados pelo utilizador.
- 4. Os utilizadores exploram a colecção de análise de conteúdos com o prospecção de análise de conteúdo.

Para exportar documentos para uma colecção de análise de conteúdos e explorar o conteúdo analisado:

- 1. Configure uma colecção de procura empresarial para suportar a exportação para uma colecção de análise de conteúdos:
 - a. Inicie sessão na consola de administração.
 - b. Na vista Colecções, expanda uma colecção de procura empresarial. Nas definições da colecção, certifique-se de que a colecção está configurada para utilizar a cache de documentos e que a sinalização dos documentos está activada. Ambas as opções têm de estar activadas quando cria a colecção.
 - c. No painel Procura, faça clique em Configurar > Sinalizadores de documentos e crie sinalizadores para identificar os documentos que planeia exportar desta colecção.
 - d. No painel Procura, faça clique em **Configurar** > **Exportar para uma colecção de análise de conteúdos**.
 - e. Na página Exportar os documentos sinalizados para uma colecção de análise de conteúdos, especifique que pretende permitir que os documentos sejam exportados.
 - Se pretende criar uma colecção para os documentos exportados, especifique o nome da colecção. Especifique outras opções para a colecção da mesma forma que cria uma colecção de análise de conteúdos na vista Colecções.
 - Se pretende exportar documentos para colecções existentes, seleccione as colecções para as quais os utilizadores estão permitidos a exportar documentos. Os utilizadores podem seleccionar uma única colecção quando submetem um pedido de exportação.
 - f. Abra a vista Segurança. Na área Segurança ao nível do sistema, adicione um utilizador ou grupo da aplicação ou especifique privilégios de utilizadores

de aplicação predefinidos. Certifique-se de que a caixa de verificação **Exportar os documentos sinalizados para uma colecção de análise de conteúdos** está seleccionada. Se atribuir privilégios a um utilizador ou grupo, especifique a aplicação de procura empresarial na qual o privilégio deve ser activado.

- 2. Exporte documentos para uma colecção de análise de conteúdos:
 - a. Abra a aplicação de procura empresarial para procurar a colecção de procura empresarial que configurou.
 - b. Execute uma consulta para aperfeiçoar os resultados. Seleccione e limpe sinalizadores para identificar os documentos que pretende exportar.
 - c. Faça clique no ícone para exportar documentos sinalizados para uma colecção de análise de conteúdos. Seleccione a colecção de destino a partir da lista de colecções de análise de conteúdos disponível e seleccione os sinalizados para os documentos que pretende exportar. Especifique se os documentos exportados anteriormente para a colecção de destino devem ser substituídos quando estes documentos forem exportados e, em seguida, faça clique em Submeter.
- **3**. Supervisione o estado do pedido de exportação e edite os correlacionamentos de campo para os documentos exportados:
 - a. Inicie sessão na consola de administração.
 - b. Na vista Colecções, expanda a colecção de procura empresarial que configurou previamente. Na área Exportar para a colecção de análise de conteúdos do painel Procura, faça clique no ícone para supervisionar o progresso do pedido de exportação. Para cada pedido de exportação, pode visualizar as informações de estado, os sinalizadores que o utilizador seleccionou, o nome das colecções de análise de conteúdos de destino e quantos documentos foram exportados. Também pode eliminar o histórico dos pedidos de exportação.
 - c. Expanda a colecção de análise de conteúdos para a qual os documentos foram exportados. Na área Pesquisadores do painel Pesquisar e importar, faça clique em Configurar > Correlacionamento de campos para documentos de chegada.
 - d. Na página Correlacionar campos para documentos de chegada, faça clique no ícone **Editar** para cada pesquisador para visualizar como os campos pesquisados foram correlacionados par campos de índice. Se um pesquisador foi eliminado da colecção de procura empresarial, o pesquisador pode ainda ser apresentado nesta lista se quaisquer documentos adicionados pelo pesquisador estiverem incluídos no conjunto de documentos exportados.
 - e. Na página Configuração de correlacionamento de campos, adicione, altere ou remova correlacionamentos de campos como for necessário para os documentos que foram exportados para esta colecção. As definições são aplicadas a documentos que são exportados após ter guardado os correlacionamentos de campo. Se pretende aplicar as respectivas alterações a documentos que foram indexados anteriormente, reconstrua o índice.
- 4. Explore o conteúdo analisado:
 - a. Abra o prospecção de análise de conteúdo e explore a colecção de análise de conteúdos para a qual os documentos foram exportados.
 - b. Para obter perspectivas adicionais ou diferentes, pode querer exportar documentos adicionais ou diferentes (passo 2). Também pode querer editar os correlacionamentos de campo para alterar a forma como o conteúdo é indexado (passo 3).

"Campos de índice" na página 191

Tarefas relacionadas:

"Criar uma colecção" na página 51

"Configurar sinalizadores de documentos" na página 303

Analisar conteúdo

Inspecção profunda

Ao explorar uma colecção com a prospecção de análise de conteúdo, pode submeter documentos que correspondam às respectivas condições de procura para uma inspecção aprofundada. As inspecções profundas podem ajudar periodicamente a detectar alterações efectuadas numa colecção ao longo do tempo.

As inspecções profundas são úteis se quiser explorar os resultados da análise e verificar valores de correlação ou o índice de desvio para consultas que envolvam um grande número de valores de faceta ou subfacetas. Por exemplo, poderá ser difícil explorar, de modo interactivo, um grande número de valores de facetas ou poderá haver demasiados valores de facetas para serem apresentados pela aplicação. Ao exportar resultados de análise para fins de inspecção profunda, poderá evitar tempos esgotados de pedidos interactivos. Também poderá agendar a execução do pedido em alturas de menor movimento, quando se podem dedicar mais recursos de sistema ao processamento do pedido.

Ao configurar as opções de exportação dos resultados da inspecção profunda na consola de administração, poderá configurar agendas para controlar o momento de exportação dos resultados. Poderá criar uma agenda geral para todos os pedidos de exportação e configurar agendas personalizadas para pedidos individuais. Ao configurar opções de inspecção profunda:

- Pode activar ou desactivar a capacidade para submeter pedidos de inspecção aprofundada a partir da prospecção de análise de conteúdo.
- Poderá especificar se pretende exportar os resultados como ficheiros XML ou CSV, ou se pretende utilizar um suplemento personalizado para controlar como e onde são exportados os resultados.
- Se exportar resultados de análise como ficheiros XML ou CSV, terá de especificar o caminho de saída de dados onde serão gravados os dados exportados. Se utilizar um suplemento personalizado, terá de especificar o caminho da classe, nome da classe e propriedades que pretende transmitir ao suplemento.

Quando um utilizador submete um pedido de inspecção profunda, as opções disponíveis para processar esse pedido dependem da vista em que o utilizador se encontrar. Estão disponíveis as seguintes opções para as vistas Facetas, Séries cronológicas, Desvios, Tendências e Pares de Facetas:

- O utilizador pode aceitar o ID de pedido predefinido ou especificar outro ID a utilizar.
- O utilizador pode especificar uma descrição. A descrição poderá ajudar um administrador a identificar o pedido, quando este se encontra em fila de processamento, em vez de identificar o pedido apenas pelo seu ID.
- O utilizador pode especificar o número máximo de resultados a analisar. Os resultados de topo, quando ordenados por contagem de frequência ou pelo índice calculado ou valor de correlação, são incluídos no relatório, e o utilizador pode especificar qual o limite de resultados de topo.

• O utilizador pode especificar o número máximo de valores de faceta ou subfacetas que devem ser contadas, de modo a assegurar que apenas as facetas de valor e subfacetas de topo e mais frequentes são inspeccionadas.

Na vista Pares de Facetas, os utilizadores podem especificar valores de contagem máxima diferentes para valores de faceta e de subfacetas representados como linhas na vista de tabela, bem como valores de facetas e subfacetas representadas como colunas na vista de tabela.

- O utilizador pode escolher ordenar os resultados da inspecção pela contagem de frequência (o número de documentos que contêm um valor de faceta ou subfaceta), pelo valor de índice calculado (quando for pedida uma inspecção profunda de desvios ou de tendências) ou pelo valor da correlação calculado (quando for pedida uma inspecção profunda de facetas ou de pares de facetas).
- O utilizador pode especificar um limiar para excluir documentos da saída de dados. Os documentos que fiquem aquém deste limiar, segundo a ordenação especificada, não são incluídos nos resultados da inspecção.
- O utilizador pode especificar se o pedido de inspecção pode ou não ser agendado. Se agendado, o pedido é processado segundo a agenda que for especificada para processamento de pedidos de inspecção profunda na consola de administração. Se o pedido não for agendado, ou se o administrador não tiver configurado nenhuma agenda, o pedido entra na fila para processamento imediato.

Conceitos relacionados:

Criar e implementar um suplemento para exportar documentos ou resultados de inspecções profundas

Exportar resultados de inspecção profunda

Quando os utilizadores exploram uma colecção de análise de conteúdo, podem pedir uma inspecção profunda do conteúdo analisado. As opções que forem especificadas aqui controlam o modo como os resultados desta inspecção profunda são exportados.

Os resultados de inspecção profunda estão disponíveis para colecções de análise de conteúdo. Esta opção não está disponível para colecções de procura.

Para configurar opções de exportação de resultados de inspecção profunda:

- Na consola administrativa, expanda uma colecção de análise de conteúdo. No painel Análise, faça clique no ícone para configurar as definições de inspecção profunda.
- 2. Na página Inspecção Profunda, especifique as suas preferências de exportação:
 - Se exportar os resultados de análise como ficheiros XML ou CSV, terá de especificar o caminho de saída onde serão gravados os dados exportados, como, por exemplo, /home/esadmin/output/.
 - Se utilizar um suplemento personalizado, terá de especificar o caminho de classes, nome da classe, e propriedades que pretende transmitir ao suplemento personalizado.
- Opcional: Para agendar a execução do programa, expanda a colecção. Na área Inspecção Profunda do painel Análise, faça clique no ícone para configurar um agendamento para processar os pedidos de inspecção profunda.
 - Poderá especificar uma agenda geral que controle a frequência com que o programa de inspecção é executado como, por exemplo, em dias da semana

ou do mês específicos, ou horas do dia específicas. Todos os pedidos agendados que utilizem a agenda geral e estejam na fila são processados à hora especificada.

- Também poderá desactivar e activar a agenda para pedidos individuais, configurar agendas personalizadas para pedidos individuais, ou eliminar pedidos da fila de pedidos de inspecção.
- 4. Submeter um pedido de inspecção profunda:
 - a. Abra o prospecção de análise de conteúdo. Seleccione facetas e especifique condições de pesquisa para reduzir os resultados a documentos que pretenda analisar.
 - b. Faça clique no ícone para executar um relatório de inspecção profunda e depois especifique opções para esse relatório como, por exemplo, se o relatório deve ser criado em formato XML ou CSV. As opções variam consoante a página que estiver a ver ao submeter o pedido.

O pedido de inspecção profunda é enviado para o servidor de controlador para fins de processamento, e é apresentado um ID de pedido.

Na vista Colecções, na consola administrativa, é possível monitorizar o estado do pedido de inspecção profunda. Por exemplo, poderá ver se o pedido foi concluído e se ocorreram erros, e poderá ver um histórico de todos os pedidos de inspecção profunda.É, também, possível configurar um agendamento personalizado para este pedido, para especificar quando é que o pedido deve ser processado.

Conceitos relacionados:

Criar e implementar um suplemento para exportar documentos ou resultados de inspecções profundas

Formato de ficheiros XML para relatórios de inspecção profunda

Quando são exportados relatórios de inspecção profunda, a saída de dados inclui elementos XML relativos aos documentos analisados e saída de dados em XML para os metadados.

Elementos de XML em relatórios de inspecção profunda

Os ficheiros XML são criados no caminho de saída de dados que especificou quando configurou as opções de inspecção profunda na consola de administração. Cada ficheiro recebe o nome report_*.xml, em que * corresponde a um número entre 0000 e 9999.

O ficheiro XML contém os seguintes elementos.

Relatório

O elemento <Relatório> é o elemento de nível de topo do ficheiro XML. Este elemento tem um atributo, **Id**, que identifica o pedido. O identificador é criado quando o utilizador submete um pedido de inspecção profunda oriundo de prospecção de análise de conteúdo.

Registo

É criado um elemento <Registo> para cada documento analisado. Este elemento tem três atributos:

Classificação

A classificação do documento, segundo a ordenação que foi especificada no pedido (contagem da frequência, valor de índice, ou valor de correlação).

Contagem

O número de documentos que contêm valor de faceta ou sub-faceta, ou um par de valores de faceta ou sub-facetas num par de facetas (contagem de frequência).

Índice

O valor de índice ou valor de correlação que foi calculado para o valor da faceta e sub-faceta.

Faceta

Cada elemento <Registo> contém elementos <Faceta> relativos às facetas que estão a ser analisadas no pedido de inspecção profunda. O elemento <Faceta> tem dois atributos:

Facet.1 de Dimensão

Identifica a faceta na linha de uma vista de facetas emparelhadas.

Facet.2(coluna) de Dimensão

Identifica a faceta na coluna de uma vista de facetas emparelhadas.

Segue-se um exemplo de um ficheiro XML criado para um pedido de inspecção profunda:

```
<?xml version="1.0" encoding="UTF-8"?>
<Report Id="Report 1251354396">
  <Record Rank="Facet.1" Count="31808" Index="4.477488456444767">
    <Facet Dimension="Facet.1">group</Facet>
    <Facet Dimension="Facet.2">group ... consisting</Facet>
  </Record>
  <Record Rank="Facet.2" Count="27448" Index="2.160072073496715">
    <Facet Dimension="Facet.1">method</Facet>
    <Facet Dimension="Facet.2">method ... comprising</Facet>
  </Record>
  <Record Rank="3" Count="24605" Index="1.0665513645282887">
    <Facet Dimension="Facet.1">claim</Facet>
    <Facet Dimension="Facet.2">group ... consisting</Facet>
  </Record>
  <Record Rank="4" Count="24081" Index="8.484712768155958">
    <Facet Dimension="Facet.1">design</Facet>
    <Facet Dimension="Facet.2">ornamental ... design</Facet>
  </Record>
  <Record Rank="5" Count="22466" Index="2.267135299905541">
    <Facet Dimension="Facet.1">step</Facet>
    <Facet Dimension="Facet.2">method ... comprising</Facet>
  </Record>
  <Record Rank="6" Count="21089" Index="1.0566644583723699">
    <Facet Dimension="Facet.1">claim</Facet>
    <Facet Dimension="Facet.2">method ... comprising</Facet>
  </Record>
  <Record Rank="7" Count="18917" Index="4.5361071335548795">
    <Facet Dimension="Facet.1">invention</Facet>
    <Facet Dimension="Facet.2">present ... invention</Facet>
  </Record>
  <Record Rank="8" Count="18131" Index="1.2265225719960702">
    <Facet Dimension="Facet.1">method</Facet>
    <Facet Dimension="Facet.2">group ... consisting</Facet>
  </Record>
  <Record Rank="9" Count="17763" Index="2.16027811189025">
    <Facet Dimension="Facet.1">method</Facet>
    <Facet Dimension="Facet.2">method ... according</Facet>
  </Record>
  <Record Rank="10" Count="15542" Index="3.884052442128364">
    <Facet Dimension="Facet.1">system</Facet>
    <Facet Dimension="Facet.2">system ... comprising</Facet>
  </Record>
</Report>
```

XML de metadados em relatórios de inspecção profunda

Este ficheiro XML de metadados inclui informações sobre o pedido de inspecção profunda submetido e os resultados do processamento do pedido. Segue-se um exemplo do ficheiro XML de metadados:

```
<?xml version="1.0" encoding="UTF-8"?>
<Document Id="report 1251354396" Type="NORMAL">
 <Content Truncated="false"
   Path="c:\tmp\content\20090827081557\0\0000 c.xml"
   Encoded="false"></Content>
 <Metadata>
   <Fields>
     <Field Name="QueryText">*:*</Field>
     <Field Name="Facet.1">General Noun</Field>
     <Field Name="Facet.2">Noun Phrase</Field>
     <Field Name="Facet.1.Id">$.tkm_base_word.noun.general</Field>
     <Field Name="Facet.2.Id">$.tkm base word.noun.phrase </Field>
     <Field Name="SortKey">frequency</Field>
     <Field Name="ViewName">2D Map</Field>
     <Field Name="Facet.1.TaxonomyType">facet values</Field>
     <Field Name="Facet.2.TaxonomyType">subfacets</Field>
     <Field Name="Expression">#2dmap_view_correlation</Field>
     <Field Name="StartedDateTime">2009.08.27 08:15:54 PDT</Field>
     <Field Name="CompletionDateTime">2009.08.27 08:15:57 PDT</Field>
     <Field Name="NumberOfRecords">10</Field>
     <Field Name="ResultCode">0</Field>
    </Fields>
    <Facets></Facets>
 </Metadata>
</Document>
```

Conceitos relacionados:

Criar e implementar um suplemento para exportar documentos ou resultados de inspecções profundas
Classificação de documentos

Quando um utilizador consulta uma colecção, os processos de procura devolvem os resultados mais relevantes para a consulta.

O servidor de procura suporta uma sintaxe da consulta excelente e utiliza várias técnicas para produzir os resultados da procura mais relevantes, tais como a classificação baseada em texto e a classificação estática:

- Na pontuação baseada em texto, é dinamicamente calculada uma pontuação para cada documento que corresponde aos termos numa consulta.
- Para certos tipos de documentos, pode configurar a colecção para utilizar um factor de classificação estática que aumente a importância desses resultados nos resultados de procura.

Pode ampliar o comportamento da classificação predefinida, configurando as opções que influenciam a importância dos documentos nos resultados da procura:

- Pode criar dicionários de palavras hierárquicas personalizados para influenciar o procedimento de como os documentos, que contêm as palavras hierárquicas especificadas, estão ordenados nos resultados da procura.
- Pode influenciar as classificações dos documentos que correspondem a um padrão URI especificado.
- Para uma colecção de procura empresarial, pode configurar as regras de expansão de consulta e especificar se os documentos que correspondem aos termos de consulta expandida devem ter uma classificação mais alta ou baixa nos resultados.

Classificação baseada em texto

Uma pontuação é calculada dinamicamente para cada documento que corresponde aos termos numa consulta.

A classificação do texto para um documento representa a importância dos termos da consulta num documento. O cálculo da pontuação final do texto para o documento depende de muitos factores como, por exemplo:

- Se os termos distinguem um documentos de outros documentos. Por exemplo, se os termos da consulta aparecem num documento e não noutros, significa que estes termos são importantes para esse documento e distinguem esse documento de outros. Os termos da consulta que aparecem na maioria dos documentos contribuem menos para uma classificação do documento do que os termos da consulta que aparecem num conjunto mais selectivo de documentos.
- O número de ocorrências dos termos da consulta num documento. A classificação de um termo é calculada com base em cada ocorrência desse termo num documento. Quanto mais vezes os termos da consulta ocorrerem num documento, tanto maior é a classificação do documento. Por exemplo, procure pelo termo thinkpad. Um documento que inclua muitas vezes este termo obtém uma classificação maior nos resultados da procura do que outros documentos que contêm menos ocorrências.
- O peso dos termos da consulta. Para personalizar a importância dos termos num documento,pode configurar os valores hierárquicos nos termos. Neste caso, associa um dicionário de palavras hierárquicas (que contém termos e respectivos valores hierárquicos) à colecção. Este dicionário é utilizado durante a pesquisa e

os valores hierárquicos dos termos no dicionário contribuem para a classificação do documento. Quanto maior o valor hierárquico, tanto maior é a contribuição do termo para a classificação do documento.

 A proximidade dos termos da consulta num documento. Se os termos da consulta aparecem fechados a si mesmo num documento, a sua autorização lexical é utilizada para calcular a classificação do texto. Por exemplo, assumindo que o utilizador tem dois documentos. Um documento fala sobre um parque de carros na cidade (carro e parque estão perto um do outro). O outro documento fala sobre um salão de exposição de carros numa cidade próxima do parque (carro e parque não estão juntos). Se procurar por parque de carros, a proximidade dos termos no primeiro documento fará com que esse documento obtenha uma classificação superior ao segundo documento.

Influenciar a pontuação final

Quando configurar as opções do servidor de pesquisa para uma colecção, pode especificar um factor adicional para influenciar a pontuação final do texto. Este factor é utilizado com a pontuação estática, que deriva da importância do documento configurado para a colecção (pelo número de ligações ao documento, pela data do documento ou pelo valor de um campo paramétrico), de forma a aumentar ou diminuir a pontuação final dos documentos que correspondem aos critérios de importância do documento.

Por exemplo, parta do princípio que a colecção determina a importância do documento pelo número de ligações ao mesmo. Quando configurar opções do servidor de pesquisa, irá verificar que a pontuação original é 10. Para assegurar que os documentos com muitas ligações são classificados como altamente relevantes nos resultados da pesquisa, pode aumentar este valor e especificar que a pontuação estática representa 80 por cento da pontuação total do documento.

Caso tenha importado pontuações de categorias do IBM Content Classification, também pode especificar se pretende usar a pontuação para influenciar a forma como os documentos são classificados. Nesse caso, especifique um factor adicional para controlar a forma como a pontuação da categoria influencia a pontuação final. A pontuação da categoria pode ajudar a assegurar que os documentos que pertencem a uma categoria do Content Classification são classificados como altamente relevantes nos resultados da pesquisa.

Conceitos relacionados:

"Influenciar o rendimento da consulta e a correcção ortográfica" na página 286

Classificação estática e importância de documentos

Em certos tipos de documentos, pode utilizar um factor de classificação estática que aumente a importância daqueles documentos nos resultados da procura.

Ao criar uma colecção ou ao editar as definições da colecção, pode especificar as opções **Importância de documento**. O tipo de importância do documento que selecciona determina se o factor de classificação estatística é associado aos documentos na colecção. Quando os utilizadores procuraram a colecção que utiliza a classificação estatística, o factor de classificação estatística influencia a classificação dos documentos devolvidos nos resultados da procura.

Para conteúdo da Web, o factor de classificação estática pode ser baseado em ligações. O número de ligações a um documento a partir de outros documentos, e as origens dessas ligações, podem aumentar a relevância desse documento nos resultados da procura.

Para documentos que incluam os campos de data ou metadados de data, o factor de classificação estática pode ser baseado na data do documento. O campo da data do documento, que é fornecido pelo pesquisador, pode ser a data em que o documento foi modificado pela última vez, dependendo da forma como define a configuração do pesquisador.

A data de um documento pode aumentar a sua relevância. Por exemplo, os artigos recentes em grupos NNTP novos podem ser mais relevantes do que os artigos antigos. Se uma origem de dados incluir valores de data múltiplos, pode escolher qual é o mais importante para determinar a relevância de documentos quando configura o pesquisador.

Para documentos que incluem campos e campos de metadados que contêm dados numéricos, o factor de classificação estática pode ser baseado num campo paramétrico, como o tamanho do documento ou a data em que o documento foi publicado.

Se utilizar a classificação estática com uma colecção, certifique-se de que não mistura origens de dados que utilizam tipos diferentes de classificação na mesma colecção. Por exemplo, se pretender utilizar as ligações a um documento como factor de classificação estática, certifique-se de que a colecção contém apenas documentos da Web. A classificação de documentos é menos precisa quando as origens com modelos de classificação diferente são combinadas na mesma colecção e a ordem dos resultados da procura pode não ser como esperada.

Deve certificar-se também de que os documentos na colecção contêm campos e valores que permitem que a classificação estática seja aplicada. Por exemplo, imagine uma colecção que é configurada para utilizar classificação estatística baseada em datas de documentos e o pesquisador na colecção é configurado para utilizar um campo específico, como a data do documento. Se um documento não contiver esse campo, a importância do documento poderá não ser correctamente classificada e a ordem dos resultados da procura poderá não ser como esperada.

Implicações da classificação baseada em ligações

A classificação estatística, juntamente com factores como atribuir uma classificação a padrões de URI hierárquicos, contribui para a classificação estatística de um documento e influencia a importância do documento. O modelo de classificação baseada em ligações é geralmente aplicado a colecções da Web, porque este modelo calcula a classificação estatística de um documento baseado no número de ligações para o documento. Um documento que está ligado a, a partir de um número maior de outros documentos é classificado como mais relevante.

Por esta razão, se configurar este modelo para uma colecção que não é da Web ou uma colecção mista (uma que contenha documentos da Web e não da Web), a qualidade da procura poderás ser degradada, porque os documentos que não são da Web não têm qualquer conceito de ligação.

Ao criar uma colecção e seleccionar a opção para classificar pelo número de ligações aos documentos, ocorre automaticamente a análise das ligações. Na análise de ligações, os documentos ligados são automaticamente recuperados e indexados

com o texto âncora na ligação. Em consequência, os documentos ligados são pesquisáveis por palavras-chave que ocorrem no texto âncora, mesmo que as palavras-chave não apareçam no documento de origem.

Classificação de documentos através de dicionários de palavras hierárquicas

Para melhorar a qualidade dos resultados da procura, pode influenciar a forma como os documentos estão classificados nos resultados da procura ao criar um dicionário de palavras hierárquicas personalizadas.

Se uma consulta especificar uma palavra contida no dicionário de palavras hierárquicas, a importância dos documentos que contêm essa palavra será aumentada ou reduzida de acordo com o factor hierárquico configurado para a palavra no dicionário.

Pode utilizar um dicionário de palavras hierárquicas para garantir que determinados documentos são devolvidos quando um utilizador especifica certos termos na consulta. Por exemplo, assumindo que o utilizador tem uma colecção que contém muitos documentos que falam sobre carros. Para tais documentos, poderá pensar que certas palavras-chave relacionadas com os modelos de carros, o nome do fabricante, entre outras, são importantes. Para influenciar a classificação dos resultados da procura, pode atribuir um grau de importância associando um valor hierárquico às palavras-chave (modelo, fabricante, etc.), num dicionário de palavras hierárquicas. Quando os utilizadores procuram a colecção e especificam uma consulta que inclua qualquer uma das palavras-chave, os documentos sobre carros são ordenados mais acima nos resultados da procura do que outros documentos.

O intervalo dos factores hierárquicos é de -10 para 10. Durante o processamento da procura, os servidores da procura aumentam a importância de documentos que contêm palavras com factores hierárquicos positivos, e reduzem a importância de documentos que contêm palavras com factores hierárquicos negativos.

Por exemplo, um documento que corresponda aos termos da consulta com factores hierárquicos elevados tem uma classificação mais elevada do que aquele que teria se o factor hierárquico não fosse aplicado. (O factor hierárquico é o único factor que contribui para a classificação do documento.)

Ao criar um dicionário, pode atribuir o mesmo factor hierárquico a qualquer número de palavras. O dicionário pode conter um termo de palavra única ou um termo de palavras múltiplas. Os termos de palavras múltiplas são correlacionados como uma frase.

Se uma palavra ponderada por um valor hierárquico for especificada na consulta que utiliza o operador OR (por exemplo: este | esse), será calculada uma média ponderada para os termos de consulta. A classificação agregada resultante é utilizada para todas as ocorrências dos operandos de consulta OR. Não são calculadas classificações diferentes para operandos de consulta OR diferentes.

Para criar um dicionário de palavras hierárquicas, um especialista em matéria de colecção necessita de criar uma lista de palavras hierárquicas em formato XML ou trabalhar com um programador de aplicações para criar o ficheiro de XML. Uma ferramenta facultada, **esboosttermdictbuilder**, deve ser utilizada para converter o ficheiro XML num ficheiro binário (.dic).

Um administrador principal carrega o ficheiro binário no sistema e atribui-lhe um nome de apresentação. Os administradores de colecção podem seleccionar um dicionário de palavras hierárquicas para procurar documentos numa colecção quando configuram opções do servidor da procura para uma colecção.

Restrição: Depois de adicionar um dicionário de palavras hierárquicas personalizadas ao sistema, não pode editá-lo. Para rever as palavras hierárquicas disponíveis para o processamento da consulta, tem de:

- 1. Actualizar o ficheiro XML de origem.
- 2. Converter a origem de XML num novo ficheiro de dicionário.
- **3**. Remover o dicionário de palavras hierárquicas antigo das colecções que o utilizam.
- 4. Eliminar o dicionário de palavras hierárquicas antigo do sistema.
- 5. Adicionar o dicionário de palavras hierárquicas novo ao sistema.
- 6. Utilizar o dicionário de palavras hierárquicas novo com as colecções que o utilizam.

É possível escrever um script que inclua estes passos e, em seguida, utilizar o script para voltar a implementar o dicionário no sistema.

Conceitos relacionados:

Personalizar dicionários de palavras hierárquicas

Tarefas relacionadas:

"Adicionar dicionários de palavras hierárquicas ao sistema"

"Associar um dicionário de palavras hierárquicas a uma colecção" na página 318

🕒 Criar um ficheiro XML para palavras hierárquicas

🕩 Criar um dicionário de palavras hierárquicas

Adicionar dicionários de palavras hierárquicas ao sistema

Se criar dicionários de palavras hierárquicas personalizadas, tem de associar os dicionários ao sistema. Mais tarde, pode seleccionar que dicionário de palavras hierárquicas pretende utilizar para procurar uma colecção.

O tamanho máximo de um dicionário de palavras hierárquicas é 8 MB.

Associar um dicionário de palavras hierárquicas a um sistema:

- 1. Faça clique em Sistema para abrir a vista Sistema.
- 2. Na página Procura, faça clique em **Configurar dicionários de palavras** hierárquicas.
- **3**. Na página Configurar dicionários de palavras hierárquicas, faça clique em **Adicionar dicionário de palavras hierárquicas**.
- 4. Na página Adicionar dicionário de palavras hierárquicas, escreva o único nome de apresentação do dicionário e, como opção, escreva uma descrição.
- 5. Especifique a localização do ficheiro .dic. Se o ficheiro se encontrar no sistema local, pode navegar até o localizar. Se o ficheiro se encontrar no servidor de controlador, escreva o caminho totalmente qualificado.
- 6. Faça clique em **OK**. O dicionário de palavras hierárquicas personalizado é adicionado ao sistema e fica disponível para associar a colecções.

Conceitos relacionados:

"Classificação de documentos através de dicionários de palavras hierárquicas" na página 316

Tarefas relacionadas:

"Associar um dicionário de palavras hierárquicas a uma colecção"

- 🕩 Criar um ficheiro XML para palavras hierárquicas
- 🕩 Criar um dicionário de palavras hierárquicas

Associar um dicionário de palavras hierárquicas a uma colecção

Se os dicionários de palavras hierárquicas estiverem associados ao sistema, pode seleccionar um a utilizar ao procurar uma colecção. Se um termo de consulta corresponder a um termo no dicionário, a importância dos documentos que contêm esse termo será aumentada ou reduzida de acordo com o factor hierárquico atribuído ao termo no dicionário.

Para utilizar um dicionário de palavras hierárquicas com uma colecção:

- Expanda a colecção que pretende configurar. No painel Procurar para uma colecção de procura empresarial ou no painel Análise para uma colecção de análise de conteúdo, faça clique em Acções > Configurar > Opções do servidor de procura.
- No campo Nome do dicionário de palavras hierárquicas, seleccione o dicionário de palavras hierárquicas que pretende utilizar quando os utilizadores consultarem esta colecção.

A lista de dicionários disponíveis inclui todos os dicionários de palavras hierárquicas que foram adicionados ao sistema.

Conceitos relacionados:

"Classificação de documentos através de dicionários de palavras hierárquicas" na página 316

Tarefas relacionadas:

"Adicionar dicionários de palavras hierárquicas ao sistema" na página 317

- 🕩 Criar um ficheiro XML para palavras hierárquicas
- 🖙 Criar um dicionário de palavras hierárquicas

Classificação de documento baseada em padrões URI

Pode aumentar ou diminuir a importância de documentos, atribuindo factores hierárquicos a padrões URI.

A todos os documentos é atribuída uma pontuação de classificação estática quando estes são adicionados ao índice. A pontuação predefinida varia de acordo com o facto de se a classificação estática foi activada para a colecção. Se a classificação estática estiver activada, a pontuação também varia de acordo com o tipo de classificação estática (pela data do documento ou pelo número de documentos ligados a um documento da Web).

Pode influenciar uma importância relativa do documento, atribuindo factores hierárquicos a padrões URI. O factor hierárquico é utilizado com a pontuação estática predefinida e outros factores para determinar a pontuação estática final do documento.

A ordem dos padrões URI que configura é importante. O servidor de pesquisa avalia os padrões de URI pela ordem em que estão listados quando calcula o valor de cada documento numa colecção. Para cada URI:

- 1. O serviço de índice examina sequencialmente os padrões URI.
- 2. Se um documento não puder ser correspondido com um padrão URI, será utilizada a pontuação final estática predefinida.

Depois de configurar um factor hierárquico para um padrão URI, tem de posicionar o padrão URI na ordem em que pretende que o serviço de índice o examine.

Tarefas relacionadas:

"Influenciar as pontuações dos documentos que correspondem aos padrões de URI"

Referências relacionadas:

"Formatos de URI no índice" na página 164

Influenciar as pontuações dos documentos que correspondem aos padrões de URI

Pode aumentar ou diminuir a importância de documentos que correspondem a um padrão URI, aplicando um factor hierárquico à pontuação da classificação estática predefinida. Para influenciar as pontuações desta forma, a influência da classificação estática tem de ser superior a 0 para a colecção quando esta for criada.

O factor hierárquico que configura é utilizado com a pontuação da classificação estática predefinida para calcular uma nova pontuação estática para todos os documentos que correspondem ao padrão URI especificado.

Os factores hierárquicos apenas aumentam as pontuações estáticas, e os factores são apenas um contribuinte para o cálculo que determina a classificação final de um documento. Por exemplo, se existir um número elevado de ligações para um documento (o que resulta numa pontuação inicial elevada), um documento sem ligações terá sempre uma classificação inferior.

Para influenciar as pontuações dos documentos que correspondem a um padrão URI:

- 1. Na página Colecções, expanda a colecção que pretende configurar.
- 2. No painel Procurar para uma colecção de procura empresarial ou no painel Análise para uma colecção de análise de conteúdo, faça clique em Configurar
 > Padrões de URI para influência de pontuações.
- **3**. Na página Influenciar Pontuações através da Marcação de Padrão URI, faça clique em **Adicionar Padrão URI**.
- Escreva um padrão URI nos documentos dos quais pretende aumentar ou diminuir a importância nos resultados da procura. Por exemplo: http://domain.org/hr/*

```
db2://*ROI*
*/afs/*
```

- 5. Escreva um valor entre -10 e 10 no factor hierárquico. A pontuação estática final para todos os documentos que correspondem ao padrão URI será calculada com base neste factor hierárquico.
- 6. Faça clique em **OK**.
- Na página Influenciar pontuações através da correspondência de padrão URI, coloque o novo padrão URI na ordem em que pretende que o serviço de índice o examine.

O serviço de índice calcula as pontuações da classificação estática na ordem em que lista os URIs. Para obter melhores resultados, liste primeiro os URIs mais específicos. No seguinte exemplo, o subdirectório /forms corresponde ao padrão URI http://www.ibm.com/hr/*. Para assegurar que as pontuações para os documentos no subdirectório /forms são calculadas correctamente, liste primeiro o padrão URI para o subdirectório /forms:

http://www.ibm.com/hr/forms/* 8
http://www.ibm.com/hr/* -2

- 8. Para alterar o padrão URI ou o factor hierárquico, seleccione o padrão URI e faça clique em 🖉 Editar.
- 9. Para remover um padrão URI da lista, seleccione o padrão URI e faça clique em 🗎 **Remover**.
- 10. Para aplicar os factores hierárquicos a documentos previamente indexados, recrie o índice.

Conceitos relacionados:

"Classificação de documento baseada em padrões URI" na página 318

Referências relacionadas:

"Formatos de URI no índice" na página 164

Regras para optimizar consultas e classificar resultados

Quando configura uma colecção de procura empresarial, é possível especificar as regras para alterar as consultas e influenciar o modo como os documentos são classificados. Quando um utilizador submete uma consulta, as regras são automaticamente avaliadas. Se os termos de consulta corresponderem à regra, a regra é aplicada antes da devolução de resultados.

Para configurar regras para aperfeiçoar consultas e classificação de resultados, abra a vista Colecções e expanda uma colecção de procura empresarial. No painel Procura, faça clique em **Configurar > Regras para aperfeiçoar consultas e resultados**. Na página Regras para aperfeiçoar consultas e resultados, é possível configurar regras para expandir consultas, filtros para classificação de documentos e regras para agregar documentos em grupos que podem ser resumidos e ordenados por prioridade nos resultados de procura.

Expansão de consultas com base em regras

Durante a expansão da consulta, a consulta original é reescrita em várias consultas de acordo com as regras que especificar. Por exemplo, os termos na consulta podem ser substituídos por termos de um dicionário personalizado, ou podem ser acrescentados termos adicionais aos termos da pesquisa original. A ordem pela qual as regras são aplicadas e a classificação atribuída pelas regras influencia a relevância de documentos correspondentes. Também é possível influenciar a relevância ao especificar se os documentos que correspondem à consulta expandida devem ter uma classificação superior ou inferior a documentos que correspondam aos termos da consulta original.

Filtros de classificação de documentos

É possível ajustar o modo como os documentos são classificados ao associar metadados a filtros específicos de classificação de documentos. Cada filtro consiste num campo de índice que é activado para procura por campos e num analisador que controla o modo como os documentos são atribuídos aos filtros quando uma consulta é processada. Quando um utilizador submete uma consulta, cada filtro é examinado para ver se os documentos que correspondem aos termos de consulta também correspondem ao valor de um campo de índice que tenha sido segmentado pelo analisador. A normalização de caracteres é aplicada aos resultados a partir do analisador.

Quando configura os filtros de classificação de documentos, especifica quais os documentos que devem aparecer na parte superior dos resultados (o filtro **Principais**), com classificação elevada nos resultados (o filtro **Topo**) ou com classificação baixa nos resultados (o filtro **Fim**).

Por exemplo, presuma que existe um filtro de classificação de documentos **Topo** que atribui um analisador denominado Cobertura a um campo de índice denominado título. O analisador segmenta o texto para gerar um token no qual todos os espaços em branco e pontuação são removidos do texto original e as palavras no texto são combinadas. Se o utilizador procurar por International Business Machines, os únicos documentos com o título "International Business Machines" são atribuídos ao filtro de classificação de documentos **Topo**. Os documentos com o título "IBM" ou "About International Business Machines Corporation" não correspondem aos critérios de filtro de classificação de documentos. Com esta configuração, é possível garantir que os documentos com títulos que correspondem exactamente ao texto de consulta obtêm uma classificação superior nos resultados.

Os filtros de classificação de documentos dispõem de uma função de classificação personalizada, a qual permite a configuração de uma classificação para cada filtro. Se associar um campo paramétrico ao filtro, o valor indexado do campo paramétrico é multiplicado pela pontuação correspondente do documento original. Esta configuração permite o controlo da importância de palavras-chave correspondentes.

O sistema IBM Content Analytics with Enterprise Search inclui dois analisadores:

Cobertura

Este analisador extrai palavras-chave do texto e gera um token que contém as palavras-chave combinadas. Por exemplo, este analisador pode produzir o token IBMUnitedStates a partir do texto "You and IBM - United States".

N-gram

Este analisador extrai palavras-chave do texto e trata as sequências de palavra-chave *n* como um token único. Por exemplo, este analisador pode produzir os tokens IBM, United, States, IBMUnited, United States, e IBMUnitedStates do texto "You and IBM - United States".

É possível implementar os seus próprios analisadores personalizados em Java ao expandir o conjunto org.apache.lucene.analysis.Analyzer e ao transferir o ficheiro ZIP para o sistema IBM Content Analytics with Enterprise Search. O conjunto contém ficheiros JAR e alguns ficheiros de amostra de configuração.

Para transferir analisadores personalizados, abra a vista Sistema e faça clique na página Analisar. Para associar um analisador personalizado a uma colecção, expanda o painel Analisar e Indexar para uma colecção de procura empresarial e faça clique em **Configura** > **Analisadores personalizados para classificação de documentos**. Quando configura o analisador para a colecção, especifica quais os campos de índice que devem ser segmentados pelo analisador para que possam ser avaliados durante o processamento de consultas.

Agregação de resultados

É possível configurar regras para agregar resultados em grupos diferentes e especificar o modo como os grupos devem ser ordenados por prioridade nos resultados. Os grupos são resumidos nos resultados, o que ajuda a assegurar que os utilizadores observam diferentes documentos quando procedem à vistoria rápida de uma página de resultados ao contrário de observarem páginas inteiras dominadas por um grupo de resultados.

Por exemplo, é especificado que os resultados que correspondem à origem:perfil_pessoa devem ser agrupados quando a consulta contém jim. Se um utilizador emitir a consulta programmer jim, os resultados que têm o campo de origem:perfil_pessoa proveniente de um documento de perfil da pessoa são agrupados.

É possível especificar o número de resultados a ser apresentado para cada grupo. Por exemplo, cada grupo pode mostrar informações para três documentos quando o grupo está resumido nos resultados. Também é possível especificar quais os grupos que devem ter uma prioridade mais elevada nos resultado.

Para obter informações completas sobre a configuração destes componentes e sobre como os combinar para alterar, dinamicamente, as consultas, e para informações sobre como criar analisadores personalizados, consulte *Expandir consultas e influenciar o modo como os documentos são classificados nos resultados*.

Conceitos relacionados:

Normalização de caracteres

Implização de caracteres

Referências relacionadas:

Expandir consultas e influenciar o modo como os documentos são classificados nos resultados

Aplicações de procura empresarial e de prospecção de conteúdo

Utilize uma aplicação de procura empresarial para consultar colecções e documentos de procura empresarial que são altamente relevantes aos seus termos de consulta. Utilize uma aplicação de exploração de conteúdo para consultar colecções de análise de conteúdo e descobrir relações, correlações e alterações nos dados ao longo do tempo.

Aplicações facultadas

O IBM Content Analytics with Enterprise Search faculta três IDs de aplicação definidos pelo sistema. Pode utilizar estes IDs de aplicação para as suas aplicações personalizadas. Poderá também criar um novo ID de aplicação e associá-lo às colecções que pretende que a sua aplicação utilize.

Procura

Este ID de aplicação permite que a aplicação aceda a todas as colecções de procura. A aplicação de amostra de procura empresarial fornecida utiliza, por predefinição, este ID de aplicação. A menos que personalize a aplicação, pode utilizar esta aplicação para consultar uma ou mais colecções de procura empresarial de uma só vez.

Análise

Este ID de aplicação permite que a aplicação aceda a todas as colecções de análise de conteúdo. O prospecção de análise de conteúdo facultado utiliza este ID de aplicação por predefinição. Pode utilizar esta aplicação para consultar uma colecção de análise de conteúdo de uma só vez.

Todas Este ID de aplicação permite que a aplicação aceda a todas as colecções de procura empresarial e de análise de conteúdo.

É possível utilizar as aplicações de procura facultadas para testar novas colecções antes de as disponibilizar a outros utilizadores. Num ambiente de produção, os administradores controlam quais as aplicações que podem aceder a diversas colecções.

Aplicações personalizadas

Para ajudar a criar aplicações personalizadas, poderá utilizar aplicações de personalizador. Uma aplicação de personalizador permite ao utilizador efectuar selecções numa interface gráfica e visualizar os efeitos das alterações à medida que as efectua. Ao guardar as alterações, o ficheiro de configuração para a aplicação é actualizado. Para aplicar as alterações à sua aplicação personalizada, reinicie a aplicação ou o sistema do IBM Content Analytics with Enterprise Search.

Conceitos relacionados:

- APIs de procura e índice
- Sintaxe de consultas
- Definir propriedades da consulta

Aceder a aplicações

Pode aceder a uma aplicação ao especificar um URL num navegador da Web.

Para iniciar uma aplicação de procura empresarial ou o prospecção de análise de conteúdo:

 Escreva o URL para a aplicação num navegador da Web. O URL inclui o nome do sistema central do servidor de procura e poderá incluir a porta. Por exemplo:

Aplicação de procura empresarial para colecções de procura empresarial http://SearchServer.com:8393/search/

Prospecção de análise de conteúdo para colecções de análise de conteúdo http://SearchServer.com:8393/analytics/

Se o IBM Content Analytics with Enterprise Search estiver configurado para utilizar o servidor de aplicação da Web incorporado, a porta predefinida é a 8393. Se utilizar o WebSphere Application Server, a porta predefinida será 80. Se o sistema não utilizar a porta predefinida, terá de indicar a porta correcta no URL.

2. Se a aplicação requerer autenticação, inicie sessão na aplicação com um ID de utilizador e palavra-passe válidos.

Se alguma das colecções disponíveis para a aplicação tiverem segurança activada, e se as colecções protegidas incluírem pesquisadores que são configurados para validar as credenciais de utilizador durante o processamento de consultas, poderá configurar um perfil de utilizador. Na página Perfil Pessoal na aplicação, especifique as credenciais para aceder aos domínios protegidos. Em seguida, pode aceder a esses domínios sem ter de iniciar sessão neles.

Se o pesquisador suportar início de sessão único (SSO), pode pesquisar domínios seguros sem criar um perfil de utilizador.

Abrir a prospecção de análise de conteúdo com condições de arranque

Pode aceder à prospecção de análise de conteúdo através da especificação de um URL que inclua condições de arranque como, por exemplo, a vista de análise a ser apresentada inicialmente.

Todos os valores de parâmetro no URL endereçável têm de ser codificados com UTF-8.

Para abrir a prospecção de análise de conteúdo com condições de arranque:

- 1. Comece com o URL base, http://localhost:8393/analytics/ search.do?action=index.
- 2. Anexe os parâmetros que pretende aplicar quando a aplicação for aberta, separados pelo carácter e comercial (&):

consulta

Especifique os termos da consulta que pretende executar.

vista

Especifique a vista de análise que pretende mostrar inicialmente. Os valores disponíveis são Facetas, Série cronológica, Desvios, Tendências ou Pares de facetas. Caso não especifique um valor, a aplicação é aberta com a vista Documentos.

facetas

Especifique o caminho para a faceta que pretende seleccionar inicialmente. Para especificar um par de facetas, separe os caminhos com uma vírgula. Caso não especifique um valor, não serão seleccionadas quaisquer facetas inicialmente.

isFacetedSearch

Tem de especificar este parâmetro com o valor definido como true. Não omita este parâmetro.

facetedCollection

Especifique o nome da colecção que pretende abrir inicialmente. Caso não especifique um valor, é aberta a colecção predefinida.

queryLanguage

Especifique o código de linguagem para a linguagem de consulta com a qual pretende publicar consultas. Caso não especifique um valor, é utilizada a linguagem de consulta predefinida.

Exemplo:

http://exampleSearchServer.com:8393/analytics/search.do?action=index &query=tire&view=facets&facets=\$.tkm_base_word,\$.tkm_base_phrase &isFacetedSearch=true&facetedCollection=Analytics&queryLanguage=en

Personalizar aplicações

Pode personalizar as suas aplicações de procura empresarial e o prospecção de análise de conteúdo ao editar ficheiros de configuração usando o personalizador e interfaces de personalizador de análise, e ao configurar a aplicação para utilizar expressões regulares para filtrar URLs.

Tarefas relacionadas:

"Personalizar aplicações através da utilização de uma interface de personalizador" na página 347

"Personalizar a vista Vista-resumo" na página 351

"Personalizar a imagem do cabeçalho de apresentação da aplicação" na página 355

"Personalizar o modo como os títulos e URLs são apresentados nos resultados" na página 356

"Personalizar URLs utilizando um filtro de URL de expressão global" na página 357

Referências relacionadas:

"Ficheiros de configuração da aplicação"

Ficheiros de configuração da aplicação

Pode editar o ficheiro de configuração de uma aplicação de forma a especificar as opções para o seu ambiente, altere o aspecto da aplicação e controle as opções que estão disponíveis para os utilizadores depois de iniciarem a aplicação.

Também pode editar propriedades utilizando aplicações de personalização da procura ou de personalização de análise. Quando efectua selecções na interface de utilizador do personalizador, pode ver o efeito das suas alterações imediatamente. Quando estiver satisfeito com as opções que especificar e guardar as alterações, o ficheiro de configuração é actualizado.

Restrição: Não é possível executar o prospecção de análise de conteúdo como uma portlet no WebSphere Portal. Se executar a aplicação da procura empresarial como portlet de procura no WebSphere Portal, não é possível utilizar o personalizador para efectuar alterações, interactivamente, à aplicação da procura empresarial. Tem de editar as propriedades no ficheiro de configuração e, em seguida, configurar a instância portlet com a interface de administração do WebSphere Portal.

O ficheiro de configuração para as aplicações facultadas é o ficheiro config.properties. O presente tópico aborda as propriedades que pode desejar personalizar; nem todas as propriedades no ficheiro de configuração são aqui documentadas.

Parâmetros do ambiente

Pode especificar opções que controlam a operação da aplicação.

applicationName

Especifica o ID da aplicação. Estão predefinidos três IDs: Procura se a aplicação for para colecções da procura empresarial, Análise se a aplicação for para colecções de análise de conteúdo, ou Tudo se a aplicação for para todas as colecções. Normalmente, não necessita de alterar este valor.O prospecção de análise de conteúdo facultado não funciona bem com colecções de procura empresarial e a aplicação exemplo facultada de procura empresarial não funciona bem com colecções de análise de conteúdo.

displayedApplicationName

O nome da aplicação apresentado. O valor predefinido é Content Analytics.

textAnalyticsEnabled

Se definido como true, o prospecção de análise de conteúdo é iniciado. Normalmente, não tem de alterar este valor porque a abertura da aplicação é controlada pelo URL. O valor predefinido é false.

sessionHeartBeat

Se definido como true, as sessões de utilizador mantêm-se abertas até que se feche o navegador da Web. Se definido como false, a sessão de utilizador é invalidada se não existir qualquer operação em 30 minutos (o período de tempo predefinido).

sessionHeartBeatTime

Especifica a quantidade de tempo a manter um pedido entre o navegador da Web e o servidor de pesquisa activos antes de o pedido expirar. O valor predefinido é de 60000 milissegundos. Normalmente, não necessita de alterar este valor.

protocol

Especifica o protocolo para comunicação com o servidor da Web: http ou https. Se estiver em branco, o valor predefinido é http.

hostname

Especifica o nome do sistema central totalmente qualificado do servidor

que configurou como o servidor de pesquisa, como por exemplo www.exemplo.com. O valor predefinido é localhost.

timeout

Especifica quantos segundos aguarda por uma resposta do servidor da Web antes de o tempo de um pedido da procura ser excedido. Este número tem de ser um número inteiro (tal como 60, não 60.5 ou sessenta). Se não especificar um valor do tempo de espera, o valor predefinido será de 30 segundos.

followRedirect

Especifica se a redirecção do URL é suportada. O valor predefinido é true.

httpVersion

Especifica a versão do protocolo HTTP.

ssoCookieName

Corresponde ao nome do cabeçalho de HTTP que contém a cadeia de LTPA para segurança de início de sessão único (SSO). A predefinição é LtpaToken.

port Especifica o número da porta do servidor da Web. A porta predefinida é 8394 para ligações de HTTP.

http.connection-manager.max-total

Determina o número máximo de ligações de uma aplicação para o servidor de pesquisa. O valor predefinido é 1000. Normalmente, não necessita de alterar esta definição.

trustStore

Se utilizar o protocolo HTTPS, especifique o caminho completo para o ficheiro de arquivo de chaves (o ficheiro de base de dados que contém as chaves públicas). Também chamada *truststore*, esta informação permite a utilização do Secure Sockets Layer (SSL) para comunicações fidedignas. Para especificar um caminho do Windows, coloque uma segunda barra invertida a seguir à barra invertida. Por exemplo: x:\\Application Server\\webserver.key

trustPassword

Se utilizar o protocolo HTTPS, especifique uma palavra-passe para o ficheiro de arquivo de chaves especificado.

validateCertificates

Se utilizar o protocolo HTTPS, especifique true ou false para activar ou desactivar a validação do certificado.

proxyHost

Especifica o nome do sistema central totalmente qualificado de um servidor proxy, caso seja necessário um servidor proxy para aceder ao servidor de procura.

proxyPort

Especifica a porta para o sistema central do servidor proxy especificado.

proxyUser

Especifica um nome de utilizador a usar para iniciar sessão no servidor proxy caso o servidor proxy requeira autenticação base.

proxyPassword

Especifica a palavra-passe para o nome do utilizador do servidor proxy especificado.

username

Especifique um nome de utilizador aqui apenas se pretende substituir o comportamento predefinido para a autenticação de utilizadores com base no nome de utilizador que ele especificar ao aceder à aplicação.

password

Se especificou um nome de utilizador, especifique a palavra-passe.

filter Especifica uma classe que deva ser utilizada para recuperar documentos listados nos resultados da procura. A classe predefinida é com.ibm.es.api.filters.SetDocumentURIFilterFetch. Não altere este valor.

logging.level

Especifica a quantidade de detalhe a registar:

OFF Nenhuma mensagem registada.

SEVERE

São registadas as mensagens que indicam uma falha grave. Esta é a predefinição.

- INFO São registadas mensagens informativas.
- FINE São registadas as mensagens de rastreio com pouco detalhe. Esta opção corresponde ao nível de registo FINE da classe de Java java.util.logging.Level.)
- ALL São registadas todas as mensagens.

Personalizador

customizerDisabled

Num ambiente de produção, especifique true se pretende impedir que os utilizadores utilizem o personalizador da aplicação para efectuar alterações à aplicação.

Icones de origem de documentos

Pode personalizar as imagens que representam o tipo de origem de dados ao qual um documento resultante pertence. Os seguintes ícones de tipo de origem, que identificam os pesquisadores que são suportados quando IBM Content Analytics with Enterprise Search estiver instalado, estão predefinidos no ficheiro config.properties.

卣

documentSource.agent.icon

Especifica o caminho e o nome de um ficheiro de imagem que indica que o documento foi pesquisado por um pesquisador do Agente para sistemas de ficheiros Windows. O ícone predefinido é /images/sourceWindowsFS.gif.

documentSource.p8.icon

Especifica o caminho e o nome de um ficheiro de imagem que indica que o documento foi pesquisado por um pesquisador Case Manager ou pesquisador FileNet P8. O ícone predefinido é /images/sourceP8.gif.



documentSource.vbr.icon

Especifica o caminho e o nome de um ficheiro de imagem que indica que o documento foi pesquisado por o pesquisador do Content Integrator. O ícone predefinido é /images/sourceVBR.gif.



documentSource.cm.icon

Especifica o caminho e o nome de um ficheiro de imagem que indica que o documento foi pesquisado por um pesquisador do Content Manager. O ícone predefinido é /images/sourceCM.gif.



documentSource.db2.icon

Especifica o caminho e o nome de um ficheiro de imagem que indica que o documento foi pesquisado por um pesquisador do DB2. O ícone predefinido é /images/sourceDB2.gif.



documentSource.dominodoc.icon

Especifica o caminho e o nome de um ficheiro de imagem que indica que o documento foi pesquisado por um pesquisador do Domino Document Manager. O ícone predefinido é /images/sourceDominoDoc.gif.



documentSource.exchange.icon

Especifica o caminho e o nome de um ficheiro de imagem que indica que o documento foi pesquisado por um pesquisador do Exchange Server. O ícone predefinido é /images/sourceExchange.gif.



documentSource.database.icon

Especifica o caminho e o nome de um ficheiro de imagem que indica que o documento foi pesquisado por um pesquisador de Base de dados JDBC. O ícone predefinido é /images/sourceJDBC.gif.



documentSource.nntp.icon

Especifica o caminho e o nome de um ficheiro de imagem que indica que o documento foi pesquisado por um pesquisador do NNTP. O ícone predefinido é /images/sourceNNTP.gif.



documentSource.notes.icon

Especifica o caminho e o nome de um ficheiro de imagem que indica que o documento foi pesquisado por um pesquisador do Notes. O ícone predefinido é /images/sourceNotes.gif.



documentSource.quickplace.icon

Especifica o caminho e o nome de um ficheiro de imagem que indica que o documento foi pesquisado por um pesquisador do Quickr for Domino. O ícone predefinido é /images/sourceWorkplace.gif.



documentSource.seedlist.icon

Especifica o caminho e o nome de ficheiro de imagem que indica um serviço do Lotus Quickr para um documento do WebSphere Portal que foi pesquisados por um pesquisador do Lista de valores geradores. O ícone predefinido é /images/sourceSeedlist.gif.



documentSource.seedlist.*.icon

Especifica que o caminho e nome de um ficheiro de imagem representa um documento do IBM Connections que foi pesquisado por um pesquisador Lista de valores geradores. O ícone predefinido é /images/ sourceConnections.gif. Se não pretender utilizar o ícone predefinido, pode especificar imagens diferentes para tipos diferentes de documentos IBM Connections (actividades, blogues, comunidades e assim por diante):

documentSource.seedlist.qr80.icon=/images/sourceSeedlist.gif documentSource.seedlist.wp61.icon=/images/sourceWP61.gif documentSource.seedlist.lc20activities.icon=/images/sourceConnectionsActivities.gif documentSource.seedlist.lc25activities.icon=/images/sourceConnectionsActivities.gif documentSource.seedlist.lc20blogs.icon=/images/sourceConnectionsBlog.gif documentSource.seedlist.lc25blogs.icon=/images/sourceConnectionsBlog.gif documentSource.seedlist.lc20communities.icon=/images/sourceConnectionsCommunities.gif documentSource.seedlist.lc25communities.icon=/images/sourceConnectionsCommunities.gif documentSource.seedlist.lc20dogear.icon=/images/sourceConnectionsDogear.gif documentSource.seedlist.lc25dogear.icon=/images/sourceConnectionsDogear.gif documentSource.seedlist.lc20profiles.icon=/images/sourceConnectionsProfiles.gif documentSource.seedlist.lc25profiles.icon=/images/sourceConnectionsProfiles.gif documentSource.seedlist.lc25wikis.icon=/images/sourceConnectionsWikis.gif documentSource.seedlist.lc25forum.icon=/images/sourceConnections.gif documentSource.seedlist.lc25files.icon=/images/sourceConnectionsFiles.gif

📕 documentSource.sp.icon

Especifica o caminho e o nome de um ficheiro de imagem que indica que o documento foi pesquisado por um pesquisador do SharePoint. O ícone predefinido é /images/sourceSharepoint.gif.



documentSource.unixfs.icon

Especifica o caminho e o nome de um ficheiro de imagem que indica que o documento foi pesquisado por um pesquisador do Sistema de ficheiros UNIX. O ícone predefinido é /images/sourceUnixFS.gif.



documentSource.web.icon

Especifica o caminho e o nome de um ficheiro de imagem que indica que o documento foi pesquisado por um pesquisador da Web. O ícone predefinido é /images/sourceWeb.gif.



documentSource.wcm.icon

Especifica o caminho e o nome de um ficheiro de imagem que indica que o documento foi pesquisado por um pesquisador do Web Content Management. O ícone predefinido é /images/sourceWPS.gif.



💮 documentSource.wp.icon

Especifica o caminho e o nome de um ficheiro de imagem que indica que o documento pertence a uma origem do WebSphere Portal Versão 6 ou anterior que foi pesquisada por um pesquisador do WebSphere Portal. O ícone predefinido é /images/sourceWPS.gif.



😵 documentSource.seedlist.wp61.icon

Especifica o caminho e o nome de um ficheiro de imagem que indica que o documento pertence a uma origem do WebSphere Portal Versão 6.1 que foi pesquisada por um pesquisador do Lista de valores geradores. O ícone predefinido é /images/sourceWP61.gif.

documentSource.winfs.icon

Especifica o caminho e o nome de um ficheiro de imagem que indica que o documento foi pesquisado por um pesquisador do Sistema de ficheiros Windows. O ícone predefinido é /images/sourceWindowsFS.gif.

Ícones do visualizador de cliente

Os documentos resultantes podem ser apresentados no navegador. Os documentos que foram submetidos a um pesquisador por pesquisadores do Notes ou por pesquisadores do Domino Document Manager que foram configuradas para utilizar o protocolo DIIOP também podem ser apresentados por uma aplicação de visualizador de cliente da Lotus.

Para permitir que os documentos sejam apresentados com uma aplicação de visualizador de cliente, certifique-se de que a seguinte propriedade é definida como verdadeira:

clientViewer.show=true

Pode personalizar as imagens que representam a aplicação do visualizador de cliente. No exemplo a seguir, o ícone do Lotus Notes indica que o documento pode ser apresentado com a aplicação do visualizador:

```
client.notes.icon=images/notes.gif
client.dominodoc.icon=images/notes.gif
```

Nos resultados da procura, o ícone e a ligação à aplicação do visualizador de cliente são apresentados do seguinte modo:



Campos dos documentos

Para os tipos de origens de dados com campos, pode controlar os campos que são apresentados nos documentos resultantes.

fields.URI prefix=lista_de_nomes_de_campos_separados_por_espaços

Tem de abandonar o carácter dois pontos (:) no prefixo URI, colocando antes uma barra invertida (\). Para continuar a lista de nomes de campos para outra linha, termine a linha anterior com um carácter barra invertida $(\).$ Por exemplo:

```
fields.db2\://=databasename tablename
fields.domino\://=databasetitle filename creator
fields.dominodoc\://=librarydbtitle documentdbtitle filename author
fields.exchange\://=from creator
fields.file\://=directory filename
fields.https\://=documentID
fields.http\://=documentID
fields.jdbc\://=databasename tablename
fields.news\://=group from
fields.quickplace\://=placetitle roomtitle creator
fields.seedlist\://=author
fields.vbr\://=itemname repositorytype revisionuser
fields.wcm\://=author owner modifier
fields.web\://=
fields.wp6\://=
fields.wps\://=
```

Icones do campo

Para os tipos de origens de dados e para documentos com campos, pode personalizar as imagens que representam campos. Todos os campos acima do resumo do documento contêm uma imagem identificadora. Os ícones do campo seguintes são predefinidos no ficheiro config.properties.



field.icon.databasetitle

Especifica o caminho e o nome de um ficheiro de imagem que indica que o campo contém o título do documento. O ícone predefinido é /images/notesdb.gif.

field.icon.databasename

Especifica o caminho e o nome de um ficheiro de imagem que indica que o ficheiro contém o nome da base de dados à qual o documento pertence. O ícone predefinido é /images/db2.gif.



field.icon.tablename

Especifica o caminho e o nome de um ficheiro de imagem que indica que o campo contém o nome da tabela à qual o documento pertence. O ícone predefinido é /images/table.gif.

field.icon.directory

Especifica o caminho e o nome de um ficheiro de imagem que indica que o campo contém o nome do directório ao qual o documento pertence. O ícone predefinido é /images/closedFolder.gif.



🗊 field.icon.filename

Especifica o caminho e o nome de um ficheiro de imagem que indica que o campo contém o nome de ficheiro do documento. O ícone predefinido é /images/document.gif.

field.icon.documentID

Especifica o caminho e o nome de um ficheiro de imagem que indica que o campo contém o identificador do documento. Pode achar conveniente utilizar esta imagem em branco com documentos da Web, por exemplo, para especificar uma imagem para o URL, mas não apresentar uma imagem ao utilizador. O ícone predefinido é /images/dot.gif.



🗎 field.icon.group

Especifica o caminho e o nome de um ficheiro de imagem que indica que o campo contém o identificador do documento. Pode achar conveniente utilizar esta imagem em branco com documentos da Web, por exemplo, para especificar uma imagem para o URL, mas não apresentar uma imagem ao utilizador. O ícone predefinido é /images/document.gif.

field.icon.from

Especifica o caminho e o nome de um ficheiro de imagem que indica que o campo identifica alguém que enviou o documento. O ícone predefinido é /images/author.gif.



field.icon.creator

Especifica o caminho e o nome de um ficheiro de imagem que indica que o campo identifica o criador do documento. O ícone predefinido é /images/author.gif.

field.icon.author

Especifica o caminho e o nome de um ficheiro de imagem que indica que o campo identifica o autor do documento. O ícone predefinido é /images/author.gif.



field.icon.revisionuser

Especifica o caminho e o nome de um ficheiro de imagem que indica que o campo identifica alguém que reviu o documento. O ícone predefinido é /images/author.gif.



field.icon.owner

Especifica o caminho e o nome de um ficheiro de imagem que indica que o campo identifica o proprietário do documento. O ícone predefinido é /images/author.gif.



field.icon.modifier

Especifica o caminho e o nome de um ficheiro de imagem que indica que o campo identifica alguém que modificou o documento. O ícone predefinido é /images/author.gif.

Ícone do campo predefinido

Pode especificar uma imagem para utilizar quando não forem configurados ícones do campo para campos apresentados nos resultados da procura. O ícone do campo predefinido seguinte é predefinido no ficheiro config.properties.



field.defaultIcon

Especifica o caminho e o nome de um ficheiro de imagem que é o ícone predefinido para os campos nos resultados da procura. O ícone predefinido é/images/database.gif.

Campos de data

Pode especificar os campos que são campos de data. Os nomes dos campos que especifica aqui estão formatados como os dados de data nos resultados da procura. O formato da data corresponde às definições locais no navegador.

date.fields=*lista_de_nomes_de_campos_separados_por_espaços*

Para continuar a lista de nomes de campos para outra linha, termine a linha anterior com um carácter barra invertida (\).

Exemplo:

date.fields=modifieddate createddate date atompublished atomupdated\
creationdate report_dt

Títulos de documentos

Pode especificar títulos alternativos para documentos, substituindo o texto do título por dados com mais significado (ou seja, pode *limpar* os títulos). Por exemplo, em vez de consultar os títulos do documento com o Dispositivo 1 com identificador não informativo, pode especificar que o Dispositivo 1 deve ser suprimido nos resultados da procura. Um campo mais significativo, tal como o nome do ficheiro, poderá ser utilizado para identificar o documento de resultados em vez disso.

Pode especificar títulos alternativos para documentos, removendo palavras sem significado dos títulos do documento (ou seja, pode *truncar* os títulos). Por exemplo, se um número de documentos de resultados começar por Microsoft Word -, pode melhorar a leitura dos resultados da procura suprimindo o texto inicial repetitivo.

titles.clean=lista_de_títulos_separados_por_vírgulas

titles.truncatePrefix=lista_de_prefixos_separados_por_vírgulas

As listas separadas por vírgulas podem conter espaços e outros caracteres, excepto para a vírgula. Para continuar a lista para outra linha, termine a linha anterior com um carácter barra invertida (\).

Por exemplo:

```
titles.clean=Slide 1, Layout 1, untitled, \
Untitled Document, PowerPoint Presentation, \
(esta página não tem título)
```

titles.truncatePrefix=Microsoft Word -, Microsoft Powerpoint -

Valores predefinidos para preferências do utilizador

Pode especificar valores predefinidos para a página Preferências na aplicação. Se um utilizador alterar as preferências, os valores novos estão activos apenas para a sessão actual do utilizador. As preferências seguintes são predefinidas no ficheiro config.properties.

preferences.resultsRange=10

Especifica que cada página dos resultados da procura pode listar 10 documentos resultantes. No prospecção de análise de conteúdo, este valor aplica-se a resultados apresentados na vista Documentos.

preferences.siteCollapsing=Yes

Especifica que os URIs da mesma origem deverão ser resumidos nos resultados da procura. O resumo de sítios só está disponível quando os resultados são ordenados por relevância. Para origens de dados da Web e NNTP, os URIs que corresponderem ao URI do sítio raiz (como, por exemplo, www.ibm.com) são automaticamente resumidos. Para outros tipos de origem de dados e para sítios da Web com níveis de caminho mais profundos (como www.ibm.com/hr), as regras de resumo de sítios têm de ser configuradas na consola de administração.

preferences.spellCorrections=Yes

Especifica que as correcções de ortografia sugeridas devem ser apresentadas quando um utilizador submeter uma consulta que contenha uma palavra possivelmente com um erro ortográfico. Note que as palavras de paragem são sempre removidas antes de as sugestões de ortografia serem calculadas.

preferences.summaryLength=n

Especifica o comprimento predefinido para resumos de documentos.Os valores válidos são 1 - 5, onde 1 indica o comprimento mínimo e 5 indica o comprimento máximo. O valor predefinido é 3. Na aplicação, os utilizadores podem controlar o comprimento dos resumos do documento ao ajustar um cursor de deslocamento na página **Preferências** > **Resultados**

preferences.summaryLength.leveln

Especifica o número de fragmentos para utilizar quando se formulam resumos de documentos, onde cada nível corresponde à definição preferences.summaryLength=n. Por exemplo, se especificar preferences.summaryLength=2 e utilizar o valor predefinido para o nível 2, preferences.summaryLength.level2=75, o comprimento do resumo do documento pode incluir até 75 fragmentos de texto. Um fragmento é uma representação interna de texto, não uma métrica exacta, como, por exemplo, uma percentagem, contagem de caracteres ou contagem de palavras. Como directriz, se duplicar o valor, o comprimento do resumo deve ser, aproximadamente, o dobro. Os valores predefinidos são:

preferences.summaryLength.level1=50

preferences.summaryLength.level2=75

preferences.summaryLength.level3=100

preferences.summaryLength.level4=150 preferences.summaryLength.level5=200

preferences.maxQueryHistorySize

Especifica o número de consultas a manter no histórico da consulta. Os utilizadores podem avançar e retroceder entre consultas emitidas previamente, até ao número que especificar aqui. O valor predefinido é 15.

preferences.extendedHighlighting=No

Especifica que os termos da consulta não serão destacados em campos extra (tal como o título do documento) em adição ao campo do resumo do documento.

Detecção duplicada

Quando os documentos forem adicionados ao índice, a análise é feita para remover duplicados para que os utilizadores não vejam o mesmo documento nos resultados da pesquisa. Para filtrar depois os resultados da procura, pode especificar uma opção para ocultar os documentos que são praticamente idênticos uns aos outros e impedir que estes sejam apresentados nos resultados de pesquisa.

preferences.nearDuplicateDetection=No

Especifica que os documentos quase duplicados não são filtrados durante o processamento de consultas.

Se especificar Sim (Yes), os documentos com títulos e resumos semelhantes são ocultados quando um utilizador visualiza os resultados da procura. Uma mensagem informa os utilizadores de que alguns documentos foram omitidos, porque são semelhantes a outros no conjunto de resultados. Os utilizadores podem fazer clique numa ligação para desactivar a supressão e visualizar todos os documentos no conjunto de resultados.

Para ocultar documentos quase duplicados, o objecto de Query da API de Procura e Índice (SIAPI) para a aplicação tem de especificar o método setProperty com a cadeia ShingleDuplicateDetection definida como Yes (por exemplo, query.setProperty("ShingleDuplicateDetection", "Yes");).

Colecções predefinidas

Pode especificar as colecções que foram seleccionadas previamente nas páginas Preferências e Pesquisa Avançada. Os utilizadores podem editar o conjunto predefinido para pesquisar menos colecções do que as que disponibiliza por predefinição. Se restringir aqui o conjunto de colecções, os utilizadores podem seleccionar colecção que esteja disponível para a aplicação, quando modificarem as suas preferências ou opções de pesquisa avançadas.

preferences.defaultCollectionIDs=*

preferences.defaultCollectionIDs=lista_de_IDs_de_colecção_separados_por_espaços Especifica um asterisco (*) para activar todas as colecções a serem pesquisadas. As colecções têm de estar associadas à aplicação na consola de administração. Esta é a definição predefinida no ficheiro config.properties.Para restringir o que os utilizadores vão procurar se não modificarem as suas preferências ou opções avançadas, especifique os IDs para as colecções que pretende que os utilizadores pesquisem por predefinição. Por exemplo:

preferences.defaultCollectionIDs=*
preferences.defaultCollectionIDs=coll_id1 coll_id2

Se omitir este parâmetro, o sistema utiliza os valores que especificar para a seguinte preferência:

preferences.defaultCollections=*

preferences.defaultCollections=*lista_separada_por_espaços_de_nomes_de_colecção* Especifica um asterisco (*) para activar todas as colecções a serem pesquisadas. As colecções têm de estar associadas à aplicação na consola de administração. Esta é a definição predefinida no ficheiro config.properties.

Para restringir o que os utilizadores vão procurar se não modificarem as suas preferências ou opções avançadas, especifique os nomes da colecção que pretende que os utilizadores procurem por predefinição. Se listar nomes de colecções, certifique-se de que os nomes não contêm espaços. Por exemplo:

preferences.defaultCollections=*
preferences.defaultCollections=collection1 collection2

Preferências predefinidas

Pode especificar valores predefinidos que estejam disponíveis na página Preferências.

preferences.defaultSearchType=facet

Para uma aplicação da procura empresarial, especifique faceta para permitir que os utilizadores pesquisem uma colecção única ou federador para permitir que os utilizadores pesquisem várias colecções de cada vez.

preferences.defaultFacetedCollectionID=*

preferences.defaultFacetedCollectionID=ID_colecção

Para uma aplicação de procura empresarial, especifique um asterisco (*) para activar todas as colecções que suportam facetas a serem pesquisadas ou especifique o ID da colecção para a colecção com faceta que pretende que os utilizadores pesquisem por predefinição. Se omitir este parâmetro, o sistema utiliza o valor que especificar para a seguinte preferência:

preferences.defaultFacetedCollection=*

preferences.defaultFacetedCollection=collection_name

Para uma aplicação da procura empresarial, especifique um asterisco (*) para activar todas as colecções que suportam facetas a serem pesquisadas ou especifique o nome da colecção para a colecção com faceta que pretende que os utilizadores pesquisem por predefinição.Se especificar uma colecção,certifique-se de que o nome não contém espaços.

preferences.queryMode=engine

Na aplicação da procura empresarial, os utilizadores podem escolher um método de análise linguística para corresponder termos não modificados nas consultas: correspondência de formato base (baseform), correspondência de formato exacto (exact) ou uma combinação de correspondência de formato base e de formato exacto (baseform_exact). O valor predefinido é engine, que especifica que os processos de pesquisa vão determinar o melhor método baseado em características da consulta.

preferences.sortBy=relevance

Na aplicação da procura empresarial, os utilizadores podem escolher ordenar os resultados por relevância, data ou nome de campo. O valor predefinido é ordenar resultados por relevância.

preferences.sortByOrder=descending

Na aplicação da procura empresarial, os utilizadores podem escolher organizar os resultados que são ordenados por data ou por nome de campo. O valor predefinido é ordenar os resultados por ordem descendente.

preferences.synonymExpansions=Automatic

Na aplicação da procura empresarial, os utilizadores podem escolher ver os documentos que contêm sinónimos dos termos da consulta nos resultados da pesquisa. Por predefinição, os sinónimos são apresentados automaticamente. Se não for especificado nenhum valor, a expansão de sinónimos não é activada.

preferences.predefinedResults=No

Na aplicação da procura empresarial, os utilizadores podem ver ligações rápidas para documentos nos resultados da pesquisa, se a colecção suportar ligações rápidas. Por predefinição, não são apresentadas ligações rápidas.

preferences.showFacet=Yes

Na aplicação da procura empresarial, os utilizadores podem optar por ver facetas apresentadas na estrutura da esquerda, se a colecção suportar facetas. Por predefinição, as facetas são apresentadas.

preferences.showTPA=Yes

Na aplicação da procura empresarial, os utilizadores podem escolher ver os resultados principais apresentados na estrutura da esquerda, se a aplicação suportar resultados principais. Por predefinição, são apresentados os resultados principais.

preferences.resultColumns=filetype,\$relevance,date,title,thumbnail

Os utilizadores podem escolher as colunas que são apresentadas na área de resultados e especificar a ordem das colunas. Na aplicação da procura empresarial facultada, as colunas são activadas e organizadas da esquerda para a direita por tipo de ficheiro, relevância, data, título de documento e miniatura. Pode especificar os nomes de outros campos que pretenda mostrar como colunas nos resultados da pesquisa. Se der seguimento a uma entrada com \$, como, por exemplo, relevância, a coluna é ocultada. Se incluir miniatura como uma coluna, tem de especificar thumbnail como a primeira ou última entrada para assegurar que as imagens em miniatura são mostradas na primeira ou última coluna nos resultados da pesquisa.

Página separadora personalizada e logótipo

Pode personalizar a imagem apresentada na área do separador no início da aplicação. Por exemplo, poderá querer substituir a imagem predefinida para IBM Content Analytics with Enterprise Search por uma imagem que reflecte a marca da sua empresa. Se não quiser apresentar um separador, torne esta linha uma linha de comentário.

banner.file=images/ibm_log.png

Imagem estática personalizada

Pode personalizar as imagens que são apresentadas em segundo plano das páginas na aplicação. Por exemplo, poderá querer substituir as imagens predefinidas por imagens que reflectem a marca da sua empresa. Se não pretender apresentar uma imagem estática numa página, marque uma dessas ou ambas as linhas de comentário. search.backgroundImage=images/edisc_bg.png bannerLeft.backgroundImage=images/edisc_banner_left.png bannerRight.backgroundImage=images/edisc_banner_right.png bannerBody.backgroundImage=images/edisc_banner_center.png

Ligações

As propriedades na área de ligações do ficheiro config.properties controlam as ligações apresentadas na barra de ferramentas da aplicação.

link.preferences.show=true

Por predefinição, é apresentada uma ligação à página Preferências.

link.myProfile.show=true

Por predefinição, é apresentada uma ligação à página Meu Perfil a especificar credenciais de início de sessão.

link.help.show=true

Por predefinição, é apresentada uma ligação a uma página que faculta ajuda para utilizar a aplicação.

link.about.show=true

Por predefinição, é apresentada uma ligação à página que mostra informações sobre o produto.

link.logOff.show=true

Por predefinição, é apresentada uma ligação à página que permite aos utilizadores encerrar sessão da aplicação.

Identificadores do tipo de documento significativos

Pode melhorar a leitura do ficheiro do tipo de documento, carregando os nomes actuais do tipo de documento para termos mais concisos e com mais significado. Os tipos de documentos que estão disponíveis para as aplicações são definidos pela classe AvailableDocumentTypes da API Procura e Índice (Search and Index API - SIAPI). Por conveniência, os tipos de documento disponíveis também são listados no fim do ficheiro config.properties.

documentType.*label=lista_de_tipos_de_documento_separados_por_espaços*

Especifica o nome apresentado na linha do filtro do tipo de documento nos resultados da procura, e uma lista de tipos de documentos actuais que serão apresentados quando um utilizador seleccionar o filtro.

Por exemplo, pode especificar o identificador html e carregar as extensões do nome do ficheiro e os tipos MIME de vários documentos da Web para esse nome. Quando um utilizador fizer clique em **html** para filtrar os resultados da procura, apenas os documentos com extensões especificadas e os tipos MIME serão apresentados.

Os carregamentos do tipo de documento seguintes são predefinidos no ficheiro de config.properties:

documentType.html=shtml text/html html xhtml htm documentType.doc=doc application/msword documentType.ppt=application/mspowerpoint ppt documentType.xls=xls application/x-excel application/msexcel \ application/x-msexcel application/excel application/vnd.ms-excel documentType.xml=xml text/xml documentType.txt=txt text/plain documentType.pdf=pdf application/pdf

Se o valor que especificar para a etiqueta de tipo de documento corresponder ao nome de uma propriedade existente no ficheiro application.properties, é apresentado o valor para a propriedade no ficheiro application.properties e não o valor especificado aqui. Por exemplo, se especificar documentType.unixfs como etiqueta para o filtro de tipo de ficheiro, o valor para a propriedade unixfs no ficheiro application.properties (**sistema de ficheiro UNIX**) é apresentado como o nome do filtro do tipo de ficheiro seleccionável.

Diálogo Pré-visualização

No prospecção de análise de conteúdo, é possível limitar o número de facetas que são apresentadas quando os utilizadores pré-visualizam documentos da vista Documentos.

limitFacets=1000

Por predefinição, 1000 facetas são apresentadas no diálogo pré-visualização.

Filtros personalizados

Pode especificar consultas personalizadas para filtrar a apresentação de documentos resultantes.

filterCustom.label=termos_consulta

Especifica o nome apresentado no menu de filtro personalizado nos resultados da pesquisa, e uma consulta que refine os resultados da pesquisa quando um utilizador seleccionar o filtro. (Enquanto vêem os resultados da procura, os utilizadores podem seleccionar um filtro personalizado para consultar apenas os documentos resultantes que correspondem à consulta predefinida.)

No exemplo seguinte, os resultados da procura são filtrados para mostrar apenas os documentos que pertencem à base de dados dos recursos humanos (hr):

filterCustom.HR_database_only=databasename::hr

Quando um utilizador faz clique em **HR_database_only** para filtrar os resultados da procura, a consulta databasename::hr é processada. Quando os resultados da procura são apresentados, apenas os documentos da base de dados hr são listados.

Se o valor que especificar para a etiqueta de filtro personalizado corresponder ao nome de uma propriedade existente no ficheiro application.properties, é apresentado o valor para a propriedade no ficheiro application.properties e não o valor especificado aqui. Por exemplo, se especificar filterCustom.hostData= como a etiqueta de filtro personalizado, o valor para a propriedade hostData no ficheiro application.properties (**Definições do servidor**) é apresentado como o nome de filtro personalizado seleccionável por clique.

Vários filtros personalizados são retirados e fornecidos como exemplos no ficheiro de config.properties.

Conjunto ordenado

Pode especificar quantos documentos podem ser colocados em conjuntos para serem armazenados nos resultados da pesquisa.

sortPoolSize=

Especifique um valor entre 0 e 3000. O valor predefinido é 3000.

Análise dos melhores resultados (gráficos de barras para campos de metadados)

Pode especificar opções para mostrar gráficos de barras que representam a análise dos melhores resultados. A predefinição é analisar os 500 melhores resultados. Cada gráfico corresponde a um só campo de metadados e cada barra num gráfico corresponde a um valor do campo. O comprimento da barra indica a frequência com que o valor do campo ocorre. Quanto maior a barra, tanto maior o número de ocorrências desse valor do campo.

Quando configura o espaço de pesquisa para um pesquisador, pode especificar opções para campos de metadados. Para configurar um gráfico dos melhores resultados para um campo de metadados, tem de especificar que o campo é um campo pesquisável e que o campo pode ser devolvido nos resultados da procura.

As propriedades que configurar para o gráfico de barras têm o seguinte formato, em que *número* é o número que identifica de forma única o gráfico de barras na aplicação, *opção* é a opção do gráfico de barras e *valor* é o valor da opção: topResultsCharts*número.opção=valor*

topResultsChartsnúmero.titleKey=chave_aplicação

Especifica um título para o gráfico, em que o *número* é um número que identifica *chave_aplicação* é uma etiqueta para o título do gráfico. Esta etiqueta pode ser a chave no ficheiro application.properties para a aplicação ou o valor que especificar aqui. Se este valor ficar em branco, é utilizada a etiqueta de sistema predefinida.

Neste exemplo, a etiqueta para o título do gráfico é especificada pelo valor para a chave topResults.mostRecentDocuments no ficheiro application.properties: topResultsCharts3.titleKey=topResults.mostRecentDocuments.

Por exemplo, o título do gráfico é o valor exacto que especificou aqui: topResultsCharts3.titleKey=Organizations

topResultsChartsnúmero.enable=true

Especifica se este gráfico de barras é para ser apresentado quando os utilizadores visualizam os resultados da procura. Se especificar falso (false), o gráfico não é apresentado.

topResultsChartsnúmero.fieldName=nome_campo

Especifica o nome do campo de metadados cujos valores são para ser analisados para este gráfico. Por exemplo, databasetitle. Tem de especificar um nome de campo diferente para cada gráfico que adiciona (o mesmo campo não pode ser utilizado em mais do que um gráfico).

topResultsChartsnúmero.maxValues=número

Especifica o número de itens a apresentar neste gráfico. Por exemplo, especifique 10 para mostrar não mais de 10 valores de campo diferentes.

topResultsChartsnúmero.fieldValueSeparator=carácter

Especifica um carácter que delimita os valores no campo a analisar. Por exemplo, se um campo contiver valores múltiplos que estão separados por um ponto e vírgula (tais como agente; vendedor; corretor), pode utilizar esta propriedade para identificar o ponto e vírgula (;) como um separador de valor do campo para que cada valor possa ser adicionado ao gráfico de barras como um item separado. Sem esta opção, todo o valor do campo é adicionado ao gráfico como um item único.

topResultsChartsnúmero.canUserChangeFieldName=true

Especifica se o utilizador pode seleccionar um campo diferente ao visualizar os resultados da procura e ver os melhores resultados para este campo. Se especificar falso (false), os utilizadores não poderão seleccionar um campo diferente a analisar, quando visualizarem os resultados da procura.

Se especificar verdadeiro (true), é apresentada uma lista de todos os campos que foram encontrados nos 500 melhores resultados iniciais assim como os resultados para o campo actual. Se o utilizador seleccionar um campo a partir desta lista, o identificador do gráfico altera para **Gráfico de campo dinâmico** e os resultados do gráfico de barras para o campo seleccionado são apresentados até que o utilizador seleccione um campo diferente ou feche o navegador. Da próxima vez que o utilizador executar a aplicação, são apresentados os resultados para o campo original.

topResultsChartsnúmero.sortKey=frequency

Especifica como é que são ordenados os itens no gráfico de barras. Valores suportados:

- none Os itens do gráfico de barras não são ordenados.
- label Os itens do gráfico de barras são ordenados de acordo com o valor do campo. Se configurar etiquetas personalizadas com a propriedade topResultsCharts.númerolistOfLabels.prefixKey=nome_campo, os itens do gráfico de barras também são ordenados pelo nome da etiqueta.

frequência

Os itens do gráfico de barras são ordenados de acordo com o número de resultados devolvidos por valor do campo.

topResultsChartsnúmero.sortOrder=descending

Para os itens que são ordenados por etiqueta ou frequência, especifica a ordem. Valores suportados:

ascending

Os itens que são ordenados por etiqueta são listados por ordem alfabética de A a Z. Para os itens que são ordenados por frequência, os valores do campo que ocorrem um maior número de vezes, aparecem menos na lista do que os valores que ocorrem com menos frequência.

descending

Os itens que são ordenados por etiqueta são listados por ordem alfabética descendente de Z a A. Para os itens que são ordenados por frequência, os valores do campo que ocorrem um maior número de vezes, aparecem mais na lista do que os valores que ocorrem com menos frequência.

Análise dos melhores resultados (número máximo de resultados)

topResult.resultSize=número

O valor predefinido e máximo para um número de resultados para devolver a partir da análise dos melhores resultados é de 500. É possível diminuir este valor, mas não é possível aumentá-lo. Por exemplo, poderá pretender especificar um número inferior se tiver problemas com pedidos com tempo de espera a esgotar durante a análise dos melhores resultados. Não pode definir esta propriedade utilizando o personalizador da aplicação.

Gestão de identidade

É possível especificar opções para controlar como o Componente de Gestão de Identidade (IMC) processa a procura protegida.

removeDisabledIdentityFromIMC=true

Se o valor estiver definido como true, a entrada do utilizador na base de dados do IMC é eliminada quando as entradas do utilizador forem validadas e o objecto pesquisável é desactivado. Se o valor estiver definido como false, o IMC não faz nada.

removeInvalidIdentityFromIMC=true

Se o valor estiver definido como true, a entrada do utilizador na base de dados do IMC é eliminada quando a validação falhar. Se o valor estiver definido como false, o IMC não faz nada.

Definições de prospecção de análise de conteúdo

Especifique os valores predefinidos para as várias vistas no prospecção de análise de conteúdo. Os utilizadores podem alterar algumas das opções para uma determinada sessão seleccionando preferências na página Preferências.

Vista Facetas

Especifique valores predefinidos para a forma como os utilizadores trabalham com a vista facetas.

analytics.facets.numberOfResults=100

Especifique o número de valores da faceta ou subfacetas para mostrar numa página quando um utilizador selecciona uma faceta.

analytics.facets.defaultSortBy=frequency

Especifique a ordem predefinida. Pode ordenar valores da faceta ou subfacetas por frequência (frequency) ou pelo valor de correlação informatizado (correlation).

analytics.facets.defaultShowTarget=keywords

Especifique se o modo predefinido deve mostrar os valores da faceta (keywords) ou subfacetas da faceta seleccionada.

Vista Série Cronológica

Especifica os valores predefinidos para a forma como os utilizadores trabalham com a vista Série Cronológica.

analytics.timeseries.defaultTimeScale=year

Especifica a escala de tempo predefinida para analisar os resultados (ano, mês, ou dia).

Vista Desvios

Especifique os valores predefinidos para a forma como os utilizadores trabalham com a vista Desvios.

analytics.deviations.numberOfCharts=3

Especifique o número predefinido de gráficos a serem apresentados.

analytics.deviations.showLineInChart=true

Especifique se pretende ver uma linha gráfica apresentada nos gráficos que mostram os valores do índice informatizado com o passar do tempo.

analytics.deviations.defaultSortBy=frequency

Especifique a ordem predefinida. Pode ordenar valores da faceta ou subfacetas pela frequência (frequency), o valor de índice informatizado (index), por nome em ordem ascendente (ascending), ou pelo nome em ordem descendente (descending).

analytics.deviations.defaultShowTarget=keywords

Especifique se o modo predefinido deve mostrar os valores da faceta (keywords) ou subfacetas da faceta seleccionada.

analytics.deviations.defaultTimeScale=year

Especifica a escala de tempo predefinida para analisar os resultados (ano, mês, ou dia).

analytics.deviations.barColorOfMultiChart=

Especifique o tema de cores para mostrar gráficos em modo de gráfico múltiplo (legend ou (index).

analytics.deviations.highlightThresholds=2,4,8,16

Especifique limiares para anomalias destacadas. Por predefinição, as células com um valor de índice superior a 2.0 são amarelas, as células com um valor de índice superior a 4.0 são laranja e as células com um valor de índice superior a 16 são vermelhas.

Vista Tendências

Especifique valores predefinidos para a forma como os utilizadores trabalham com a vista Tendências.

analytics.trends.numberOfCharts=3

Especifique o número predefinido de gráficos a serem apresentados.

analytics.trends.showLineInChart=true

Especifique se pretende ver uma linha gráfica apresentada nos gráficos que mostram os valores do índice informatizado com o passar do tempo.

analytics.trends.defaultSortBy=frequency

Especifique a ordem predefinida. Pode ordenar valores da faceta ou subfacetas pela frequência (frequency), o valor de índice informatizado (index), por nome em ordem ascendente (ascending), ou pelo nome em ordem descendente (descending).

analytics.trends.defaultShowTarget=keywords

Especifique se o modo predefinido deve mostrar os valores da faceta (keywords) ou subfacetas da faceta seleccionada.

analytics.trends.defaultTimeScale=year

Especifica a escala de tempo predefinida para analisar os resultados (ano, mês, ou dia).

analytics.trends.barColorOfMultiChart=

Especifique o tema de cores para mostrar gráficos em modo de gráfico múltiplo (legend ou (index).

analytics.trends.highlightThresholds=2,4,8,16

Especifique limiares para anomalias destacadas. Por predefinição,

as células com um valor de índice superior a 2.0 são amarelas, as células com um valor de índice superior a 4.0 são laranja e as células com um valor de índice superior a 16 são vermelhas.

Vista Pares de Facetas

Especifique valores predefinidos para a forma como os utilizadores trabalham com a vista Pares de Facetas.

analytics.facetPairs.numberOfRowResults=

Especifique o número predefinido de linhas a serem apresentadas na vista Tabela.

analytics.facetPairs.numberOfRowsInTable=

Especifique o número predefinido de linhas a serem apresentadas na vista Grelha.

analytics.facetPairs.numberOfColumnResults=

Especifique o número predefinido de colunas a serem apresentadas na vista Tabela.

analytics.facetPairs.numberOfColumnsInTable=

Especifique o número predefinido de colunas a serem apresentadas na vista Grelha.

analytics.facetPairs.defaultRowShowTarget=

Especifique o número predefinido de linhas a serem apresentadas.

analytics.facetPairs.defaultColumnShowTarget=

Especifique o número predefinido de colunas a serem apresentadas.

analytics.facetPairs.defaultListSortBy=frequency

Especifique a ordem predefinida para a vista Tabela. Pode ordenar valores da faceta ou subfacetas pela frequência (frequency), pelo valor de correlação informatizado (correlation), ou pelo valor da faceta ou nome da subfaceta por ordem ascendente ou descendente (ascending) ou (descending).

analytics.facetPairs.defaultRowSortBy=alphabetical

Especifique a ordem predefinida para linhas na vista Grelha. Pode ordenar os valores da faceta ou subfacetas por ordem alfabética pelo nome (alphabetical, pela frequência (frequency) ou pelo valor de correlação informatizado (correlation).

analytics.facetPairs.defaultColumnSortBy=alphabetical

Especifique a ordem predefinida para as colunas na vista Grelha. Pode ordenar os valores da faceta ou subfacetas por ordem alfabética pelo nome, pela frequência (frequency), ou pelo valor de correlação informatizado (correlation).

analytics.facetPairs.defaultView=flat

Especifique a subvista: flat, grid, ou table.

analytics.facetPairs.highlightThresholds=2,4,8,16

Especifique limiares para anomalias destacadas. Por predefinição, as células com um valor de correlação superior a 2.0 são amarelas, as células com um valor de correlação superior a 4.0 são laranja e as células com valores de correlação superiores a 16 são vermelhas.

Configuração da faceta

Pode especificar opções predefinidas para mostrar a árvore da faceta e gráficos da faceta na aplicação de procura.

facetTree.enable=true

Mostra (true) ou oculta (false) a árvore da faceta na aplicação da procura empresarial.

facetTree.titleKey=Facet Tree

Especifique uma etiqueta para a árvore de faceta. A etiqueta pode ser uma chave no ficheiro application.properties para a aplicação ou o valor que especificar aqui. Se omitida, é utilizada a etiqueta predefinida, Árvore da Faceta.

facetTree.maxValues.default=15

Especifica o número máximo de valores da faceta que pode ser apresentado na árvore da faceta. Se o número de valores disponíveis for superior a este número, a ligação **More** é apresentada para que os utilizadores possam ver os valores da faceta adicionais.

facetTree.maxValues.all=500

Especifique o número máximo de facetas que podem ser apresentadas na árvore da faceta.

facetTree.open=true

Especifique se a árvore da faceta está aberta (expandida) por predefinição.

facetChartsnúmero.enable=true

Mostra (true) ou oculta (false) este gráfico da faceta na aplicação da procura empresarial.

facetChartsnúmero.maxValues.all=500

Especifique o número máximo de valores da faceta que podem ser apresentados neste gráfico da faceta.

facetChartsnúmero.facetName=model

Especifique um nome para este gráfico da faceta.

facetChartsnúmero.canUserChangeFacetName=true

Especifique se o utilizador pode seleccionar uma faceta diferente deste gráfico ao ver os resultados da pesquisa. Se definido como verdadeiro, os utilizadores podem seleccionar uma faceta de uma lista de facetas disponíveis.

facetChartsnúmero.titleKey=

Especifique uma etiqueta para este gráfico da faceta. A etiqueta pode ser uma chave no ficheiro application.properties para a aplicação ou o valor que especificar aqui.

facetChartsnúmero.open=false

Especifique se este gráfico da faceta está aberto (expandido) por predefinição.

topResultsChartsnúmero.open=true

Especifique se o gráfico de resultados principais está aberto (expandido) por predefinição.

Configuração de introdução adiantada

Pode especificar opções que permitam aos utilizadores ver sugestões de potenciais correspondências de termos de pesquisa como tipos de utilizador ou selecciona

uma faceta na árvore de facetas. Os utilizadores podem seleccionar uma das correspondências sugeridas para executar a consulta ou explorar a faceta.

typeAhead.search.numberOfResults=n

Especifique o número máximo de sugestões potenciais a serem apresentadas de cada origem de sugestão. O valor predefinido é 10.

typeAhead.search.mode=mode

Especifique a origem predefinida para gerar correspondências potenciais de termos da consulta:

off Desactiva as sugestões de introdução adiantada.

queryLog

Especifica que as sugestões vão surgir do índice do registo da consulta. São mostradas sugestões de consultas anteriores aos utilizadores.

term Especifica que as sugestões vão surgir do índice de texto. São mostradas sugestões de termos no índice.

queryLog,term

Especifica que as sugestões vão surgir no índice do registo da consulta e do índice de texto. São mostradas sugestões de consultas anteriores primeiro, e depois sugestões de termos no índice. Esta opção é o valor predefinido.

term,queryLog

Especifica que as sugestões vão surgir no índice de texto e do índice do registo de consultas. São mostradas sugestões de termos no índice primeiro, e depois sugestões de consultas anteriores.

typeAhead.facet.numberOfResults=n

Especifique o número máximo de valores de facetas potencialmente correspondentes a serem sugeridas quando um utilizador seleccionar uma faceta na árvore de facetas. Os utilizadores podem seleccionar um dos valores de facetas sugeridos a explorar. O valor predefinido é 15.

Conceitos relacionados:

"Campos de data e formatos de data personalizados" na página 223

"Personalizar aplicações" na página 325

Definir propriedades da consulta

Tarefas relacionadas:

"Personalizar aplicações editando propriedades"

Personalizar aplicações editando propriedades

Uma aplicação de procura empresarial pode procurar em todas as colecções de procura activas no sistema. Uma aplicação de prospecção de conteúdos pode procurar uma única colecção de análise de conteúdos de cada vez.

O programa de instalação instala as aplicações no servidor de procura. Para configurar uma aplicação, edite um ficheiro de configuração, config.properties, que é implementado com a aplicação. Por exemplo, pode especificar opções para o ambiente do servidor Web, utilizar uma aplicação diferente da aplicação predefinida, ou controlar quais as opções apresentadas quando a aplicação é iniciada.

Para editar propriedades de aplicações:

- 1. Inicie sessão para procurar o servidor como o administrador predefinido do IBM Content Analytics with Enterprise Search.
- 2. Edite o ficheiro config.properties com um editor de texto padrão.

Os ficheiros para aplicações config.properties são instalados na localização seguinte, em que ES_INSTALL_ROOT é o directório de instalação do IBM Content Analytics with Enterprise Search em cada servidor de procura:

- Se utilizar o servidor de aplicação da Web incorporado: ES_INSTALL_ROOT/jetty/searchapp/search/WEB-INF/config.properties ES_INSTALL_ROOT/jetty/searchapp/analytics/WEB-INF/config.properties
- Se utilizar WebSphere Application Server:

ES_INSTALL_ROOT/installedApps/search.ear/search.war/WEB-INF/config.properties ES_INSTALL_ROOT/installedApps/analytics.ear/analytics.war/WEB-INF/ config.properties

- **3**. Edite as propriedades para especificar preferências, depois guarde e feche o ficheiro. No ficheiro, o carácter cardinal (#) indica uma linha de comentário.
- 4. Para aplicar as alterações à sua aplicação, reinicie o sistema IBM Content Analytics with Enterprise Search ou a aplicação. Para reiniciar a aplicação:
 - Se utilizar o servidor de aplicação da Web incorporado, introduza estes comandos, onde *ID_nó* identifica o servidor de procura:

esadmin searchapp.*ID_nó* stop esadmin searchapp.*ID nó* start

Para determinar o ID do nó para o servidor de procura, execute o comando esadmin check para ver a lista de IDs de sessão. Procure o ID do nó que está listado para a sessão searchapp.

• Se utilizar o WebSphere Application Server, pare e reinicie a aplicação de procura empresarial ou o prospecção de análise de conteúdo.

Conceitos relacionados:

Definir propriedades da consulta

Referências relacionadas:

"Ficheiros de configuração da aplicação" na página 325

Personalizar aplicações através da utilização de uma interface de personalizador

Um personalizador de aplicação é uma interface gráfica que pode ser utilizada para personalizar aplicações de procura empresarial e o prospecção de análise de conteúdo para a sua empresa sem editar um ficheiro de configuração.

Por exemplo, pode alterar as imagens de faixa e pano de fundo, alterar o esquema da interface e especificar opções para mostrar os resultados.

Quando faz selecções num personalizador de aplicações, é possível pré-visualizar os efeitos das alterações. Quando guarda as alterações, actualiza o ficheiro de configuração para a aplicação.Para aplicar as alterações à sua aplicação, pode reiniciar a aplicação ou o sistema IBM Content Analytics with Enterprise Search.

Para personalizar uma aplicação:

- 1. Abra o personalizador de aplicações.
 - Se utilizar o servidor de aplicação da Web incorporado, faça clique em Personalizador de Procura ou Personalizador de Análise na consola administrativa.

• Se utilizar o WebSphere Application Server, introduza o URL da aplicação que pretende personalizar, em que *ServidorProcura* é o nome do sistema central do servidor de procura, e porta é a porta do servidor de procura como, por exemplo, 9080:

http://ServidorProcura.com:porta/search/search.do?action=customize
http://ServidorProcura.com:porta/analytics/search.do?action=customize

- 2. Se o personalizador não estiver aberto, faça clique em **Abrir Personalizador** nos Controlos de Personalizador.
- Seleccione as opções que pretende personalizar. As seguintes páginas estão disponíveis para as aplicações de procura empresarial e para o prospecção de análise de conteúdo.
 - **Server** Especifique preferências do ambiente, como, por exemplo, o nome do sistema central e porta do servidor de pesquisa, o nome de aplicação (normalmente, Procura para uma aplicação que suporte colecções de procura empresarial e Análise para uma aplicação que suporte colecções de análise de conteúdo), o nível das mensagens que pretende que os utilizadores vejam, o período de tempo que um pedido de procura pode aguardar por resposta antes de esgotar o tempo de espera e a colecção ou colecções predefinidas a serem procuradas.
 - **Ecrã** Especifique caminhos para as imagens que pretende apresentar na área de cabeçalho de apresentação, o texto de título de aplicação e as ligações que pretende que apareçam na barra de ferramentas. Para o prospecção de análise de conteúdo, também é possível especificar quais as vistas de análise que serão apresentadas e a ordem em que devem ser apresentadas.

Opções de Consulta

Especifique as suas preferências para o modo como as consultas são processadas. Por exemplo, pode permitir que os utilizadores vejam sugestões de ortografia para termos de consulta, especificar o comprimento máximo de resumos de documentos e especificar os tipos de documentos que os utilizadores podem seleccionar quando querem limitar a consulta a certos tipos de documentos.

Resultados

Especifique as suas preferências para o modo como os resultados são apresentados. Por exemplo, pode permitir aos utilizadores contraírem resultados a partir da mesma fonte e mostrar ou esconder documentos duplicados nos resultados. Também é possível seleccionar as colunas predefinidas para a página de resultados, eliminar ou cortar texto de título pouco importante e activar documentos Domino nos resultados para serem abertos com uma aplicação de aplicação do cliente.

Filtros de Resultados

Especifique as suas preferências para filtrar resultados por título ou URL. Os documentos que são devolvidos em resposta a um pedido de procura são actualizados pelos filtros que especifica. Por exemplo, se utilizar uma expressão global para substituir caracteres nos URLs para documentos nos resultados, os URLs revistos são apresentados nos resultados. Caso prefira, é, também, possível alterar o modo como os URLs são apresentados nos resultados ao editar um ficheiro de configuração, especificando as regras de filtragem e configurando a aplicação para utilizar o filtro de URL de expressão global.
Imagens

Especifique as imagens que pretende utilizar para identificar a fonte de dados a que um documento pertence nos resultados de procura. Por exemplo, imagens diferentes podem ajudar os utilizadores a distinguir documentos oriundos de uma base de dados Lotus Notes e de um sistema de ficheiros Windows.

4. Se estiver a personalizar uma aplicação de procura empresarial, seleccione as opções predefinidas para os painéis que pretende incluir e especifique se quer activar a colecção de procura empresarial para mostrar dados de análise.

Painéis de Faceta

Especifique as suas preferências para mostrar diferentes tipos de informações na área de navegação de facetas da aplicação. Diferentes gráficos podem mostrar facetas, facetas dinâmicas, categorias e sinalizadores. Pode adicionar os painéis que entender, atribuir um nome único a cada painel e controlar tanto o tamanho do painel como a quantidade de valores que são mostradas por predefinição.

Outros Painéis

Especifique as suas preferências para mostrar outros tipos de informações, como, por exemplo, uma área de consola para mostrar a sintaxe de consulta detalhada ou um gráfico cronológico que reflicta as datas dos documentos nos resultados. As opções de visualização dependem do tipo de painel que adiciona.

Esquema

Especifique o esquema predefinido para os painéis que incluir na aplicação de procura empresarial. É possível arrastar painéis para a esquerda, direita, topo ou fundo das áreas do contentor. Para cada contentor, pode especificar a altura e largura predefinidas. Diferentes acções estão disponíveis dependendo se a aplicação permitir ou não mostrar as análises.

Análise

É possível activar o modo de análise para a aplicação de procura empresarial. Neste modo, ligações de valor de faceta nos painéis de faceta mostram a classificação da correlação, o que pode ajudar os utilizadores a identificar documentos que talvez sejam mais relevantes do que outros documentos que correspondem às condições de consulta. Neste modo, os utilizadores também podem limitar os resultados explorando a linha cronológica do documento.

5. Se estiver a personalizar um prospecção de análise de conteúdo, seleccione as opções predefinidas para as diferentes vistas de análise.

Facetas

Especifique quantos resultados são apresentados por predefinição, se os valores da faceta ou sub-facetas devem ser apresentados por predefinição e a sequência de ordenação predefinida.

Série Cronológica

Especifique a escala de tempo predefinida.

Desvios e Tendências

Especifique quantos resultados são apresentados por predefinição, se os valores da faceta ou sub-facetas devem ser apresentados por predefinição, a sequência de ordenação predefinida e a escala de tempo predefinida. Também pode especificar se pretende que os utilizadores vejam uma linha gráfica a mostrar os valores de índice calculados com o tempo, e seleccionar uma cor de esquema para identificar os diferentes gráficos na vista multi-gráfico.

Pares de Facetas

Especifique a sub-vista predefinida, se os valores da faceta ou sub-facetas devem ser apresentados por predefinição, a sequência de ordenação predefinida e quantos resultados são apresentados por predefinição. Pode especificar diferentes preferências para resultados horizontais e verticais.

Ligações

Especifique quantos resultados são apresentados por predefinição e se os valores da faceta ou sub-facetas devem ser apresentados por predefinição. Pode especificar diferentes preferências para resultados de cada faceta. É, também, possível especificar se os nós no diagrama de relação podem sobrepor-se e se os valores de frequência e de correlação devem ser visualmente representados por alterações no tamanho do nó e os comprimentos de ligações entre nós.

Vista-Resumo

Especifique as preferências para vários ficheiros de esquema. Cada ficheiro de esquema pode mostrar estatísticas de análise em diferentes formatos como, por exemplo, gráficos de barras, gráficos circulares ou gráficos de linhas cronológicas. Especifica os gráficos a incluir, as facetas a analisar, a consulta predefinida, a ordem predefinida e outras opções para apresentar vários gráficos que os utilizadores podem explorar numa única vista.

Relatórios

Especifique um ID de utilizador e palavra-passe que permitam à aplicação aceder ao servidor IBM Cognos Business Intelligence, caso seja necessário iniciar sessão. Também especifique se comentários de cabeçalho devem ser descarregados por predefinição quando os utilizadores criam um relatório em formato CSV.

Postura

Especifique quantos resultados são apresentados por predefinição na vista Postura e o número máximo de resultados que podem ser apresentados. Especifique, também, o número máximo e predefinido de expressões que podem ser apresentadas quando um utilizador explora uma faceta, o número máximo e predefinido de documentos que podem ser disponibilizados para pré-visualização, se os valores da faceta ou sub-facetas devem ser apresentados por predefinição e a escala de tempo predefinida.

- 6. Faça clique em **Fechar** para fechar o personalizador e pré-visualizar os efeitos das alterações na aplicação. Introduza uma consulta para ver como as alterações modificam os resultados. Se pretender continuar a fazer alterações, faça clique em **Abrir Personalizador** nos Controlos do Personalizador.
- 7. Quando estiver satisfeito com as opções que fez, faça clique em Guardar Alterações nos Controlos de Personalizador para actualizar os ficheiros de configuração. Se fizer clique em Cancelar Alterações, as definições são restauradas com valores da última versão guardada do ficheiro de configuração.
- 8. Faça clique em **Sair** nos Controlos de Personalizador para sair do personalizador da aplicação.

- Para impedir que outros utilizadores façam alterações na aplicação quando o sistema está em modo de produção, desactive o personalizador de aplicações. Edite o ficheiro de configuração da aplicação e defina a propriedade customizerDisabled como true.
- 10. Para aplicar as alterações à sua aplicação, reinicie o sistema IBM Content Analytics with Enterprise Search ou a aplicação. Para reiniciar a aplicação:
 - Se utilizar o servidor de aplicação da Web incorporado, introduza estes comandos, onde *ID_nó* identifica o servidor de procura:

esadmin searchapp.*ID_nó* stop esadmin searchapp.*ID_nó* start

Para determinar o ID do nó para o servidor de procura, execute o comando esadmin check para ver a lista de IDs de sessão. Procure o ID do nó que está listado para a sessão searchapp.

• Se utilizar o WebSphere Application Server, pare e reinicie a aplicação de procura empresarial ou o prospecção de análise de conteúdo.

Conceitos relacionados:

"Campos de data e formatos de data personalizados" na página 223

"Personalizar aplicações" na página 325

Tarefas relacionadas:

"Personalizar a vista Vista-resumo"

"Personalizar a imagem do cabeçalho de apresentação da aplicação" na página 355 "Personalizar o modo como os títulos e URLs são apresentados nos resultados" na página 356

Personalizar a vista Vista-resumo

Pode configurar a vista Vista-resumo no prospecção de análise de conteúdo para apresentar resultados de análise múltipla em vários gráficos e tabelas.

O IBM Content Analytics with Enterprise Search fornece dois esquemas predefinidos para apresentar gráficos e tabelas. Pode utilizar o personalizador de análise para criar esquemas personalizados para propósitos de exploração de conteúdo. Os utilizadores podem comutar para um esquema diferente quando exploram uma colecção e podem estabelecer preferências para especificar qual o esquema que deve ser apresentado como o esquema predefinido. Os utilizadores também podem salvar os quadros e os gráficos como imagens na forma de ficheiros Bitmap, PNG e JPEG.

Quando personaliza a Vista-resumo, especifica:

- Os tipos de gráficos a mostrar e como estes são dispostos na vista
- Quais as facetas mostradas em cada gráfico e quais as estatísticas que o gráfico apresenta como, por exemplo, frequência, correlação, o índice de desvios ou o índice de tendências
- A ordem predefinida para os resultados devolvidos em cada gráfico
- A faceta de data predefinida e os valores predefinidos para especificar unidades de tempo.

Quando faz selecções no personalizador de análise, pode pré-visualizar os efeitos das alterações. Quando guarda as alterações, actualiza um ficheiro XML de esquema. Para aplicar as alterações na sua aplicação, reinicie a aplicação ou o sistema IBM Content Analytics with Enterprise Search.

Para personalizar a Vista-Resumo:

- Na página Vista-resumo no personalizador de análise, seleccione uma colecção de análise de conteúdo para a qual pretende criar um esquema de vista-resumo personalizado. A lista de colecções incluem aquelas que podem ser procuradas.
- 2. Para o campo **Esquema**, seleccione [**Novo**] e, depois, dê um nome com significado para o esquema no campo **Título**.
- 3. Especifique quantos contentores horizontais e verticais devem ser apresentados na Vista-resumo e, depois, faça clique nos botões para adicionar contentores ao esquema. Vários contentores permitem-lhe mostrar mais do que um gráfico e tabela por esquema. Quando adiciona dois ou mais contentores horizontais, são criados lado a lado, como colunas numa tabela. Quando adiciona contentores verticais, são criados como filas numa tabela.
- 4. Seleccione um contentor. Pode adicionar mais contentores ou pode adicionar um painel. Deve adicionar um painel para cada contentor no esquema.
- 5. Seleccione um painel e faça clique em **Editar** para configurar o gráfico ou tabela a apresentar.

Título Especifique um título descritivo para o painel.

Tipo Seleccione o tipo de gráfico que	pretende criar.
--	-----------------

Tipo de gráfico	Descrição
Gráfico de barras	Este gráfico apresenta valores de faceta como barras separadas no gráfico. O comprimento de cada barra representa estatísticas para o tipo de análise especificado. Por exemplo, a barra maior poderá representar o valor de faceta com o valor de correlação mais elevado.
Gráfico de colunas	Este gráfico apresenta valores de faceta como colunas discretas de dados. O eixo vertical representa o tipo de análise especificado.
Gráfico circular	Este gráfico apresenta valores de faceta como divisões separadas no gráfico. O tamanho de cada divisão representa as estatísticas para o tipo de análise especificado. Por exemplo, a maior divisão poderá representar o valor de faceta com o maior valor de índice de desvio.
Tabela de facetas	Este gráfico lista um valor de faceta em cada linha. As colunas da tabela contêm os valores de frequência e correlação correspondentes.
Gráfico de barras de série cronológica	Este gráfico apresenta valores de faceta como colunas ao longo da escala de tempo especificada. O eixo vertical representa o tipo de análise especificado.
Gráfico de linhas de série cronológica	Este gráfico apresenta valores de faceta como linhas distintas que atravessam a escala de tempo especificada. O eixo vertical representa o tipo de análise especificado.
Tabela de marca de verificação de pares de facetas	Este gráfico faculta uma visualização útil da vista Pares de facetas. Caso a contagem de frequências seja superior a um, é apresentada uma marca de confirmação onde os valores da faceta se cruzam. Por exemplo, poderá utilizar este gráfico para ver quais os documentos sinalizados como Importantes configurando o gráfico para mostrar sistemas de ficheiros como a faceta horizontal (linha) e sinalizadores para a faceta vertical (coluna). Os utilizadores podem ver rapidamente quais os sistemas de ficheiros que contêm documentos importantes quando abrem a Vista-Resumo.

Tipo de análise

Seleccione o tipo de análise estatística a ser mostrada no gráfico. As escolhas disponíveis dependem do tipo de gráfico que cria.

Tipo de análise	Descrição
Correlação	Devolve os valores de frequência e correlação (as mesmas estatísticas devolvidas na vista Facetas).
Desvios	Devolve a frequência e o valor de índice de desvios. Para utilizar esta opção, tem de especificar a escala de tempo a utilizar.
Tendências	Devolve a frequência e o valor de índice de tendências. Para utilizar esta opção, tem de especificar a escala de tempo a utilizar.

ID de faceta

Especifique o ID da faceta a analisar como, por exemplo, **\$._**word.noun. Para uma tabela de marca de confirmação de pares de facetas, este parâmetro aplica-se às filas na tabela.

Tipo de faceta

Especifique o tipo de facetas, assim como valores de facetas, sub-facetas, categorias baseadas em regras, sinalizadores ou facetas baseadas no alcance. Para uma tabela de marca de confirmação de pares de facetas, este parâmetro aplica-se às filas na tabela. Se o **ID da faceta** e o **Tipo da faceta** não corresponderem, não é apresentado qualquer dado no painel.

Vista contextual

Se a colecção suportar vistas contextuais, seleccione a vista que será analisada e mostrada no gráfico, se existir.

Contar para analisar

Especificar quantos valores de facetas de topo N (ordenadas por frequência) serão seleccionadas para análise. Para uma tabela de marca de confirmação de pares de facetas, este parâmetro aplica-se às filas na tabela.

Contar para apresentar

Especifique quantos dos primeiros valores de facetas *N* serão mostradas no gráfico. Para uma tabela de marca de confirmação de pares de facetas, este parâmetro aplica-se às filas na tabela.

Mostrar tipo de dados

Especifique os dados estatísticos para mostrar por predefinição. Cada faceta tem dois valores: frequência e correlação ou valor de índice. Os utilizadores podem alterar os dados que são apresentados por predefinição quando vêem a Vista-resumo.

Tipo de ordem

Especifique a ordem predefinida. Pode escolher ordenar por frequência ou por correlação. Para os tipos de análises de Desvio e Tendência, também pode escolher ordenar pelo valor de índice mais baixo ou mais alto.

Por exemplo, se o gráfico mostrar a faceta Tipo de Carro, o tipo de análise é Tendências, a sequência de ordenação é índice mais baixo e o gráfico é configurado para mostrar os primeiros cinco valores de faceta, depois, o gráfico mostra os primeiros cinco valores de faceta que aumentaram acentuadamente durante um dado período de tempo.

Valores de faceta

Especifique os valores de faceta que pretende analisar apenas se não pretender que o gráfico mostre os primeiros valores de faceta *N*. Por predefinição, a Vista-resumo mostra os primeiros valores de faceta *N*.

Para uma tabela de marca de confirmação de pares de facetas, este parâmetro aplica-se às filas na tabela.

Por exemplo, um gráfico pode mostrar a faceta Estado e os três primeiros valores de frequência. Neste caso, os valores apresentados podem variar com base na consulta actual. Contudo, se especificar valores de faceta explícitos como, por exemplo, AK, AL e AZ, o gráfico mostra estatísticas para estes três estados e não para os estados com a maior frequência.

Explicitar consultas para análise

Se especificar uma consulta explícita, a Vista-resumo mostra resultados para o documento que corresponde a essa consulta, independentemente de qual é a consulta actual.

Consulta adicional para análise

Se quiser ver os resultados que são limitados dos resultados actuais, especifique uma consulta neste campo. Esta consulta adicional aplica-se às condições de consulta actuais para limitar o conjunto de documentos mostrados nos resultados.

Largura e Altura

Especifique a largura e a altura, como uma percentagem, que o painel ocupa com respeito ao contentor ascendente, e não à totalidade da vista-resumo.

Quando configura um gráfico de barras de série cronológica ou um gráfico de linhas de série cronológica, especifique estes parâmetros adicionais:

Escala de Tempo

Especifique a escala de tempo para análise temporal, tal como mês, ano ou dia da semana.

Faceta de data

Seleccione uma faceta da lista de tipos de facetas de data. Este valor especifica qual a faceta que é utilizada como o X-axis no gráfico da escala de tempo. Por exemplo, se a faceta Data de Receber contém dados sobre a data, pode especificar a Data de Receber como um valor para este campo para mostrar esse dado como o X-axis do gráfico da escala de tempo. Se não existirem facetas de tipo de data configuradas, então apenas a faceta de data predefinida, **data**, se encontra disponível na lista.

Data final de escala de tempo

Especifique a última data para os dados serem mostrados na escala de tempo. Por exemplo, se a escala de tempo for Mês e o valor deste campo for 12, o final da escala de tempo é 12 meses anterior ao último valor de dados. Se este campo estiver por preencher, então não limita a data final do período de tempo que é apresentado no gráfico e o último valor de dados é apresentado.

Data inicial de escala de tempo

Especifique a duração da escala de tempo. Por exemplo, se a escala de tempo for Mês e o valor deste campo for 36, o gráfico mostra dados com três anos a partir do valor **Data final de escala de tempo**.

Valores de facetas de data

Se quiser limitar os valores de faceta de data a valores específicos, então especifique os valores. Por predefinição, nenhum valor de faceta de data é estabelecido, para que o gráfico apresente todos os valores para a faceta de data. Quando configura uma tabela de marca de confirmação de pares de facetas, especifique esses parâmetros adicionais para a faceta horizontal (coluna):

ID da faceta de coluna

Especifique o ID da faceta para os valores que são apresentados como colunas na tabela. O ID da faceta de coluna é automaticamente estabelecido para \$._word a não ser que especifique explicitamente outro valor de faceta para analisar.

Tipo de faceta de coluna

Especifique o tipo de faceta para a faceta de coluna, tal como valores de faceta, sub-facetas, categorias baseadas em regras, sinalizadores ou facetas baseadas no alcance.

Contar colunas para apresentar

Especifique quantos dos primeiros valores de facetas *N* serão mostrados para a faceta de coluna no gráfico.

Valores de faceta de coluna

Especifique os valores de facetas que pretende analisar apenas se não pretender que o gráfico mostre os primeiros valores de faceta *N* para a faceta de coluna.

- 6. Faça clique em Guardar para guardar as suas alterações.
- 7. Opcional: Faça clique em **Estabelecer como predefinição** caso pretenda utilizar o actual ficheiro de esquema como predefinição.
- 8. Com quaisquer alterações que faça com o personalizador, pode fechar as páginas do personalizador e pré-visualizar os efeitos das suas alterações, continuar a fazê-las, guardá-las ou cancelá-las.
- 9. Se guardou as suas alterações, aplique as mesmas à sua aplicação reiniciando a aplicação ou o sistema IBM Content Analytics with Enterprise Search.

Conceitos relacionados:

"Personalizar aplicações" na página 325

Tarefas relacionadas:

"Personalizar aplicações através da utilização de uma interface de personalizador" na página 347

Personalizar a imagem do cabeçalho de apresentação da aplicação

Se criar uma aplicação de procura empresarial personalizada, é possível alterar a imagem apresentada no cabeçalho de apresentação.

Por predefinição, o personalizador da procura empresarial é executado em http://host_name:8390/searchCustomizer e o caminho do ficheiro do personalizador é jetty/adminapp/search. A aplicação de procura é executada em http://host_name:8393/search e o respectivo caminho do ficheiro é jetty/searchapp/search. Tem de copiar a sua imagem personalizada para ambas as localizações.

Se utilizar o servidor de aplicação da Web incorporado, utilize o seguinte procedimento para personalizar o cabeçalho de apresentação para a sua aplicação de procura empresarial. Se utilizar o WebSphere Application Server, o directório da aplicação (ES_INSTALL_ROOT/installedApps/search.ear/search.war) e a porta são diferentes.

Para personalizar a aplicação da imagem do cabeçalho de apresentação:

- Copie o seu ficheiro de imagem para os seguintes directórios: ES_INSTALL_ROOT/jetty/adminapp/search/images ES_INSTALL_ROOT/jetty/searchapp/search/images
- 2. Na consola administrativa do IBM Content Analytics with Enterprise Search, abra a aplicação Personalizador de Procura.
- No campo Cabeçalho de Apresentação no separador Ecrã, introduza o caminho do ficheiro para o seu ficheiro de imagem. Por exemplo: images/my_image.PNG
- 4. Faça clique em **Fechar** para fechar o personalizador e pré-visualizar os efeitos das alterações na aplicação.
- Quando estiver satisfeito com as opções que fez, faça clique em Guardar Alterações nos Controlos de Personalizador para actualizar os ficheiros de configuração.
- 6. Faça clique em **Sair** nos Controlos de Personalizador para sair do personalizador da aplicação.
- 7. Para aplicar as alterações à sua aplicação, reinicie o sistema IBM Content Analytics with Enterprise Search ou a aplicação.

Conceitos relacionados:

"Personalizar aplicações" na página 325

Tarefas relacionadas:

"Personalizar aplicações através da utilização de uma interface de personalizador" na página 347

Personalizar o modo como os títulos e URLs são apresentados nos resultados

É possível utilizar o personalizador de procura e o personalizador de análise para alterar o modo como os títulos e URLs são apresentados nos resultados.

Os filtros de resultados podem actualizar os títulos e URLs para documentos que são devolvidos em resposta a um pedido de procura ao substituir os valores no título e URL com informações especificadas. Os títulos e URLs actualizados são apresentados nos resultados.

Sugestão: Em vez de utilizar uma interface de personalizador, é possível alterar o modo como os URLs são apresentados nos resultados ao editar o ficheiro de configuração para a aplicação e utilizando o filtro de URL de expressão global.

Para personalizar o modo como os títulos e URLs são apresentados quando os utilizadores observam os resultados:

- Na página Filtros de Resultados no personalizador de procura ou personalizador de análise, faça clique no botão para adicionar um novo filtro de título ou para adicionar um novo filtro de URL.
- 2. Seleccione a caixa de verificação para activar o filtro.
- **3**. Seleccione a colecção a que pretende aplicar o filtro ou especifique que pretende aplicar o filtro a todas as colecções. A lista de colecções incluem aquelas que podem ser procuradas.
- 4. Seleccione o tipo de dados a que pretende aplicar o filtro.
- 5. Seleccione o tipo de filtro:

Valor de campo

Se seleccionar esta opção, tem de especificar o nome do campo.

Quando este filtro é aplicado, o título ou URL nos resultados é alterado para o valor do campo especificado.

Filtro de expressão global

Se seleccionar esta opção, tem de especificar a fórmula de expressão global a ser aplicada. Quando este filtro é aplicado, o título ou URL nos resultados é alterado de acordo com os valores de substituição especificados na fórmula.

Também é possível especificar um nome de campo para o filtro de expressão global. Se o nome de campo for especificado, o valor do campo especificado é utilizado como o valor alvo a que o filtro se aplica. Se não for especificado um nome de campo (vazio), o título original é utilizado como o alvo de substituição.

O seguinte exemplo de filtro de título remove extensões do nome do ficheiro:

Nome de campo: "nome de ficheiro" Fórmula: "/\..*\$//"

No seguinte exemplo de filtro de URL, | pode ser utilizado como separador de expressões e \$1 pode ser utilizado para correspondência de padrão:

Nome de campo: (vazio) Fórmula: "|IBM_(.*)_|http://server.example.com/\$1.html|" ()

- Guarde as suas alterações. Aplique as alterações à sua aplicação ao reiniciar a aplicação ou ao reiniciar o sistema IBM Content Analytics with Enterprise Search.
- Execute uma consulta para verificar se os títulos e URLs são apresentados nos resultados com os valores de substituição especificados.

Conceitos relacionados:

"Personalizar aplicações" na página 325

Tarefas relacionadas:

"Personalizar aplicações através da utilização de uma interface de personalizador" na página 347

"Personalizar URLs utilizando um filtro de URL de expressão global"

Personalizar URLs utilizando um filtro de URL de expressão global

É possível alterar a forma como os URL são apresentados nos resultados configurando a aplicação para utilizar o filtro URL de expressão regular.

O filtro de URL da expressão regular actualiza os URLs do documento que são devolvidos pelo servidor de pesquisa em resposta a um pedido de pesquisa substituindo caracteres no URL por caracteres que especificar. O URL actualizado é apresentado nos resultados da procura.

Sugestão: É, também, possível alterar a forma como os URLs são apresentados nos resultados utilizando o personalizador de procura ou o personalizador de análise, em vez de editar um ficheiro de configuração. Na interface do personalizador, pode especificar as regras de filtragem e aplicá-las a colecções, origens de dados e campos específicos.É, também, possível especificar regras para personalizar o modo como os títulos são mostrados nos resultados de procura.

Para aplicar as substituições de expressões regulares a URLs de documentos:

- 1. Inicie sessão como administrador predefinido do IBM Content Analytics with Enterprise Search no servidor de procura.
- 2. Edite o ficheiro de configuração para a aplicação.

Os ficheiros config.properties para a aplicação de procura empresarial e prospecção de análise de conteúdo são instalados nas seguintes localizações do servidor de procura:

- Se utilizar o servidor de aplicação da Web incorporado: ES_NODE_ROOT/master_config/searchapp/search/ ES_NODE_ROOT/master_config/searchapp/analytics
- Se utilizar WebSphere Application Server: ES_INSTALL_ROOT/installedApps/search.ear/search.war/WEB-INF/
 - ES_INSTALL_ROOT/installedApps/analytics.ear/analytics.war/WEB-INF/
- **3**. Localize a linha seguinte no ficheiro:

filter=com.ibm.es.api.filters.SetDocumentURIFilterFetch

4. Insira com.ibm.es.api.filters.SetDocumentURIFilterRegex; como o primeiro filtro neste parâmetro e, de seguida, guarde o ficheiro. O exemplo seguinte é apresentado em duas por uma questão de legibilidade; especifique o valor do filtro numa única linha:

filter=com.ibm.es.api.filters.SetDocumentURIFilterRegex; com.ibm.es.api.filters.SetDocumentURIFilterFetch

5. Crie um ficheiro de configuração que especifique as expressões regulares que pretende substituir e o texto de substituição pretendido. É possível adicionar várias linhas para as regras de substituição de processamento. Por exemplo:

```
/zig/zag/
/\?/questionmark/g
|aaa/bbb|aaaslashbbb|
```

Neste exemplo, quando os URLs são devolvidos nos resultados da procura, a primeira ocorrência da palavra-chave zig é substituída por zag, todas as ocorrências do carácter ? transformam-se em questionmark (a opção g indica a substituição global) e a primeira ocorrência de aaa/bbb transforma-se em aaaslashbbb.

Sugestões:

- É possível criar este ficheiro com qualquer nome e guardá-lo em qualquer local onde o servidor de procura o possa ler, como, por exemplo, C:\IBM\temp. O sistema é configurado para utilizar este ficheiro no passo 6 na página 359.
- Para substituir um carácter especial, tem de substituir o carácter fazendo-o preceder de um carácter de barra invertida (\), conforme se mostra no segundo em exemplo anterior.
- Se pretende especificar uma barra (/) como destino de substituição, pode utilizar outro carácter como separador, conforme se mostra no último exemplo anterior.Os caracteres delimitadores são definidos como caracteres não-palavra e não-espaço em branco. Por exemplo, é possível utilizar os seguintes caracteres como delimitadores em vez do carácter /, embora / seja o carácter delimitador padrão e recomendado:

!"#\$%&'()*+,-./:;<=>?@[\]^`{|}~

 O filtro de URL de expressão regular suporta quaisquer caracteres que sejam suportados pela classe java.util.regex.Pattern. Por exemplo, pode utilizar meta-caracteres como o asterisco (*) ou o ponto final (.) como caracteres globais.

- Se uma URI do documento iniciar com http, https, ou notes, então o título é apresentado como uma ligação passível de ter um clique. Caso pretenda assegurar que o título é sempre apresentado como uma ligação, sem especificar os caracteres iniciais, tem de definir forceTitleLink=true no ficheiro config.properties.
- 6. Configure a variável de ambiente do sistema Java.
 - Se utilizar o servidor de aplicação da Web incorporado, insira a linha seguinte no ficheiro %ES_INSTALL_ROOT%/configurations/interfaces/ searchapp_interface.ini, especifique o caminho para o ficheiro que criou no passo 5 na página 358 e guarde as alterações:

JVMOptions=-DRegexFilterFilename=configuração_ficheiro_caminho

Por exemplo: JVMOptions=-DRegexFilterFilename=C:\IBM\temp

- Se utilizar o WebSphere Application Server, execute os passos seguintes. Estes passos podem diferir da sua versão de WebSphere Application Server:
- a. Abra a consola administrativa do WebSphere Application Server num navegador da Web.
- b. Seleccione Servidores da aplicação > ESSearchServer e, em seguida, expanda Java e Gestão de Processos e faça clique em Definição de Processo.
- c. Em Propriedade Adicional, abra Java Virtual Machine, adicione o seguinte parâmetro ao campo Argumentos JVM genéricos e especifique o caminho para o ficheiro que criou no passo 5 na página 358.
 -DRegexFilterFilename=configuração_ficheiro_caminho

Por exemplo: -DRegexFilterFilename=C:\IBM\temp

- d. Guarde as alterações e saia da consola administrativa.
- 7. No servidor de pesquisa, inicie sessão como o administrador do IBM Content Analytics with Enterprise Search e reinicie a aplicação:
 - Se utilizar o servidor de aplicação da Web incorporado, introduza estes comandos, onde *ID_nó* identifica o servidor de procura:

esadmin searchapp.*ID_nó* stop esadmin searchapp.*ID nó* start

Para determinar o ID do nó para o servidor de procura, execute o comando esadmin check para ver a lista de IDs de sessão. Procure o ID do nó que está listado para a sessão searchapp.

- Se utilizar o WebSphere Application Server, pare e reinicie a aplicação de procura empresarial ou o prospecção de análise de conteúdo.
- 8. Execute uma consulta para verificar se os URLs aparecem nos resultados da procura com os caracteres de substituição que especificou.

Conceitos relacionados:

"Personalizar aplicações" na página 325

Tarefas relacionadas:

"Personalizar o modo como os títulos e URLs são apresentados nos resultados" na página 356

Configurar opções de expansão de grupo para melhorar tempos de resposta da procura protegida

Para reduzir o tempo necessário para extrair informações de grupo de uma origem de dados quando uma aplicação valida as credenciais do utilizador, é possível especificar uma propriedade de configuração, sameGroupDataSources, para especificar qual o servidor a utilizar como repositório do utilizador.

Por exemplo, se uma colecção incluir vários servidores do Lotus Notes e todos esses servidores tiverem informações do utilizador idênticas (ou seja, todos utilizam o mesmo repositório de utilizador), é possível configurar o sistema para obter a informação do utilizador a partir de um servidor, em vez de verificar todos os servidores.

Para configurar as opções de expansão de grupo que podem melhorar o desempenho da procura protegida:

- 1. Inicie sessão como administrador predefinido do IBM Content Analytics with Enterprise Search no servidor de procura.
- 2. Edite o ficheiro de configuração para a aplicação.

Os ficheiros config.properties para as aplicações são instalados na seguinte localização, onde ES_INSTALL_ROOT é o directório de instalação de IBM Content Analytics with Enterprise Search no servidor de procura:

• Se utilizar o servidor de aplicação da Web incorporado:

ES_INSTALL_ROOT/jetty/searchapp/search/WEB-INF/config.properties ES_INSTALL_ROOT/jetty/searchapp/analytics/WEB-INF/config.properties

• Se utilizar WebSphere Application Server:

ES_INSTALL_ROOT/installedApps/search.ear/search.war/WEB-INF/config.properties ES_INSTALL_ROOT/installedApps/analytics.ear/analytics.war/WEB-INF/ config.properties

- 3. Integre uma ou mais origens de dados que partilhem o mesmo repositório de utilizador:
 - a. Adicione a propriedade sameGroupDataSourcesnúmero= ao ficheiro, onde número é um número inteiro que começa por 1 e aumenta 1 em cada novo grupo das origens de dados.
 - b. Para o valor, especifique as origens de dados que partilham a informação do grupo no formato *datasourceDomain_datasourceType*. Certifique-se de que a primeira origem de dados no valor é aquela que pretende utilizar para extrair as informações do grupo e utilize dois pontos (:) para separar as entradas da origem de dados.

É possível determinar o nome do domínio do elemento <DataSources> no ficheiro XML do pesquisador como, por exemplo, o ficheiro notescrawler.xml.Por exemplo:

```
<DataSources>
  <Server Name="ExampleCo.db1.com!!test.nsf" DirectoryMode="False"
  ID="NS123456789" Domain="db1!!ExampleCo.db1.com!!NOTES"
  SS0Enabled="False">
```

Neste exemplo, iria adicionar a propriedade sameGroupDataSources*número*= da seguinte forma:

sameGroupDataSources1=db1!!ExampleCo.db1.com!!NOTES_notes;

Para outro exemplo (a primeira linha nesta exemplo é apresentada em duas linhas por uma questão de legibilidade):

sameGroupDataSources1=db1!!ExampleCo.db1.com!!NOTES_notes; db2!!ExampleCo.db2.com!!NOTES_notes; sameGroupDataSources2=ExampleCo1 winfs;ExampleCo2 winfs

4. Se utilizar o WebSphere Application Server, e pretender executar processos de validação do utilizador e de expansão de grupo em paralelo, inclua a propriedade **useWorkManagerInIMC=true** no ficheiro.

Esta funcionalidade utiliza o WorkManager, que é uma função do WebSphere Application Server. Para especificar o WorkManager como a Java Naming and Directory Interface (JNDI), utilize o parâmetro **workManagerJNDIName**. Para especificar um valor de tempo de espera, utilize o parâmetro **workManagerTimeout**.

- 5. Para aplicar as alterações à sua aplicação, reinicie o sistema IBM Content Analytics with Enterprise Search ou a aplicação. Para reiniciar a aplicação:
 - Se utilizar o servidor de aplicação da Web incorporado, introduza estes comandos, onde *ID_nó* identifica o servidor de procura:

esadmin searchapp.*ID_nó* stop esadmin searchapp.*ID nó* start

Para determinar o ID do nó para o servidor de procura, execute o comando esadmin check para ver a lista de IDs de sessão. Procure o ID do nó que está listado para a sessão searchapp.

• Se utilizar o WebSphere Application Server, pare e reinicie a aplicação de procura empresarial ou o prospecção de análise de conteúdo.

Reter dados da sessão para deslocar mais depressa através de resultados protegidos

Para melhorar o tempo de resposta da procura protegida, é possível configurar a aplicação de procura para reter informação da sessão que permite aos utilizadores deslocarem-se mais depressa nos resultados da procura.

É possível configurar aplicações de procura empresarial e o prospecção de análise de conteúdo para deslocar as páginas de resultados de procura ao utilizar páginas obtidas previamente para a mesma consulta. As novas páginas são obtidas de acordo com a classificação dos resultados que foram incluídas nas páginas anteriormente obtidas. Assim, se os dados da Lista de Controlo de Acesso (ACL) no índice ou na origem de dados forem alterados enquanto as páginas de resultados da procura estiverem a ser deslocadas, a classificação dos resultados poderá não ser exacta.

Para melhorar o tempo de resposta da procura protegida:

- 1. Inicie sessão como administrador predefinido do IBM Content Analytics with Enterprise Search no servidor de procura.
- 2. Edite o ficheiro de configuração para a aplicação.

Os ficheiros para aplicações config.properties são instalados na localização seguinte, em que ES_INSTALL_ROOT é o directório de instalação do IBM Content Analytics with Enterprise Search em cada servidor de procura:

- Se utilizar o servidor de aplicação da Web incorporado: ES_INSTALL_ROOT/jetty/searchapp/search/WEB-INF/config.properties ES_INSTALL_ROOT/jetty/searchapp/analytics/WEB-INF/config.properties
- Se utilizar WebSphere Application Server: ES_INSTALL_ROOT/installedApps/search.ear/search.war/WEB-INF/config.properties ES_INSTALL_ROOT/installedApps/analytics.ear/analytics.war/WEB-INF/ config.properties
- 3. Adicione a propriedade searchable.sessional=true ao ficheiro. Se esta opção não existir no ficheiro, a função não será activada.
- 4. Para aplicar as alterações à sua aplicação, reinicie o sistema IBM Content Analytics with Enterprise Search ou a aplicação. Para reiniciar a aplicação:
 - Se utilizar o servidor de aplicação da Web incorporado, introduza estes comandos, onde *ID_nó* identifica o servidor de procura:

esadmin searchapp.*ID_nó* stop esadmin searchapp.*ID_nó* start

Para determinar o ID do nó para o servidor de procura, execute o comando esadmin check para ver a lista de IDs de sessão. Procure o ID do nó que está listado para a sessão searchapp.

• Se utilizar o WebSphere Application Server, pare e reinicie a aplicação de procura empresarial ou o prospecção de análise de conteúdo.

Segurança no IBM Content Analytics with Enterprise Search

Os mecanismos de segurança permitem proteger origens de procuras não autorizadas e restringir funções administrativas a utilizadores específicos.

Com IBM Content Analytics with Enterprise Search, os utilizadores podem procurar um vasto leque de origens de dados. Para assegurar que apenas os utilizadores que estão autorizados a acederem ao o conteúdo o fazem, e, para assegurar que apenas os utilizadores autorizados podem aceder à consola de administração, o sistema coordena e faz cumprir a segurança em vários níveis.

Servidor de aplicação da Web

O primeiro nível de segurança é o servidor de aplicação da Web, seja através do servidor de aplicação da Web incorporado ou através do WebSphere Application Server de definições de segurança global. Pode configurar o sistema para utilizar um registo LDAP e permitir apenas aos utilizadores registados iniciar sessão nas aplicações ou na consola de administração. Pode, também, configurar o sistema para utilizar um ficheiro de chave LTPA para fornecer suporte de autenticação de início de sessão único (SSO) aos utilizadores da aplicação.

Quando definir os controlos de segurança, são necessários diferentes procedimentos caso as suas aplicações sejam suportadas através do servidor de aplicação da Web incorporado ou WebSphere Application Server, especialmente se pretender implementar suporte para autenticação de início de sessão único (SSO).

Segurança ao nível do sistema

Ao nível do sistema, pode atribuir utilizadores a funções administrativas e autenticar utilizadores que administram o sistema. Quando um utilizador inicia sessão na consola de administração, apenas as funções e colecções que o utilizador está autorizado a gerir estão disponíveis para esse utilizador. Também pode atribuir privilégios a utilizadores e grupos para controlar funções da aplicação. Por exemplo, pode limitar a capacidade de exportar documentos de uma aplicação a utilizadores específicos.

Pode também configurar credenciais que permitem aos pesquisadores aceder a origens de dados que inclui nas colecções. Outros componentes do sistema também necessitam de estas credenciais. Por exemplo, para verificar se os utilizadores estão autorizados a consultar documentos nos resultados da procura, os servidores de procura podem utilizar as credenciais para ligação a uma origem de dados e verificar as listas actuais de controlo de acesso.

Segurança ao nível da colecção

Ao criar uma colecção, pode activar a segurança ao nível de colecção. Não pode alterar esta definição depois da colecção ser criada. Se não activar a segurança de nível de colecção, não pode, mais tarde, especificar controlos de segurança de nível de documento.

Quando a segurança ao nível de colecção está activada:

- Os processos de análise global aplicam regras diferentes para indexar documentos duplicados.
- Pode configurar opções para impor segurança ao nível de documento.

- Pode impor segurança através do correlacionamento de aplicações (não utilizadores individuais) com a colecção à qual podem aceder. Em seguida, pode utilizar mecanismos de controlo de acesso padrão para permitir ou recusar o acesso a aplicações por parte dos utilizadores.
- Pode configurar o sistema para utilizar o componente de gestão de identidade, que permite aos utilizadores da aplicação serem autenticados sem configurar um perfil de aplicação.

Existe um compromisso entre activar a segurança de colecção e a qualidade da procura. Activar a segurança da colecção reduz as informações que são indexadas para cada documento. Um efeito secundário é o de que serão encontrados menos resultados para algumas consultas.

Segurança ao nível do documento

Ao configurar pesquisadores para uma colecção, pode activar a segurança de nível de documento. Por exemplo, pode especificar opções para associar tokens de segurança a dados à medida que estes são recolhidos pelos pesquisadores. As suas aplicações podem utilizar estes tokens, que são armazenados com documentos no índice, para realizar a filtragem prévia dos resultados e assegurar que apenas os utilizadores com as credenciais adequadas podem consultar os dados e ver os documentos.

Para determinados tipos de origens de dados, pode configurar opções para validar as credenciais de um início de sessão de um utilizador com controlos de acesso actuais durante o processamento da consulta. Esta camada extra de segurança de filtragem posterior assegura que os privilégios de utilizador são validados em tempo real com a origem de dados. Esta capacidade pode proteger contra ocorrências nas quais as credenciais de um utilizador sejam alteradas depois de um documento e respectivos tokens de segurança serem indexados.

A fase de processamento de texto âncora da análise global normalmente associa texto que aparece num documento (o texto de origem) a outro documentos (o documento de destino) no qual esse não aparece necessariamente. Quando configurar um pesquisador da Web, pode especificar se pretende excluir o texto âncora do índice, caso o texto tenha uma ligação a um documento que o pesquisador Web não esteja autorizado a pesquisar.

Codificação

Para proteger dados sensíveis, é utilizada codificação para a parte dos dados de autenticação de todas as mensagens que são transmitidas através do sistema. A palavra-passe para o administrador IBM Content Analytics with Enterprise Search predefinido é armazenada em formato codificado. As palavras-passe que os utilizadores especificam em perfis de utilizador e palavras-passe que são armazenadas pelo sistema (em ficheiros de configuração, em bases de dados internas, etc.) também são codificadas. Este processo consome pouco tempo porque apenas os IDs e palavras-passe de autenticação são codificados.

A segurança para as colecções do utilizador expande-se para além dos mecanismos de autenticação e controlo de acesso que o sistema pode utilizar para proteger conteúdo indexado. Também existem salvaguardas para evitar que um utilizador malicioso e não autorizado ganhe acesso aos dados enquanto estes estão em trânsito. Por exemplo, os servidores de procura utilizam protocolos como o Secure Sockets Layer (SSL), o Secure Shell (SSH) e o Secure Hypertext Transfer Protocol (HTTPS) para comunicarem com o servidor de controlador e as aplicações.

Para obter uma segurança aumentada, tem de assegurar que o hardware do servidor está isolado adequadamente e protegido de qualquer interferência de terceiros não autorizada. Ao instalar uma firewall, pode proteger os servidores da interferência de terceiros através de outra parte da rede. Certifique-se também de que não existem portas abertas nos servidores. Configure o sistema de maneira a que este escute por pedidos apenas em portas que estejam explicitamente atribuídas a actividades e aplicações de IBM Content Analytics with Enterprise Search.

Referências relacionadas:

Configurar suporte para autenticação de início de sessão único (SSO)

Activar segurança no servidor de aplicação da Web incorporado

Se utilizar o servidor de aplicação da Web incorporado, e planear impor a segurança de início de sessão quando os utilizadores administram o sistema IBM Content Analytics with Enterprise Search ou acedem a colecções, tem de configurar os controlos de segurança no sistema, colecções e níveis de origem de dados.

Para configurar o servidor de aplicação da Web incorporado para suportar pesquisa protegida:

- Na consola administrativa, configure a segurança de início de sessão da aplicação. Consulte "Configurar a segurança do início de sessão da aplicação no servidor de aplicação da Web incorporado".
- 2. Na consola administrativa, configure o suporte para autenticação de início de sessão único (SSO) ao configurar um ficheiro de chave Lightweight Third-Party Authentication (LTPA). Tem de verificar se a autenticação de início de sessão único LTPA está configurada correctamente. Consulte "Configurar suporte de início de sessão único no servidor de aplicação da Web incorporado" na página 367.
- 3. Active a segurança quando criar colecções e configurar pesquisadores que suportem a autenticação de início de sessão único. Pesquisadores diferentes têm requisitos de configuração diferentes. Reveja os requisitos do pesquisador para suportar a segurança e conclua as tarefas adequadas aos tipos de documentos que tenciona incluir nas colecções.Consulte "Requisitos de configuração do pesquisador para suporte de segurança" na página 397.
- Configure o componente de gestão de identidade para activar o início de sessão único para procura segura de origens de dados específicos. Consulte "Configurar a gestão de identidade" na página 394.

Configurar a segurança do início de sessão da aplicação no servidor de aplicação da Web incorporado

Se o sistema estiver configurado para utilizar o servidor de aplicação da Web incorporado, configure as definições de segurança do início de sessão da aplicação para especificar as informações que permitem aos utilizadores serem autenticados quando iniciam sessão para administrar o sistema ou quando iniciam sessão em aplicações. Se o sistema estiver configurado para utilizar o WebSphere Application Server, a autenticação tem de ser configurada através das definições de segurança global de WebSphere Application Server.

Se instalou um servidor Lightweight Directory Access Protocol (LDAP), pode configurar o sistema para utilizar o registo de utilizador LDAP para autenticar utilizadores. Para se certificar que a configuração está definida correctamente, pode testar a capacidade do sistema de se conectar ao servidor LDAP. Pode também executar assistentes para ver listas de possíveis atributos de nome exclusivo (DN), e rapidamente testar a validade das entradas do ID de utilizador e do ID de grupo no registo de utilizador LDAP.

Antes de configurar a segurança do início de sessão de aplicação incorporada, tem de recolher informações sobre o servidor LDAP, como, por exemplo, o nome do servidor, porta, credenciais para aceder a LDAP, DN de base para procurar utilizadores e grupos, o nome de atributo que pretende utilizar como nome de início de sessão, as classes de objecto para entradas de utilizador e de grupo e um nome de atributo que represente um membro de grupos. Estes dados de configuração dependem do tipo de produto LDAP que utiliza e de como é configurado no seu ambiente.

Restrição: Todos os servidores de origem de dados que incluir no sistema IBM Content Analytics with Enterprise Search devem consultar o mesmo servidor LDAP.

Para configurar a segurança do início de sessão da aplicação:

- 1. Faça clique em Segurança para abrir a vista Segurança.
- Na área de segurança do início de sessão da aplicação, clique em Acções > Configurar definições de início de sessão da aplicação.
- Na página Configurar definições de início de sessão da aplicação, seleccione as caixas de verificação para obrigar a que os utilizadores iniciem sessão e utilizem o servidor LDAP para autenticar utilizadores.

Sugestão: Pode seleccionar a caixa de verificação para permitir que o ID do administrador predefinido do IBM Content Analytics with Enterprise Search aceda a todas as aplicações sem adicionar o ID ao seu registo de utilizador LDAP.

- 4. Especifique a informação sobre o servidor LDAP que instalou para suportar a autenticação de início de servidor. Deve identificar o nome, porta e credenciais do sistema central para aceder ao servidor, se forem requeridas credenciais.Se utilizar o Active Directory como o seu servidor LDAP, um formato típico para o nome de utilizador é Administrador e a porta típica é a 389. Se utilizar o Tivoli Directory Server como o seu servidor LDAP, um formato típico para o nome de utilizador é uid=administrator,o=analytics e a porta típica é a 389.
- 5. Clique em **Testar Conexão de Servidor LDAP**. Para ser capaz de descobrir os atributos DN no registo LDAP, o servidor de procura deve conectar-se com sucesso ao servidor LDAP.
- 6. Se tiver conhecimentos sobre os seus dados de configuração de LDAP, especifique as propriedades LDAP para as entradas de utilizador e de grupo e teste se consegue iniciar sessão no servidor LDAP:
 - a. Insira o DN de base para entradas de utilizador, atributo de ID de utilizador e classe de ID de objecto na página Configurar definições de início de sessão da aplicação.
 - b. Faça clique em Testar Entradas de Utilizador, introduza um ID de utilizador e palavra-passe válidos e faça clique em Início de Sessão de Teste. Se o início de sessão for bem sucedido, a configuração dos atributos de entrada de utilizador para o seu servidor LDAP será estará, em princípio, correcta.
 - c. Insira o DN de base para entradas de grupo, atributo de ID de grupo, atributo de membro de grupo e classe de ID de objecto na página Configurar definições de início de sessão da aplicação.

- d. Faça clique em Testar Entradas de Grupo, introduza um DN de utilizador válido e faça clique em Obter Informações do Grupo. Se o início de sessão for bem sucedido, é apresentada a informação sobre os grupos aos quais o DN de utilizador pertence e a configuração de atributos de grupo para o seu servidor LDAP estará, em princípio, correcta.
- 7. Se não tiver conhecimentos sobre os seus dados de configuração de LDAP, configure as propriedades LDAP para entradas de utilizador e de grupo utilizando um assistente para descobrir possíveis atributos:
 - a. Clique em **Descobrir Entradas**. Insira um DN de base de exemplo para uma entrada de utilizador, como o=ibm, e depois clique em **Descobrir entradas de utilizador**.
 - b. Em **Possíveis entradas de utilizador**, seleccione um item como exemplo de entrada de utilizador, como uid=esuser01,o=ibm.
 - **c**. Seleccione um DN base, um atributo de ID de utilizador, e uma classe de objecto do ID das listas de possíveis atributos DN que o sistema descobre.
 - d. Faça clique em Testar Entradas de Utilizador, introduza um ID de utilizador e palavra-passe válidos e faça clique em Início de Sessão de Teste. Se o início de sessão for bem sucedido, a configuração dos atributos de entrada de utilizador para o seu servidor LDAP será estará, em princípio, correcta. O teste de configuração para entradas de utilizador deve ser bem sucedido antes de o assistente poder descobrir possíveis entradas de grupo.
 - e. Insira um DN de base de exemplo para uma entrada de grupo, como o=ibm, e depois clique em **Descobrir Entradas de Grupo**.
 - f. Em **Possíveis entradas de grupo**, seleccione um item como exemplo de entrada de grupo, como cn=ibmusers,o=analytics.
 - g. Seleccione um DN base, um atributo de ID de grupo, atributo de membro de grupo e classe de objecto do ID das listas de possíveis atributos DN que o sistema descobre.
 - h. Faça clique em Testar Entradas de Grupo, introduza um DN de utilizador válido e faça clique em Obter Informações do Grupo. Se o início de sessão for bem sucedido, é apresentada a informação sobre os grupos aos quais o DN de utilizador pertence e a configuração de atributos de grupo para o seu servidor LDAP estará, em princípio, correcta.
 - i. Faça clique em **OK**. Os atributos que foram testados para entradas de utilizador e de grupo são copiados para a página Configurar definições de início de sessão da aplicação.
- 8. Opcional: Configure os tokens LTPA e o ficheiro de chave para suportar a autenticação de início de sessão único (SSO) para procurar origens que são pesquisadas por pesquisadores que suportam início de sessão único.
- 9. Reinicie o sistema IBM Content Analytics with Enterprise Search:

```
esadmin system stopall
esadmin system startall
```

Configurar suporte de início de sessão único no servidor de aplicação da Web incorporado

Quando activa as definições de segurança de início de sessão da aplicação para o servidor de aplicação da Web incorporado, é possível especificar as informações que permitam aos utilizadores procurar origens protegidas que suportem autenticação de início de sessão único (SSO).

Para expandir a autenticação de início de sessão único para utilizadores que consultam colecções activadas por início de sessão único, pode configurar o sistema para utilizar tokens de Lightweight Third-Party Authentication (LTPA). Pode gerar num novo ficheiro de chave LTPA, exportar o ficheiro de chave para ser partilhado por outras aplicações na sua empresa, ou importar um ficheiro de chave existente para o servidor de procura. O mesmo ficheiro de chave, que é utilizado para descodificar um token LTPA, deve existir em todos os servidores que partilham a sessão SSO.

Se as suas colecções incluírem origens que necessitem de WebSphere Application Server, como, por exemplo, origens alojadas nos servidores de WebSphere Portal, é possível importar o ficheiro de chave LTPA no formato que é necessário no WebSphere Application Server. Por exemplo, para activar o início de sessão único para procura protegida, é possível importar o ficheiro de chave de WebSphere Application Server para IBM Content Analytics with Enterprise Search, ou exportar o ficheiro de chave de IBM Content Analytics with Enterprise Search para WebSphere Application Server.

Para configurar o sistema para suportar a autenticação de início de sessão único:

- 1. Faça clique em **Segurança** para abrir a vista Segurança.
- Na área de segurança do início de sessão da aplicação, clique em Acções > Configurar definições de início de sessão da aplicação.
- 3. Configure os tokens LTPA e o ficheiro de chave para suportar a segurança SSO. Recolha dados de configuração de início de sessão único a partir do sistema que utiliza para a autenticação de início de sessão único. Todos os parâmetros que especificar na consola administrativa do IBM Content Analytics with Enterprise Search têm de corresponder aos parâmetros de configuração no seu sistema de início de sessão único primário.
 - a. Especifique um limite de tempo para sessões SSO (o número de minutos antes de o token LTPA expirar).
 - b. Especifique um domínio onde o cookie LTPA possa ser acedido por todos os servidores que partilhem a sessão de início de sessão único, como ltpa.example.com.
 - c. Se partilhar o ficheiro de chave LTPA com um servidor WebSphere Application Server que utilize um repositório federado, copie o valor de sufixo do nome de utilizador adicional da entrada de base do repositório federado em WebSphere Application Server.
 - d. Se partilhar o ficheiro de chave LTPA com uma aplicação que não suporta a propagação de atributos, como as versões mais antigas do WebSphere Application Server, seleccione a caixa de verificação para permitir a interoperabilidade.
 - e. Se o ficheiro de chave existir no servidor de procura, especifique o nome do caminho e do ficheiro. A localização padrão para o ficheiro de chave é ES_NODE_ROOT/master_config/.
 - f. Se um ficheiro de chave não existir, pode gerar um ou importá-lo de outra localização. Se importar o ficheiro de chave e especificar apenas o nome do ficheiro, a localização padrão é ES_NODE_ROOT/master_config/.
 - g. Para assegurar que o mesmo ficheiro de chave é partilhado por todas as aplicações que partilham a sessão SSO, pode exportar o ficheiro de chave e depois importá-lo para as suas aplicações. Ao exportar o ficheiro de chave, este é criado como ES_NODE_ROOT/master_config/export.*hostname*.key.

- h. Se partilhar o ficheiro de LTPA com um servidor WebSphere Application Server que utilize um repositório federado, copie o valor do nome do nicho adicional da configuração de repositório federado em WebSphere Application Server.
- 4. Reinicie o sistema IBM Content Analytics with Enterprise Search:

```
esadmin system stopall
esadmin
system startall
```

- 5. Certifique-se de que o LTPA SSO está correctamente configurado:
 - a. Inicie sessão num servidor de origem de dados que suporte o início de sessão único e, de seguida, passe para a aplicação IBM Content Analytics with Enterprise Search utilizando a mesma janela do navegador.
 - b. Se consegue alcançar a interface da aplicação sem que lhe seja solicitado o início de sessão, então a configuração foi concluída com êxito. A página Meu Perfil não é um pedido de início de sessão. Não tome a página de perfil por um ecrã de início de sessão.
 - c. Tente iniciar sessão utilizando a ordem oposta. Primeiro, inicie sessão na aplicação de procura empresarial ou prospecção de análise de conteúdo e, em seguida, desloque-se para o servidor de origem de dados na mesma janela de navegador.

Estas são duas definições de configuração para configurar pesquisadores para suportar a autenticação de início de sessão único:

- Quando configurar o componente de gestão de identidade na página Segurança, seleccione a caixa de verificação para cada tipo de pesquisador na colecção que pretende activar para suportar o início de sessão único.
- Quando configurar definições de segurança para um pesquisador individual, activa o início de sessão único.

Quando ambas estas definições estiverem configuradas para suportar a autenticação de início de sessão único, a procura protegida de início de sessão único está em vigor. A aplicação pára de pedir aos utilizadores que introduzam as credenciais na caixa de diálogo Perfil Pessoal.

Para obter informações completas sobre suporte para o início de sessão único no IBM Content Analytics with Enterprise Search, consulte *Configurar suporte para autenticação de início de sessão único (SSO)*.

Conceitos relacionados:

"Validação de utilizador através de autenticação de início de sessão único" na página 392

Tarefas relacionadas:

"Configurar a gestão de identidade" na página 394

Referências relacionadas:

Configurar suporte para autenticação de início de sessão único (SSO)

Activar a segurança no WebSphere Application Server

Se utilizar o WebSphere Application Server e tencionar fazer cumprir a segurança no início de sessão quando os utilizadores administrarem o sistema IBM Content Analytics with Enterprise Search ou acederem a colecções, tem de configurar a segurança global no WebSphere Application Server. Também tem de configurar informações de segurança nos ficheiros de configuração IBM Content Analytics with Enterprise Search, na consola administrativa e nas aplicações.

Para activar a segurança no WebSphere Application Server:

- 1. Decida qual o tipo de registo de utilizadores que pretende utilizar para autenticar utilizadores. Por exemplo, muitos administradores do WebSphere Application Server optam por utilizar o registo de utilizadores do Lightweight Directory Access Protocol (LDAP).
- Na consola de administração do IBM Content Analytics with Enterprise Search, faça clique em Segurança para abrir a vista-resumo Segurança e atribua pelo menos um dos utilizadores no registo de utilizadores do WebSphere Application Server para o cargo de administrador principal.

Em alternativa, adicione o ID de administrador predefinido IBM Content Analytics with Enterprise Search ao registo de utilizadores WebSphere Application Server.

- 3. Se activar a segurança global depois de o IBM Content Analytics with Enterprise Search ter sido instalado, tem de fornecer o ID de utilizador e palavra-passe do WebSphere Application Server ao sistema. Para fornecer estas informações, deve utilizar o comando **eschangewaspw**. Siga o procedimento adequado para o modo como instalou o IBM Content Analytics with Enterprise Search, quer seja num servidor único ou numa configuração de servidor distribuído.
- 4. No WebSphere Application Server, configure o suporte para a autenticação de início de sessão único (SSO, single sign-on) ao configurar um ficheiro de chave Lightweight Third-Party Authentication (LTPA). Tem de verificar se a autenticação de início de sessão único LTPA está configurada correctamente. Consulte "Configurar o suporte SSO no WebSphere Application Server".
- 5. Active a segurança quando criar colecções e configurar pesquisadores que suportem a autenticação de início de sessão único. Pesquisadores diferentes têm requisitos de configuração diferentes. Reveja os requisitos do pesquisador para suportar a segurança e conclua as tarefas adequadas aos tipos de documentos que tenciona incluir nas colecções.Consulte "Requisitos de configuração do pesquisador para suporte de segurança" na página 397.
- Configure o componente de gestão de identidade para activar o início de sessão único para procura segura de origens de dados específicos. Consulte "Configurar a gestão de identidade" na página 394.

Tarefas relacionadas:

"Activar segurança para um sistema de servidor único no WebSphere Application Server" na página 372

"Activar a segurança para um sistema de servidor distribuído no WebSphere Application Server" na página 373

Referências relacionadas:

"Requisitos de configuração do pesquisador para suporte de segurança" na página 397

Activar a segurança no servidor WebSphere Application 8.0

Configurar o suporte SSO no WebSphere Application Server

Pode activar as capacidades de suporte do início de sessão único (SSO, single sign-on) que permitem aos utilizadores ignorar o diálogo de correlação de credencial de utilizador do Meu perfil numa aplicação de procura empresarial ou prospecção de análise de conteúdo. Este procedimento é necessário para expandir o suporte de SSO a quaisquer origens que implementa num ambiente do WebSphere Application Server. Isto inclui origens que são implementadas através do WebSphere Portal, tais como origens que são pesquisadas pelo pesquisador da Lista de valores geradores. Para suportar a autenticação de SSO quando os utilizadores acedem às aplicações do IBM Content Analytics with Enterprise Search, tem de confirmar primeiro que o SSO está configurado da forma correcta no WebSphere Application Server.

Para configurar o suporte de SSO no WebSphere Application Server e verificar se está activado para procurar origens seguras:

- 1. Configurar capacidades de LTPA SSO em WebSphere Application Server:
 - a. No servidor que aloja o WebSphere Application Server, tal como um servidor do IBM Connections que pretende adicionar a uma colecção, exporte o ficheiro de chave Lightweight Third-Party Authentication (LTPA). Na consola de administração do WebSphere Application Server, utilize a configuração de início de sessão único de Célula cruzada da WebSphere Application Server Integrated Solutions Console.
 - b. No servidor do IBM Content Analytics with Enterprise Search que aloja o WebSphere Application Server, importe o ficheiro de chave LTPA que exportou.
 - c. Certifique-se de que os dados de configuração de SSO são os mesmos nos dois servidores do WebSphere Application Server, tal como o Nome de domínio e o nome da área de SSO.
 - d. Reinicie ambos os servidores.
- 2. Certifique-se de que o LTPA SSO está correctamente configurado:
 - a. Inicie sessão num servidor de origem de dados que suporte SSO, tal como um servidor do IBM Connections de destino, em seguida, desloque-se para a aplicação do IBM Content Analytics with Enterprise Search através da utilização da mesma janela de navegador.
 - b. Se consegue alcançar a interface da aplicação sem que lhe seja solicitado o início de sessão, então a configuração foi concluída com êxito. A página Meu Perfil não é um pedido de início de sessão. Não tome a página de perfil por um ecrã de início de sessão.
 - **c.** Tente iniciar sessão utilizando a ordem oposta. Primeiro, inicie sessão na aplicação de procura empresarial ou prospecção de análise de conteúdo e, em seguida, desloque-se para o servidor de origem de dados na mesma janela de navegador.
- **3**. Configure pesquisadores para suportarem a autenticação de SSO. Execute estes passos para cada pesquisador que pretende activar para utilizar os dados de configuração do WebSphere Application Server:
 - a. Aceda ao directório ES_NODE_ROOT/master_config/ ID_colecção.ID_pesquisador, sendo que ID_pesquisador identifica um pesquisador que tenha criado.
 - b. Abra o ficheiro *tipo_pesquisador*.xml (tais como seedlistcrawler.xml) com um editor com noção de UTF-8. Adicione o seguinte atributo ao elemento do /Pesquisador/DataSources/Servidor: SS0Enabled="true"
 - Sound real crue
 - c. Abra o ficheiro ES_NODE_ROOT/master_config/sso.properties e adicione a linha seguinte: imc.sso.tipo_origem=true Por exemplo: imc.sso.connections=true
 - d. Reinicie o sistema IBM Content Analytics with Enterprise Search:

```
esadmin system stopall
esadmin
system startall
```

Estas são duas definições de configuração para configurar pesquisadores para suportar a autenticação de início de sessão único:

- Quando configurar o componente de gestão de identidade na página Segurança, seleccione a caixa de verificação para cada tipo de pesquisador na colecção que pretende activar para suportar o início de sessão único.
- Quando configurar definições de segurança para um pesquisador individual, activa o início de sessão único.

Quando ambas estas definições estiverem configuradas para suportar a autenticação de início de sessão único, a procura protegida de início de sessão único está em vigor. A aplicação pára de pedir aos utilizadores que introduzam as credenciais na caixa de diálogo Perfil Pessoal.

Para obter informações completas sobre suporte para o início de sessão único no IBM Content Analytics with Enterprise Search, consulte *Configurar suporte para autenticação de início de sessão único (SSO)*.

Conceitos relacionados:

"Validação de utilizador através de autenticação de início de sessão único" na página 392

Tarefas relacionadas:

"Configurar a gestão de identidade" na página 394

Referências relacionadas:

Configurar suporte para autenticação de início de sessão único (SSO)

Activar segurança para um sistema de servidor único no WebSphere Application Server

Se activar segurança global do WebSphere Application Server depois de instalar o IBM Content Analytics with Enterprise Search, ou se alterar a palavra-passe do utilizador do WebSphere Application Server que é utilizada para o IBM Content Analytics with Enterprise Search, tem de executar o comando **eschangewaspw** para propagar a alteração ao sistema IBM Content Analytics with Enterprise Search.

No ficheiro de configuração ES_INSTALL_ROOT/nodeinfo/es.cfg, certifique-se de que o WASUser especifica um nome de utilizador WebSphere Application Server válido. Para ser válido:

- O nome de utilizador tem de ser um ID de utilizador administrativo ou ID do servidor ou se configurar funções administrativas, um ID que pertença à função do operador ou à função de administrador principal.
- O nome de utilizador tem de estar registado no repositório de utilizador que especificar quando activar a segurança global no WebSphere Application Server.

O comando **eschangewaspw** codifica a palavra-passe para o utilizador WebSphere Application Server e depois armazena a palavra-passe codificada no ficheiro ES_INSTALL_ROOT/nodeinfo/es.cfg. O nome de utilizador e palavra-passe são utilizados principalmente para parar a aplicação ESSearchServer no WebSphere Application Server. Para activar a segurança num sistema de servidor único existente, ou para propagar uma nova palavra-passe ao utilizador do WebSphere Application Server:

1. Inicie sessão como administrador predefinido do IBM Content Analytics with Enterprise Search e pare o sistema:

esadmin system stopall

Importante: Quando o sistema é parado, os utilizadores não podem submeter pedidos de procura.

2. Execute o script seguinte, onde *WAS_password* é a palavra-passe para o utilizador do WebSphere Application Server:

AIX ou Linux

eschangewaspw.sh WAS_password

Windows

eschangewaspw WAS_password

3. No Windows, execute o comando **WASService** para actualizar o registo de palavras-passe para o serviço do Windows. O serviço tem de ser executado como LocalSystemAccount. Por exemplo:

WASService -add ESSearchServer -serverName ESSearchServer -profilePath "WASDIR\profiles\default" -stopArgs "-username WASADMIN -password PASSWORD" -startType automatic△-logRoot "WASDIR\profiles\default\logs\ ESSearchServer"

em que:

WASDIR

É o directório onde o WebSphere Application Server está instalado. Certifique-se de que especifica o caminho para o perfil WebSphere Application Server em vez do caminho predefinido, WASDIR\profiles\ default.

WASADMIN

É o nome de utilizador especificado no ficheiro ES_INSTALL_ROOT/nodeinfo/ es.cfg.

PASSWORD

É a palavra-passe do utilizador especificado no ficheiro ES_INSTALL_ROOT/nodeinfo/es.cfg.

4. Reinicie o sistema IBM Content Analytics with Enterprise Search:

esadmin system startall

Tarefas relacionadas:

"Activar a segurança no WebSphere Application Server" na página 369 **Referências relacionadas**:

- 🕩 Activar a segurança no servidor WebSphere Application 8.0
- Comando WASService, Servidor WebSphere Application 8.0

Activar a segurança para um sistema de servidor distribuído no WebSphere Application Server

Se activar a segurança global do WebSphere Application Server depois de instalar o IBM Content Analytics with Enterprise Search, ou se alterar a palavra-passe do utilizador do WebSphere Application Server, tem de executar o comando **eschangewaspw** para propagar a alteração ao sistema IBM Content Analytics with Enterprise Search. No ficheiro de configuração ES_INSTALL_ROOT/nodeinfo/es.cfg, certifique-se de que o WASUser especifica um nome de utilizador WebSphere Application Server válido. Para ser válido:

- O nome de utilizador tem de ser um ID de utilizador administrativo ou ID do servidor ou se configurar funções administrativas, um ID que pertença à função do operador ou à função de administrador principal.
- O nome de utilizador tem de estar registado no repositório de utilizador que especificar quando activar a segurança global no WebSphere Application Server.

O comando **eschangewaspw** codifica a palavra-passe para o utilizador WebSphere Application Server e depois armazena a palavra-passe codificada no ficheiro ES_INSTALL_ROOT/nodeinfo/es.cfg. O nome de utilizador e palavra-passe são utilizados principalmente para parar a aplicação ESSearchServer no WebSphere Application Server.

Para activar a segurança num sistema de servidor distribuído existente ou para propagar uma nova palavra-passe ao utilizador WebSphere Application Server:

- 1. Execute os seguintes passos no servidor do controlador:
 - a. Inicie sessão como administrador predefinido do IBM Content Analytics with Enterprise Search e pare o sistema: esadmin system stopall

Importante: Quando o sistema é parado, os utilizadores não podem submeter pedidos de procura.

b. Execute o script seguinte, em que WAS_password é a palavra-passe para o utilizador WebSphere Application Server que é especificado no ficheiro ES INSTALL ROOT/nodeinfo/es.cfg.

AIX ou Linux

eschangewaspw.sh WAS_password

Windows

eschangewaspw WAS_password

- 2. Execute os passos seguintes nos outros servidores do sistema.
 - a. Inicie sessão como administrador predefinido do IBM Content Analytics with Enterprise Search.
 - b. Execute o script seguinte, em que WAS_password é a palavra-passe para o utilizador WebSphere Application Server que é especificado no ficheiro ES_INSTALL_ROOT/nodeinfo/es.cfg.

AIX ou Linux

eschangewaspw.sh WAS_password

Windows

eschangewaspw WAS_password

3. Nos servidores de procura que são executados na plataforma Windows execute o comando **WASService** para actualizar o registo de palavras-passe para o serviço do Windows. O serviço tem de ser executado como LocalSystemAccount. Por exemplo:

WASService -add ESSearchServer -serverName ESSearchServer -profilePath "WASDIR\profiles\default" -stopArgs "-username WASADMIN -password PASSWORD" -startType automatic△-logRoot "WASDIR\profiles\default\logs\ ESSearchServer"

em que:

WASDIR

É o directório onde o WebSphere Application Server está instalado. Certifique-se de que especifica o caminho para o perfil WebSphere Application Server em vez do caminho predefinido, WASDIR\profiles\ default.

WASADMIN

É o nome de utilizador especificado no ficheiro ES_INSTALL_ROOT/nodeinfo/ es.cfg.

PASSWORD

É a palavra-passe do utilizador especificado no ficheiro ES_INSTALL_ROOT/nodeinfo/es.cfg.

4. No servidor do controlador, reinicie o sistema IBM Content Analytics with Enterprise Search:

esadmin system startall

Tarefas relacionadas:

"Activar a segurança no WebSphere Application Server" na página 369

Referências relacionadas:

- Activar a segurança no servidor WebSphere Application 8.0
- Comando WASService, Servidor WebSphere Application 8.0

Segurança ao nível do sistema

Os controlos de segurança ao nível do sistema incluem a capacidade de atribuir funções administrativas, atribuir privilégios de utilizador e especificar credenciais que permitem aos componentes do sistema aceder a origens de dados.

Conceitos relacionados:

"Autenticação versus controlo de acesso"

"ID de administrador predefinido" na página 376

Tarefas relacionadas:

"Configurar funções administrativas" na página 376

"Configurar privilégios de utilizador da aplicação" na página 378

"Configurar credenciais de ligação" na página 380

Autenticação versus controlo de acesso

Para proteger o conteúdo de utilizadores de aplicação não autorizados e para controlar o acesso a funções administrativas, o IBM Content Analytics with Enterprise Search suporta autenticação autorização de utilizadores (controlos de acesso).

Autenticação

Autenticação é qualquer processo pelo qual um sistema verifica a identidade de um utilizador que pretenda obter acesso ao sistema. Como o controlo de acesso é normalmente baseado na identidade do utilizador que pede acesso a um recurso, a autenticação é essencial para uma segurança efectiva.

A autenticação de utilizadores é implementada através de credenciais que, no mínimo, consistem num ID de utilizador e palavra-passe.

Autorização (controlo de acesso)

Autorização é qualquer mecanismo através do qual um sistema concede ou revoga os direitos de acesso a certos dados ou executa alguma acção. Muitas vezes, um utilizador tem de iniciar sessão utilizando qualquer forma de autenticação. Os mecanismos de controlo de acesso determinam as operações que o utilizador pode ou não executar comparando a identidade do utilizador com uma lista de controlo de acessos (ACL). Os controlos de acesso englobam:

- Permissões de ficheiros, tais como o direito a criar, ler, editar ou eliminar um ficheiro.
- Permissões de programas, tais como o direito a executar um programa.
- Permissões de dados, tais como o direito a obter ou actualizar informações numa base de dados.

Conceitos relacionados:

"Segurança ao nível do sistema" na página 375

ID de administrador predefinido

O programa de instalação para IBM Content Analytics with Enterprise Search estabelece um ambiente para fazer cumprir a segurança quando os utilizadores administram o sistema.

Durante o processo de instalação, é pedido ao instalador um ID de utilizador e palavra-passe a utilizar para o administrador predefinido do IBM Content Analytics with Enterprise Search. O sistema utiliza as credenciais especificadas para:

- Autenticar o administrador quando são executadas tarefas administrativas.
- Criar as bases de dados internas.
- · Iniciar todas as sessões ou processos de sistema.

O ID de utilizador que é especificado durante a instalação tem de ser um ID de utilizador válido do sistema operativo com privilégios administrativos do sistema. O programa de instalação armazena as credenciais, devidamente codificadas, num ficheiro de propriedades em cada servidor de sistema.

Conceitos relacionados:

"Segurança ao nível do sistema" na página 375

Configurar funções administrativas

Através da configuração de funções administrativas, pode restringir o acesso a colecções e controlar as funções que cada utilizador administrativo pode efectuar. Após iniciar sessão na consola administrativa, o utilizador acede apenas às funções que estiverem associadas à função administrativa atribuída, retendo essas permissões ao longo da sessão de administração.

Quando o IBM Content Analytics with Enterprise Search é instalado, o instalador configura um ID de utilizador e palavra-passe para o administrador predefinido. Na primeira vez que aceder à consola de administração, tem de iniciar sessão como este utilizador.

Antes de adicionar outros utilizadores administrativos, certifique-se de que os utilizadores estão registados no registo de utilizadores do Lightweight Directory Access Protocol (LDAP) que o sistema do IBM Content Analytics with Enterprise Search está configurado para utilizar. Se utilizar o WebSphere Application Server, active a segurança global para configurar o servidor de LDAP e registar utilizadores. Todos os utilizadores administrativos que adiciona ao IBM Content Analytics with Enterprise Search têm de estar registados no registo de utilizadores para o servidor do WebSphere Application Server LDAP.

Pode inscrever utilizadores nas seguintes funções administrativas:

Administrador principal

Estes utilizadores criam colecções e têm autoridade para gerirem todos os aspectos do sistema, incluindo a adição de utilizadores administrativos. Até que o administrador predefinido nomeie outros utilizadores à função de administrador principal, só o administrador predefinido poderá criar colecções ou adicionar utilizadores administrativos.

Administrador da colecção

Estes utilizadores podem editar, supervisionar e controlar o funcionamento de colecções que estejam autorizados a administrar. Estes utilizadores não podem criar colecções. Os administradores de colecção podem supervisionar e exercer actividades ao nível do sistema apenas se lhes for concedida autoridade por um administrador principal.

Operador

Estes utilizadores podem supervisionar e controlar o funcionamento de colecções que estejam autorizados a administrar. Estes utilizadores podem iniciar ou parar actividades de colecção, por exemplo, mas não podem criar nem editar colecções. Um operador pode supervisionar e exercer actividades ao nível do sistema apenas se lhes for concedida autoridade por um administrador principal.

Supervisionar

Estes utilizadores podem supervisionar colecções que estejam autorizados a administrar. Estes utilizadores não podem controlar operações (tais como iniciar e parar servidores), nem criar nem editar colecções. Um supervisor pode observar, mas não exercer, actividades ao nível de sistema, apenas se essa autoridade lhe for concedida por um administrador principal.

Administrador de análise de texto

Estes utilizadores podem criar, modificar e supervisionar recursos de análise de texto relativos às colecções de análise de texto que estão autorizados a administrar (facetas, dicionários e regras de análise de texto personalizadas).

Editor de árvore da faceta

Estes utilizadores podem configurar a árvore da faceta para colecções de análise de texto que estejam autorizados a administrar.

Editor do dicionário do utilizador

Estes utilizadores podem configurar os dicionários do utilizador personalizados para colecções de análise de conteúdos que estão autorizados a administrar.

Categorizador com base em regras

Estes utilizadores pode configurar categorias com base em regras para colecções de procura empresarial e de análise de conteúdos.

Personalizador de aplicações

Estes utilizadores podem personalizar aplicações de procura empresarial e a prospecção de análise de conteúdo através da utilização de aplicações do personalizador. Na consola administrativa, as únicas acções que estes utilizadores podem realizar é fazer clique em **Procurar no Personalizador** e no **Personalizador de Análise**.

Para nomear um utilizador para uma função administrativa:

- 1. Faça clique em **Segurança** para abrir a vista Segurança.
- 2. Na área de segurança ao nível do sistema, faça clique em Acções > Adicionar utilizador administrativo.
- **3.** Escreva o ID de utilizador do utilizador que pretende inscrever e seleccione uma função administrativa adequada.
- 4. Se não estiver a inscrever este utilizador como administrador principal, especifique se este utilizador pode aceder às funções da barra de ferramentas do **Sistema**.

Por exemplo, pode querer permitir que alguns operadores ou administradores de colecções iniciem e parem os servidores do sistema.

5. Se não estiver a inscrever este utilizador como administrador principal, seleccione as colecções que este utilizador pode gerir. Pode seleccionar as caixas de verificação para colecções individuais ou permitir que o utilizador efectue a gestão de todas as colecções, todas as colecções de procura empresarial, ou todas as colecções de análise de conteúdos.

Pode editar a função de um utilizador, alterar as colecções que o utilizador está autorizado a administrar, ou eliminar o utilizador da lista de administradores. Não pode editar nem eliminar o administrador predefinido IBM Content Analytics with Enterprise Search.

Conceitos relacionados:

"Segurança ao nível do sistema" na página 375

Tarefas relacionadas:

"Iniciar sessão na consola de administração" na página 25

Configurar privilégios de utilizador da aplicação

Ao configurar os privilégios de utilizadores de aplicações, pode controlar as funções que disponíveis para utilizadores de aplicações.

O modo como configura os privilégios de utilizador depende de as definições de início de sessão da aplicação estarem, ou não, configuradas para o sistema IBM Content Analytics with Enterprise Search:

- Caso não tenha activado as definições de início de sessão da aplicação, ou caso pretenda conceder privilégios a utilizadores de aplicações sem adicionar IDs de utilizador ou nomes de grupos ao sistema, pode especificar privilégios predefinidos.
- Se activou as definições de início de sessão na aplicação e configurou o sistema para utilizar um servidor Lightweight Directory Access Protocol (LDAP) para autenticar utilizadores, certifique-se de que os nomes de utilizador e nomes de grupo que pretende autorizar existem no registo de utilizadores. Para que uma função seja disponibilizada, o ID especificado pelo utilizador para iniciar sessão na aplicação deve corresponder ao ID de utilizador indicado no registo.
- Se tiver activado as definições de início de sessão e utilizar o WebSphere Application Server para gerir funções e autenticações de utilizadores, tem de utilizar a consola administrativa WebSphere Application Server para correlacionar funções de segurança específicas com utilizadores ou grupos. A tabela seguinte apresenta os privilégios de utilizador e as funções que têm de

estar atribuídas no WebSphere Application Server se não processar a autenticação através do servidor de aplicação da Web incorporado.

Privilégios de utilizadores de aplicações	Funções necessárias, caso utilize o WebSphere Application Server		
Privilégios para a procura empresarial e a prospecção de conteúdos			
Guardar procuras.	SAVE_SEARCH		
Exportar documentos que correspondem às condições de consulta actuais.	EXPORT		
Defina e limpe sinalizadores para documentos seleccionados ou para todos os documentos nos resultados.	DOCUMENT_FLAGGING		
Utilize o construtor de consultas para construir consultas destacando texto e facetas num documento seleccionado.	QUERY_BUILDER		
Adicione regras à categoria baseadas na consulta actual.	CATEGORY_RULE_REGISTRATION		
Reconstrua o índice de categorias após adicionar regras de categoria. Necessita do privilégio de adicionar regras que se baseiam na consulta actual a uma categoria.	CATEGORY_REBUILD		
Privilégios para a procura empresarial			
Modifique o esquema dos painéis apresentados na aplicação de procura empresarial.	MODIFY_LAYOUT		
Exporte documentos sinalizados para uma colecção de análise de conteúdos.	EXPORT_FLAGGED_DOCUMENT		
Privilégios para prospecção de conteúdos			
Crie relatórios de inspecção aprofundada para explorar estatísticas de análise com outras ferramentas em vez de o fazer interactivamente por intermédio da prospecção de análise de conteúdo.	DEEP_INSPECTION		
Crie relatórios que possam ser abertos com ferramentas de análise de IBM Cognos Business Intelligence.	COGNOS_REPORT		

Para atribuir privilégios de utilizadores de aplicações:

- 1. Faça clique em **Segurança** para abrir a vista-resumo Segurança e aceda à área de segurança ao nível do sistema.
- 2. Caso as definições de início de sessão na aplicação estiverem activadas e os utilizadores tenham de iniciar sessão na aplicação:
 - a. Faça clique em Acções > Adicionar um utilizador ou grupo da aplicação.
 - b. Especifique um ID de utilizador para autorizar um utilizador individual, ou um nome de grupo para autorizar um grupo de utilizadores, e de seguida seleccione as caixas de verificação para as funções que podem ser executadas pelo utilizador.
 - **c.** Caso tenha especificado um ID de utilizador, especifique o ID de aplicação para as colecções para as quais este utilizador tem privilégios. Pode

seleccionar um ID de aplicação existente ou criar um novo ID de aplicação e seleccionar as colecções para as quais pretende que este utilizador tenha privilégios.

- **3**. Se os utilizadores não tiverem de iniciar sessão na aplicação, ou se quiser especificar funções de aplicação personalizadas para utilizadores sem ter de as registar explicitamente:
 - a. Faça clique em Acções > Especificar privilégios de utilizadores da aplicação predefinidos.
 - b. Seleccione as caixas de verificação relativas às funções que podem ser efectuadas por todos os utilizadores.

A aplicação à qual se aplica estes privilégios é controlada pelas definições de configuração da aplicação. Por exemplo, pode utilizar as aplicações do personalizador para identificar o ID da aplicação e as colecções aos quais se aplicam estes privilégios.

Conceitos relacionados:

"Segurança ao nível do sistema" na página 375

Configurar credenciais de ligação

Na vista segurança, pode especificar credenciais que permitam aos pesquisadores aceder às origens dos dados de acesso. As credenciais são armazenadas numa localização comum, denominada *cofre de credenciais*.

O cofre de credenciais é um repositório em que todas as informações destino do pesquisador e as credenciais para pesquisar as origens de dados destino são federadas num ficheiro. Ao ter estas informações numa localização, os pesquisadores em servidores diferentes num ambiente de servidores múltiplos podem todos aceder a credenciais de que necessitam para aceder à origem de dados.

O componente de gestão de identidade remete ao cofre para obter as informações de que necessita para aceder a origens de dados seguras. O cofre de credenciais inclui informações acerca da segurança ao nível de documentos estar ou não activada para a origem de dados e, se for o caso, se as credenciais de utilizador têm de ser validadas quando os utilizadores pesquisam a colecção.

Os servidores de pesquisa também acedem ao cofre de credenciais ao determinar se deve ser pedido ao utilizador as credenciais e quando executam a pós-filtragem dos resultados da pesquisa para assegurar que os documentos que o utilizador não está autorizado a ver são excluídos.

Para configurar credenciais de ligação de origem de dados

- 1. Faça clique em Segurança para abrir a vista Segurança.
- Na área de segurança a nível do sistema, clique em Acções > Especificar credenciais de ligação de origem de dados.
- 3. Na página Especificar Credenciais de Ligação, seleccione o tipo de origem de dados ou protocolo de ligação que pretende configurar e especifique um ID para a ligação. Pesquisadores, o componente de gestão de identidade e os servidores da pesquisa utilizam o ID da ligação para identificar as credenciais que são necessárias para aceder à origem de dados.
- 4. Especifique credenciais que permitam acesso à origem de dados. As informações que deve facultar dependem do tipo de origem de dados que seleccionar.

Pode editar as credenciais de ligação se os requisitos para aceder a uma origem de dados mudarem. Por exemplo, poderá necessitar de actualizar uma palavra-passe ou alterar informações acerca do servidor de proxy.

Conceitos relacionados:

"Segurança ao nível do sistema" na página 375

Segurança ao nível da colecção

Ao criar uma colecção, pode escolher uma opção para permitir a segurança de colecção. Se escolher esta opção, pode, mais tarde, configurar controlos de segurança de nível de documento. Não é possível activar a segurança a nível da colecção depois de esta ter sido criada.

Conceitos relacionados:

"Componente de gestão de identidade" na página 388

"Análise de documentos duplicados e segurança de colecção" na página 382

"Identificadores de aplicações" na página 382

"Segurança ao nível do documento" na página 383

Tarefas relacionadas:

"Configurar a gestão de identidade" na página 394

"Associar aplicações a colecções"

Associar aplicações a colecções

Antes de poder utilizar uma aplicação personalizada para procura ou análise, deve associá-la às colecções às quais pode aceder.

Uma aplicação de procura empresarial pode pesquisar várias colecções. Pode configurar um prospecção de análise de conteúdo para explorar várias colecções, mas os utilizadores só podem explorar uma colecção de cada vez.

Para associar uma aplicação a uma ou mais colecções:

- 1. Faça clique em Segurança para abrir a vista Segurança.
- Na área de segurança ao nível da colecção, faça clique em Acções > Associar aplicações com colecções.
- 3. Na página Configurar IDs de Aplicações, clique em Adicionar IDs de Aplicação.
- 4. Introduza o ID da aplicação de procura empresarial ou prospecção de análise de conteúdo. O ID é especificado no parâmetro **applicationName** do ficheiro de configuração para a aplicação. O ID para a aplicação de procura empresarial facultada é Procura e o ID para o prospecção de análise de conteúdo facultado é Análise.
- 5. Seleccione as colecções às quais a aplicação pode aceder:
 - Faça clique em Todas as colecções para permitir que aplicação aceda a todas as colecções que adicionou ao sistema.
 - Faça clique em Todas as colecções de procura ou em Todas as colecções de análise de conteúdo se quiser limitar a aplicação a certos tipos de colecções.
 - Faça clique em **Colecções específicas** se pretender que a aplicação aceda somente às colecções que especificar.

Quando selecciona esta opção, é apresentada uma lista de nomes de colecções. Seleccione a caixa de verificação para cada colecção à qual a aplicação pode aceder.

Conceitos relacionados:

"Segurança ao nível da colecção" na página 381

Identificadores de aplicações

A capacidade para aceder a colecções diferentes é controlada através da correlação de aplicações de procura e análise com colecções a que as mesmas possam aceder.

Todas as aplicações têm de transmitir um nome de aplicação válido (APPID) à interface de programação de aplicações (API) do IBM Content Analytics with Enterprise Search. Apenas as colecções associadas a este APPID podem ser acedidas pela aplicação.

Cada aplicação tem um ficheiro de configuração que especifica o respectivo nome. Antes de uma aplicação poder aceder a uma colecção, um administrador tem de associar a aplicação às colecções específicas que esta pode aceder. Uma aplicação pode aceder a todas as colecções no sistema, ou aceder apenas às colecções que especificar.

Para controlar quais os utilizadores que podem consultar colecções, deverá associar utilizadores (ou grupos de utilizadores) à aplicação do cliente utilizando funcionalidades do controlo de acesso do servidor de aplicação da Web incorporado ou WebSphere Application Server, da mesma forma como poderá utilizar estas funcionalidades para restringir o acesso a um URL. Por exemplo, pode restringir o acesso ao URL que inicia a aplicação.

Conceitos relacionados:

"Segurança ao nível da colecção" na página 381

- APIs de procura e índice
- 🕞 Segurança da aplicação

Análise de documentos duplicados e segurança de colecção

Se activar a segurança de colecção, os processos de análise global não identificam documentos duplicados na colecção.

Durante a análise global, os processos de indexação identificam documentos que são duplicados (ou quase duplicados) uns dos outros. Em seguida, associam todos esses documentos a uma representação mais canónica do conteúdo. Ao permitir que os documentos duplicados sejam identificados, pode assegurar que os resultados da consulta não contêm vários documentos com o mesmo (ou quase o mesmo) conteúdo.

Se activar a segurança de colecção ao criar uma colecção, os documentos duplicados não são identificados e, dessa forma, não são associados a uma representação canónica comum. Em vez disso, cada documento é indexado independentemente. Isto assegura que os controlos de segurança para cada documento são avaliados de modo a que os utilizadores pesquisem apenas os documentos com categorias de segurança que correspondam às próprias credenciais. Dois documentos podem ser praticamente idênticos no conteúdo, mas utilizarem listas de controlo de acessos diferentes para fazerem cumprir a segurança.

Por exemplo, para dois documentos duplicados, documento_A e documento_B, assuma que um utilizador só possui direitos de acesso para o documento_B. Se o

documento_B for eliminado pela detecção duplicada, o utilizador não pode ver o documento nos resultados da consulta, por causa das restrições de acesso que estão em vigor para o documento_A.

Desactivar a análise de documentos duplicados pode melhorar a segurança de documentos numa colecção, mas a qualidade da procura pode degradar-se se os utilizadores receberem várias cópias do mesmo documento nos resultados da consulta.

Conceitos relacionados:

"Detecção de documentos duplicados" na página 228

"Segurança ao nível da colecção" na página 381

Segurança ao nível do documento

Se a segurança estiver activada para uma colecção quando esta for criada, pode configurar controlos de segurança de nível de documento. O controlo ao nível do documento assegura que os resultados da procura só contêm os documentos que o utilizador que submeteu o pedido de procura está autorizado a visualizar.

O IBM Content Analytics with Enterprise Search suporta muitas abordagens para configuração dos controlos de segurança ao nível do documento:

- Os documentos podem ser pré-filtrados e associados a categorias de segurança antes de serem adicionados ao índice. Todos os pesquisadores suportam a capacidade de associar documentos com categorias de segurança, para que os dados de controlo de acesso possam ser armazenados com os documentos no índice.
- Para alguns pesquisadores e tipos de origem de dados, os resultados da procura podem ser pós-filtrados para validar as credenciais de início de sessão do utilizador contra os dados de controlo de acesso actuais. A pós-filtragem também pode ser implementada por meio de um suplemento personalizado criado pelo utilizador.
- O componente de gestão de identidade pode codificar as várias credenciais de que os utilizadores necessitam para aceder a diferentes repositórios e armazenar as credenciais codificadas em perfis. Se as origens a pesquisar estiverem protegidas por um produto que forneça a segurança de início de sessão único (SSO), o componente de gestão de identidade pode controlar o acesso a documentos sem exigir que os utilizadores criem perfis.
- Para a maioria dos pesquisadores, pode ser utilizada uma classe Java (suplemento) para associar categorias de segurança a documentos no índice.
- Para os documentos pesquisados por um pesquisador Web, o texto âncora nos documentos que contêm ligações a documentos proibidos pode ser excluído do índice.

Conceitos relacionados:

"Segurança ao nível da colecção" na página 381

"Pré e pós-filtragem de resultados da procura"

"Facetas em colecções protegidas" na página 385

"Validação por tokens de segurança armazenados" na página 386

"Validação das credenciais actuais durante o processamento de consultas" na página 387

"Análise de texto âncora" na página 395

🕩 Segurança da aplicação

Referências relacionadas:

L Origens de dados suportadas

Pré e pós-filtragem de resultados da procura

Estas são duas abordagens distintas da filtragem de documentos para assegurar que os resultados da procura só contêm os documentos que o utilizador que submeteu o pedido de procura está autorizado a visualizar.

- A primeira abordagem é replicar as Listas de Controlo de Acesso (ACL) da origem do documento no momento da pesquisa para o índice e basear-se no motor de procura para comprar credenciais do utilizador para ACL de documentos replicados. A pré-filtragem dos documentos e o controlo de quais os documentos que serão adicionados ao índice, resultam num melhor desempenho. No entanto, é difícil modelar todas as políticas de segurança das várias origens back-end no índice e implementar a lógica de comparação de maneira uniforme. Esta abordagem também não fornece uma resposta tão rápida quanto a quaisquer alterações que possam ocorrer nas ACLs de origem.
- A segunda abordagem consiste e pós-filtrar documentos no conjunto de resultados procurando nas origens back-end para dados de segurança actuais. Esta abordagem permite que as origens back-end contribuintes sejam os árbitros finais dos documentos devolvidos ao utilizador e assegura que o conjunto de resultados reflecte os controlos de acesso actuais. No entanto, esta abordagem resulta numa degradação do desempenho da procura, uma vez que requer a existência de ligações a todas as origens back-end. Se uma origem não estiver acessível, as ligações aos documentos têm de ser filtradas fora do conjunto de resultados juntamente com os documentos que o utilizador não está autorizado a visualizar.

O suporte para fazer cumprir os controlos de acesso baseia-se numa combinação destas duas abordagens. A concepção permite um desempenho optimizado enquanto mantém as políticas de segurança exactas dos repositórios de documentos de origem. Ao armazenar dados de controlo de acesso de alto nível no índice, o sistema pode fornecer um conjunto de resultados provisório (potencialmente menor) que pode, em seguida, ser pós-filtrado para verificação dos controlos de acesso actuais. Assume-se que, se o utilizador tiver acesso ao repositório ao qual pertence o documento, é muito provável que também tenha acesso ao documento.

Os dados de controlo de acesso que estão armazenados no índice variam de acordo com o tipo de pesquisador. Por exemplo, o pesquisador do Notes pode armazenar controlos de acesso de nível de bases de dados e de servidor, e o pesquisador do Quickr for Domino pode armazenar controlos de acesso para servidores, locais e salas.
Todos os pesquisadores e tipos de origem de dados suportam a capacidade de indexar dados de controlo de acesso de origem durante o tempo de pesquisa. Alguns pesquisadores e tipos de origem de dados também suportam a capacidade de pós-filtrar o conjunto de resultados e verificar as credenciais actuais do utilizador. Este tipo de suporte é fornecido através dos mecanismos de segurança de origem dos dados ou o componente de gestão de identidade.

Esta concepção de segurança ambivalente engloba as seguintes tarefas:

- Extrair informação de ACL durante tempo da pesquisa.
- Armazenamento de informações de ACL do servidor e base de dados no índice.
- Criação do contexto de segurança do utilizador quando este inicia sessão ou quando a sessão é inicializada. Esta tarefa tem de ter em conta os diferentes identificadores que um único utilizador tem de utilizar para aceder às várias origens back-end.
- Processamento da procura com o contexto de segurança do utilizador e produção de um conjunto de resultados provisório que contém apenas os documentos aos quais o utilizador pode aceder ao nível do repositório.
- Pós-filtragem do conjunto de resultados provisório consultando as origens back-end que contribuíram com documentos para o conjunto de resultados para obter informação de ACL actual.

Exportar documentos

A pós-filtragem de resultados da pesquisa não é suportada pela função de exportação. Se um utilizador exportar documentos após a pesquisa de uma colecção protegida, os documentos que são excluídos dos resultados da pesquisa através de pós-filtragem serão incluídos no conjunto de documentos que são exportados. Os documentos que são excluídos dos resultados da pesquisa através de pré-filtragem não serão exportados.

Conceitos relacionados:

"Segurança ao nível do documento" na página 383

"Facetas em colecções protegidas"

Facetas em colecções protegidas

Se a segurança estiver activada para uma colecção e se a colecção estiver configurada para aplicar a pré-filtragem de resultados, os resultados podem incluir todas as facetas configuradas para a colecção. Se a colecção estiver configurada para aplicar a pós-filtragem de resultados, tem de especificar quais as facetas que pretende disponibilizar para procura.

Para calcular um número exacto de documentos quando executar a pós-filtragem de resultados, o sistema deve executar uma verificação em tempo real para cada documento que pertença a cada faceta, o que aumenta bastante o tempo de resposta da consulta. Ao limitar as facetas disponíveis para procura, é possível alcançar um equilíbrio entre procura facetada protegida e tempos de resposta de consulta aceitáveis.

Na vista Segurança da consola administrativa, é possível configurar as definições de segurança ao nível de documento para todos os documentos na colecção. Se a caixa de verificação **Não validar as credenciais actuais durante o processamento da consulta** não estiver seleccionada, a pós-filtragem de resultados de procura está activada. Na lista de facetas disponíveis para a colecção, seleccione as facetas a serem contadas quando os utilizadores consultam esta colecção.

Importante: Existe uma possível exposição de segurança quando a pós-filtragem de resultados e a contagem de facetas estão activadas para uma colecção. Devido aos documentos serem largados dos resultados durante a etapa de pós-filtragem, o número de documentos nos resultados pode diferir das contagens apresentadas para uma faceta. Por exemplo, a contagem de faceta pode ser superior ao número de documentos que o utilizador está autorizado a procurar ou a ver nos resultados de procura.

Conceitos relacionados:

"Segurança ao nível do documento" na página 383

"Pré e pós-filtragem de resultados da procura" na página 384

"Validação das credenciais actuais durante o processamento de consultas" na página 387

Validação por tokens de segurança armazenados

Se a segurança estiver activada para uma colecção quando esta é criada, pode configurar controlos de segurança de nível de documento armazenando dados de segurança no índice.

Por predefinição, cada documento é atribuído a um token público que faz com que o documento fique disponível para todos. Se a segurança estiver activada para a colecção, o token público pode ser substituído por um valor que seja fornecido pelo administrador ou que seja extraído de um campo no documento pesquisado. Ao configurar um pesquisador, irá especificar que pretende utilizar tokens de segurança para limitar quais os utilizadores que podem aceder aos documentos pesquisados por esse pesquisador.

Quando um administrador de colecção configura um pesquisador, o administrador pode especificar opções de segurança para tabelas individuais, sistemas de ficheiros, etc. (ou seja, podem ser configuradas regras de segurança diferentes para origens de dados diferentes no espaço da pesquisa). O administrador pode:

- Especificar que os documentos são públicos (todos os utilizadores podem procurar nos documentos)
- Atribuir tokens de segurança definidos pelo utilizador a cada documento
- Extrair tokens de segurança de um campo nos dados pesquisados e atribuir o token extraído a cada documento

Os tokens de segurança (à excepção do token público predefinido) são totalmente definidos pelo utilizador. Um token de segurança pode representar um ID de utilizador, um ID de grupo, uma função de utilizador ou qualquer outro valor determinado por si como sendo válido para a origem de dados.

Por exemplo, um administrador pode especificar que o campo hrDeptName deverá ser utilizado para controlar o acesso a documentos que são pesquisados por um pesquisador do Notes. O administrador também pode especificar que, se esse campo não existir num documento ou se não contiver dados de segurança, deverão ser utilizados dois tokens definidos pelo utilizador, hrgroup1 e hrgroup2, para controlar o acesso a documentos.

Os tokens de segurança são disponibilizados ao pesquisador através do respectivo ficheiro de configuração. Para cada documento, o pesquisador fornece o valor do token de segurança como metadados. O componente de indexação lê o token de segurança e aplica-o às informações de edição do documento no índice. Se o

administrador da origem de dados actualizar a lista de controlo de acesso, os controlos de segurança actualizados ficam disponíveis da próxima vez que ocorrer uma construção de índice.

Pode aplicar regras de negócio personalizadas para determinar o valor dos tokens de segurança, codificando as regras numa classe Java. Quando configurar as propriedades do pesquisador, especifique o nome do suplemento que pretende que o pesquisador utilize quando pesquisar documentos. Os tokens de segurança que o respectivo suplemento adiciona estão armazenados no índice e podem ser utilizados para controlar o acesso aos documentos.

Como as aplicações de procura empresarial utilizam tokens de segurança

É da responsabilidade da aplicação de procura empresarial do cliente fornecer os tokens de segurança no momento da procura para que os documentos possam ser adequadamente filtrados. Se não for fornecido nenhum token de segurança, então os tokens públicos predefinidos são automaticamente aplicados durante o processamento de procuras.

A aplicação de procura empresarial facultada demonstra como poderá implementar a segurança ao nível do documento. Para este exemplo, pressupõe-se que o administrador atribuiu um valor de token de segurança a um grupo de documentos (por oposição à extracção do token de segurança de um campo num documento pesquisado). A aplicação de procura utiliza o ID de início de sessão do utilizador para determinar quais são os documentos aos quais o utilizador pode aceder. Em vez de utilizar o ID de utilizador real, a aplicação de procura baseia-se no ID de grupo ao qual pertence o utilizador. Ao utilizar um ID de grupo como token de segurança, os utilizadores podem ser adicionados e removidos do grupo sem ser necessária a recriação do índice.

A token de segurança atribuído pelo administrador a um conjunto de documentos representa um ID de grupo de sistema operativo válido. São atribuídos IDs de grupo diferentes a documentos diferentes no espaço de pesquisa. Por exemplo:

Documento1-5:token de segurança = Grupo1 Documento6-10: token de segurança = Grupo2

Conceitos relacionados:

"Segurança ao nível do documento" na página 383

Validação das credenciais actuais durante o processamento de consultas

Se a segurança for activada para uma colecção quando esta é criada, determinados tipos de domínios permitem-lhe validar as credenciais actuais de utilizador quando um utilizador submete uma consulta.

Antes de responder a uma consulta, os servidores de procura estabelecem interface com os repositórios de destino para validar as permissões actuais do utilizador, em seguida, removem todos os documentos que o utilizador não tem permissão para visualizar nos resultados da procura.

Quando configurar os seguintes tipos de pesquisadores, é possível seleccionar uma opção para validar as credenciais de utilizador comparando as credenciais para controlos de acesso actuais que são geridas pelo repositório. Após os documentos serem pesquisados e indexados, o componente de gestão de identidade é utilizado para validar os utilizadores que tentarem procurar colecções seguras.

- Pesquisador Agente para sistemas de ficheiros Windows
- Pesquisador Content Integrator (apenas tipos de repositório Documentum e Hummingbird DM)
- Pesquisador do Content Manager
- Pesquisador do Domino Document Manager
- Pesquisador do FileNet P8
- Pesquisador do Notes
- Pesquisador do Quickr for Domino
- · Pesquisador do SharePoint
- Pesquisador do Sistema de ficheiros Windows

As credenciais de utilizador actuais também podem ser validadas quando os utilizadores utilizam a portlet Procura para procurar colecções em WebSphere Portal. A validação da credencial não é suportada quando os utilizadores utilizam a aplicação de procura empresarial autónoma para procurar origens num servidor do WebSphere Portal. Pode configurar os seguintes pesquisadores para validar credenciais actuais através da portlet Procura:

- Pesquisador da Lista de valores geradores
- Pesquisador do Web Content Management
- Pesquisador do WebSphere Portal

Conceitos relacionados:

"Segurança ao nível do documento" na página 383

"Componente de gestão de identidade"

"Validação de utilizadores com perfis de utilizador" na página 391

"Validação de utilizador através de autenticação de início de sessão único" na página 392

"Facetas em colecções protegidas" na página 385

"Cumprimento da segurança ao nível do documento para documentos do sistema de ficheiros do Windows" na página 411

"Cumprimento da segurança ao nível do documento para documentos do Lotus Domino" na página 405

Tarefas relacionadas:

"Configurar a gestão de identidade" na página 394

"Configurar Servidores Fidedignos Lotus Domino para validar credenciais de utilizador" na página 405

Componente de gestão de identidade

A gestão de múltiplas credenciais de utilizador é um problema comum numa empresa. O IBM Content Analytics with Enterprise Search resolve o problema fornecendo um componente opcional de gestão de identidade.

As informações inerentes a uma empresa podem existir em vários formatos. Podem ser distribuídas pela empresa e geridas pelo software mais apropriado para a tarefa a executar. Por exemplo, os utilizadores da empresa podem utilizar uma aplicação de SQL para aceder a bases de dados relacionais ou a um sistema de gestão de documentos para obterem documentos relevantes para o respectivo trabalho.

O controlo do acesso a informações sensíveis nestes repositórios é, normalmente, aplicado pelo software de gestão. Os utilizadores identificam-se ao sistema central através de uma combinação de ID de utilizador e palavra-passe. Após ser

autenticado pelo sistema, o software de gestão controla os documentos que o utilizador está autorizado a visualizar e actua com base nos direitos de acesso definidos para o utilizador.

É comum os utilizadores terem IDs de utilizador e palavras-passe diferentes associados a cada repositório. Da mesma forma que é pedido aos utilizadores que se identifiquem junto dos repositórios originais da empresa, os utilizadores têm de fornecer credenciais antes de visualizarem documentos numa colecção que requeiram a validação de credenciais actuais. Os utilizadores que tiverem múltiplas identidades têm de apresentar credenciais para cada identidade.

Se especificar que pretende utilizar o componente de gestão de identidade na consola de administração, os servidores de procura poderão utilizar as seguintes abordagens para validar as credenciais actuais de um utilizador durante o processamento de consultas:

- A aplicação de procura empresarial pode pedir ao utilizador que registe as credenciais de que necessita para aceder a vários domínios de um perfil de utilizador. O perfil, que é codificado e armazenado num arquivo de dados seguro, permite que o utilizador pesquise os domínios seguros. Se não forem especificadas credenciais para um domínio que requeira a validação das credenciais actuais, os documentos desse domínio serão excluídos dos resultados da procura.
- Se os documentos de uma colecção tiverem sido pesquisados por um pesquisador que forneça suporte para segurança de início de sessão único (SSO), e o utilizador especificar que pretende utilizar a segurança de SSO para controlar o acesso a documentos, o sistema utilizará métodos de segurança de SSO para autenticar utilizadores enquanto durar uma sessão de procura. O utilizador não necessita de criar um perfil na página Meu Perfil que especifique credenciais ou forneça um ID e palavra-passe de utilizador ao procurar domínios seguros.
- Se utilizar o portlet Procurar no WebSphere Portal, as procuras de fontes seguras através do portlet são suportadas apenas a partir do portlet Procurar. Com a procura protegida relacionada com as portlets, o utilizador não necessita de criar um perfil na página **Meu Perfil** que especifique credenciais.

Quando os utilizadores consultam colecções que requerem a validação das credenciais actuais ao submeter uma consulta, o sistema pode utilizar o perfil ou métodos de segurança de SSO para recusar ou permitir o acesso a documentos.

Obter as informações de grupo do utilizador

Para validar as credenciais de um utilizador, o componente de gestão de identidade tem de obter as informações de grupo do utilizador para cada uma das identidades do utilizador e adicionar estas informações a uma cadeia do contexto de segurança do utilizador (USC, user security context). Estas informações de grupo são utilizadas para filtrar resultados de acordo com os dados de controlo de acesso que estão armazenados no índice, ou de acordo com os dados de autenticação de SSO. O componente de gestão de identidade executa esta operação utilizando categorias de SSO ou as credenciais do utilizador para estabelecer ligação com o sistema back-end e para pedir os grupos dos quais o utilizador é membro.

Quando configurar opções de gestão de identidade na consola de administração, pode especificar a frequência com que estas informações de grupo deverão ser renovadas. Pode extrair novos dados do grupo sempre que o utilizador iniciar sessão na aplicação de procura empresarial, podendo também extrair os dados do grupo regularmente como, por exemplo, de três em três dias.

Segurança sem o componente de gestão de identidade

Nem todas as empresas pretendem gerir as múltiplas identidades das respectivas comunidades de utilizadores com o componente de gestão de identidade. Se desactivar o componente de gestão de identidade na consola de administração, é da responsabilidade da sua aplicação de procura gerar a cadeia de contexto de segurança do utilizador. Após gerada, a cadeia de USC sé utilizada para definir o valor de restrições de ACL em cada consulta. Por exemplo:

```
Query q = factory.createQuery("IBM");
q.setACLConstraints("Contexto de Segurança do Utilizador em XML");
```

Sugestão: Para ajudar a criar a sua própria funcionalidade de gestão de identidades, o IBM Content Analytics with Enterprise Search faculta uma API que fornece controlo programático na base de dados de gestão de identidade. Esta API permite-lhe gerar o USC com objectos de Java e a cadeia de XML é, em seguida, criada automaticamente.c

A cadeia da consulta de XML tem de estar no seguinte formato, em que ... contém a cadeia de XML formada integralmente:

```
@SecurityContext::'...'
```

O formato da cadeia de XML é o seguinte:

```
<identities id="login_UserName">
    <ssoToken>token_value</ssoToken>
    <identity id="security_domain">
    <type>Notes</type>
    <username>domain_userName</username>
    <password encrypt="no">domain_userPW</password>
    <groups>
        <group id="g1" />
        <group id="g2" />
        </groups>
        <properties>
            <property name="property_name">property_value</property>
            ...
        </properties>
    </identity>
```

</identities>

identities

O valor do atributo id é o ID de utilizador que o utilizador fornece quando inicia sessão no sistema.

ssoToken

Opcional: especifica a categoria de Lightweight Third-Party Authentication (LTPA) que é criada para o utilizador enquanto durar a sessão do navegador. Este parâmetro só é utilizado se o domínio de destino estiver activado para SSO e se o pesquisador estiver configurado para utilizar a segurança de SSO.

identity

Contém as credenciais de utilizador para uma origem de dados em particular. O valor do atributo id é o domínio que armazena as informações das credenciais do utilizador (no caso do Domino, este é o nome dom domínio do Domino).

type

Identifica o tipo de dados e corresponde ao tipo de pesquisador (Notes, DB2, Exchange Server, etc.).

username

Especifica o nome de utilizador que deverá ser utilizado para procurar no domínio.

password

Especifica a palavra-passe do nome do utilizador especificado. O atributo encrypt tem de ser definido como no (o sistema não fornece um método de codificação exterior ao componente de gestão de identidade).

groups

Especifica os nomes de grupos aos quais pertence o utilizador. É utilizado um elemento de grupo separado para cada nome de grupo.

properties

Especifica uma lista de propriedades específicas da ligação, tais como o ID de administrador e palavra-passe codificada que foram utilizados para criar o pesquisador, ou se o SSO está activado para a origem.

property_name

O nome da propriedade.

property_value

O valor da propriedade.

Conceitos relacionados:

"Segurança ao nível da colecção" na página 381

"Validação das credenciais actuais durante o processamento de consultas" na página 387

Validação de utilizadores com perfis de utilizador

As aplicações podem pedir aos utilizadores que registem as credenciais de que necessitam para aceder a vários domínios de um perfil de utilizador.

Para procurar um domínio que necessita que as credenciais do utilizador sejam validadas quando é submetida uma consulta, os utilizadores têm de fornecer à aplicação as credenciais que utilizam para iniciar sessão no domínio. Com a gestão de identidade, os utilizadores podem armazenar credenciais para qualquer quantidade de domínios num perfil de utilizador. As credenciais são codificadas e armazenadas de forma segura no sistema IBM Content Analytics with Enterprise Search.

Se não forem especificadas credenciais para um domínio que requeira a validação das credenciais actuais, os documentos desse domínio serão excluídos dos resultados da procura.

Os utilizadores podem criar um perfil e registar as credenciais enquanto utilizam uma aplicação. Nas aplicações facultadas, esta capacidade é fornecida pela opção **O Meu Perfil**. As aplicações personalizadas podem implementar esta capacidade de modo diferente.

As colecções podem conter documentos de vários tipos de origens diferentes. Por exemplo, uma colecção pode conter documentos que foram pesquisados a partir de um sistema de ficheiros do Windows e de várias bases de dados do Lotus Notes. O

componente de gestão de identidade distingue os diferentes tipos de origens e pede apenas as credenciais que são necessárias para aceder a domínios que requerem validação.

Por predefinição, cada credencial é activada para procura e, por isso, requer que o utilizador forneça o ID de utilizador e palavra-passe que correspondem aos domínios da procura. Se o utilizador se tiver esquecido do ID de utilizador ou da palavra-passe para um domínio em particular, o domínio pode ser desactivado para pesquisa se a caixa de verificação for limpa. A desactivação de um domínio impede que os documentos seguros existentes nesses domínios sejam devolvidos num conjunto de resultados.

Após criar um perfil, o utilizador pode submeter um pedido de procura. O componente de gestão de identidade tem as informações necessárias para criar a cadeia do contexto de segurança do utilizador (USC, User Security Context) em pedidos de procura subsequentes. Se não utilizar o componente de gestão de identidade, a aplicação tem de fornecer a cadeia de USC quando os utilizadores consultam domínios que requerem a validação das credenciais actuais.

Da próxima vez que o utilizador tentar procurar colecções, o componente de gestão de identidade repete o processo de verificação de credenciais mas, desta vez, consegue localizar o perfil do utilizador. Se nenhum valor tiver sido alterado, o utilizador é posicionado automaticamente onde os pedidos de procura podem ser submetidos e não lhe é pedido que crie um perfil.

Se o componente de gestão de identidade detectar uma alteração em qualquer uma das credenciais de utilizador, será automaticamente apresentada ao utilizador a página de perfis quando a aplicação for acedida. Esta situação ocorre, por exemplo, quando uma palavra-passe para qualquer um dos domínios que estão activados para procura for alterada ou quando um domínio que requer autenticação for adicionado a uma colecção.

Os utilizadores podem ignorar a recomendação para actualizar o perfil mas, ao fazê-lo, podem causar a exclusão desses documentos dos resultados da procura.

Conceitos relacionados:

"Validação das credenciais actuais durante o processamento de consultas" na página 387

Validação de utilizador através de autenticação de início de sessão único

Se os documentos de uma colecção tiverem sido pesquisados por um pesquisador que forneça suporte para segurança de início de sessão único (SSO), pode especificar que pretende utilizar a segurança de SSO para controlar o acesso a documentos quando configurar opções de gestão de identidade.

Activação de SSO

A autenticação de início de sessão único permite que um utilizador seja autenticado uma vez e obtenha a acesso a muitos recursos sem lhe ser pedido que apresente credenciais de novo. A autenticação de SSO simplifica o esforço inerente à gestão de muitos nomes de utilizador e palavras-passe que os utilizadores têm de especificar para aceder a documentos em colecções seguras.

O servidor de aplicação da Web incorporado, WebSphere Application Server, e o Lotus Domino suportam uma forma de início de sessão único conhecida como Lightweight Third-Party Authentication (LTPA). Quando um utilizador tenta aceder a uma aplicação, é-lhe pedido que se autentique com um nome de utilizador e palavra-passe. Este nome de utilizador e palavra-passe são comparados com um repositório de LDAP que é partilhado por ambos os produtos. Após o utilizador ser autenticado, é criado um cookie de sessão para conter o token de LTPA. Em seguida, o utilizador pode aceder a outros recursos em qualquer servidor que tenha a mesma configuração de autenticação sem que lhe seja pedido que especifique credenciais de novo. Este token persiste enquanto a sessão do navegador for válida.

Se utilizar o servidor de aplicação de Web incorporado, o ficheiro de chave LTPA tem de estar no servidor de procura. Quando configurar as definições de inicio de sessão da aplicação, pode gerar um ficheiro de chave de LTPA. Se criar o ficheiro de chave noutro sistema, tal como o servidor de origem de dados, pode importar o ficheiro de chave para o IBM Content Analytics with Enterprise Search. Também pode exportar o ficheiro de chave a partir de IBM Content Analytics with Enterprise Search e depois importar o ficheiro de chave para os outros sistemas.

Se utilizar o WebSphere Application Server, utilize as directrizes seguintes para activar o suporte de SSO para as colecções:

- Certifique-se de que a segurança global do WebSphere Application Server e um registo de LDAP válido estão activados nos servidores de procura. O registo de LDAP pode ser qualquer produto de LDAP válido suportado por um WebSphere Application Server.
- Certifique-se de que o mecanismo de autenticação do WebSphere está configurado para utilizar um mecanismo de autenticação de LDAP activo. Quando configura o LDAP, especifica um nome de domínio válido, mas flexível como, por exemplo, seu.servidor.com.
- Certifique-se de que a chave de LTPA foi exportada do WebSphere Application Server e importada para outros produtos no mesmo domínio em que pretende activar o suporte para LTPA.

Após utilizar um navegador para verificar se a configuração de segurança anterior está a funcionar correctamente, pode utilizar a consola de administração para configurar pesquisadores que suportem a autenticação de SSO.

SSO e gestão de identidade

Quando os utilizadores pesquisam colecções que requerem a validação das credenciais actuais, o sistema pode utilizar métodos de segurança SSO para recusar ou permitir o acesso a documentos. Não são pedidas credenciais aos utilizadores quando pesquisam origens que suportam autenticação SSO. O componente de gestão de identidade é utilizado se todas as seguintes condições forem verdadeiras:

- O SSO está devidamente activado no servidor de aplicações Web e nos domínios destino, de modo a que se possam partilhar tokens LTPA e nomes de domínio.
- A segurança está activada em, pelo menos, uma das colecções que a aplicação pode pesquisar.
- As opções para utilizar o componente de gestão de identidade e a segurança de SSO estão configuradas na consola de administração.
- A opção para utilizar a segurança de SSO e as opções para fazer cumprir a segurança ao nível do documento (tal como a indexação de controlos de acesso ou a validação de credenciais actuais durante o processamento de consultas) tiverem sido seleccionadas quando foram configurados os seguintes tipos de pesquisadores:

- Domino Document Manager (disponível apenas para os pesquisadores que utilizem o protocolo DIIOP)
- Notes (disponível apenas para os pesquisadores que utilizem o protocolo DIIOP)
- Quickr for Domino (disponível apenas para os pesquisadores que utilizem o protocolo DIIOP)
- Lista de valores geradores (disponível apenas para origens de IBM Connections)

Se utilizar o portlet Procurar no WebSphere Portal, as procuras de fontes seguras através do portlet são suportadas apenas a partir do portlet Procurar. Tal como a autenticação SSO que é activada através da aplicação de procura autónoma, os utilizadores do portlet não necessitam de criar um perfil na página de uma aplicaçãoO Meu Perfil para especificar credenciais.

Conceitos relacionados:

"Validação das credenciais actuais durante o processamento de consultas" na página 387

Tarefas relacionadas:

"Configurar suporte de início de sessão único no servidor de aplicação da Web incorporado" na página 367

"Configurar o suporte SSO no WebSphere Application Server" na página 370

Referências relacionadas:

Configurar suporte para autenticação de início de sessão único (SSO)

Configurar a gestão de identidade

Quando os utilizadores consultam colecções que requerem a validação das credenciais actuais durante o processamento de consultas, o componente de gestão de identidade pode usar um perfil de utilizador ou os métodos de segurança de início de sessão único (SSO) para recusar ou permitir o acesso a documentos.

Para configurar a gestão de identidade:

- 1. Faça clique em Segurança para abrir a vista Segurança.
- 2. Na área de segurança ao nível da colecção, clique em Acções > Configurar gestão de identidade.
- 3. Na página Configurar Gestão de Identidade, seleccione a caixa de verificação que permite utilizar o componente de gestão de identidade para controlar o modo como as credenciais de utilizador são validadas durante o processamento de consultas. Se esta caixa de verificação não estiver marcada, a aplicação tem de fornecer a cadeia do contexto de segurança do utilizador (USC, user security context) quando os utilizadores consultam domínios que requerem a validação de credenciais durante o processamento de consultas.
- 4. Especifique com que regularidade o componente de gestão da identidade irá extrair as credenciais de utilizador de registos de grupo no registo do utilizador. Pode renovar os dados de credenciais sempre que o utilizador acede à aplicação ou após decorrer um número de dias especificado.
- 5. Para os tipos de pesquisadores que suportam a autenticação de SSO, especifique se pretende que o componente de gestão de identidade utilize categorias de segurança de SSO em vez de perfis de utilizador para validar utilizadores. Pode seleccionar a caixa de verificação para utilizar a autenticação de SSO com todos os tipos de pesquisadores ou seleccionar as caixas de verificação para tipos de pesquisadores individuais.

Importante: O componente de gestão de identidade utiliza os mecanismos de segurança de SSO apenas se o ficheiro de chave LTPA para segurança de SSO estiver adequadamente configurado no servidor de aplicação da Web incorporado ou WebSphere Application Server e nos domínios de destino.

Conceitos relacionados:

"Segurança ao nível da colecção" na página 381

"Validação das credenciais actuais durante o processamento de consultas" na página 387

Tarefas relacionadas:

"Configurar suporte de início de sessão único no servidor de aplicação da Web incorporado" na página 367

"Configurar o suporte SSO no WebSphere Application Server" na página 370

Análise de texto âncora

Se activar a segurança de colecção, os processos de análise globais aplicam regras especiais para indexar o texto âncora em documentos pesquisados por pesquisadores da Web. Se não activar a segurança de colecção, pode especificar se pretende indexar o texto âncora em ligações para documentos proibidos ao configurar pesquisadores da Web individuais.

O texto âncora é as informações dentro de uma ligação de hipertexto que descreve a página à qual a ligação se refere. Por exemplo, na seguinte ligação, o texto Sintaxe da Consulta é o texto âncora numa ligação que liga à página syntax.htm:

Sintaxe da Consulta

Regra geral, o pesquisador Web segue ligações em documentos para pesquisar documentos adicionais e inclui essas páginas ligadas no índice. Durante a análise global, os processos de indexação associam o texto âncora não apenas ao documento no qual está incorporado (o documento origem), mas também ao documento de destino. No exemplo acima, o texto âncora Sintaxe da Consulta está associado à pagina de destino syntax.htm e à página de origem que contém a construção âncora. Esta associação permite que o documento de destino seja obtido por consultas que especificam texto que aparece no documento de origem. No entanto, a associação apresenta um risco de segurança, se os utilizadores forem autorizados a visualizar o documento de destino, mas não o documento de origem.

Se activar a segurança de colecção ao criar uma colecção, o processamento de texto âncora é desactivado. O texto âncora já não é indexado com um documento a não ser que realmente apareça no documento ou nos metadados do documento. Este controlo de segurança assegura que os utilizadores não são expostos a informações em documentos às quais não estão autorizados a aceder; só é devolvido um documento nos resultados da procura se o respectivo conteúdo ou metadados corresponderem à consulta.

Activar a segurança de colecção pode melhorar a segurança de documentos da Web, permitindo aos utilizadores procurarem apenas documentos com categorias de segurança a que correspondam às próprias credenciais. No entanto, ao não processar o texto âncora, os resultados da procura podem não incluir todos os documentos que são potencialmente relevantes para uma consulta.

Se não activar a segurança de colecção, pode especificar se pretende indexar o texto âncora em ligações a documentos proibidos ao configurar propriedades avançadas de pesquisadores da Web.

"Segurança ao nível do documento" na página 383

Tarefas relacionadas:

"Indexar o texto âncora em ligações para documentos proibidos"

Indexar o texto âncora em ligações para documentos proibidos

Se um documento incluir documentos que o pesquisador Web esteja proibido de pesquisar, pode especificar se pretende reter o texto âncora para essas ligações ao configurar um pesquisador Web.

As directivas num ficheiro robots.txt ou nos metadados de documentos da Web podem impedir o pesquisador de aceder a documentos num sítio da Web. Se um documento autorizado ao pesquisador Web incluir ligações para documentos proibidos, pode especificar a forma como pretende tratar o texto âncora para essas ligações.

Pode especificar se pretende indexar o texto âncora para documentos proibidos quando configura o pesquisador Web. Para segurança máxima, especifique que não pretende indexar o texto âncora em ligações para documentos proibidos. Porém, ao não indexar texto âncora, os resultados da procura podem não incluir documentos potencialmente relevantes para uma consulta.

Para activar ou desactivar a indexação de texto âncora em ligações para documentos proibidos:

- 1. No painel Pesquisar e Importar da vista Colecções, localize o pesquisador Web que pretende configura e faça clique em **Editar propriedades do pesquisador**.
- 2. Expanda as opções avançadas e faça clique em Editar propriedades avançadas do pesquisador Web.
- 3. Para indexar o texto âncora em todos os documentos que este pesquisador pesquisa, seleccione a caixa de verificação **Indexar o texto âncora em ligações para documentos proibidos**. Os utilizadores ficarão a saber que existem páginas em que o pesquisador Web não está autorizado a pesquisar no texto âncora de ligações que apontam para essas páginas.

Para excluir o texto âncora em ligações para documentos proibidos do índice, desmarque esta caixa de verificação. Os utilizadores não terão a capacidade de saber que existem páginas que o pesquisador não tem autorização para pesquisar. O texto âncora será excluído do índice adicionalmente aos documentos proibidos.

- 4. Faça clique em **OK** e, em seguida, na página Propriedades do pesquisador, faça clique em **OK** novamente.
- 5. Para que as alterações se tornem efectivas, pare e reinicie o pesquisador.

Para aplicar as alterações a documentos indexados anteriormente, tem de voltar a pesquisar os documentos de forma a poderem ser indexados novamente. Se uma pesquisa anterior tiver adicionado informações sobre documentos proibidos ao índice, essas informações serão removidas do índice.

"Análise de texto âncora" na página 395

Requisitos de configuração do pesquisador para suporte de segurança

Para recolher informações que permitam a aplicação da segurança ao nível do documento, os pesquisadores têm de ter permissão para ler os dados de controlo de acesso na origem de dados de destino para que seja possível indexar os dados de segurança com os documentos.

Todos os pesquisadores podem ser configurados para associar documento a categorias de segurança, para que os dados de controlo de acesso possam ser armazenados com os documentos no índice. Para alguns pesquisadores e tipos de origens de dados, são necessários passos adicionais para configurar um ambiente seguro, especialmente para origens que suportam a capacidade de validar as credenciais de um utilizador com dados de controlo de acesso que são mantidos pela origem de dados.

Tabela 21. Requisitos de configuração do pesquisador para suporte de segurança

Pesquisadores do Case Manager Pesquisadores do Agente para sistemas de ficheiros Windows		
Antes de configurar o pesquisador, tem de instalar o IBM FileNet Content Engine Client no servidor do pesquisador. Também terá de executar um script (escrfilenet.sh em AIX ou Linux, ou escrfilenet.vbs em Windows) para configurar o servidor do pesquisador.	 Tópicos relacionados: "Pesquisadores do Case Manager" na página 63 "Configurar o servidor de pesquisadores para suportar pesquisadores de Case Manager e FileNet P8" na página 65 	
Pesquisadores do Content Integrator		
Antes de criar o pesquisador, execute um script (escrvbr.sh em AIX, Linux ou escrvbr.vbs em Windows) para configurar o servidor do pesquisador. Quando configurar o pesquisador, especifique um ID de utilizador e palavra-passe que permita ao pesquisador aceder a cada repositório a ser pesquisado. Pode especificar um ID de utilizador e palavra-passe diferente, conforme pecessário, para	 Tópicos relacionados: "Configurar um conector para aceder a repositórios do Content Integrator" na página 67 "Configurar o servidor do pesquisador para pesquisadores do Content Integrator" na página 67 	
cada repositório no espaço da pesquisa.		

Antes de criar o pesquisador, execute um script (escrdb2.sh em AIX, Linux ou escrdb2.vbs em Windows) para configurar o servidor do pesquisador.	 Tópicos relacionados: "Configurar o servidor do pesquisador para DB2 pesquisadores" na página 76 	
Quando configurar o pesquisador para pesquisar bases de dados remotas e não catalogadas, especifique um ID de utilizador e palavra-passe que permita a pesquisa de cada base de dados do servidor de bases de dados destino. Pode especificar um ID de utilizador e palavra-passe diferente, conforme necessário, para cada base de dados no espaço da pesquisa.		
Pesquisadores Content Manager		
Antes de criar o pesquisador, execute um script (escrcm.sh em AIX, Linux ou escrcm.vbs em Windows) para configurar o servidor do pesquisador.	 Tópicos relacionados: "Configurar o servidor do pesquisador no UNIX para pesquisadores Content Manager" na página 70 "Configurar o servidor do pesquisador no Windows para pesquisadores Content Manager" na página 71 	
Quando configurar o pesquisador, especifique um ID de utilizador e palavra-passe que permita ao pesquisador aceder a cada servidor a ser pesquisado. Pode especificar um ID de utilizador e palavra-passe diferente, conforme necessário, para cada servidor no espaço da pesquisa.		
Pesquisadores do Domino Document Manager, Notes e Quickr for Domino		

Tabela 21. Requisitos de configuração do pesquisador para suporte de segurança (continuação)

Tabela 21. Requisitos de configuração do pesquisador para suporte de segurança (continuação)

Para pesquisar servidores Lotus Domino que utilizem o protocolo de chamada de procedimento remoto (NRPC) de Notes:

- Num sistema AIX, certifique-se de que o módulo Porta de Conclusão de E/S está instalado e disponível no servidor do pesquisador.
- Antes de criar o pesquisador, execute um script (escrnote.sh em AIX, Linux ou escrnote.vbs em Windows) para configurar o servidor do pesquisador.
- Tem de ser instalado um servidor Domino no servidor do pesquisador e este servidor Domino deve ser membro do domínio de Domino a ser pesquisado.
- Para validar credenciais de utilizador actuais quando um utilizador submete um pedido de procura, o servidor Domino a ser pesquisado tem de ser configurado como um Servidor Fidedigno Lotus Domino.
- Quando configurar o pesquisador, especifique o caminho para um ficheiro de ID de utilizador do Lotus Notes que esteja autorizado a aceder ao servidor, como o c:\Program Files\lotus\notes\ data\name.id ou /local/notesdata/name.id e a palavra-passe para este ficheiro de ID.

Para pesquisar servidores Lotus Domino que utilizem o Domino Internet Inter-ORB Protocol (DIIOP):

- Num sistema AIX, certifique-se de que o módulo Porta de Conclusão de E/S está instalado e disponível no servidor do pesquisador.
- Configure o servidor do pesquisador para que possa utilizar o protocolo.
- Quando configurar o pesquisador, especifique um ID de utilizador completo do Lotus Notes que esteja autorizado a aceder ao servidor, como Nome de Utilizador/Qualquer Cidade/Minha Empresa, e a palavra-passe para este ID de utilizador.

Para pesquisar o Quickr for Domino, deve configurar o servidor para suportar a segurança do Utilizador Local ou Assistência de Directório, dependendo do tipo de segurança que pretende utilizar.

Pesquisadores Exchange Server

Tópicos relacionados:

- "Configurar a porta de conclusão E/S no AIX para pesquisar origens no Lotus Domino" na página 118
- "Configurar o servidor do pesquisador no UNIX para pesquisar origens do Lotus Domino" na página 114
- "Configurar o servidor do pesquisador no Windows para pesquisar origens do Lotus Domino" na página 116
- "Configurar Servidores Fidedignos Lotus Domino para validar credenciais de utilizador" na página 405
- "Configurar servidores que utilizam o protocolo DIIOP" na página 117
- "Configurar o Quickr para o servidor do Domino para utilizar a segurança Utilizador Local" na página 407
- "Configurar Assistência de Directório no Quickr para servidor do Domino" na página 408

Tabela 21. Requisitos de configuração do pesquisador para suporte de segurança (continuação)

5 3 1	
Quando configurar o pesquisador, especifique um ID de utilizador que esteja autorizado a aceder a pastas públicas no Exchange Server a ser pesquisado, e a palavra-passe para este ID de utilizador.	 Tópico relacionado: "Verificar o acesso aos documentos Exchange Server protegidos" na página 96
Para que o pesquisador utilize gestão de chaves do Exchange Server e o protocolo Secure Sockets Layer (SSL) ao pesquisar dados, especifique também o caminho totalmente qualificado para o ficheiro de arquivo de chaves e uma palavra-passe que permita ao pesquisador aceder a este ficheiro. O ficheiro de arquivo de chaves tem de existir no servidor do pesquisador.	
Pesquisadores FileNet P8	
Antes de configurar o pesquisador, tem de instalar o IBM FileNet Content Engine Client no servidor do pesquisador.	 Tópicos relacionados: "Pesquisadores do FileNet P8" na página 96 "Configurar o servidor de pesquisadores para suportar pesquisadores de Case Manager e FileNet P8" na página 65
em AIX ou Linux, ou escrfilenet.vbs em Windows) para configurar o servidor do pesquisador.	Leedwarene en eveneren 200 e meret en Leidew en
Pesquisadores Base de dados JDBC	
Quando configurar o pesquisador, pode especificar um ID de utilizador e palavra-passe que permita a pesquisa das tabelas existentes na base de dados destino. Pode especificar um ID de utilizador e palavra-passe diferente, conforme necessário, para cada base de dados no espaço da pesquisa.	
Pesquisadores NNTP	
Os servidores de NNTP a pesquisar têm de permitir que o servidor do pesquisador leia dados.	
Pesquisadores SharePoint	
Antes de instalar o pesquisador, deverá implementar serviços web fornecidos no servidor do SharePoint.	Tópico relacionado:"Implementar serviços da Web para pesquisadores SharePoint" na página 131
Pesquisadores Sistema de ficheiros UNIX	
Os subdirectórios AIX e Linux a pesquisar têm de permitir que o servidor do pesquisador leia os dados.	
Pesquisadores da Web	

Tabela 21. Requisitos de configuração do pesquisador	para suporte de segurança (continuação)
O pesquisador orienta-se pelo protocolo Robots Exclusion. Se um servidor da Web incluir um ficheiro robots.txt no nível superior do directório do servidor, o pesquisador analisa o ficheiro e só pesquisa sítios da Web nesse servidor se tiver autorização para o fazer. Para obter informações sobre este protocolo, consulte http:// www.robotstxt.org/wc/exclusion.html.	 Tópicos relacionados: "Sítios da Web protegidos por autenticação base HTTP" na página 147 "Sítios da Web protegidos por autenticação baseada em formulário" na página 148
 Quando configurar o pesquisador da Web: Tem de especificar um nome de agente de utilizador para o pesquisador. As regras existentes nos ficheiros robots.txt dos servidores a serem pesquisados podem especificar este nome, de modo a permitir ou recusar acesso. Opcional: se um servidor Web utilizar a autenticação base de HTTP para restringir o acesso a sítios da Web, pode especificar credenciais de autenticação que permitam que o pesquisador aceda a páginas protegidas por palavra-passe. Opcional: se um servidor Web utilizar formulários de HTML para restringir o acesso a sítios da Web, pode especificar credenciais de autenticação que permitam que o pesquisador Web aceda a páginas protegidas por palavra-passe. 	
Pesquisadores do Lista de valores geradores, Web Co	ontent Management e WebSphere Portal
Quando configurar o pesquisador para pesquisar origens num servidor WebSphere Portal, especifique um nome distinto completo (ND) que permita que o pesquisador obtenha páginas do servidor a ser pesquisado, tais como uid=admin,cn=RegularEmployees,ou=Software Group,o=IBM,c=US, e especifique a palavra-passe para este ND. O ND tem de corresponder a um ND que esteja configurado no WebSphere Portal.	
Pesquisadores de Sistema de ficheiros Windows	
Os subdirectórios a pesquisar têm de permitir que o servidor do pesquisador leia os dados. Quando configurar o pesquisador para pesquisar sistemas de ficheiros remotos, especifique um ID de utilizador que permita ao pesquisador aceder aos dados remotos e especifique uma palavra-passe para este ID de utilizador.	 Tópicos relacionados: "Cumprimento da segurança ao nível do documento para documentos do sistema de ficheiros do Windows" na página 411 "Procura protegida de domínios do Windows fidedignos" na página 413
Para validar as credenciais de utilizador actuais quando um utilizador submete um pedido de procura, certifique-se de que as contas de domínio estão correctamente configuradas. Os requisitos para definir contas de domínio para ficheiros que tenham sido pesquisados no computador local são diferentes dos requisitos para ficheiros pesquisados num servidor do Windows remoto.	

"Cumprimento da segurança ao nível do documento para documentos do Lotus Domino" na página 405

"Cumprimento da segurança ao nível do documento para documentos do sistema de ficheiros do Windows" na página 411

Tarefas relacionadas:

"Suportar procura protegida quando o servidor do pesquisador não está disponível"

"Verificar o acesso aos documentos Exchange Server protegidos" na página 96

Referências relacionadas:

🕒 Origens de dados suportadas

Suportar procura protegida quando o servidor do pesquisador não está disponível

É possível configurar o sistema para continuar a processar consultas seguras se o servidor do pesquisador ficar indisponível.

Para suportar a pesquisa segura sem o servidor do pesquisador, tem de configurar as bibliotecas do cliente para as origens de dados que pretende procurar no servidor de procura. Também tem de executar os scripts de configuração do pesquisador. Numa configuração de servidores múltiplos, tem de efectuar estes passos de configuração em todos os servidores de procura.

Restrição: Para activar o suporte para o pesquisador do Sistema de ficheiros Windows, o servidor do pesquisador e o servidor de procura têm de utilizar uma conta de domínio do Windows.

Esta configuração é útil para permitir procuras protegidas para continuar caso o servidor do pesquisador pare a execução. Quando o servidor do pesquisador fica disponível, os servidores de procura regressam à operação normal e começar a comunicar novamente com o servidor do pesquisador e o servidor do controlador.

Para verificar se as origens dos dados forem configurados correctamente, poderá criar um ficheiro de propriedades que especifica os parâmetros de ligação para as origens de dados que pretende para procurar e executar o comando **esadmin verifySecureCrawler** para testar as ligações. Um ficheiro de configuração de exemplo é como se segue:

```
# Content Manager EE
cm.server=icmnlsdb
cm.user=icmadmin
cm.password=password
```

```
# DB2
db2.url=jdbc:db2//localhost:50000/dbname
db2.username=user
db2.password=password
db2.driver=com.ibm.db2.jcc.DB2Driver
```

```
# Lotus Notes
#notes.protocol=BOTH/NRPC/DIIOP
notes.protocol=DIIOP
notes.password=password
notes.server=localhost
notes.username=username
# method to get IOR: this should be one of "HTTP", "HTTPS", "IOR" or "None"
notes.iorMethod=None
```

```
notes.iorParams=user:password
notes.iorParams=ior text
# TrustedCerts class path. If you don't use SSL, you don't need to specify tcPath.
notes.tcPath=
# If domino user ID is differenent with username and password
#notes.dominoUser=
#notes.dominoUserPassword=
# Content Integrator
iice.repository=rep
iice.user=user ID
iice.password=password
iice.jndiFactory=
```

iice.jndiProvider=

iice.optional=

Para activar e desactivar o suporte para processar consultas seguras quando o pesquisador não está disponível, e para verificar se as origens de dados que pretende suportar estão configuradas correctamente:

- 1. Activar o suporte para processar consultas protegidas quando o servidor do pesquisador não está disponível:
 - a. Inicie sessão como o administrador predefinido do IBM Content Analytics with Enterprise Search no servidor de controlador.
 - b. Introduza este comando:

esadmin configmanager separateSearchNode

Sugestão: Também é possível executar este comando iniciando sessão no servidor do controlador e especificando o ID do nó do servidor de procura que pretende activar. Por exemplo:

esadmin configmanager separateSearchNode -nid ID_nó

c. Reinicie o sistema IBM Content Analytics with Enterprise Search:

```
esadmin system
stopall
esadmin system startall
```

- 2. Desactive o suporte para processar consultas protegidas quando o servidor do pesquisador não estiver disponível:
 - a. Inicie sessão como o administrador predefinido do IBM Content Analytics with Enterprise Search no servidor de controlador.
 - b. Introduza este comando:

esadmin configmanager separateSearchNode -flags false

Sugestão: Também é possível executar este comando no servidor do controlador e especificar o ID de nó do servidor de procura que pretende para desactivar. Por exemplo:

esadmin configmanager separateSearchNode -nid ID_nó -flags false

Para determinar o ID do nó para o servidor de procura, execute o comando esadmin check para ver a lista de IDs de sessão. Procure o ID do nó que está listado para a sessão searchapp.

c. Reinicie o sistema IBM Content Analytics with Enterprise Search:

```
esadmin system
stopall
esadmin system startall
```

- **3**. Verifique se as origens de dados que pretende procurar estão configuradas correctamente no servidor local:
 - a. Inicie sessão como administrador predefinido do IBM Content Analytics with Enterprise Search no servidor de procura.
 - b. Crie um ficheiro de propriedades que especifica as propriedades de ligação do origens de dados como um ficheiro denominado securecrawler.properties.
 - **c**. Insira este comando e certifique-se de que especifica o caminho absoluto para o ficheiro de propriedades que criou:

esadmin verifySecureCrawler -fname absolute_path_to_file.properties

Sugestão: Também é possível executar este comando iniciando sessão no servidor do controlador e especificando o ID do nó do servidor de procura que pretende verificar. Por exemplo:

```
esadmin verifySecureCrawler -fname absolute_path_to_file.properties
-nid server ID
```

Referências relacionadas:

"Requisitos de configuração do pesquisador para suporte de segurança" na página 397

Verificar o acesso aos documentos Exchange Server protegidos

Para usar um pesquisador Exchange Server 2000 e 2003 para pesquisar documentos protegidos por uma firewall, terá de verificar se o pesquisados é capaz de aceder ao servidor de pastas públicas Microsoft Exchange Server.

Se o servidor do pesquisador não for capaz de aceder a um servidor Exchange Server 2000 ou 2003 seguro, receberá uma mensagem Código 501 de HTTP (Não implementado) do servidor. Também pode ver as mensagens que indicam que foi recebida uma resposta HTTP inesperada.

Para assegurar que o servidor do pesquisador possa aceder aos documentos atrás da firewall:

- 1. Inicie um navegador da Web no servidor do pesquisador.
- 2. Avance para o URL no servidor de pastas públicas Exchange Server que pretende pesquisar. Por exemplo: http://exchange.Empresa.com/public/
- 3. Verifique se pode abrir a página Exchange Server.

Se não conseguir aceder ao servidor do Exchange Server, contacte o administrador do Exchange Server da sua empresa.

"Pesquisadores Exchange Server 2000 e 2003" na página 95

Referências relacionadas:

"Requisitos de configuração do pesquisador para suporte de segurança" na página 397

Cumprimento da segurança ao nível do documento para documentos do Lotus Domino

Se o servidor Domino a ser pesquisado utilizar o protocolo de chamada de procedimento remoto do Notes (Notes remote procedure call), tem de configurar o servidor do pesquisador para que os controlos de acesso ao nível do documento possam ser cumpridos.

Para fazer cumprir a segurança ao nível de documento para documentos num servidor Domino que utilize o protocolo NRPC, tem de instalar um servidor Domino no servidor do pesquisador. Este servidor Domino tem de ser membro do domínio Domino. Siga as instruções na documentação do Lotus Domino para instalar e configurar o servidor Domino.

Tem, também, de configurar o Servidor Fidedigno do Lotus Domino para validar as credenciais do utilizador. Através da pós-filtragem, os servidores de procura podem verificar se um utilizador que procure uma colecção protegida está autorizado a ver documentos que correspondam aos critérios de procura.

Conceitos relacionados:

"Validação das credenciais actuais durante o processamento de consultas" na página 387

Tarefas relacionadas:

"Configurar Servidores Fidedignos Lotus Domino para validar credenciais de utilizador"

"Configurar o Quickr para o servidor do Domino para utilizar a segurança Utilizador Local" na página 407

"Configurar Assistência de Directório no Quickr para servidor do Domino" na página 408

"Indexar segurança de nível de documento do Lotus Notes para melhorar o desempenho da procura" na página 408

Referências relacionadas:

"Requisitos de configuração do pesquisador para suporte de segurança" na página 397

Configurar Servidores Fidedignos Lotus Domino para validar credenciais de utilizador

Para impor segurança para documentos que foram pesquisados por um pesquisador Notes que utiliza o protocolo Notes remote procedure call (NRPC), os servidores do Domino que serão pesquisados devem ser configurados para serem Servidores FidedignosLotus Domino.

Este procedimento é necessário se pretender fazer cumprir a segurança de nível de documento ao procurar bases de dados remotas. Para procurar bases de dados locais para o servidor do pesquisador, este procedimento não é necessário.

Para configurar Servidores Fidedignos, um servidor do Domino deve ser instalado no pesquisador. Este servidor do Domino deve ser membro do seu domínio Domino.

Ao configurar opções de segurança ao nível do documento para um pesquisador do Notes, especifique se pretende fazer cumprir controlos de acesso validando as credenciais actuais do utilizador quando o utilizador submeter uma consulta. Para impor este tipo de segurança, os servidores do Domino a serem pesquisados devem ser Servidores FidedignosLotus Domino.

Quando os utilizadores procuram um domínio que necessita que as credenciais actuais sejam validadas, o Servidor Fidedigno permite que o ID do servidorDomino comute o contexto para o ID do utilizador actual. A base de dados Domino é aberta como se o utilizador actual o tivesse aberto e todas as informações da lista de controlo de acesso à base de dados para esse utilizador são aplicadas.

A capacidade de comutar contextos desta maneira está tipicamente disponível apenas para bases de dados que estejam armazenadas no directório dados do servidor do Domino local. Tendo começado com a versão 6.5.1 do Lotus Domino, esta capacidade é fornecida através do Servidor Fidedigno. Para configurar o Servidor Fidedigno, um administrador do Domino especifica quais os servidores do Domino que são fidedignos para executarem operações delicadas, tal como fazer-se passar por outro utilizador quando a base de dados é acedida a partir de um computar remoto.

Para configurar um Servidor Fidedigno, conclua os seguintes passos em todos os servidores do Domino que são pesquisados pelo pesquisador Notes:

- 1. Num servidor do Domino, utilize o ficheiro do ID de administrador de domínios Domino para abrir o cliente Lotus Domino Administrator.
- 2. Faça clique em Ficheiro e, em seguida, seleccione Abrir servidor.
- Escreva o nome do servidor Domino para o qual pretende activas as funções do Servidor Fidedigno.
- 4. Seleccione o separador Configuração.
- 5. Expanda o objecto **Servidor**, seleccione o documento **Servidor Actual** e faça clique em **Editar Servidor**.
- 6. Seleccione o separador **Segurança**, desloque-se para o fim do documento, localize a entrada **Servidores Fidedignos** e faça clique na seta para baixo.
- 7. Especifique uma das seguintes opções:

ServidoresDomínioLocal

Seleccione esta opção se todos os servidores no domínio Domino devem ser considerados Servidores Fidedignos.

nome_servidor

Especifique o nome de um servidor do Domino que pretenda pesquisar e procurar como um Servidor Fidedigno.

Se o servidor do Domino a ser pesquisado estiver num domínio Domino diferente, nesse caso, deve especificar o nome do servidor ou seleccionar o grupo **OtherDomainServers**. Também deve seguir os procedimentos do Domino para efectuar a certificação cruzada do ficheiro de ID do servidor do Domino com o outro domínio do Domino. Consulte a documentação de servidor do Domino para obter informações sobre estes procedimentos.

- 8. Faça clique em Guardar e Fechar para guardar as alterações.
- **9**. Pare e reinicie os servidores remotos do Domino que activou para actuarem como Servidores Remotos.

"Cumprimento da segurança ao nível do documento para documentos do Lotus Domino" na página 405

"Validação das credenciais actuais durante o processamento de consultas" na página 387

Configurar o Quickr para o servidor do Domino para utilizar a segurança Utilizador Local

Se pretender configurar um pesquisador do Quickr for Domino para utilizar a opção Utilizador Local para implementar segurança, tem de configurar o Domino Directory no servidor Lotus Quickr for Domino antes de criar o pesquisador.

Quando configurar um pesquisador de Quickr for Domino, seleccione um modo de segurança para o pesquisador utilizar com vista a aplicar segurança ao nível dos documentos. Se seleccionar o modo Utilizador Local, tem de se certificar de que todos os IDs do utilizador local e todos os grupos locais estão registados no Domino Directory (a hierarquia do Domino Directory tem de corresponder ao Quickr para a hierarquia Domino).

Também tem de se certificar de que o ID do utilizador e a palavra-passe que especificar para o pesquisador a utilizar estão registados no Domino Directory e que tem permissão para ler a base de dados a ser pesquisada.

Para utilizar o Quickr para Domino, apenas é requerido o nome do utilizador. Para pesquisar Quickr para fontes Domino, todavia, é requerido o ID do utilizador totalmente expandido. O ID de utilizador expandido está no seguinte formato: nome utilizador/nome local/QP/nomedomínio

Utilize este procedimento para determinar a versão totalmente expandida do ID do utilizador, certifique-se de que este ID está autorizado a ler Quickr para as bases de dados Domino e adicione o ID do utilizador ao Domino Directory. O Domino Directory tem de conter o ID do utilizador que será utilizado para pesquisar Quickr para as bases de dados Domino e todos os utilizadores e grupos locais (a hierarquia do Domino Directory tem de corresponder à hierarquia do Quickr para Domino).

Para configurar o Quickr para o servidor Domino para utilizar a segurança Utilizador Local:

- 1. Confirme as permissões do ID de utilizador:
 - a. Abra o documento do servidor no Quickr para o servidor do Domino.
 - b. Abra a página Ficheiros e depois abra a lista de controlo de acessos (ALC) na base de dados que pretende pesquisar.
 - **c.** Confirme que o ID de Utilizador Local que o pesquisador configurará para utilização existe na ACL e que este ID de utilizador tem permissão para ler a base de dados.

Tem de especificar o formato totalmente expandido deste ID de utilizador no passo 2.

- 2. Adicione o utilizador ao Domino Directory:
 - a. Abra o documento do servidor no Quickr para o servidor do Domino.

- b. Na página Pessoas e Grupos, no item em árvore de pessoas, adicione o ID de utilizador totalmente expandido que confirmou no passo 1.
- **c.** No campo **Palavra-passe para Internet**, especifique a palavra-passe para ID de utilizador.

"Cumprimento da segurança ao nível do documento para documentos do Lotus Domino" na página 405

Configurar Assistência de Directório no Quickr para servidor do Domino

Se pretender configurar um pesquisador de Quickr for Domino para utilizar um directório LDAP com vista a implementar segurança, terá de criar uma base de dados da Assistência de Directório no servidor do Lotus Quickr for Domino antes de configurar o pesquisador.

O servidor que pretende para efectuar a pesquisa tem de estar a executar as tarefas DIIOP e HTTP.

Para configurar Assistência de Directório LDAP Directory Assistance no Quickr para o servidor do Domino:

- 1. Crie uma base de dados da Assistência de Directório:
 - a. Abra o documento do servidor no Quickr para o servidor do Domino.
 - b. Crie uma base de dados, utilizando o modelo **Assistência de Directório(6)**. Este modelo está no servidor.
 - c. Faça clique em Adicionar Assistência de Directório para criar um documento na base de dados.
 - d. Abra a tabulação Básica e no campo **TipoDomínio**, seleccione LDAP.
 - e. Abra a tabulação Contextos Nomeados e certifique-se de que a caixa de verificação Fidedigno para credenciais é seleccionada.
 - f. Abra a tabulação LDAP e especifique a informação acerca do servidor LDAP.
 - g. Guarde e feche o documento Servidor.
- 2. Configure o Quickr para o servidor do Domino para utilizar a base de dados de Assistência de Directório:
 - a. Abra o documento do servidor no Quickr para o servidor do Domino.
 - b. Abra a tabulação Básica e no campo Nome da base de dados Assistência de Directório, especifique o nome da base de dados que criou no passo 1.
 - c. Guarde e feche o documento Servidor.

O Quickr para o servidor do Domino pode agora utilizar o servidor LDAP como um directório Domino secundário.

Conceitos relacionados:

"Cumprimento da segurança ao nível do documento para documentos do Lotus Domino" na página 405

Indexar segurança de nível de documento do Lotus Notes para melhorar o desempenho da procura

Quando configura o pesquisador Notes, pode escolher armazenar o documento, base de dados e listas de controlo de acesso ao servidor (ACLs) no índice. Se escolher apenas armazenar base de dados e ACLs de servidor, pode melhorar o desempenho de procura protegida ao criar um ficheiro de configuração para também armazenar ACLs de nível de documento. Quando as ACLs de nível de documento são indexadas, a filtragem posterior dos resultados da procura (o tempo necessário para validar as credenciais de utilizador em tempo real) é amplamente reduzida e o desempenho da procura é altamente melhorado para casos onde existem muitos utilizadores com privilégios de leitor ou autor na base de dados de destino. Se não existirem valores nos campos de leitor ou autor, ou se esses campos não estiverem definidos na base de dados, o pesquisador armazena apenas os dados de ACL de nível de base de dados.

Importante: Nos casos em que um utilizador é impedido de aceder à base de dados e as entradas de ACL de nível de documento indicam que o utilizador pode ler o documento, o documento será eliminado dos resultados da procura através da filtragem posterior. Esta filtragem posterior poderá reduzir o impacto da melhoria do desempenho da procura.

Para activar o suporte para indexar ACLs de nível de documento:

- 1. Inicie sessão como administrador predefinido do IBM Content Analytics with Enterprise Search no servidor do pesquisador.
- Com um editor de texto capaz de ler UTF-8, crie um ficheiro de configuração XML denominado notescrawler_ext.xml e adicione o seguinte conteúdo:

3. Abra a consola administrativa e crie um pesquisador Notes ou edite o espaço de pesquisa para um pesquisador existente Notes. Na página Segurança de nível de documento para uma origem de dados do Notes, seleccione **Base de dados de índice e listas de controlo de acesso**.

Restrição: Não é possível seleccionar **Documento do índice, base de dados e listas de controlo de acesso ao servidor** e configurar o pesquisador para utilizar um ficheiro notescrawler_ext.xml personalizado.

- 4. Pare o pesquisador Notes.
- Copie o ficheiro notescrawler_ext.xml que criou no directório ES_NODE_ROOT/master_config/crawler_session/ para o pesquisador do Notes. O pesquisador utiliza este ficheiro de extensão juntamente com o ficheiro notescrawler.xml.
- 6. Reinicie o pesquisador do Notes.

Conceitos relacionados:

"Cumprimento da segurança ao nível do documento para documentos do Lotus Domino" na página 405

Configurar o pesquisador de Notes para utilizar um servidor LDAP como directório do utilizador

Para melhorar o desempenho da expansão de grupo quando utilizadores procuram em origens Lotus Notes protegidas, pode configurar pesquisadores de Notes que utilizem o protocolo de chamada de procedimento remoto (NRPC) de Notes para utilizar um servidor LDAP como Directório do Utilizador. Utilize as seguintes informações ao configurar o pesquisador para utilizar um servidor LDAP:

ServerName

O nome do sistema central do servidor LDAP.

Port A porta do servidor LDAP.

BaseDN

Opcional. O nome exclusivo base (DN) para o directório LDAP.

SSL Opcional. Caso utilize SSL para ligações LDAP (LDAPS), especifique True para activar este atributo.

TrustPath

Opcional. Caso o servidor LDAP necessite do armazenamento de certificados fidedignos para ligações SSL; especifique o caminho do ficheiro de armazenamento de certificados fidedignos.

TrustPassword

Opcional. Caso seja necessário, a palavra-passe codificada para o armazenamento de certificados fidedignos.

Credential

Opcional. Caso necessite de autenticação para o servidor LDAP, especifique True para activar este atributo.

DirectoryUser

Opcional. Caso active Credential, o nome do utilizador que pode aceder ao servidor LDAP.

DirectoryPassword

Opcional. Caso active Credential, a palavra-passe codificada para o utilizador que pode aceder ao servidor LDAP. Copie este valor do elemento <Password> (Crawler/DataSources/Server/Password) no mesmo ficheiro.

Restrição:

- O servidor LDAP tem de ser uma tarefa do servidor Domino para ser pesquisado.
- A palavra-passe do ficheiro de ID e a palavra-passe da Internet do utilizador do pesquisador têm de ser iguais.

Para configurar um pesquisador de NRPC de Notes para utilizar o servidor LDAP como Directório do Utilizador:

- 1. Utilize a consola de administração para criar um pesquisador de Notes que utilize o protocolo de NRPC.
- Inicie a sessão como administrador predefinido do IBM Content Analytics with Enterprise Search no servidor do pesquisador e edite o ficheiro notescrawler.xml. Este ficheiro encontra-se no directório ES_NODE_ROOT/ master_config/collection_ID.NOTES_ID/.
- **3**. Actualize o elemento do Directório de Utilizador (Crawler/DataSources/ Server/UserDirectory). Por exemplo:

```
<ExcludedPatterns />
<UserDirectory Type="LDAP">
   <ServerName>ldap.example.com</ServerName>
   <Port>389</Port>
   <BaseDN></BaseDN>
   <SSL Enabling="False">
    <TrustPath></TrustPath>
```

```
<TrustPassword></TrustPassword>
</SSL>
<Credential Enabling="True">
<DirectoryUser>LdapAdmin</DirectoryUser>
<DirectoryPassword>AbcDefGhiJkl==</DirectoryPassword>
</Credential>
</UserDirectory>
<ScheduleList>
```

4. Reinicie o sistema IBM Content Analytics with Enterprise Search:

```
esadmin system
stopall
esadmin system startall
```

Cumprimento da segurança ao nível do documento para documentos do sistema de ficheiros do Windows

Para permitir que as credenciais actuais sejam validadas quando um utilizador procura documentos que tenham sido pesquisados por um pesquisador do Sistema de ficheiros Windows, tem de configurar informações de conta de domínio no servidor do pesquisador e no servidor do Microsoft Windows.

Ao configurar um pesquisador do Sistema de ficheiros Windows, especifica se pretende pesquisar sub-directórios no computador ou sub-directórios num computador remoto. Se a segurança estiver activada para a colecção, também pode especificar opções para controlar o acesso a documentos nos sub-directórios pesquisados.

Se escolher fazer cumprir controlos de acesso validando as credenciais actuais do utilizador quando o mesmo submete uma consulta, tem de se assegurar de que as contas de domínio estão configuradas correctamente. Os requisitos para definir contas de domínio para ficheiros que tenham sido pesquisados no computador local são diferentes dos requisitos para ficheiros pesquisados num servidor do Windows remoto.

Importante: As credenciais de utilizador não podem ser validadas durante o processamento de consultas se ambas as seguintes condições forem verdadeiras:

- O servidor do Windows a ser pesquisado não for um membro de um domínio.
- O directório a pesquisar for um directório remoto como, por exemplo, \\servername\hostname.

Validação com dados de controlo de acesso local

Para validar as credenciais de utilizador actuais, o sistema utiliza a informação de conta de utilizador local e a informação de conta de domínio (se o computador pertencer a um domínio do Windows). Para validar as credenciais durante o processamento da consulta, ambos os nomes de utilizador têm de estar listados nas informações de segurança para documentos a serem procurados.

Para uma conta local, o nome do utilizador tem o seguinte formato: *NOME DO COMPUTADOR\NOME DO UTILIZADOR*

Para uma conta de domínio, o nome de utilizador está no seguinte formato: NOME DO DOMÍNIO\NOME DO UTILIZADOR Para iniciar sessão, os utilizadores especificam apenas o nome do utilizador, mas a atribuição de direitos do utilizador do Windows especificada correctamente utiliza o nome completo. Por exemplo, se o nome do utilizador da conta local for utilizadorabc, o nome da conta completo pode ser SERVIDORWIN1\utilizadorabc.

Quando os utilizadores acedem a uma aplicação e configuram um perfil para procurar documentos protegidos num sistema local, têm de especificar o nome de utilizador que utilizam para iniciar sessão no Windows (por exemplo, utilizadorabc).

Para fazer cumprir a validação de credencial actual em computadores locais, as contas de utilizador utilizadas pelo servidor do pesquisador têm de ter os seguintes direitos de utilizador do Windows. Para atribuir direitos de utilizador, utilize as Ferramentas Administrativas do Windows: Ferramentas Administrativas > Política de Segurança Local > Políticas Locais > Atribuição de Direitos de Utilizador Local.

- O ID de utilizador que o servidor do pesquisador está a executar tem de ter o direito Agir como componente do sistema operativo. Este direito é configurado para o utilizador administrativo no servidor do pesquisador quando se instala o IBM Content Analytics with Enterprise Search.)
- Os utilizadores têm de ter o direito de utilizador Iniciar Sessão Localmente.

Validação com dados de controlo de acesso de domínio remoto

Para o sistema operativo Windows, qualquer directório que comece por \\nome de servidor é considerado um directório remoto. Por exemplo: \\software\utilities\IBM

Para acederem a um directório remoto, os utilizadores especificam os nomes de utilizador no seguinte formato: NOME DO UTILIZADOR@NOME DO DOMÍNIO

Quando os utilizadores acedem a uma aplicação e configuram um perfil que lhes permite procurar documentos num sistema remoto, têm de especificar o nome de utilizador que utilizam para aceder ao sistema remoto do Windows (por exemplo, utilizadorabc@win1.empresa.com).

Para fazer cumprir a validação de credencial actual em computadores remotos, as contas de utilizador têm de ter os seguintes direitos de utilizador do Windows. Para atribuir direitos de utilizador, utilize as Ferramentas Administrativas do Windows: Ferramentas Administrativas > Política de Segurança de Domínio.

- O servidor do pesquisador e o servidor do Windows a serem procurados têm de ser membros do mesmo domínio.
- O ID de utilizador que o servidor do pesquisador está a executar tem de ter o direito Agir como componente do sistema operativo. Este direito é configurado para o utilizador administrativo no servidor do pesquisador quando se instala o IBM Content Analytics with Enterprise Search.)
- Os utilizadores têm de ter o direito de utilizador **Iniciar sessão como trabalho não interactivo**.

Utilize as seguintes directrizes para sistemas de ficheiros Windows remotos:

• Quando o IBM Content Analytics with Enterprise Search pesquisa um servidor Windows como um sistema de ficheiros remoto, recolhe o ACL do domínio configurado, mas não recolhe ACLs de domínios locais ou vários domínios. Apenas grupos de domínio e utilizadores são suportados para a pesquisa protegida das pastas remotas. Os grupos e utilizadores locais não são suportados, independentemente de os grupos serem definidos pelo utilizador ou incorporados (como Administradores, Utilizadores, Utilizadores de Domínios, Todos e Utilizadores Autenticados).

- Se utilizar o Identity Management Component (IMC), especifique nomes de utilizadores sem o domínio (por exemplo, nomeutilizador, e não nomeutilizador@domínio).
- Se criar a cadeia de XML de contexto da segurança dos utilizadores em vez de utilizar o IMC facultado, adicione os grupos de domínio a que o utilizador pertence na cadeia XML USC. Os utilizadores podem depois ser capazes de pesquisar ficheiros num servidor Windows remoto.

Conceitos relacionados:

"Procura protegida de domínios do Windows fidedignos"

"Validação das credenciais actuais durante o processamento de consultas" na página 387

Referências relacionadas:

"Requisitos de configuração do pesquisador para suporte de segurança" na página 397

Procura protegida de domínios do Windows fidedignos

Para fazer cumprir a segurança ao nível dos documentos nos sistemas de ficheiros Windows remotos, o sistema IBM Content Analytics with Enterprise Search suporta a verificação de listas de controlo de acessos (ACL, Access Control List) ao longo de domínios fidedignos.

Configurar o pesquisador

Para configurar o pesquisador de Sistema de ficheiros Windows para suportar domínios fidedignos,tem de especificar opções num novo ficheiro de configuração. Não existe suporte para configurar esta capacidade na consola de administração.

1. Para suportar a segurança ao nível dos documentos ao longo de domínios do Windows fidedignos, edite o ficheiro seguinte:

ES_NODE_ROOT/master_config/session_ID/winfscrawler_ext.xml

Sugestão: Para determinar o ID da sessão para o pesquisador Sistema de ficheiros Windows que pretende configurar, pode supervisionar os detalhes do pesquisador na consola administrativa ou utilizar o comando **esadmin report collections**.

2. Especifique o nome do domínio do Windows e o nome da NETBIOS do Active Directory. Por exemplo:

```
<ExtendedProperties>
<SetAttribute XPath="/Crawler/DataSources/Server/Target"
Name="Domain">ExampleCo.com
</SetAttribute>
<SetAttribute XPath="/Crawler/DataSources/Server/Target"
Name="NetBIOSDomain>EXC1
</SetAttribute>
</ExtendedProperties>
```

3. Pare e reinicie o pesquisador para que as alterações se tornem efectivas.

Restrições

- Os documentos não podem incluir ACLs de vários domínios do Windows. Os utilizadores e grupos de domínios devem pertencer a um domínio do Windows por colecção.
- Para suportar a verificação de acesso a sistemas de ficheiros remotos, os servidores do Windows devem ser executados no mesmo domínio do Windows ou em domínios do Windows fidedignos.
- O pesquisador de sistemas de ficheiros do Windows lê o nome da NETBIOS do Active Directory associado ao servidor do Windows a pesquisar e utiliza o nome da NETBIOS para filtrar o ACL do ficheiro. O Active Directory a que o servidor do pesquisador se junta confia no outro Active Directory que define as contas de utilizador e contas de grupo.
- A conta de utilizador que especificar para ser utilizada pelo pesquisador ao aceder a um servidor do Windows remoto tem de pertencer ao domínio do Windows em que pretende fazer cumprir e verificar o controlo de acesso.
- O sistema operativo Windows só permite uma conta para ligar ficheiros de rede num único servidor de ficheiros. A outras contas não podem estabelecer ligação ao mesmo servidor de ficheiros ao mesmo tempo. Por conseguinte, não é possível configurar contas diferentes para pesquisadores diferente para pesquisar no mesmo servidor do Windows, mesmo que os pesquisadores estejam em colecções diferentes.

Conceitos relacionados:

"Cumprimento da segurança ao nível do documento para documentos do sistema de ficheiros do Windows" na página 411

Desactivar a segurança

É possível desactivar a segurança para uma aplicação. Se tiver configurado anteriormente controlos de segurança ao nível do documento, pode especificar que os controlos devem ser ignorados. As definições de segurança também afectam o modo como os resultados totalmente qualificados são apresentados nos resultados da procura.

Tarefas relacionadas:

"Desactivar a segurança de nível de documento"

"Desactivar a segurança para resultados da procura contraídos" na página 416

Desactivar a segurança de nível de documento

Pode permitir aos utilizadores consultarem uma colecção independentemente de quaisquer controlos de acesso estarem associados aos documentos no índice. Para pesquisadores que suportem validação de credenciais actuais, também pode permitir aos utilizadores consultarem uma colecção sem validar os controlos de acesso actuais durante o processamento da consulta.

Pode especificar opções de nível de documento apenas se a segurança tiver sido activada para a colecção quando esta foi criada.

É possível configurar todos os pesquisadores para associar categorias de segurança a documentos à medida que são pesquisados. As aplicações podem utilizar estas categorias, que são armazenadas no índice, para fazer cumprir controlos de acesso quando os utilizadores consultam a colecção. Este tipo de segurança é conhecido como pré-filtragem dos resultados.

Para alguns pesquisadores, também é possível especificar que pretende validar os controlos de acesso actuais que estão associados aos documentos na origem dos dados quando os utilizadores submetem consultas.Este tipo de segurança é conhecido como pós-filtragem dos resultados.

Para remover estas restrições de segurança, pode especificar que os servidores de procura devem ignorar quaisquer categorias de segurança que sejam transmitidas com uma consulta. Também pode permitir aos utilizadores consultarem documentos sem terem as credenciais comparadas a controlos de acesso actuais.

Poderá optar por desactivar temporariamente a segurança de nível de documentos, se estiver a testar uma colecção nova, ou se necessitar de solucionar problemas com uma aplicação.

Para desactivar a segurança ao nível do documento:

- 1. Tome uma das seguintes acções:
 - Na vista Colecções, localize a colecção segura e faça clique em Acções > Definições > Editar segurança ao nível do documento.
 - Na vista Segurança, seleccione a colecção na secção **Definições de procura** segura da área de segurança ao nível da colecção e faça clique em Editar segurança ao nível do documento.
- Na página Segurança ao Nível do Documento para Todos os Documentos, seleccione uma ou ambas as caixas de verificação, dependendo das restrições de segurança que pretende desactivar:
 - Seleccione a caixa de verificação Ignorar controlos de acesso ao nível do documento no índice se não pretende que os tokens de segurança que os pesquisadores associaram aos documentos sejam utilizados quando os utilizadores consultarem a colecção. Os pesquisadores continuam a adicionar categorias de segurança a documentos, mas os servidores de procura ignoram as categorias e permitem aos utilizadores procurarem os documentos protegidos anteriormente.

Caso tenha configurado quaisquer pesquisadores para validarem as credenciais actuais, os servidores de procura continuarão a fazer a pós-filtragem dos resultados e a validar os controlos de acesso a documentos actuais durante o processamento da consulta. Para remover todas as restrições de segurança e permitir que todos os utilizadores procurem todos os documentos, seleccione também a caixa de verificação **Não validar as credenciais actuais durante o processamento da consulta**.

• Seleccione a caixa de verificação **Não validar as credenciais actuais durante** o processamento da consulta se não pretender validar os controlos de acesso actuais quando os utilizadores submetem as consultas. Esta caixa de verificação está disponível apenas para documentos que tenham sido pesquisados por pesquisadores que suportem esta função.

Se seleccionar apenas esta caixa de verificação, os controlos de segurança pré-filtragem mantêm-se activos. Por exemplo, se tiver especificado opções para armazenar controlos de acesso no índice quando configurou o pesquisador, esses controlos de segurança continuam a ser aplicados, a não ser que também seleccione a caixa de verificação **Ignorar controlos de acesso de nível de documento no índice**.

3. Reinicie os servidores de procura para uma colecção.

"Desactivar a segurança" na página 414

Desactivar a segurança para resultados da procura contraídos

Se a segurança de colecção for activada, os resultados da procura provenientes do mesmo sítio não podem ser contraídos nos resultados da procura, a menos que especifique que não pretende validar credenciais de utilizador durante o processamento de consultas.

Pode especificar opções de nível de documento apenas se a segurança tiver sido activada para a colecção quando esta foi criada.

Quando as credenciais actuais são validadas, a origem de cada documento é inspeccionada e possivelmente encaminhada para validação. Tem de desactivar a validação de credenciais actuais se pretender que os documentos que tenham o mesmo prefixo de URI, ou que pertençam a um grupo de URIs contraído anteriormente configurado, sejam contraídos nos resultados da procura.

Para desactivar a validação de credenciais actuais para que os documentos possam ser contraídos nos resultados da procura:

- 1. Tome uma das seguintes acções:
 - Na vista Colecções, localize a colecção segura e faça clique em Acções > Definições > Editar segurança ao nível do documento.
 - Na vista Segurança, seleccione a colecção na secção Definições de procura segura da área de segurança ao nível da colecção e faça clique em Editar segurança ao nível do documento.
- Na página Segurança ao Nível do Documento para Todos os Documentos, seleccione a caixa de verificação Não validar credenciais actuais durante o processamento de consultas.
- 3. Reinicie os servidores de procura para uma colecção.

Quando os utilizadores consultarem a colecção, os documentos que tiverem o mesmo prefixo de URI, ou que pertençam a sítios que estão configurados para serem contraídos, são contraídos nos resultados da consulta. Na aplicação de procura empresarial, os utilizadores podem ver os resultados contraídos fazendo clique na ligação **Mais resultados na mesma origem**.

Conceitos relacionados:

"Desactivar a segurança" na página 414

Iniciar e parar o sistema

Depois de criar uma colecção, deve iniciar os servidores para pesquisarem conteúdos. Os servidores de processamento de documentos e servidores de procura são iniciados automaticamente. Pare e reinicie os servidores depois de efectuar alterações à colecção.

A maioria dos servidores do IBM Content Analytics with Enterprise Search pode ser executada de forma contínua ou de acordo com as agendas que especifica. Depois de iniciar o sistema, normalmente, só é necessário parar e reiniciar os processos do servidor apenas quando altera as definições de configuração.

Por exemplo, se alterar a configuração do pesquisador para recolher o conteúdo de diferentes origens de dados, deve parar e reiniciar o pesquisador para que as alterações se tornem efectivas. Se não alterar as regras de pesquisa, o pesquisador da Web funciona continuamente e os outros pesquisadores funcionam de acordo com as agendas que especificar.

Para melhorar a disponibilidade dos servidores de procura quando o servidor de controlador e a consola de administração não estão disponíveis, pode especificar comandos para iniciar os servidores de procura para uma colecção no modo autónomo.

Tarefas relacionadas:

"Iniciar o sistema"

"Parar os servidores de colecção ou todo o sistema" na página 419

- "Controlar quais os componentes iniciados ou parados" na página 420
- "Administrar os servidores de procura no modo autónomo" na página 421

"Iniciar e parar serviços locais" na página 422

Iniciar o sistema

Se o seu sistema incluir pesquisadores, tem de iniciar as sessões do pesquisador. Os servidores para processar documentos, construir o índice e processar pedidos de consulta são iniciados automaticamente quando há dados disponíveis para analisar, indexar ou consultar.

Para iniciar o sistema, tem de utilizar uma conta de utilizador que possa ser autenticada com autenticação local. Se tentar iniciar o sistema com uma conta do Andrew File System (AFS), ocorrerão erros.

Para iniciar o sistema:

1. Inicie sessão como administrador predefinido do IBM Content Analytics with Enterprise Search e inicie os componentes do sistema:

esadmin system startall

Este comando inicia todos os componentes e sessões do sistema como, por exemplo :

- O CCL no computador em que executa o comando, se o CCL ainda não estiver em execução
- O servidor de rede da base de dados no servidor do pesquisador
- A sessão ESAdmin para a consola administrativa no servidor do controlador

- As aplicações de procura e a prospecção de análise de conteúdo nos servidores de procura
- O Centro de Informações do IBM Content Analytics with Enterprise Search
- Inicie a consola de administração e inicie sessão como administrador. Se utilizar as funções administrativas, pode iniciar sessão como um administrador de colecção ou como um operador que tem autoridade para a colecção que pretende iniciar.
- 3. Na vista Colecções, expanda a colecção que pretende administrar.
- 4. No painel Pesquisar e Importar, faça clique num ícone para iniciar um pesquisador específico, ou faça clique no ícone para seleccionar e iniciar vários pesquisadores ao mesmo tempo.
 - Se iniciar um pesquisador Web, o pesquisador iniciará imediatamente a pesquisa dos dados. Estes tipos de pesquisadores funcionam continuamente para pesquisar e voltar pesquisar documentos da Web.
 - Se iniciar um dos outros tipos de pesquisadores, é iniciada a sessão do pesquisador. O pesquisador começa por pesquisar documentos de acordo com a acção que tiver seleccionado nas propriedades do pesquisador (pesquisa completa, todas as actualizações ou apenas documentos novos e modificados). Pode controlar o tipo de pesquisa que é feita editando as propriedades do pesquisador, configurando uma agenda de pesquisador ou supervisionando detalhes sobre os conteúdos pesquisados por um pesquisador específico.

Depois de o pesquisador iniciar, pode deixá-lo a executar continuamente. Se tiver agendado o pesquisador, este será executado de novo nas datas e horas agendadas.

O serviço de analisador é automaticamente iniciado. Regra geral, não é preciso parar o analisador, a menos que efectue alterações ao procedimento de interpretação e análise dos dados. A indexação também inicia automaticamente e o índice é actualizado gradualmente quando os documentos estão disponíveis para o índice.

Os servidores de procura são iniciados automaticamente e pode deixá-los funcionar continuamente. Normalmente, não tem necessidade de parar os servidores da procura a menos que efectue as alterações na memória cache de procura ou nas definições do resumo dos documentos.

Se o servidor do controlador não estiver em execução, a consola administrativa não fica disponível. Para melhorar a disponibilidade dos servidores de procura quando a consola de administração não está disponível, pode especificar comandos para iniciar os servidores de procura para uma colecção no modo autónomo.

"Iniciar e parar o sistema" na página 417

Tarefas relacionadas:

"Iniciar sessão na consola de administração" na página 25

"Configurar funções administrativas" na página 376

"Controlar quais os componentes iniciados ou parados" na página 420

"Administrar os servidores de procura no modo autónomo" na página 421

Parar os servidores de colecção ou todo o sistema

Se fizer alterações a uma colecção, ou se precisar de resolver problemas, poderá ter de parar e reiniciar os servidores de uma colecção ou todo o sistema IBM Content Analytics with Enterprise Search.

Pode parar os servidores independentemente uns dos outros. Por exemplo, se parar e reiniciar um pesquisador para incorporar alterações que tenha efectuado à respectiva configuração, não necessita de parar e reiniciar os servidores de procura.

Para parar os servidores de uma colecção:

- 1. Na vista Colecções, expanda a colecção que pretende administrar.
- Para parar um ou mais pesquisadores, aceda ao painel Pesquisar e Importar, faça clique num ícone para parar um pesquisador específico, ou faça clique no ícone para seleccionar e parar vários pesquisadores ao mesmo tempo.

Se alterar o espaço de pesquisa ou as propriedades do pesquisador, pare-o e reinicie-o para incorporar as alterações. Se alterar o espaço de pesquisa e pretender aplicar as alterações aos documentos que já estão indexados, terá também de pesquisar os documentos de novo.

3. Para parar os serviços do analisador, aceda ao painel Analisar e Indexar e faça clique no ícone para parar os servidores de processamento de documentos. Se fizer clique em **Detalhes**, pode parar e reiniciar servidores de processamento de documentos individuais e reiniciar construções de índices.

Quando alterar as regras de análise de dados, pare e reinicie o analisador para incorporar as alterações. As alterações aplicam-se somente aos documentos recentemente pesquisados. Para aplicar as alterações a documentos que já se encontrem no índice, tem de iniciar uma pesquisa completa para pesquisar de novo todos os documentos, ou reconstruir o índice a partir da memória cache de documentos.

4. Para parar os servidores de procura, aceda ao painel Procura ou ao painel Análise e faça clique no ícone para parar os servidores de procura. Se fizer clique em **Detalhes**, pode parar e reiniciar servidores de procura individuais. Normalmente, só necessita de parar e reiniciar os servidores da procura quando alterar a memória cache de procura ou as definições do resumo do documento.

"Iniciar e parar o sistema" na página 417

Tarefas relacionadas:

"Iniciar sessão na consola de administração" na página 25

"Controlar quais os componentes iniciados ou parados"

"Administrar os servidores de procura no modo autónomo" na página 421

Parar todo o sistema

Para parar todo o sistema em vez dos servidores de uma colecção:

- Inicie sessão como administrador predefinido do IBM Content Analytics with Enterprise Search. Num sistema com vários servidores, inicie sessão em qualquer servidor do sistema.
- 2. Pare todos os componentes do sistema: esadmin system stopall

Controlar quais os componentes iniciados ou parados

Pode controlar quais os componentes iniciados ou parados pelos comandos esadmin system startall e esadmin system stopall.

O ficheiro ES_INSTALL_ROOT/default_config/AutoRunComponents.properties contém uma lista de componentes que podem ser iniciados ou parados pelos comandos **esadmin system startall** e **esadmin system stopall**. Por predefinição, todos os componentes listados são iniciados e parados por estes comandos.

Se pretender impedir que certos componentes sejam iniciados ou parados, pode editar o ficheiro de propriedades.

Para especificar os componentes que serão iniciados ou parados quando iniciar ou parar o sistema:

- 1. Inicie sessão como administrador predefinido do IBM Content Analytics with Enterprise Search no servidor em que tenciona executar o comando **esadmin system startall** ou **esadmin system stopall**.
- Edite o ficheiro ES_INSTALL_ROOT/default_config/ AutoRunComponents.properties.
- Para impedir que um componente seja iniciado, adicione um campo relativo ao componente, chamado Component.startable.ID_componente=false, em que ID_componente é o componente que não pretende iniciar.
- Para impedir que um componente seja parado, adicione um campo relativo ao componente, chamado Component.stopable.*IDd_componente*=false, em que *ID_componente* é o componente que não pretende parar.
- 5. Guarde e saia do ficheiro.

Da próxima vez que utilizar o comando **esadmin system startall** ou **esadmin system stopall**, o componente que modificou não será iniciado ou parado, de acordo com as alterações que efectuou no ficheiro de propriedades.

Exemplo: Neste exemplo, o servidor de HTTP será iniciado pelo comando **esadmin system startall** (a predefinição), mas não será parado pelo comando **esadmin system stopall** (tal como controlado pela linha evidenciada no exemplo):
Component.impl.class.3=com.ibm.es.control.util.component.impl.HTTPControlImpl
Component.nodes.3=search
Por predefinição, todos os componentes são iniciáveis
Component.stopable.3=false

Conceitos relacionados:

"Iniciar e parar o sistema" na página 417

Administrar os servidores de procura no modo autónomo

Para garantir a elevada disponibilidade dos servidores de procura, pode iniciá-los para colecções individuais mesmo que o servidor do controlador não esteja em execução.

A capacidade para iniciar e parar servidores de procura no modo autónomo não está disponível na consola de administração. Se o servidor de indexação não estiver em execução, não será possível aceder à consola de administração.

Antes de poder iniciar e parar os servidores de procura no modo autónomo, certifique-se de que o pesquisador, o processamento de documentos e os servidores de procura para a colecção foram todos iniciados, pelo menos, uma vez. Isto é necessário para assegurar que os ficheiros necessários são sincronizados nos servidores de procura.

Se a segurança de nível de documento estiver activada para a colecção, certifique-se de que o servidor do pesquisador também foi iniciado. Isto é necessário para assegurar que os controlos da segurança de nível de documento podem ser aplicados. Se o servidor do pesquisador não estiver em execução, só serão devolvidos nos resultados da consulta os documentos que não requeiram autenticação.

Se o servidor do controlador não estiver disponível, pode assegurar que os utilizadores podem continuar a procurar no sistema iniciando os servidores de procura no modo autónomo. Pode executar os comandos para iniciar e parar os servidores de procura a partir de qualquer servidor, numa instalação de vários servidores. Os comandos tentam iniciar ou parar os servidores de procura para a colecção especificada em todos os servidores de procura disponíveis.

Para iniciar ou parar os servidores de procura no modo autónomo:

- 1. Inicie os servidores de procura para uma colecção quando o servidor de indexação não estiver a executar.
 - a. No servidor do pesquisador e nos servidores de procura, inicie sessão como administrador predefinido do IBM Content Analytics with Enterprise Search e depois inicie o serviço CCL (common communication layer):

AIX ou Linux

startccl.sh -bg

Linha de comandos do Windows startccl

Ferramenta administrativa de Serviços Windows

Para iniciar o CCL em segundo plano:

- Inicie os Serviços do Windows: Iniciar > Programas > Ferramentas Administrativas > Serviços.
- 2) Faça clique com o botão direito do rato em IBM Content Analytics with Enterprise Search e faça clique em Iniciar.

- b. Nos servidores de procura, introduza o seguinte comando: esadmin system startLocal
- c. Introduza o seguinte comando, em que *id_colecção* identifica a colecção à qual pertencem os servidores de procura que pretende iniciar:
 esadmin startSearch –cid *id_colecção*
- 2. Pare os servidores de procura para uma colecção quando o servidor do controlador não estiver em execução:
 - a. Inicie sessão como administrador predefinido do IBM Content Analytics with Enterprise Search em qualquer servidor.
 - b. Introduza o seguinte comando, em que *id_colecção* identifica a colecção à qual pertencem os servidores de procura que pretende parar:

esadmin stopSearch -cid id_colecção

Conceitos relacionados:

"Iniciar e parar o sistema" na página 417

Iniciar e parar serviços locais

Numa configuração de vários servidores de IBM Content Analytics with Enterprise Search, é possível executar comandos para iniciar e parar serviços na máquina local, sem iniciar ou parar todo o sistema.

Quando executa o comando **startLocal** para iniciar sessões na máquina local, são iniciadas as sessões mínimas necessárias para a procura. O comando **stopLocal** pára todas as sessões, mas não pára o Common Communication Layer (CCL). Se o WebSphere Application Server estiver instalado no servidor de procura, o comando stopLocal não pára o WebSphere Application Server.

Para iniciar ou parar serviços apenas na máquina local:

- 1. Inicie sessão como administrador predefinido do IBM Content Analytics with Enterprise Search no servidor que quer iniciar ou parar.
- Iniciar serviços apenas nesta máquina: esadmin system startLocal
- 3. Parar serviços apenas nesta máquina: esadmin system stopLocal

Conceitos relacionados:

"Iniciar e parar o sistema" na página 417

Supervisionar o sistema

Quando supervisiona actividades de sistema e de colecção, pode ver o estado de vários processos, estar atento a problemas potenciais ou ajustar as definições de configuração para melhorar o desempenho.

Na consola de administração, pode supervisionar o sistema e ajustar as operações segundo as necessidades. Pode ver estatísticas detalhadas para cada actividade principal (pesquisar, analisar e processamento de pesquisas). As estatísticas incluem tempos de resposta médios e informações sobre o progresso, tais como a quantidade de documentos que foram pesquisados ou indexados durante uma sessão.

Ao fazer clique em ícones, pode parar e iniciar a maioria das actividades. Por exemplo, poderá ter de parar um servidor para efectuar alterações à respectiva configuração ou detectar e corrigir um problema, bem como reiniciar o processamento quando estiver pronto a permitir que a actividade continue.

Pode ver a topologia do sistema na página Configuração do Servidor da vista Sistema. Nesta vista, pode rapidamente ver os nomes do sistema central dos servidores e endereços IP, a função de cada servidor no sistema (como, por exemplo, se é um servidor principal ou um servidor dedicado à procura ou ao processamento de documentos) como os servidores se interligam.

Utilize a página Configuração do Servidor para iniciar e interromper os servidores individuais ou múltiplos servidores ao mesmo tempo, adicionar servidores ao sistema ou remover servidores do sistema. Para um servidor de procura, pode também supervisionar as estatísticas de processamento de consulta, como o desempenho e histórico do processamento de pedidos.

Supervisionar pesquisadores

Pode ver informações gerais sobre o estado de cada pesquisador numa colecção ou seleccionar opções para ver informações detalhadas sobre a actividade de um pesquisador.

Se a função administrativa limitar o utilizador à supervisão de colecções, pode ver as estatísticas do pesquisador mas não pode alterar o comportamento de um pesquisador (como iniciar e parar).

Para supervisionar um pesquisador:

- 1. Na vista Colecções, expanda a colecção que pretende supervisionar e aceda ao painel Pesquisar e Importar.
- 2. Caso queira parar ou iniciar vários pesquisadores, faça clique no ícone para seleccionar os pesquisadores que pretende iniciar ou parar.
- **3**. Se um pesquisador em particular estiver em execução ou parado e se pretender visualizar as informações detalhadas do estado sobre o pesquisador, faça clique no ícone para supervisionar os detalhes sobre o conteúdo pesquisado pelo pesquisador. Os tipos de estatísticas encontrados variam consoante o tipo de pesquisador.

Se a sua função administrativa lhe permitir administrar processos para uma colecção, poderá iniciar e parar o pesquisador enquanto visualiza detalhes

sobre a actividade do pesquisador. Se for possível agendar o pesquisador, pode também activar e desactivar a agenda de pesquisa.

4. Se o pesquisador estiver parado e caso queira iniciar uma sessão do pesquisador, faça clique no ícone para começar ou retomar o pesquisador.

Para pesquisadores da Web:

Se o pesquisador tiver sido parado, começará a pesquisar de novo todo o espaço de pesquisa. Se o pesquisador tiver sido interrompido, retoma a pesquisa no início do destino onde foi interrompido.

Se pretender forçar o pesquisador a iniciar um nova pesquisa completa imediatamente, faça clique no ícone para ver os detalhes sobre o pesquisador e, de seguida, faça clique no ícone **Iniciar uma nova pesquisa completa**. O pesquisador inicia a pesquisa de todo o espaço de pesquisa, incluindo páginas que não foram alteradas desde a última vez que foram pesquisadas. Pode achar conveniente pesquisar de novo todos os documentos, por exemplo, se alterar as regras para analisar documentos e pretender aplicá-las aos documentos que foram anteriormente indexados.

Para todos os outros tipos de pesquisadores:

Se o pesquisador tiver sido parado, iniciará a pesquisa na data e hora agendadas. Da primeira vez que o pesquisador pesquisa uma origem de dados, executa uma pesquisa completa. Quando uma pesquisa agendada é repetida, o pesquisador pesquisa ou todas as actualizações para a origem de dados (adições de documentos, eliminações e modificações) ou apenas documentos novos e modificados. Pode especificar o tipo de pesquisa a executar quando uma sessão do pesquisador é iniciada quando configura as propriedades do pesquisador ou a agenda do pesquisador.

Se seleccionar a opção de pesquisar apenas documentos novos e modificados, o pesquisador procura documentos com datas de modificação que são posteriores à hora da pesquisa anterior. Se copiar ficheiros para um recurso, a data de modificação pode não se alterar, o que significa que o pesquisador pode não detectar que os ficheiros foram adicionados ao recurso. Por exemplo, se copiar os ficheiros para uma pasta Windows, o Windows não muda automaticamente a data de modificação dos ficheiros. Para assegurar que esses ficheiros são pesquisados, seleccione a opção para pesquisar todas as actualizações ou uma pesquisa completa.

Se pretender forçar um pesquisador a iniciar a pesquisa, faça clique no ícone para ver os detalhes sobre o pesquisador. Em seguida, na área de detalhes do espaço de pesquisa, faça clique no ícone para o tipo de pesquisa que pretende iniciar: uma pesquisa completa, todas as actualizações ou apenas documentos novos e modificados. Tem de fazer clique no ícone to para começar a pesquisar cada origem de dados que pretende pesquisar (tais como um servidor, uma base de dados ou uma subpasta).

5. Se o pesquisador estiver em execução e pretender pará-lo, faça clique no ícone para o parar. O pesquisador pára a pesquisa dos dados até que o utilizador o reinicie.

Conceitos relacionados:

"Como o pesquisador do Notes determina o número de documentos pesquisados" Tarefas relacionadas:

"Visualizar detalhes sobre a actividade de pesquisadores da Web" na página 426

Como o pesquisador do Notes determina o número de documentos pesquisados

Ao supervisionar o pesquisador Notes, o número que é mostrado no campo **Documentos Pesquisados** na página Detalhes para Vistas e Pastas do Notes pode parecer maior do que o número de documentos na base de dados do Lotus Notes.

Um anexo de um documento do Lotus Notes é considerado como um documento em separado do seu documento principal. Se o anexo foi um ficheiro de arquivo (como .zip, .tar, .gz e assim por diante), os ficheiros no arquivo também serão contados em separado.

Na página Detalhes para Vistas e Pastas do Notes, o valor que é apresentado para o número de documentos pesquisados é igual à saída do seguinte comando:

esadmin Notes_crawler_session_ID getCrawlSpaceStatusDetail -ts
target_Server

O valor de documentos pesquisados é basicamente a soma dos seguintes números, e não corresponde necessariamente o número de documentos que se encontram armazenados na base de dados de origem:

- O número de documentos que foram acedidos.
- O número de anexos
- O número de ficheiros nos ficheiros do arquivo

Além disso, alguns documentos podem ser encontrados pelo pesquisador e, em seguida, abandonados (por exemplo, os documentos corrompidos são abandonados). Neste caso, é possível confirmar quais os documentos que foram pesquisados consultando os ficheiros de registo de auditoria do pesquisador no directório ES_NODE_ROOT/logs/audit/.

Se um documento não estiver acessível para o utilizador do pesquisador, não será pesquisado nem irá aparecer no ficheiro de registo de auditoria do pesquisador.

Ao determinar se um documento ou anexo tiver sido actualizado, o pesquisador do Notes utilize a seguinte lógica:

- Quando um anexo de um documento Notes é actualizado, a data alterada do documento Notes também é actualizada. O pesquisador do Notes assume que tanto o texto do corpo do documento e o anexo foram actualizados, e serão contabilizados individualmente como documentos actualizados.
- Se apenas o corpo de texto de um documento Notes for actualizado, o corpo do texto e o anexo são considerados como documentos actualizados.
- Se um anexo estiver num ficheiro de arquivo, os ficheiros actualizados no ficheiro de arquivo serão contabilizados individualmente.
- Se apenas um dos ficheiros num arquivo anexado for actualizado, o corpo do texto do documento Notes e o ficheiro alterado no ficheiro de arquivo são contados como documentos individualmente actualizados. Outros ficheiros no ficheiro de arquivo são processados como documentos não alterados.

Referências relacionadas:

"Comandos, códigos de retorno e IDs da sessão" na página 511

Visualizar detalhes sobre a actividade de pesquisadores da Web

Ao visualizar detalhes sobre a actividade dos pesquisadores da Web, pode avaliar o desempenho geral e ajustar as propriedades dos pesquisadores da Web e as definições do espaço de pesquisa conforme necessário.

Para ver detalhes sobre a actividade de um pesquisador Web:

- 1. Na vista Colecções, expanda a colecção que pretende supervisionar e aceda ao painel Pesquisar e Importar.
- Se o pesquisador da Web que pretende supervisionar estiver em execução ou em pausa, faça clique no ícone para supervisionar os detalhes sobre o conteúdo pesquisado pelo pesquisador.
- **3**. Na página de detalhes do pesquisador Web, visualize ou seleccione as seguintes opções para ver estatísticas detalhadas sobre a actividade actual e anterior do pesquisador:
 - Faça clique em **Detalhes do módulo** para ver quantos módulos estão a pesquisar activamente sítios da Web e quantos se encontram num estado inactivo.
 - Faça clique em **Sítios activos** para ver informações sobre os sítios da Web que o pesquisador está a pesquisar activamente.
 - Faça clique em **URLs recentemente pesquisados**. Estas informações mostram o que o pesquisador pesquisou recentemente. Se os itens da lista não forem alterados quando renovar a vista, não está a ocorrer nenhuma pesquisa.
 - Faça clique em **Histórico do pesquisador** para ver relatórios sobre a actividade anterior do pesquisador.
 - Na área **Estado do URL**, escreva um URL sobre o qual pretende visualizar informações.
 - a. Faça clique em **Detalhes do URL** para ver informações de estado sobre o URL. Pode pedir detalhes de URL apenas para os URLs que foram anteriormente pesquisados.
 - b. Faça clique em Detalhes do sítio para especificar informações que pretende incluir num relatório acerca do sítio da Web ao qual o URL pertence. Pode pedir detalhes de sítio para um sítio da Web anteriormente pesquisado ou para um sítio da Web que ainda não tenha sido pesquisado.

Por exemplo, utilize esta opção para ver se um URL se encontra no espaço da pesquisa, se foi pesquisado ou apenas identificado, quando deverá ser pesquisado de novo e informações sobre a última tentativa de pesquisa do sítio da Web. Também pode pedir a visualização do conteúdo do ficheiro robots.txt do sítio da Web, o que poderá ajudá-lo a determinar por que motivo o sítio não está a ser pesquisado.

Conceitos relacionados:

"Detalhes do módulo do pesquisador da Web"

"Sítios activos do pesquisador da Web"

"Velocidade da pesquisa do pesquisador da Web" na página 428

Tarefas relacionadas:

"Criar relatórios de pesquisadores da Web" na página 429

Referências relacionadas:

"Códigos de estado de HTTP devolvidos ao pesquisador da Web" na página 430

Detalhes do módulo do pesquisador da Web

Pode supervisionar o pesquisador da Web para ver quantos módulos estão a pesquisar activamente sítios e quantos estão num estado inactivo.

Quando visualiza detalhes sobre um pesquisador da Web enquanto supervisiona uma colecção, pode ver o estado dos módulos do pesquisador. Os estados mais prováveis de ver incluem:

Em espera

Indica que o módulo não tem um URL para pesquisar. Esta condição pode ocorrer quando um módulo termina uma pesquisa e o pesquisador não consegue encontrar URLs para pesquisar suficientemente rápido. Por exemplo, se a propriedade do pesquisador que controla quanto tempo este tem de esperar antes de poder obter outra página do mesmo servidor for demasiado alta, pode impedir que os URLs sejam fornecidos de forma suficientemente rápida.

Carregamento em memória

Indica que o módulo está a transferir uma página de um sítio da Web.

Concluído

Indica que o módulo está a enviar as páginas que pesquisou para o resto do pesquisador, mas ainda não está pronto para pesquisar outro URL.

Suspenso

Indica que o pesquisador está em pausa

Idealmente, todos os módulos estão a carregar páginas em memória a toda a hora. Se os módulos estiverem frequentemente num estado de conclusão, podem existir problemas de débito com a base de dados.

Se os módulos estiverem frequentemente num estado de espera, reveja o valor especificado para o campo **Número máximo de sistemas centrais activos** nas propriedades do pesquisador. Se o valor for baixo, pode não haver sítios suficientes no espaço de pesquisa para manter os módulos ocupados, ou pode não haver suficientes URLs elegíveis para serem pesquisados. Entre as condições que podem provocar tal actividade baixa incluem-se as falhas na procura de DNS e as falhas de procura de robôs.

Tarefas relacionadas:

"Visualizar detalhes sobre a actividade de pesquisadores da Web" na página 426

Sítios activos do pesquisador da Web

Pode supervisionar o pesquisador Web para ver informações sobre os sítios da Web que o pesquisador estiver a pesquisar activamente. Quando visualiza detalhes sobre um pesquisador enquanto supervisiona uma colecção, pode ver estatísticas sobre sítios activos. As estatísticas mostram:

- Quantos URLs o pesquisador importou para a memória a partir da sua base de dados interna para pesquisar neste momento
- · Quantos URLs o pesquisador tentou pesquisar até ao momento
- Quanto tempo resta antes de um sítio ser desactivado e removido da memória para esta iteração do pesquisador
- Quanto tempo um sítio esteve na memória até ao momento

Estas informações alteram-se de momento para momento à medida que o pesquisador progride através das regras de pesquisa configuradas para ele. O ideal é que o número de URLs activados se aproxime ao valor configurado para o campo **Número máximo de sistemas centrais activos** nas propriedades avançadas do pesquisador.

Se o número de URLs activados for próximo do zero, o pesquisador não está a encontrar URLs elegíveis. Entre as condições que podem provocar tal actividade baixa incluem-se as falhas na procura de DNS, questões de conectividade de rede, erros de base de dados e problemas de definição de espaço de pesquisa. Por exemplo:

- Se muitos sítios tiverem estado durante muito tempo na memória, e poucos URLs tiverem sido pesquisados, espere problemas de conectividade de rede.
- Se não existirem sítios suficientes na lista, espere problemas de definição de espaço de pesquisa ou problemas de busca de DNS.

Tarefas relacionadas:

"Visualizar detalhes sobre a actividade de pesquisadores da Web" na página 426

Velocidade da pesquisa do pesquisador da Web

Pode supervisionar o pesquisador da Web para consultar informações sobre a velocidade a que o pesquisador está a transferir páginas de sítios da Web.

Quando visualiza detalhes sobre um pesquisador enquanto supervisiona uma colecção, pode ver estatísticas sobre a velocidade a que o pesquisador está a pesquisar dados (a velocidade da pesquisa). Também pode ver estatísticas sobre a quantidade de URLs que o pesquisador pesquisou desde que a sessão actual começou.

A velocidade da pesquisa é o número de páginas que estão a ser pesquisadas por segundo. Este número está relacionado com diversas propriedades que pode configurar para o pesquisador Web:

- O número de módulos do pesquisador
- O número de sítios activos
- A quantidade de tempo que o pesquisador tem de aguardar antes de poder obter outra página do mesmo servidor da Web

Se o pesquisador tiver um sítio activo por cada módulo e tiver de esperar dois segundos antes de poder obter outra página do mesmo servidor Web, o pesquisador não consegue pesquisar mais rapidamente que a uma página por cada módulo por cada dois segundos. Por exemplo, se o pesquisador utilizar o número predefinido de módulos (200), o pesquisador pode pesquisar 100 páginas por cada 200 módulos. Se existir o dobro de sítios activos do número de módulos e o pesquisador tiver de esperar dois segundos antes de obter outra página do mesmo servidor da Web, o pesquisador pode alcançar uma página por cada módulo por cada segundo. Porém, as velocidades de transferência de rede e débito da base de dados tornar-se-iam factores restritivos. Uma indicação de um forte desempenho do pesquisador é visível quando o ritmo de pesquisa se põe a par com o número de módulos do pesquisador, sítios activos e o tempo de espera do pesquisador.

Outro factor a rever durante a supervisão do desempenho do pesquisador Web é o número de URLs que o pesquisador pesquisou desde o início da sessão do pesquisador actual. Divida esse número pelo montante total de tempo de execução do pesquisador para calcular a média do débito a longo prazo. Se este número não estiver a aumentar, o pesquisador terminou ou não consegue continuar. Por exemplo, erros de conectividade de rede, erros de base de dados e falhas de procura de DNS podem bloquear o progresso do pesquisador.

Tarefas relacionadas:

"Visualizar detalhes sobre a actividade de pesquisadores da Web" na página 426

Criar relatórios de pesquisadores da Web

Ao visualizar relatórios sobre a actividade passada do pesquisador da Web, pode avaliar o desempenho geral e ajustar as propriedades do pesquisador da Web e as propriedades do espaço de pesquisa conforme necessário.

Se a sua função administrativa limitar a supervisão de colecções, pode visualizar estatísticas do pesquisador e criar relatórios sobre a actividade do pesquisador, mas não pode alterar o comportamento de um pesquisador (tal como iniciar ou parar o pesquisador).

Tipos diferentes de relatórios podem fornecer informações sobre a actividade do pesquisador Web. Para determinados tipos de relatórios, as informações são devolvidas tão rapidamente quanto for possível recolhê-las da base de dados interna do pesquisador. O relatório Sítio e os relatórios de códigos de estado de HTTP demoram a ser criados. Se criar estes tipos de relatórios, pode especificar um endereço de correio electrónico para receber o relatório em vez de ter de esperar que os resultados sejam devolvidos à consola de administração.

Para criar relatórios de pesquisadores Web:

- 1. Expanda a colecção que possui o pesquisador da Web que pretende supervisionar e aceda ao painel Pesquisar e Importar.
- 2. Se o pesquisador da Web para o qual pretende criar relatórios estiver em execução ou em pausa, faça clique no ícone para supervisionar os detalhes sobre o conteúdo pesquisado pelo pesquisador.
- **3**. Na página dos detalhes para o pesquisador Web, seleccione uma opção para o tipo de relatório que pretende criar:
 - Na área **Resumo de estado do pesquisador**, faça clique em **Histórico do pesquisador** para criar relatórios sobre o pesquisador e todos os sítios que este descobre ou pesquisa.
 - Na área **Estado do URL**, especifique o URL do sítio específico para o qual pretende criar um relatório e, em seguida, faça clique em **Detalhes do sítio**.
- 4. Para ambos os relatórios do histórico de pesquisadores e do sítio, seleccione a caixa de verificação de cada estatística que pretende ver num relatório e, em seguida, faça clique em **Visualizar relatório**.

Para estes tipos de estatística, o pesquisador devolve um relatório à consola de administração tão rapidamente quanto possível, depois de obter informações da própria base de dados interna.

5. Se estiver a criar um relatório do histórico do pesquisador, pode especificar opções para criar um relatório Sítio e, em seguida, fazer clique em **Executar Relatório**.

Este relatório é criado com as estatísticas que escolheu incluir e é guardado num ficheiro que especificou (o nome do ficheiro tem de ser absoluto). Pode especificar que pretende receber mensagens de correio electrónico após a criação do relatório.

6. Se estiver a criar um relatório do histórico do pesquisador, pode especificar opções para criar um relatório de códigos de estado de HTTP e, em seguida, fazer clique em **Executar Relatório**.

Este relatório fornece informações sobre o número de códigos de estado de HTTP distribuídos por sítio. O relatório é guardado num ficheiro especificado pelo utilizador (o nome do ficheiro tem de ser absoluto). Pode especificar que pretende receber mensagens de correio electrónico após a criação do relatório.

Utilize este relatório para ver quais são os sítios que devolvem um grande número de códigos de estado 4xx (que indicam que não foram encontradas páginas), códigos de estado 5xx (que indicam um problema no servidor), códigos de estado 6xx (que indicam problemas de ligação), etc.

Este relatório é mais útil se o pesquisador tiver estado activo durante algum tempo (por exemplo, um pesquisador que tenha estado activo durante semanas). Pode ajudar a identificar sítios desaparecidos, sítios recentes, sítios com grande número de URLs (o que pode indicar uma pesquisa redundante de uma base de dados do Lotus Notes), e sítios com um sistema de ficheiros recursivo vindos do servidor de HTTP. Se os sítios com grandes números de códigos de estado de HTTP não contribuírem para o índice, pode melhorar o desempenho do pesquisador removendo os sítios do espaço de pesquisa.

Tarefas relacionadas:

"Visualizar detalhes sobre a actividade de pesquisadores da Web" na página 426

Códigos de estado de HTTP devolvidos ao pesquisador da Web

Quando supervisiona um pesquisador Web, pode visualizar informações sobre os códigos de estado de HTTP que o pesquisador recebe das páginas que tenta pesquisar.

Resumo em tabela

Quando supervisiona o histórico do pesquisador Web, ou quando supervisiona o estado de um URL específico, pode visualizar informações sobre os códigos de estado de HTTP que foram devolvidos ao pesquisador. Pode utilizar estas informações para gerir o espaço de pesquisa e optimizar o desempenho do pesquisador. Por exemplo, se o pesquisador receber um grande número de códigos de estado de HTTP para um URL, e se os códigos de estado indicarem que não é possível pesquisar páginas nessa localização, pode melhorar o desempenho removendo esse URL do espaço de pesquisa.

A tabela que se segue lista os códigos de estado de HTTP e a forma como o pesquisador os interpreta. Valores de 100 a 505 são códigos de estado HTTP padrão (consulte a norma Hypertext Transfer Protocol para obter mais

informações). Os códigos de estado de HTTP restantes são da propriedade do IBM Content Analytics with Enterprise Search e do pesquisador Web.

Tabela 22.	Códigos de	estado d	le HTTP d	o pesquisador	Web
	0			, ,	

Código	Descrição	Código	Descrição	Código	Descrição	Código	Descrição
NULO	Não pesquisado	400	Pedido Incorrecto	500	Erro interno de servidor	693	Falha de selecção (URLFetcher)
100	Continuar	401	Não autorizado	501	Não implementado	694	Erro de escrita (URLFetcher)
101	Protocolos de comutação	402	É necessário pagamento	502	Porta de ligação errada	695	Cabeçalho de bloco incompleto (URLFetcher)
200	Com êxito	403	Proibido	503	Serviço não disponível	699	Erro inesperado (URLFetcher)
201	Criado	404	Não encontrado	504	Tempo de espera esgotado da porta de ligação	700	Erro de análise (fim sem cabeçalho)
202	Aceite	405	Método não permitido	505	Versão HTTP não suportada	710	Erro de análise (cabeçalho)
203	Informações sem ser de autoridade	406	Não aceitável	611	Erro de leitura	720	Erro de análise (sem código HTTP)
204	Sem conteúdo	407	Autenticação proxy necessária	612	Erro de ligação	730	Erro de análise (corpo)
205	Repor conteúdo	408	Tempo de espera esgotado do pedido	613	Tempo de espera esgotado de leitura	740 ou 4044	Excluído por ficheiro robots.txt
206	Conteúdo parcial	409	Conflito	614	Protocolo de reconhecimento SSL falhou	741	Robôs temporariamente indisponíveis
300	Opções múltiplas	410	Desaparecido	615	Outro erro de leitura	760	Excluído pela definição do espaço de pesquisa
301	Permanentem. movido	411	Comprimento necessário	616	Anomalia FBA	761	Desautorizado pelo espaço de pesquisa local; permitido pelo global
302	Encontrado	412	Pré-condição falhou	617	Erro de codificação	770	Protocol incorrecto ou porta de sistema não padrão
303	Consultar outro	413	Entidade de pedido demasiado grande	618	Redireccionar sem URL de redireccionamen	780 to	Excluído pelas exclusões de tipo de ficheiros

Código	Descrição	Código	Descrição	Código	Descrição	Código	Descrição
304	Não modificado	414	O URI de pedido é demasiado comprido	680	Falha de procura de DNS	786	URL inválido
305	Utilizar proxy	415	Tipo de suporte não suportado	690	URL mal formato	2004	Sem etiqueta META de índice
306	(Não utilizado)	416	Intervalo pedido não satisfatório	691	Ligação perdida (URLFetcher)	3020	Reencaminham. permanente
307	Reencaminham. temporário	417	Expectativa falhou	692	Tempo de espera esgotado de escrita (URLFetcher)		

Tabela 22. Códigos de estado de HTTP do pesquisador Web (continuação)

Notas de tabela

Códigos de estado 4xx

Raramente verá um código 400 (pedido incorrecto). De acordo com a norma de códigos de estado de HTTP, os códigos 4xx devem indicar que o cliente (o pesquisador) falhou. Porém, o problema geralmente está no servidor ou no URL que o pesquisador recebeu como ligação. Por exemplo, alguns servidores da Web não toleram URLs que tentem navegar para cima a partir da raiz do sítio (como, por exemplo, http://xyz.ibm.com/../../ sales). Outros servidores da Web não têm problemas com esta navegação ascendente e ignoram o operador do directório ascendente (..) quando o pesquisador já está na raiz (root).

Alguns servidores tratam o pedido pela raiz do servidor como um erro, e algumas ligações obsoletas podem solicitar operações que já não são reconhecidas ou implementadas. Quando lhe são pedidas informações acerca de uma página à qual já não tem acesso, o servidor da aplicação emite uma excepção, o que faz com que o servidor da Web devolva o código de estado de HTTP 400 porque o pedido já não é considerado válido.

615 Indica que o servidor do pesquisador que transfere dados de sítios da Web encontrou uma excepção inesperada. Um grande número deste tipo de códigos de estado pode indicar que existe um problema com o pesquisador.

Códigos de estado 61x

Excepto para o 615, os códigos de estado 61x indicam problemas que podem ocorrer na pesquisa como, por exemplo, tempo de espera esgotado. Os códigos de estado que se seguem podem requerer uma acção correctiva:

611, 612 e 613

Sítios lentos ou um desempenho fraco da rede podem ser a causa destes problemas.

- **611** Indica que ocorreu um erro quando o pesquisador obteve um documento.
- 612 Indica que ocorreu um erro quando o pesquisador tentou ligar a um servidor Web.

- **613** Indica que ocorreu uma situação de tempo esgotado enquanto o pesquisador estava a obter um documento.
- 614 Indica que o pesquisador não consegue pesquisar sítios seguros (HTTPS). Se achar que estes sítios deviam estar acessíveis, verifique se os certificados estão definidos correctamente no servidor do pesquisador e no servidor da Web de destino. Por exemplo, se um sítio for certificado por uma entidade certificadora (CAs), pode adicionar novas CAs ao armazenamento de certificados fidedignos utilizado pelo pesquisador.

Também pode consultar a forma como certificados de autenticação própria são configurados nos sítios que pretende pesquisar. O pesquisador é configurado para aceitar certificados de autenticação própria. Alguns sítios criam um certificado de autenticação própria para um URL raiz (como, por exemplo, http://sales.ibm.com/) e, em seguida, tentam utilizar esse certificado em subdomínios (como, por exemplo, http://internal.sales.ibm.com/). O pesquisador não pode aceitar certificados que são utilizados desta maneira. O pesquisador aceita certificados de autenticação própria apenas se o nome do domínio do assunto (sales.ibm.com) e o certificador corresponderem ao nome do domínio da página que está a ser solicitada.

616 Indica que o formulário de início de sessão para a autenticação baseada e formulário (FBA, form-based authentication) ainda aparece na transferência depois de voltar a autenticar.

Se as informações fornecidas no ficheiro de configuração de FBA (formulário de início de sessão, além de dados de autenticação, como o nome de utilizador, palavra-passe, etc.) não conseguirem autenticar o pesquisador, o código de estado 616 é atribuído a todas as páginas dependentes da autenticação baseada em formulário. O administrador deve investigar para determinar por que motivo a configuração de FBA não está a funcionar.

- 617 Indica a incapacidade de criar uma Cadeia a partir do conteúdo de bytes de um documento porque a cadeia de codificação (charset) é inválida ou o documento contém bytes inválidos.
- 618 Indica que o URL de redireccionamento não é válido quando o pesquisador recebe os códigos de estado de HTTP que se seguem. É possível que a localização do cabeçalho de resposta de HTTP não seja válido.
 - 301 Movido Permanentemente 302 Encontrado
- **680** Indica que o pesquisador não conseguiu obter endereços de IP para sistemas centrais no espaço de pesquisa, talvez devido a problemas de acesso de rede. Este tipo de erro significa que o pesquisador não consegue pesquisar sítios completos, não apenas que não foi capaz de pesquisar alguns URLs. Um grande número deste tipo de códigos de estado reduz significativamente o débito.

Códigos de estado 69x

Os códigos de estado 690 e 699 nunca são gravados na base de dados persistente do pesquisador. Estes códigos representam resultados que não reflectem o resultado real de uma transferência de um sistema central remoto, mas uma condição temporária no pesquisador, tal como um componente que é encerrado enquanto outro está a aguardar um resultado ou o envio de um resultado. Estes códigos de estado aparecem em certos registos, mas não no registo persistente e, deste modo, não devem ser utilizados como valores definidos por selecção.

Códigos de estado 7xx

Os códigos 7xx devem-se, sobretudo, a regras no espaço de pesquisa:

```
710 - 730
```

Indica que houve problemas que impediram que o pesquisador concluísse a transferência, ou que o pesquisador encontrou dados HTML invadidos num sítio. Se visualizar grande quantidade deste tipo de códigos de estado, contacte o técnico dos serviços de assistência para obter ajuda.

740 ou 4044

Indica que o conteúdo de um ficheiro não pode ser indexado porque o documento é excluído por restrições no ficheiro robots.txt do sítio.

- 740 Indica que as ligações âncora que apontam para o documento excluído podem ser incluídos no índice.
- **4044** Indica que as ligações âncora em documentos que apontem para o documento excluído também são excluídas do índice.
- 741 Indica que um sítio tem um ficheiro robots.txt que permite pesquisar, mas a transferência falhou. Se não conseguir pesquisar o URL repetidamente, este é removido do espaço de pesquisa. Se tiver visto um grande número deste tipo de códigos de estado, verifique se o sítio de destino está indisponível temporária ou permanentemente. Se o sítio de destino já não estiver disponível, remova-o do espaço de pesquisa.

Os restantes códigos de estado 7xx ocorrem na maior parte das vezes quando efectua alterações ao espaço de pesquisa, depois de o pesquisador ter estado em execução durante algum tempo. Normalmente, estes códigos de estado não indicam problemas que necessite de considerar.

3020 Indica que um documento com o código de estado 200 contém um cabeçalho de localização que faz referência ao agente utilizador para outro URL.

Tarefas relacionadas:

"Visualizar detalhes sobre a actividade de pesquisadores da Web" na página 426

Supervisionar os serviços de análise e de indexação

Supervisione os serviços de análise e indexação para ver informações de estado e de erro sobre documentos à medida que são interpretados e analisados. Supervisione serviços de índice para ver o tamanho do índice e o estado de vários serviços de indexação.

Se a função administrativa limitar o utilizador à supervisão de colecções, pode ver o estado do analisador ou índice, mas não pode iniciar ou parar construções de analisador ou índice.

Quando supervisiona detalhes de analisador ou índice, visualiza um instantâneo com estatísticas acerca das actividades de análise e indexação num determinado momento. Por exemplo, as estatísticas mostram o número de documentos que foram analisados e adicionados ao índice, e o tamanho total do índice. As estatísticas também mostram se alguns documentos foram abandonados porque não foi possível indexá-los ou analisá-los, e faculte informações sobre o motivo de os documentos serem abandonados.

Quando o analisador está activo, as mensagens fornecem informações adicionais sobre o estado do analisador. Por exemplo:

- O analisador pode estar a analisar documentos de forma activa.
- O analisador poderá estar a aguardar por documentos pesquisados adicionais para analisar. O analisador fica inactivo até estarem disponíveis mais documentos para analisar.

Para supervisionar os serviços de análise e índice para uma colecção:

- 1. Expanda a colecção que pretende supervisionar e aceda ao painel Analisar e Indexar.
- 2. Se o analisador tiver sido parado e pretender iniciá-lo, faça clique em **Inicia**r.

Quando criar uma colecção pela primeira vez, inicie o analisador apenas depois de o pesquisador começar a pesquisar dados. Isto garante que o analisador terá dados para analisar e categorizar. A menos que efectue alterações nas regras de análise, pode deixar que o analisador funcione continuamente.

3. Se o analisador estiver em execução e pretender pará-lo, faça clique em Parar. Tem de parar e reiniciar o analisador quando faz alterações às regras de análise. Por exemplo, se alterar a configuração do analisador, tem de pará-lo e reiniciá-lo antes de as alterações se tornarem efectivas.

Se supervisionar uma colecção que não utilize uma cache de documentos para suportar a reconstrução do índice dos documentos na cache, também tem de voltar a pesquisar ou a importar os documentos para aplicar alterações de análise a documentos que já estão armazenados no índice.

4. Se os serviços de análise e indexação estiverem em execução e quiser ver informações de estado detalhadas sobre a actividade de análise e indexação, faça clique na área que pretende analisar, como, por exemplo **Detalhes**, **Processos Globais**, ou **Anotadores**.

Pode ver informações detalhadas sobre a actividade de análise e o estado actuais das construções do índice:

- Se foram abandonados documentos antes de serem analisados ou indexados, pode fazer clique no ícone associado para abrir uma página com mais detalhes que o possam ajudar a saber o motivo de serem abandonados.
- Se o analisador estiver configurado para detectar documentos de chegada e para os analisar imediatamente após terem sido adicionados a uma colecção, as estatísticas indicam que esta actividade está em vigor. Se não estiver em vigor, pode fazer clique em Configurar > Opções de análise e seleccione a opção para activar a detecção automática analisar os documentos de chegada.
- Se a função administrativa lhe permitir administrar processos para uma colecção, pode iniciar e parar os serviços de análise e de indexação. Também pode reorganizar o índice. A reorganização do índice desfragmenta o índice mas não altera os ficheiros indexados. Adicionalmente, a reorganização do índice não reduz imediatamente o tamanho do índice. O tamanho do índice é reduzido quando reinicia os serviços de indexação para a colecção.
- Se a empresa incluir vários servidores de processamento de documentos, pode ver o estado de cada servidor, e iniciar e parar a actividade em cada servidor.

- Parar e reiniciar uma criação de índice aplica as definições de configuração de análise e índice actuais a documentos novos. Aplicar as definições de análise e índice actuais a documentos indexados anteriormente depende se a colecção está configurada para utilizar uma cache de documento:
 - Se a colecção não utilizar uma cache de documento, pode aplicar as definições de análise e índice actuais para indexar documentos ao voltar a pesquisar ou a importar os documentos após a paragem ou o reinício da criação do índice.
 - Se a colecção utiliza uma cache de documento, não necessita de voltar a pesquisar ou a importar os documentos. Pode aplicar as definições de análise e índice actuais a documento indexados ao fazer clique em **Reiniciar uma construção de índice completa**. A reconstrução do índice não elimina o índice actual e constrói um novo. Em vez disso, reconstrói valores para os documentos no índice.
- Se a análise da ligação, geração de miniaturas ou detecção de documentos duplicados estiver activa para a colecção, pode supervisionar esses processos globais. Por exemplo, pode ver quantos documentos foram processados. Se tiver sido configurada uma agenda para estas actividades, pode ver quando está agendado o processo para ser executado novamente ou pode alterar o agendamento.
- Pode reconstruir o índice de categorias para que sejam aplicadas novas regras e alterações a regras. Caso os utilizadores da aplicação tenham autorização para iniciar um processo de reconstrução do índice de categorias, pode supervisionar o progresso do processo. Pode reconstruir o índice de categorias sempre que efectuar alterações à árvore de categorias na consola de administração ou quando utilizadores adicionam regras a categorias a partir da aplicação.
- Caso a colecção esteja configurada para categorizar documentos com base na análise de palavras extraídas de agrupamentos de documentos no índice, pode supervisionar o progresso do processo de agrupamento de documentos. Esta tarefa é executada em segundo plano para categorizar todos os documentos do índice. Caso um documento não corresponda aos critérios especificados para implementar a análise de agrupamentos, o documento não é categorizado. Quando se adicionam novos documentos, a categorização de novos documentos é realizada de forma assíncrona. Caso não existam documento novos, a tarefa é suspensa.

Visualizar detalhes sobre um URI

Pode visualizar informações detalhadas sobre um URI. Pode ver informações actuais e históricas sobre a forma como o documento representado por este URI é pesquisado, indexado e procurado.

Antes de submeter um pedido para visualizar um relatório de URI ou enviar um relatório para um endereço de correio electrónico, certifique-se de que o componente do qual pretende receber informações está activo. Por exemplo, para ver detalhes sobre o modo como um documento é pesquisado, indexado e procurado, certifique-se de que o pesquisador da Web, o servidor de indexação e os servidores de procura estão em execução.

A recolha de informações acerca de um URI é um processo moroso. Pode escolher uma opção para visualizar as informações que pedir e aguardar a respectiva apresentação. Uma opção mais eficiente é enviar o relatório para um endereço de correio electrónico especificado. Antes de poder receber correio electrónico, também tem de se assegurar de que as informações sobre o servidor de correio electrónico estão configuradas. Estas informações são especificadas ao configurar as opções do correio electrónico na página Registo da vista Sistema.

O servidor de indexação e os servidores de procura podem fornecer informações sobre todos os URIs, tais como se um URI se encontra no índice e se foi copiado para os servidores de procura. Para visualizar informações sobre como um documento foi pesquisado, tem de especificar o URI para um documento que tenha sido pesquisado por um pesquisador Web.

Para ver detalhes sobre um URI:

- 1. Na vista Colecções, expanda a colecção que pretende supervisionar.
- 2. No painel Analisar e Indexar, faça clique em Configurar > Detalhes URI.
- 3. Na página Detalhes URI, escreva o URI cujas informações pretende visualizar.
- 4. Seleccione as caixas de verificação correspondentes ao tipo de informações que pretende visualizar:
 - Detalhes do pesquisador (disponível apenas para pesquisadores da Web) Seleccione esta caixa de verificação para ver informações sobre como um documento foi pesquisado por um pesquisador da Web e

informações sobre o respectivo estado actual no espaço de pesquisa.

Detalhes do índice

Seleccione esta caixa de verificação para verificar se um documento foi indexado e copiado para os servidores de procura.

Detalhes da procura

Seleccione esta caixa de verificação para ver informações sobre como o documento pode ser pesquisado e se está disponível para pesquisa.

- 5. Para aguardar que o relatório seja apresentado, faça clique em **Visualizar relatório**.
- 6. Para enviar o relatório para um endereço de correio electrónico, de modo a que o possa visualizar posteriormente, faça clique em **Enviar relatório**.
 - a. Na página Enviar um Relatório de URI Detalhado, escreva um endereço de correio electrónico para receber o relatório no campo **Endereço de correio** electrónico para notificação.
 - b. Faça clique em Enviar Relatório.

Referências relacionadas:

"Formatos de URI no índice" na página 164

Supervisionar documentos abandonados

Quando supervisionar serviços de análise e indexação, as estatísticas mostram se quaisquer documentos foram abandonados por não ser possível analisar ou indexá-los. Pode ver detalhes para saber o motivo de os documentos terem sido abandonados.

Os documentos podem ser abandonados no sistema em várias etapas do processamento. Pode supervisionar uma colecção para ver quando um documento é largado e os problemas que implicaram que fosse largado.Se o analisador encontrar um erro que impeça o documento de ser analisado, é registada uma mensagem com um código da razão sobre o documento desactivado. Este tipo de erro não faz com que as versões mais antigas do documento sejam removidas do índice. Os documentos podem ser abandonados durante as etapas de indexação e estas informações também são registadas. Por exemplo, é possível eliminar explicitamente URIs e padrões de URI. Um documento poderá ter sido pesquisado por um pesquisador posteriormente eliminado. O documento origem poderá já não existir (é associado ao documento um código HTTP negativo) ou o código HTTP associado ao documento poderá ser desconhecido. Os documentos também podem ser abandonados se as informações de classificação estiverem em falta para um documento que requeira uma análise global.

Se tiver conhecimento de que um documento foi pesquisado, mas o documento não aparece no índice, pode usar a consola de administração para rastrear o fluxo do documento através do sistema. As mensagens detalhadas podem mostrar quando, onde e porque é que o documento foi largado. Por exemplo, poderá ficar a saber que o documento foi largado de forma inesperada durante a análise global ou que um administrador removeu o URI do índice.

Ao obter informações sobre documentos abandonados, poderá ser capaz de resolver o problema para assegurar que o documento é adicionado à colecção. Por exemplo, se um documento estiver num tipo de documento não suportado, poderá ter de configurar o extractor de texto para adicionar o suporte desse tipo de documento.

Para rever informações sobre documentos abandonados numa colecção da pesquisa:

- Expanda a colecção que pretende supervisionar. No painel Analisar e Indexar, certifique-se de que os serviços do analisador e de índice estão a ser executados e faça clique em Detalhes. O campo Número de documentos largados mostra se foram largados alguns documentos antes de poderem ser analisados ou indexados.
- 2. Para saber por que motivo foram largados os documentos, faça clique no ícone para ver detalhes sobre documentos largados.
- **3**. Na página Estado de Erro do Documento Abandonado, especifique o padrão de URI ou URI para os documentos sobre os quais pretende saber informações.
- 4. Faça clique em **Aplicar** para ver uma lista de documentos que correspondem ao padrão de URI ou URI que especificou. Pode especificar o número de resultados que pretende ver por página e utilizar ícones de setas para mover entre várias páginas de resultados.

Os resultados mostram códigos de erros que correspondem a motivos para os documentos terem sido abandonados. Também são facultadas explicações para os códigos de erro. Por exemplo, o documento pode ser um tipo de documento não suportado ou pode estar danificado ou protegido por palavra-passe.

5. Para ver mais informações sobre o motivo de um determinado documento ser largado, faça clique no ícone **Detalhes** correspondente. A página Mensagens de Erro de Documentos Largado mostra mensagens de erro que descrevem o problema. Reveja as mensagens para obter informações que podem ajudar a resolver o problema, caso tencione incluir o documento largado na colecção.

Supervisionar os servidores de procura

Pode visualizar informações de estado detalhadas sobre a actividade do servidor de procura para uma colecção específica ou para os servidores de procura em todo o sistema.

Para supervisionar os servidores de procura:

Utilize os seguintes métodos para iniciar, parar ou supervisionar os servidores de procura para uma colecção específica ou para todos os servidores de procura no sistema:

 Para parar e iniciar os servidores de procura para uma única colecção, abra a vista Colecções e expanda a colecção que pretende supervisionar. No painel Procura ou no painel Análise, pode iniciar e parar múltiplos servidores de procura para esta colecção ou todos os servidores de procura de uma vez. Pode também fazer clique em **Detalhes** para iniciar e parar servidores de procura individuais.

Por exemplo, ser activar ou desactivar a cache de procuras, fizer alterações ao tamanho da cache de procuras, ou às ligações rápidas, tem de parar e reiniciar os servidores de procura para que as alterações se tornem efectivas.

• Para supervisionar todos os servidores de procura no sistema, abra a vista Sistema. Na página Procurar, pode ver as estatísticas do processamento de consultas por colecção, iniciar servidores de procura que estejam parados, bem como parar servidores de procura que estejam em execução.

Pode também fazer clique em **Estatísticas de consulta** para ver estatísticas detalhadas acerca do processamento da consulta para colecções específicas ou para todas as colecções. Os gráficos de linhas, gráficos de barras e as tabelas estatísticas permitem-lhe supervisionar o desempenho do processamento de consultas durante um período de tempo especificado pelo utilizador.

Pode ver estatísticas das consultas mais frequentes e de consultas que não tenham devolvido resultados, bem como informações detalhadas sobre consultas individuais, tais como termos de consulta, quem fez a consulta e quando foi feita a consulta. Para uma análise estatística mais aprofundada, pode exportar as estatísticas da consulta como um ficheiro de valor separado por vírgula (CSV), que pode depois importar para outra aplicação.

• Para supervisionar todos os servidores de procura no sistema, abra a vista Sistema. Na página Configuração do servidor, pode iniciar e parar servidores de procura individuais ou seleccionar vários servidores para que iniciem ou parem ao mesmo tempo. Caso um servidor de procura esteja em execução, pode ver estatísticas da consulta detalhadas acerca dos pedidos processados por esse servidor.

Ver estatísticas de consultas detalhadas

Quando supervisionar os servidores de procura para uma colecção, pode ver estatísticas detalhadas sobre processamento de consulta actual e anterior.

As informações nesta página dividem-se em gráficos que pode configurar de forma independente. Para focar nas estatísticas de interesse, pode fazer clique nos ícones de controlo da janela para minimizar e maximizar cada gráfico conforme seja necessário:

- O Gráfico de Linhas permite-lhe supervisionar o rendimento do processamento da consulta num período de tempo que especificar. Pode ver uma linha de tempo que mostra quantas consultas foram processadas e quantos utilizadores submeteram consultas ao longo do tempo. Também pode ver linhas cronológicas que mostram a média e os tempos de resposta mais lentos.
- O gráfico Consultas Populares apresenta gráficos de barras que facultam informações sobre as consultas submetidas com mais frequência e informações sobre consultas que não devolveram resultados. As consultas mais populares e as consultas que não devolveram resultados com mais frequência, surgem no topo dos gráficos.

Ao visualizar as consultas mais frequentes, pode identificar candidatos para ligações rápidas. Ao criar ligações rápidas, pode melhorar a qualidade da procura para muitos utilizadores. Pode assegura que os elementos muito importantes são sempre devolvidos nos resultados da procura.

Também pode achar conveniente criar ligações aos recursos que respondem às consultas do portal da empresa. Por exemplo, se os utilizadores procurarem frequentemente informações sobre contas de despesas, inclua uma ligação à página que explica os procedimentos de contas de despesas na página inicial da intranet.

O gráfico Histórico da Consulta faculta estatísticas detalhadas sobre o
processamento da consulta, como ternos da consulta, o ID do utilizador que
submeteu a consulta e quanto tempo foi necessário para avaliar a consulta e
devolver uma resposta. Se quiser efectuar uma análise estatística adicional numa
folha de cálculo ou outra aplicação, pode exportar as estatísticas da consulta
para um ficheiro de valor separado por vírgula (CSV).

Pode ordenar as estatísticas de formas diferentes. Por exemplo, poderá ordenar as estatísticas por marca de tempo para que possa ver informações sobre as consultas mais recentes. Ao visualizar as consultas mais recentes, pode identificar as tendências actuais e as situações urgentes na organização. Esse aumento de interesse pode indicar que é necessária uma ligação rápida para esse tópico e que é necessário torná-lo disponível para os utilizadores de outras formas (como facultar uma ligação num portal empresarial).

Para ver informações detalhadas sobre o processamento de consultas:

- 1. Especifique as suas preferências para como as consultas serão registadas na base de dados de estatísticas da consulta:
 - a. Faça clique em Sistema para abrir a vista Sistema.
 - b. Seleccione a página Pesquisar e faça clique em **Configurar a base de dados da estatística da consulta**.
 - **c**. Abra a página Base de Dados de Estatística da Consulta, especifique as suas preferências para registar consultas que os utilizadores submeterem. Por exemplo, pode especificar por quanto tempo as consultas serão retidas na base de dados e a frequência com que as actualizações à base de dados serão consolidadas quando os utilizadores submeterem novas consultas.
- 2. Supervisione estatísticas da consulta para uma única colecção:
 - a. Na vista Colecções, expanda a colecção que pretende supervisionar.
 - b. No painel Procura ou no painel Análise, faça clique no ícone para ver as estatísticas da consulta detalhadas.
- 3. Supervisione estatísticas da consulta para todas as colecções no sistema:
 - a. Faça clique em **Sistema** para abrir a vista Sistema.
 - b. Abra a página Procura e faça clique no ícone para visualizar as estatísticas da consulta detalhadas.
- 4. Seleccione o período de tempo que pretende supervisionar, como hoje, esta semana, ou este mês, ou um intervalo de data específico. Quando seleccionar um período de tempo, o gráfico de linhas é renovado automaticamente para mostrar dados para o período de tempo especificado.
- **5**. Faça clique em **Opções de Apresentação** para especificar as suas preferências para visualizar e renovar as estatísticas. Por exemplo:
 - Especifique se pretende que o ecrã seja renovado com novos dados da consulta automaticamente e, se assim for, a frequência com que pretende que os dados sejam renovados.
 - Para um gráfico de linhas, pode:

- Optar por ver uma cronologia que mostra o número de consultas que foram processadas durante o período de tempo que especificar ou uma cronologia que mostra quantos utilizadores submeteram consultas durante esse período de tempo. Também pode ver estatísticas para o tempo de reposta de consulta médio, em milissegundos, e os tempos de resposta da consulta mais lentos, em milissegundos.
- Seleccione a escala de tempo para mostrar no gráfico cronológico. Por exemplo, ao ver estatísticas para o dia actual, poderá escolher mostrar estatísticas por hora. Ao ver estatísticas durante uma semana ou um mês, pode escolher ver as estatísticas por dia.
- Seleccione os servidores de pesquisa em que pretende ver as estatísticas. Pode seleccionar servidores individuais em todos os servidores da pesquisa no sistema.
- Para as consultas populares e os gráficos de histórico das consultas, especifique o número de consultas para que pretende ver estatísticas.
- 6. Opcional: Faça clique em **Refinar Resultados** se quiser limitar as estatísticas a informações sobre consultas que foram submetidas por um determinado utilizador. Especifique o ID que o utilizador usa para iniciar a sessão na aplicação de procura empresarial.
- 7. Opcional: Faça clique em **Colecção** para limitar as estatísticas a uma ou mais colecções ou ver estatísticas para todas as colecções no sistema.
- 8. Opcional: No gráfico **Histórico da Consulta**, faça clique no cabeçalho da coluna em que pretende ordenar os dados. Pode ordenar os dados por ordem ascendente ou descendente segundo os IDs de utilizador dos utilizadores que submeteram consultas, a data e hora em que as consultas foram submetidas, o número de resultados em reposta à consulta ou a quantidade de tempo que demorou a avaliar a consulta e a devolver resultados. Faça clique nos cabeçalhos da coluna novamente para ordenar os dados pela ordem inversa.
- 9. Opcional: No gráfico Histórico da Consulta, faça clique em Guardar como ficheiro CSV para exportar as estatísticas da consulta para que possa analisá-las fora da consola de administração, como com uma folha de cálculo ou outra aplicação. Pode optar por exportar todos os dados no histórico da consulta ou as informações sobre consultas que foram processadas durante um período de tempo específico. Também pode optar por descarregar o ficheiro ou especificar um caminho no servidor onde pretende guardar o ficheiro.

Supervisionar recursos de análise

Ao fazer alterações à árvore de facetas, a dicionários de utilizador, ou a regras de análise de texto personalizadas numa colecção de análise de conteúdo, terá de reimplementar os recursos de análise de texto para que as alterações entrem em vigor. Caso a colecção utilize o índice opcional de facetas para armazenar dados de facetas, pode também iniciar e supervisionar as construções de índice de facetas.

Ao reimplementar os recursos:

- Definições de árvore da faceta novas e alteradas ficam disponíveis na prospecção de análise de conteúdo.
- São utilizados dicionários e regras de análise de texto novos e alterados para analisar documentos novos. Caso pretenda aplicar os dicionários e regras novos ou alterados a documentos que já se encontrem no índice, terá de reconstruir o índice.

Restrição: Ao reimplementar recursos, o IBM Content Analytics with Enterprise Search tem de poder gravar em determinadas pastas. O sistema operativo

Windows não permite substituir directórios abertos. Para garantir que a implementação de recursos é concluída com sucesso no Windows, feche todas as pastas no IBM Content Analytics with Enterprise Search antes de iniciar a tarefa de reimplementação dos recursos.

Para reimplementar os recursos de análise de texto para uma colecção:

- Na vista Colecções, expanda a colecção de análise de conteúdos que pretende supervisionar. No painel Analisar e Indexar, faça clique em Recursos de Análise.
- 2. Para propagar as alterações que tenha efectuado à árvore de facetas, a dicionários de utilizador, ou a regras de análise de texto do índice, faça clique no ícone para iniciar a tarefa de implementação do recurso. Utilize a barra de estado para supervisionar o progresso da tarefa.
- Caso esta colecção utilize um índice da faceta, faça clique no ícone para reconstruir o índice da faceta opcional. Utilize a barra de estado para supervisionar o progresso da construção do índice.

Conceitos relacionados:

"Índice da faceta opcional para colecções de análise de conteúdo" na página 261

Tarefas relacionadas:

"Configurar a árvore da faceta para colecções de análise de conteúdo" na página 251

"Configurar facetas de intervalo decimal para uma colecção de análise de conteúdos" na página 252

"Configurar facetas de intervalo de data para uma colecção de análise de conteúdos" na página 255

"Configurar dicionários do utilizador para colecções de análise de conteúdo" na página 257

"Configurar ficheiros de regras para colecções de análise de conteúdo" na página 260

"Configurar análise de postura para colecções de análise de conteúdo" na página 262

Supervisionar Data Listener

Supervisione o Data Listener para ver o respectivo estado e detalhes sobre a actividade da aplicação Data Listener de cliente.

Importante: O Data Listener não será suportado em edições futuras. Utilize as APIs de procura e indexação (SIAPI - search and index) em vez das APIs Data Listener para desenvolver aplicações cliente. As informações seguintes são fornecidas para utilizadores que tenham criado previamente aplicações de Data Listener.

Para supervisionar o Data Listener:

- 1. Faça clique em Sistema para abrir a vista Sistema.
- 2. Na página Data Listener, visualize os ícones de estado para ver se o Data Listener está activo ou parado.
- 3. Se o Data Listener estiver em execução e se quiser ver informações de estado detalhadas sobre a actividade da aplicação de cliente, faça clique em Detalhes. Os ícones de estado da página Detalhes do Data Listener indicam se o Data Listener está em execução ou parado. As estatísticas mostram quantos pedidos estão a aguardar processamento, o estado actual de cada módulo em

funcionamento nos pedidos de aplicação de cliente e quantos módulos estão activos para um estado de módulo específico.

4. Se alterar o número de porta do Data Listener, ou se mudar o número máximo de documentos que podem ser mantidos no armazenamento temporário, clique em **Reiniciar**.

O Data Listener é iniciado ao mesmo tempo que o sistema IBM Content Analytics with Enterprise Search. Só é necessário reiniciar o Data Listener se alterar uma ou ambas estas opções de configuração.

Tarefas relacionadas:

"Configurar suporte para aplicações de Data Listener" na página 160

Ficheiros de registo e alertas

Pode escolher os tipos de mensagens que quiser registar numa colecção e num sistema e especificar opções para criar e ver os ficheiros de registo, receber alertas e mensagens de correio electrónico sobre as mesmas.

Durante o funcionamento normal, os componentes escrevem mensagens de registo num ficheiro de registo comum. Este ficheiro de registo está no directório ES_NODE_ROOT/logs, no servidor do controlador. Pode utilizar a consola de administração para ver estes dados de registo comuns.

Se ocorrer um problema, como uma falha de comunicação de rede, os componentes escrevem mensagens de registo num directório de registos no servidor onde o componente está instalado. Para ver esses ficheiros de registo locais, utilize um visualizador de ficheiros nesse computador, como o utilitário tail num sistema UNIX. Não pode utilizar a consola de administração para visualizar estes tipos de ficheiros de registo.

Ao configurar ficheiros de registo, pode escolher os tipos de mensagens que pretende registar (como, por exemplo, mensagens de erro ou aviso), pode especificar a frequência com que os ficheiros de registo antigos devem ser eliminados para deixarem mais espaço para novos ficheiros de registo, pode especificar um tamanho máximo para os ficheiros de registo e, pode seleccionar o idioma das mensagens. Pode também especificar opções para receber mensagens de correio electrónico sempre que ocorrerem certos eventos ou sempre que certas mensagens ou tipos de mensagens são registados.

Ao supervisionar ficheiros de registo, pode escolher qual o ficheiro de registo que pretende abrir. Pode filtrar o conteúdo do ficheiro de registo para que veja apenas mensagens de um determinado nível de gravidade (como apenas mensagens de erro) ou mensagens que foram produzidas por uma sessão específica. Ao visualizar um ficheiro de registo, pode ver detalhes sobre mensagens individuais. Por exemplo, pode querer ver o nome da função que produziu a mensagem e outras informações que possam ajudá-lo a tomar uma acção de correcção, se necessário.

Conceitos relacionados:

"Alertas"

Improvement Analytics with Enterprise Search

Tarefas relacionadas:

"Configurar ficheiros de registo" na página 448

"Visualizar ficheiros de registo" na página 451

"Visualizar ficheiros de registo utilizando a consola de administração" na página 453

Alertas

Pode configurar o sistema para escrever mensagens no ficheiro de registo sempre que detectar que ocorreram determinados eventos.

As mensagens que são activadas por eventos, chamados alertas, informam-no sobre as condições que pode pretender abordar, tais como um recurso que esteja a ficar sem espaço livre. Quando configurar alertas, especifique as condições que pretende que o sistema supervisione. Sempre que a condição ocorrer, o sistema escreve uma mensagem automaticamente no ficheiro de registo.

Se quer ser directamente notificado sobre uma condição, pode especificar opções para receber uma mensagem de correio electrónico sempre que uma das mensagens monitorizadas estiver registada.

Pode configurar alertas para eventos de nível de colecção e para eventos que ocorram ao nível de sistema. No nível de colecção, o sistema pode:

- Supervisionar o número de documentos que cada pesquisador pesquisa e emitir uma mensagem de alerta quando o número máximo permitido de documentos estiver quase a ser alcançado.
- Supervisionar o número de documentos a serem adicionados ao índice para as colecções e emitir uma mensagem de alerta quando o número máximo permitido de documentos estiver quase a ser alcançado.
- Informá-lo de quando o tempo necessário para responder a pedidos de procura estiver a exceder um limite especificado.

Ao nível do sistema, o sistema pode supervisionar o espaço em disco em cada servidor e emitir uma mensagem de alerta quando a quantidade de espaço em disco livre for baixa.

Conceitos relacionados:

"Ficheiros de registo e alertas" na página 445

Tarefas relacionadas:

"Configurar alertas ao nível da colecção"

"Configurar alertas ao nível do sistema" na página 447

Configurar alertas ao nível da colecção

Ao configurar alertas, pode assegurar que as mensagens são escritas num ficheiro de registo sempre que determinados eventos ao nível de colecção ocorrerem. Também pode receber correio electrónico sempre que forem registadas mensagens sobre estes eventos.

Para configurar alertas de nível de colecção:

- Na vista Colecções, localize a colecção que pretende configurar e faça clique em Acções > Registo > Configurar alertas.
- Se pretender que o sistema supervisione o número de documentos que cada pesquisador está a pesquisar, execute os seguintes passos:
 - a. Seleccione a caixa de verificação Quando o número de documentos pesquisados por qualquer pesquisador atingir uma percentagem do máximo permitido.
 - b. No campo Percentagem, especifique quando pretende que uma mensagem seja registada. Especifique este número como uma percentagem do número máximo de documentos que os pesquisadores podem pesquisar (o Número máximo de documentos que pesquisadores podem pesquisar é configurado quando se configura as propriedades do pesquisador). O valor predefinido é 90 por cento.

Devido a poder configurar pesquisadores diferentes, as mensagens separadas são registadas para cada pesquisador. Por exemplo, se utilizar o limiar de alerta predefinido, permita que um pesquisador DB2 pesquise 2,000,000 documentos e permita a um pesquisador Notes que pesquise 1,000,000 documentos, será registada uma mensagem quando o pesquisador DB2 pesquisa 1,800,000 documentos e outra mensagem será registada quando o pesquisador Notes pesquisar 900,000 documentos.

- **3**. Se pretender que o sistema supervisione o número de documentos que estão a ser adicionados ao sistema, execute os seguintes passos:
 - a. Seleccione a caixa de verificação **Quando o número de documentos na colecção atingir uma percentagem do tamanho calculado**.
 - b. No campo Percentagem, especifique quando pretende que uma mensagem seja registada. Especifique este número como uma percentagem da estimativa do número de documentos que espera que a colecção tenha. O valor predefinido é 85 por cento.
- 4. Se pretender que o sistema o avise quando o tempo necessário para responder aos pedidos da procura exceder um limite, execute os seguintes passos:
 - a. Seleccione a caixa de verificação **Quando o tempo de resposta excede um limite** .
 - b. No campo **Limite**, escreva o número de segundos que considera aceitável como tempo de resposta máximo da procura.

Quando este número é excedido, o sistema escreve uma mensagem de registo sobre o evento. Por exemplo, se mantiver o valor predefinido, o sistema cria uma mensagem de registo sempre que um servidor de procura tiver cinco segundos ou mais como média de resposta a pedidos de procura.

Os tempos de resposta típicos são inferiores a meio segundo. Médias superiores a um segundo podem indicar que o sistema operativo necessita de ser optimizado para obter melhor desempenho, ou, podem indicar que existe um problema nas definições da configuração do servidor de procura. Por exemplo, pode querer aumentar o tamanho do espaço alocado para a memória cache da procura.

Se pretender receber correio electrónico quando o sistema registar mensagens sobre estes eventos, faça clique em **Acções > Registo > Configurar opções de correio electrónico para mensagens** para que possa especificar em endereço de correio electrónico. Os IDs da mensagem para os alertas que activou são adicionados automaticamente à lista de IDs de mensagem para os quais se pretende enviar correio electrónico.

Antes de poder receber correio electrónico, deve certificar-se de que a informação sobre o seu servidor de correio SMTP está configurada. Para o fazer, faça clique em **Sistema**, abra a página Registo e depois faça clique em **Configurar opções de correio electrónico para mensagens**.

Conceitos relacionados:

"Alertas" na página 445

Tarefas relacionadas:

"Receber correio electrónico sobre mensagens registadas" na página 450

Configurar alertas ao nível do sistema

Ao configurar alertas, pode assegurar que as mensagens são escritas num ficheiro de registo sempre que determinados eventos ao nível de sistema ocorrerem. Também pode receber correio electrónico sempre que forem registadas mensagens sobre estes eventos.

Para configurar alertas de nível de sistema:

1. Na vista Sistema, abra a página Registo e faça clique em Configurar alertas.

- 2. Se pretender que o sistema supervisione a quantidade de espaço livre disponível nos servidores, seleccione a caixa de verificação Quando a quantidade de espaço disponível do sistema de ficheiros atingir uma percentagem do espaço total.
- **3**. No campo **Percentagem**, especifique quando pretende que o sistema o notifique sobre a pequena quantidade de espaço livre num servidor. Especifique este número como uma percentagem do espaço total do sistema de ficheiros. O valor predefinido é 80 por cento.

Se o sistema estiver configurado em vários servidores, o sistema cria uma mensagem de registo separada para cada servidor. Por exemplo, uma mensagem informa quando o espaço no servidor do pesquisador é reduzido; e mensagens separadas informam sobre as restrições de espaço nos servidores de controlador e de procura.

Se pretender receber correio electrónico sempre que o sistema registar uma mensagem sobre este evento, abra a página Registo e, em seguida, faça clique em **Configurar opções de correio electrónico para mensagens** de forma a poder especificar um endereço de correio electrónico e informações sobre o servidor de correio electrónico SMTP.

Conceitos relacionados:

"Alertas" na página 445

Tarefas relacionadas:

"Receber correio electrónico sobre mensagens registadas" na página 450

Configurar ficheiros de registo

Pode especificar os tipos de mensagens que pretende registar e especificar opções para criação de ficheiros de registo.

Para evitar que os ficheiros de registo ocupem demasiado espaço, o sistema roda ficheiros de registo e inicia um ficheiro de registo novo sempre que a data actual muda. Se um ficheiro de registo aumentar até atingir o tamanho máximo permitido e a data não tiver mudado, o sistema cria um ficheiro de registo novo. Quando é atingido o número máximo de ficheiros de registo, o ficheiro de registo mais antigo é eliminado para poder ser criado um novo.

Para receber correio electrónico sobre mensagens de registo, deve primeiro especificar informação sobre o modo como o correio electrónico deve ser entregue. Em seguida, deve especificar sobre que mensagens quer receber uma mensagem de correio electrónico.

Para configurar ficheiros de registo:

- 1. Execute uma das seguintes acções:
 - Para especificar opções para a criação e rotação de ficheiros de registo de nível de colecção, abra a vista Colecções, localize a colecção que pretende configurar e faça clique em Acções > Registo > Configurar opções de ficheiro de registo.
 - Para especificar opções para a criação e rotação de ficheiros de registo de nível de sistema, abra a vista Sistema. Na página Registo, faça clique em Configurar opções de ficheiros de registo.
- 2. Nos campos **Tipo de informação a registar** e **Tipo de informação a rastrear para auditoria**, seleccione os tipos de mensagem que pretende registar:

Mensagens de erro apenas

As mensagens de erro indicam que ocorreu uma situação indesejável ou um comportamento inesperado e que o processo não pode continuar. Tem de agir para corrigir o problema.

Mensagens de erro e de aviso

As mensagens de aviso indicam um possível conflito ou inconsistência, mas não param os processos. Esta opção é a predefinida.

Todas as mensagens

As mensagens informativas fornecem informações gerais sobre o sistema ou tarefa actual e não necessitam de qualquer acção de correcção.

Importante: Seleccionar esta opção pode afectar negativamente o desempenho do sistema. Registe todas as mensagens apenas quando necessitar de solucionar problemas ou se lhe for pedido que o faça pelo Suporte para Software IBM.

 No campo Tamanho máximo de cada ficheiro de registo, escreva o número máximo de megabytes para cada ficheiro de registo. O valor predefinido e máximo é 16 MB.

Quando o ficheiro de registo crescer até esse tamanho, é criado um ficheiro de registo novo até ao número máximo de ficheiros de registo permitido. Mantendo os ficheiros de registo relativamente pequenos, pode visualizá-los mais eficientemente.

4. No campo **Tamanho máximo de ficheiros de registo**, escreva o número máximo de ficheiros de registo que pretende criar. O valor predefinido é 10.

Se pretender assegurar que as mensagens mais antigas estejam disponíveis para revisão, aumente este valor. Se estiver mais interessado em mensagens recentes e não necessitar de manter um histórico longo das actividades, reduza este valor.

- 5. No campo **Local predefinido**, seleccione o idioma que pretende utilizar para registar mensagens. O valor predefinido é inglês.
- 6. Para que as alterações se tornem efectivas, reinicie o sistema IBM Content Analytics with Enterprise Search:

esdmin system stopall esdmin system startall

Conceitos relacionados:

"Ficheiros de registo e alertas" na página 445

Tarefas relacionadas:

"Configurar informações de servidor SMTP"

"Receber correio electrónico sobre mensagens registadas" na página 450

Configurar informações de servidor SMTP

Antes de poder receber mensagens de correio electrónico sobre actividades de sistema, tem de configurar informações sobre o servidor Simple Mail Transfer Protocol (SMTP).

Várias funções administrativas permitem receber correio electrónico. Antes de poder receber correio electrónico de qualquer uma destas funções, tem de especificar informações sobre o servidor de SMTP:

- Se configurar alertas de nível de colecção ou alertas de nível de sistema, pode receber correio electrónico sempre que essas mensagens forem registadas. Pode também receber correio electrónico quando são registadas outras mensagens, e não apenas as mensagens que forem accionadas por eventos supervisionados.
- Se supervisionar um pesquisador da Web e especificar que pretende criar relatórios de histórico do pesquisador Web, pode ser notificado por correio electrónico após a criação de um relatório.

Para configurar informações sobre o servidor SMTP:

- 1. Abra a vista Sistema. Na página Registo, faça clique em **Configurar opções de correio electrónico para mensagens**.
- 2. Na página Opções de Correio Electrónico para Mensagens do Sistema, no campo Servidor SMTP de correio electrónico para enviar para enviar correio electrónico, escreva o nome do sistema central totalmente qualificado ou o endereço de IP do servidor SMTP que pretende utilizar. O sistema utiliza este servidor para enviar correio electrónico para os endereços que especificar.
- No campo Frequência de verificação de correio electrónico, especifique a frequência com que pretende que o sistema verifique se existem mensagens elegíveis e envie correio electrónico sobre as mesmas.

O sistema combina todas as mensagens para um endereço de correio electrónico específico numa mensagem e envia essa mensagem com a frequência especificada.

Tarefas relacionadas:

"Configurar ficheiros de registo" na página 448

Receber correio electrónico sobre mensagens registadas

Pode especificar opções para receber correio electrónico sempre que determinadas mensagens, ou determinado tipo de mensagens, sejam registadas.

Antes de poder receber correio electrónico, é necessário configurar previamente informação sobre o servidor de Simple Mail Transfer Protocol (SMTP) para que o correio electrónico possa ser entregue.

Ao configurar alertas, pode escolher uma opção para registar mensagens quando determinadas mensagens ocorrerem. Se activar estas opções, pode, em seguida, configurar opções para receber correio electrónico automaticamente sempre que essas mensagens sejam registadas. Também pode especificar opções para receber correio electrónico quando forem registadas outras mensagens, não apenas quando forem registadas por eventos.

Para configurar opções de correio electrónico para mensagens:

- 1. Se pretender receber correio electrónico sobre mensagens de sistema:
 - a. Faça clique em Sistema para abrir a vista Sistema.
 - b. Abra a página Registo e faça clique em **Configurar opções de correio** electrónico para mensagens.
 - c. Na página Opções de correio electrónico para mensagens de sistema, seleccione a caixa de verificação Enviar correio electrónico sobre mensagens de nível de sistema.
 - d. No campo **Endereço de correio electrónico para receber correio electrónico**, escreva um ou mais endereços de correio electrónico. Em regra geral, um administrador deve receber informações sobre as mensagens do sistema.

Separe cada endereço com uma vírgula. Por exemplo:

steinbeck@us.ibm.com, yeats@ireland.ibm.com, dante@it.ibm.com.

- e. Se pretender receber correio electrónico sobre todas as mensagens de erro registadas, seleccione a caixa de verificação **Enviar correio electrónico sobre todas as mensagens de erro**.
- f. Se pretender receber correio electrónico apenas quando são registadas determinadas mensagens de nível de sistema, escreva os IDs das mensagens para essas mensagens na área Enviar correio electrónico determinadas mensagens. Escreva um ID de mensagem por cada linha. Por exemplo: FFQC4819E FFQ00005E
- 2. Se pretender receber correio electrónico sobre mensagens para uma colecção:
 - a. Faça clique em Colecções para abrir a vista Colecções.
 - b. Para a colecção que pretende configurar, faça clique em Acções > Registo > Configurar opções de correio electrónico para mensagens.
 - c. Na página Opções de correio electrónico para mensagens de colecção, seleccione a caixa de verificação Enviar correio electrónico sobre mensagens de nível de colecção.
 - d. No campo **Endereço de correio electrónico para receber correio electrónico**, escreva um ou mais endereços de correio electrónico. Tipicamente, um administrador de colecção deve receber informações sobre mensagens de nível de colecção.

Separe cada endereço com uma vírgula. Por exemplo:

steinbeck@us.ibm.com, yeats@ireland.ibm.com, dante@it.ibm.com.

- e. Se pretender receber correio electrónico sobre todas as mensagens de erro registadas, seleccione a caixa de verificação Enviar correio electrónico sobre todas as mensagens de erro.
- f. Se pretender receber correio electrónico apenas quando são registadas determinadas mensagens de nível de colecção, escreva os IDs das mensagens para essas mensagens na área Enviar correio electrónico sobre determinadas mensagens. Escreva um ID de mensagem por cada linha. Por exemplo: FFQC4819E
 - FFQ00005E

Conceitos relacionados:

Mensagens para o IBM Content Analytics with Enterprise Search

Tarefas relacionadas:

"Configurar ficheiros de registo" na página 448

"Configurar alertas ao nível da colecção" na página 446

"Configurar alertas ao nível do sistema" na página 447

Visualizar ficheiros de registo

Pode visualizar mensagens de registo que os componentes de colecção e sistema escrevem num ficheiro de registo comum. Também pode especificar filtros para ver mensagens de um nível de gravidade específico e mensagens de sessões de sistema específicas.

Para ver ficheiros de registo:

1. Execute uma das seguintes acções:

- Para ver os ficheiros de registo de uma única colecção, abra a vista Colecções, localize a colecção que pretende supervisionar e faça clique em Acções > Registo > Ver ficheiros de registo.
- Para ver ficheiros de registo de nível de sistema, abra a vista Sistema e, de seguida, abra a página Registar.
- 2. No campo **Ficheiro de registo**, seleccione o ficheiro de registo que pretende visualizar.

O nome de cada ficheiro de registo inclui o tipo de ficheiro de registo (tal como um nome de colecção ou de sistema) e a data em que o ficheiro foi criado. Se for criado mais do que um ficheiro de registo do mesmo tipo na mesma data, um sufixo numérico indica a ordem pela qual o ficheiro foi criado. Por exemplo:

```
log_file_type_20060526.log (contains the most recent entries on this date)
log_file_type_20060526.log.1
log_file_type_20060526.log.2 (contains the oldest entries on this date)
log_file_type_20060525.log (contains the most recent entries on this date)
log_file_type_20060525.log.1
log_file_type_20060525.log.2
log_file_type_20060525.log.3 (contains oldest entries on this date)
```

- **3.** Para visualizar apenas mensagens de níveis de segurança específicos, seleccione as caixas de verificação adequadas no campo **Gravidade**. Por exemplo, para ver apenas as mensagens que exigem acção correctora, seleccione a caixa de verificação para **Erro** e limpe as caixas de verificação para mensagens de aviso e de informação.
- 4. Para visualizar apenas mensagens de sessões específicas, seleccione as caixas de verificação adequadas no campo Sessão. Por exemplo, para ver apenas as mensagens emitidas durante uma sessão de análise, seleccione as caixas de verificação relacionadas com a análise e limpe todas as outras caixas de verificação.
- 5. Faça clique em Visualizar registo.

Para cada mensagem na página Conteúdo de Ficheiro de Registo, pode ver a data e hora em que a mensagem foi emitida, o nível de gravidade da mensagem, o nome da sessão que emitiu a mensagem e o ID da mensagem e texto de erro.

Especifique quantas mensagens devem ser apresentadas em cada página. Pode fazer clique sobre botões para aceder à primeira página, à última página, à página anterior ou à seguinte do ficheiro de registo. Também pode especificar um número de página e aceder directamente a essa página.

6. Para ver informação mais detalhada sobre a mensagem, faça clique em **Detalhes**.

Na página Registar Detalhes de Mensagens, pode ver o nome do sistema central do servidor onde a mensagem ocorreu, o nome do ficheiro que produziu o erro, o nome da função e número da linha onde ocorreu o erro, bem como o ID do processo e o ID do módulo.

Pode fazer clique nos botões para se deslocar para a mensagem seguinte ou para a anterior no ficheiro de registo.

Conceitos relacionados:

"Ficheiros de registo e alertas" na página 445

Tarefas relacionadas:

"Visualizar ficheiros de registo utilizando a consola de administração"

Visualizar ficheiros de registo utilizando a consola de administração

Em alguns casos, poderá optar por executar scripts para visualizar ficheiros de registo e auditar ficheiros sem utilizar a consola de administração. Por exemplo, se os ficheiros de registo contêm cadeias com caracteres de DBCS, os caracteres podem aparecer alterados na consola. Ao executar o script, pode especificar que o conteúdo será apresentado no formato UTF-8.

Pode utilizar o comando **esviewlog** para ver o conteúdo dos ficheiros de registo no directório ES_NODE_ROOT/logs e ficheiros de auditoria no directório ES_NODE_ROOT/logs/audit.

Este script é particularmente útil para ver ficheiros que contenham caracteres DBCS, como ficheiros de registo em japonês ou chinês.

Quando executar o script, especifique o caminho do ficheiro de registo que pretende visualizar e o caminho para um ficheiro onde serão escritos os dados de saída formatados. Também é possível especificar a opção **-e UTF-8**, que utiliza a codificação UTF-8 para formatar os dados de saída.

Se omitir a opção **-e UTF-8**, serão utilizado o local do sistema e a codificação predefinida do sistema. Por exemplo, se utilizar o Microsoft Windows Server 2003 num ambiente japonês, a codificação de caracteres é Shift-JIS. Se o ficheiro de registo contém caracteres japoneses e chineses, os caracteres chineses não serão apresentados num formato legível.

Para ver ficheiros de registo ou de auditoria sem utilizar a consola de administração:

- 1. Inicie sessão como administrador predefinido do IBM Content Analytics with Enterprise Search. Para uma instalação de servidores múltiplos, inicie sessão no servidor do controlador.
- 2. Altere o directório ES_NODE_ROOT/logs ou o directório ES_NODE_ROOT/logs/audit.
- 3. Introduza o comando seguinte para visualizar um ficheiro de registo ou um ficheiro de auditoria e utilize a codificação UTF-8 para formatar a saída. O nome do ficheiro está no formato *col_ID_data* ou system_*data*, no qual *col_ID* é o ID da colecção e *data* é a data do ficheiro de registo.

```
AIX ou Linux
```

esviewlog.sh -e UTF-8 -f col_5255_20090512.log -o
col_5255_utf8.log

Windows

esviewlog.bat -e UTF-8 -f col_5255_20090512.log -o
col 5255 utf8.log

Conceitos relacionados:

"Ficheiros de registo e alertas" na página 445

Tarefas relacionadas:

"Visualizar ficheiros de registo" na página 451

Efectuar cópia de segurança e restaurar o sistema

A cópia de segurança e o restauro de scripts permitem-lhe efectuar cópia de segurança e restauro do sistema IBM Content Analytics with Enterprise Search.

De que é que os scripts fazem cópia de segurança

Os scripts fazem a cópia de segurança e o restauro dos seguintes ficheiros:

- Ficheiros de configuração do directório ES_NODE_ROOT/master_config
- Bases de dados do sistema, como bases de dados para dados pesquisados, gestão de identidade e assim por diante.
- Todos os ficheiros do directório ES_NODE_ROOT/data
- Ficheiros índice para colecções que estão configuradas com directórios de dados não predefinidos

Restrição: Não utilize comandos do sistema operativo para copiar ficheiros entre o servidor IBM Content Analytics with Enterprise Search de nível de produção e um servidor de cópia de segurança. O sistema não vai funcionar como esperado porque as informações do nome do sistema central e de endereço de IP, exigidas pelo IBM Content Analytics with Enterprise Search, não são facultadas através de mecanismos de cópia de ficheiro. Além disso, excepto se o sistema IBM Content Analytics with Enterprise Search de produção estiver completamente parado antes de copiar os ficheiro, a base de dados interna pode ser copiada num estado inconsistente, que pode impedir que o IBM Content Analytics with Enterprise Search inicie no servidor destino.

Estrutura do directório de segurança

O script da cópia de segurança cria os seguintes subdirectórios sob um directório especificado quando executa o script. O ID do administrador deverá ter permissão para escrever no directório que especificar.

master_config

Contém os ficheiros de configuração do directório ES_NODE_ROOT/ master_config

base de dados

Contém os ficheiros de base de dados do servidor do pesquisador

dados Contém os ficheiros de índice do servidor do controlador

Directrizes de utilização

- Pode fazer a cópia de segurança dos dados a partir de um computador e restaurá-los para outro computador. No entanto:
 - Não é possível restaurar ficheiros cuja cópia de segurança foi feita a partir de uma versão diferente do IBM Content Analytics with Enterprise Search.
 - Tem de restaurar dados para um sistema que tenha o mesmo número de servidores. Por exemplo, se efectuar cópia de segurança do sistema que é executado em dois servidores, deverá restaurar os dados num sistema que utilize dois servidores.
 - Não é possível restaurar ficheiros cuja cópia de segurança tenha sido feita a partir de um sistema operativo para um sistema que esteja a executar um

sistema operativo diferente. Por exemplo, caso tenha instalado o produto em AIX e agora pretende executá-lo em Linux, deverá instalar um novo sistema nos servidores Linux.

- Todas as definições para o directório de instalação (ES_INSTALL_ROOT), o directório de dados (ES_NODE_ROOT) e o ID de administrador e a palavra-passe deve ser a mesma entre o sistema do qual foi feita cópia de segurança e o sistema para o qual os dados são restaurados.
- Para configurar vários servidores, crie uma cópia de segurança e restaure o sistema a partir do servidor do controlador. Os scripts executam comandos remotos para cria uma cópia de segurança e restaurar os dados do pesquisador.
- Tem de ter espaço em disco suficiente disponível para efectuar cópia de segurança de ficheiros do sistema para outro directório. A cópia de segurança e a restauração de scripts não verificam os ficheiros.
- Todas as sessões do sistema são interrompidas enquanto os scripts da cópia de segurança e restauro estiverem em execução. Para evitar consultar informações incorrectas ou inconsistente do sistema, não utilize a consola de administração enquanto os scripts estiverem a ser executados.
- Se o sistema falhar devido a um erro irrecuperável, terá de reinstalar o IBM Content Analytics with Enterprise Search.

Tarefas relacionadas:

"Efectuar a cópia de segurança do sistema"

"Restaurar o sistema" na página 457

Referências relacionadas:

"Exportar e importar configurações da colecção" na página 458

Efectuar a cópia de segurança do sistema

Pode efectuar uma cópia de segurança do sistema utilizando o script esbackup.sh para AIX ou Linux, o script de esbackup.bat para Microsoft Windows.

O programa esbackup não cria uma imagem de cópia de segurança dos dados da colecção nos seguintes casos:

- Os dados estão numa localização personalizada.
- Para colecções que utilizam o IBM InfoSphere BigInsights, os dados da colecção em sistemas de armazenamento de ficheiros Hadoop Distributed File System (HDFS).

Nestes casos, tem de copiar, manualmente, o directório onde os dados da colecção estão localizados para criar uma cópia de segurança dos dados.

O ID do administrador deve ter permissão para escrever no directório que especificar quando executar o script da cópia de segurança.

Todas as sessões do sistema são interrompidas enquanto os scripts da cópia de segurança e restauro estiverem em execução. Para evitar consultar informações incorrectas ou inconsistente do sistema, não utilize a consola de administração enquanto os scripts estiverem a ser executados.
Aviso: Se premir Ctrl+C para interromper o script de cópia de segurança, o sistema entra num estado inconsistente. Tem de introduzir o seguinte comando para iniciar todas as sessões de inicialização e quaisquer sessões em execução que tenham sido paradas durante o processo de cópia de segurança:

esadmin system startall

Para efectuar cópia de segurança no sistema:

- 1. No servidor do controlador, inicie sessão como administrador predefinido do IBM Content Analytics with Enterprise Search.
- 2. Se o nível de comunicação comum (CCL) não estiver em execução, inicie-o:

AIX ou Linux

startccl.sh -bg

Linha de comandos do Windows startccl

Ferramenta administrativa de Serviços Windows

Para iniciar o CCL em segundo plano:

- a. Inicie os Serviços do Windows: Iniciar > Programas > Ferramentas Administrativas > Serviços.
- b. Faça clique com o botão direito do rato em **IBM IBM Content Analytics with Enterprise Search** e faça clique em **Iniciar**.
- Para uma configuração de múltiplos servidores, certifique-se de que o CCL é iniciado em cada servidor. Se necessário, inicie sessão como administrador predefinido do IBM Content Analytics with Enterprise Search e inicie o CCL.
- 4. Inicie a cópia de segurança, em que *directório_cópia_segurança* é um directório onde colocar os dados sujeitos a cópia de segurança:

AIX ou Linux

esbackup.sh directório_cópia_segurança

Linha de comandos do Windows esbackup.bat directório_cópia_segurança

Tarefas relacionadas:

"Restaurar o sistema"

Referências relacionadas:

"Exportar e importar configurações da colecção" na página 458

"Comandos, códigos de retorno e IDs da sessão" na página 511

Restaurar o sistema

Depois de re-instalar o IBM Content Analytics with Enterprise Search, pode utilizar o script **esrestore.sh** para AIX ou Linux, ou o script **esrestore.bat** para Microsoft Windows de forma a restaurar o sistema.

Todas as sessões do sistema são interrompidas enquanto os scripts da cópia de segurança e restauro estiverem em execução. Para evitar consultar informações incorrectas ou inconsistente do sistema, não utilize a consola de administração enquanto os scripts estiverem a ser executados.

Não pode restaurar ficheiros dos quais tenha sido efectuada cópia de segurança a partir de uma versão da IBM Content Analytics with Enterprise Search para um sistema que esteja a executar uma versão diferente da IBM Content Analytics with

Enterprise Search. Além disso, o sistema para o qual está a restaurar dados tem de ter um número igual ou superior de servidores do que o sistema do qual foi feita a cópia de dados.

Ao restaurar o sistema o sistema, este é restaurado para a versão de cópia de segurança. Se criou colecções desde a altura em que criou a cópia de segurança, as novas colecções não estarão presentes no seu sistema depois de executar o comando **esrestore**.

Para restaurar o sistema:

- 1. No servidor do controlador, inicie sessão como administrador do IBM Content Analytics with Enterprise Search.
- 2. Se o nível de comunicação comum (CCL) não estiver em execução, inicie-o:

AIX ou Linux startccl.sh -bg

Linha de comandos do Windows startccl

Ferramenta administrativa de Serviços Windows

Para iniciar o CCL em segundo plano:

- a. Inicie os Serviços do Windows: Iniciar > Programas > Ferramentas Administrativas > Serviços.
- b. Faça clique com o botão direito do rato em **IBM IBM Content Analytics with Enterprise Search** e faça clique em **Iniciar**.
- Para uma configuração de múltiplos servidores, certifique-se de que o CCL é iniciado em cada servidor. Se necessário, inicie sessão como administrador predefinido do IBM Content Analytics with Enterprise Search e inicie o CCL.
- 4. Pare o controlador:

esadmin stop

- 5. Restaure os dados, em que *backup_directory* é o directório onde efectuou a cópia de segurança dos ficheiros:
 - AIX ou Linux

esrestore.sh directório_cópia_segurança

Linha de comandos do Windows esrestore.bat directório cópia segurança

Tarefas relacionadas:

"Efectuar a cópia de segurança do sistema" na página 456

Referências relacionadas:

"Exportar e importar configurações da colecção"

"Comandos, códigos de retorno e IDs da sessão" na página 511

Exportar e importar configurações da colecção

Pode exportar ou importar configurações de colecções individuais. Apenas os ficheiros de configuração da colecção são exportados e importados, não os dados da colecção.

Pode exportar uma colecção de um sistema, e depois importar a colecção para um sistema diferente. No entanto, pode exportar e importar colecções apenas entre sistemas que estejam em execução na mesma versão do produto.

Se exportar uma colecção, pode importá-la apenas para um sistema que utilize o mesmo sistema operativo. Por exemplo, não pode exportar uma colecção de um sistema que esteja instalado em Linux e importá-lo para um sistema que esteja instalado em Windows.

Exportar colecções

Para exportar uma colecção, pode utilizar o comando **esadmin export** para exportar uma colecção de cada vez. Não existe suporte para exportar colecções na consola administrativa.

- Inicie sessão como administrador predefinido do IBM Content Analytics with Enterprise Search. Numa configuração de vários servidores, tem de iniciar sessão no servidor do controlador. O comando falha se o tentar executar de outro servidor.
- 2. Introduza os seguintes comandos:

esadmin export -help esadmin export -cid collection_id [-fname export_filename.zip] [-verbose]Em que:

-help

Fornece informações de ajuda para o comando.

-cid ID colecção

Especifica o ID de colecção para a colecção a exportar.

Sugestão: Para determinar o identificador da colecção para a colecção que seja exportar, faça clique em **Acções** > **Visualizar definições da colecção** na consola de administração ou utilize o comando **esadmin report collections**.

-fname nomeficheiro_exportado

Especifica o caminho para o ficheiro exportado. Se o nome do ficheiro não for absoluto, o directório ES_NODE_ROOT/dump é assumido. Se omitir esta opção, é criado um ficheiro que utiliza a seguinte convenção de nomenclatura no directório ES_NODE_ROOT/dump:

export*aaaaMMdd_HHmmssz*.zip

em que:

aaaaaMMdd

É o ano, o mês e o dia actual quando o comando exportado é executado.

HHmmss

É a hora, o minuto e o segundo actual quando o comando exportado é executado.

z É a distância em fuso horário do GMT quando o comando exportado é executado. Por exemplo, a Hora Local do Pacífico é GMT -0800.

-verbose

Apresenta informações que podem ajudar a solucionar problemas na operação de exportação.

Importar colecções

Para importar uma colecção, pode utilizar o comando **esadmin import** para importar uma colecção de cada vez. Não existe suporte para importar colecções na consola de administração.

- Inicie sessão como administrador predefinido do IBM Content Analytics with Enterprise Search. Numa configuração de vários servidores, tem de iniciar sessão no servidor do controlador. O comando falha se o tentar executar de outro servidor.
- 2. Introduza os seguintes comandos:

```
esadmin import -help
esadmin import -fname export_filename
    [-cid new_collection_id]
    [-name new_collection_name]
    [-colDataDir new_collection_data_directory]
    [-force]
    [-verbose]
```

Em que:

-help

Fornece informações de ajuda para o comando.

-fname nomeficheiro_exportado

Especifica o caminho para o ficheiro que pretende importar. Trata-se do ficheiro que foi criado quando executou o comando **esadmin export** para exportar configurações da colecção. Se o nome do ficheiro não for absoluto, o directório ES_NODE_ROOT/dump é assumido.

-cid new_collection_id

Especifica um novo ID de colecção se a colecção necessitar de ser importada com um ID diferente. Se especificar o mesmo ID de colecção ao importar uma colecção, a maior parte dos IDs de sessão não será alterada. O ID de sessão baseia-se no ID de colecção como, por exemplo, *collectionID*.indexer. No entanto, os IDs de sessão dos pesquisadores e os suplementos do pesquisadores podem alterar-se. Além disso, a autenticação da escuta de dados e os dados de autorização podem não ser guardados ao importar uma colecção.

Sugestão: Para determinar o identificador da colecção para a colecção que deseja importar, faça clique em **Acções** > **Visualizar definições da colecção** na consola de administração ou utilize o comando **esadmin report collections**.

-name new_collection_name

Especifica um nome de apresentação para a colecção. Se não especificar um nome de apresentação ao importar a colecção, é utilizado o valor predefinido, Collection*collection_ID*.

-colDataDir new_collection_data_directory

Especifica o directório de dados para a colecção. Se for omitido, é atribuído um directório predefinido.

-force

Força a importação de uma colecção que já existe no sistema de destino. O sistema utiliza o ID de colecção para a colecção importada para determinar que colecção substituir no sistema de destino.

Importante:

 Depois de importar uma colecção, não é possível importar novamente uma colecção com o mesmo ID de colecção, a não ser que utilize esta opção para forçar a colecção a ser importada.

- Uma colecção que é importada pela força não retém os motores de análise de texto, dicionários e suplementos do pesquisador. Estes são sobrepostos pelas informações contidas na colecção importada.
- Quaisquer pesquisadores associados à colecção são removidos e substituídos por pesquisadores que são especificados no ficheiro importado. Devido à remoção dos pesquisadores, todos os metadados do pesquisador e os documentos que ainda não tiverem sido analisados são removidos. Depois da importação da colecção, tem de voltar a pesquisar todos os documentos para gerar novamente estes dados para os pesquisadores importados.

-verbose

Apresenta informações que podem ajudar a solucionar problemas na operação de importação.

Directrizes de utilização

Motores de análise de texto e dicionários

Quaisquer motores de análise de texto e dicionários (tais como um dicionário de sinónimos, palavras de paragem e palavras hierárquicas) que são associados à colecção no sistema exportado não são exportados com os dados de configuração da colecção. Só são exportadas as associações de nome a motores de análise d texto e dicionários.

Na importação, se um motor de análise de texto ou um dicionário com o mesmo nome existir no sistema de destino, é associado à colecção importada. Caso contrário, é apresentada uma mensagem de aviso para indicar que as associações não puderam ser estabelecidas. Por este motivo, deve carregar os motores de análise de texto e dicionários para o sistema de destino e utilizar os mesmos nomes que foram utilizados na colecção importada.

Se um motor de análise de texto ou um dicionário com o mesmo nome não existir no sistema de destino, a associação será quebrada. Se a associação se quebrar, a colecção irá funcionar apropriadamente no sistema de destino, mas a colecção não utilizará os motores de análise de texto ou dicionários correspondentes.

Pesquisadores

Tem de se certificar de que os pré-requisitos do pesquisador estão instalados e configurados no sistema de destino.

Os pesquisadores importados não funcionam a não ser que as origens dos dados que os pesquisadores pesquisam estejam acessíveis. Por exemplo, se a colecção inclui um pesquisador de sistemas de ficheiros que pesquisa um sistema de ficheiros local, o pesquisador não pode pesquisar, excepto se a estrutura do mesmo directório existir no sistema de destino.

Os suplemento do pesquisador não são exportados. Quando importa uma colecção, é apresentada uma mensagem de aviso e, em seguida, a importação continua até à conclusão. Depois de importar a colecção, actualize as propriedades do pesquisador e associe os suplementos do pesquisador. Tem de se certificar de que os suplementos do pesquisador são implementados no sistema de destino de forma a garantir o funcionamento correcto do pesquisador.

Reiniciar o sistema

Se especificar o mesmo ID da colecção quando importar uma colecção e não especificar a opção **colDataDir**, ou se especificar a opção **colDataDir**

com um directório de dados diferente do que foi usado no sistema original, tem de reiniciar o novo sistema antes de iniciar pesquisadores na colecção importada.

Executar os comandos **esadmin stop** e **esadmin start** no novo sistema assegura que o valor alterado para o directório de dados da colecção é reflectido de forma correcta.

Se especificar o mesmo ID da colecção quando importar uma colecção, e especificar o mesmo directório de dados na opção **colDataDir**, não necessita de reiniciar o sistema.

Tarefas relacionadas:

"Efectuar a cópia de segurança do sistema" na página 456

"Restaurar o sistema" na página 457

Referências relacionadas:

"Comandos, códigos de retorno e IDs da sessão" na página 511

Integração com o IBM Cognos BI

Para ajudar o utilizador a ter um maior conhecimento do conteúdo analisado, pode gerar relatórios que contenham estatísticas de análise acerca de conteúdo em colecções do IBM Content Analytics with Enterprise Search e ver relatórios com ferramentas de análise do IBM Cognos Business Intelligence. Também é possível exportar documentos pesquisados, documentos analisados ou documentos que correspondam directamente às condições de consulta para uma base de dados relacional IBM Cognos BI.

IBM Cognos BI é um conjunto de informações empresarias integradas que faculta uma vasta gama de funções para ajudar analistas empresariais a tomar decisões empresariais eficazes analisando dados e vendo ou criando relatórios de análise.

Normalmente, o IBM Cognos BI ajuda os utilizadores empresariais a compreender dados estruturados representados como campos como, por exemplo, registos em tabelas de bases de dados relacionais. O IBM Content Analytics with Enterprise Search extrai informações de dados não estruturados, que normalmente se encontram em formato de texto livre. Uma vez que os dados extraídos podem ser processados como dados estruturados, o IBM Cognos BI pode consumir informações do IBM Content Analytics with Enterprise Search.

Pode transmitir informações do IBM Content Analytics with Enterprise Search para o IBM Cognos BI de duas formas:

Exportar documentos com resultados de análise de conteúdo

Ao utilizar as funções de exportação na consola administrativa ou no prospecção de análise de conteúdo, é possível exportar documentos para tabelas de esquema em estrela numa base de dados relacional. De seguida pode executar consultas de processamento analítico online (OLAP) em relação aos relatórios para efectuar uma análise mais profunda de dados estruturados e não estruturados.

Gerar relatórios que contenham resultados de prospecção

Ao utilizar o prospecção de análise de conteúdo, é possível explorar diferentes facetas do conteúdo analisado numa colecção. Para analisar os resultados de diferentes formas, é possível gerar um relatório que possa ser aberto com o IBM Cognos BI. O relatório fica disponível na página do portal do IBM Cognos BI, facilitando a partilha do relatório entre utilizadores.

O relatório ao ser visto por utilizadores empresariais pode ser impresso, guardado em formato PDF, editado com o Cognos Report Studio, para além de outras acções. Ao fazer clique numa ligação facultada no relatório, os utilizadores podem abrir o prospecção de análise de conteúdo com as mesmas condições de procura utilizadas para criar o relatório. Esta função ajuda os analistas a explorar documentos, por exemplo, antes de efectuar decisões empresariais.

No prospecção de análise de conteúdo, utilize a vista Relatórios para visualizar uma lista de todos os relatórios. Quando selecciona um relatório na lista, pode abri-lo com o IBM Cognos BI. Também pode copiar o URL para o relatório e, em seguida, colar o URL noutro documento ou na mensagem de correio electrónico para partilhar o relatório com outros utilizadores. Para automatizar a exportação de documentos para uma base de dados relacional e permitir a utilizadores gerar relatórios de análise, é possível executar assistentes ao configurar uma colecção de análise de conteúdo na consola de administração de IBM Content Analytics with Enterprise Search para especificar as preferências de exportação e de relatório. Também deve conceder aos utilizadores a autoridade de exportar documentos de uma aplicação e criar relatórios de IBM Cognos BI. Tem de configurar a integração entre o IBM Content Analytics with Enterprise Search e IBM Cognos BI antes de executar os assistentes.

Para obter as informações mais recentes acerca da versão do IBM Cognos BI suportadas pelo IBM Content Analytics with Enterprise Search, consulte as fontes de dados suportadas.

Uma visita guiada pode mostrar-lhe como configurar a integração entre o IBM Content Analytics with Enterprise Search e o IBM Cognos BI, exportar dados analisados e importar dados para relatórios, bases de dados e modelos do IBM Cognos BI. A visita também lhe mostra como utilizar o IBM Cognos BI para efectuar análises mais aprofundadas dos dados importados. Para ver esta visita, consulte Integração IBM Cognos Business Intelligence com IBM Content Analytics with Enterprise Search.

Conceitos relacionados:

"Exportar documentos para utilização noutras aplicações" na página 469

Tarefas relacionadas:

"Exportar documentos pesquisados ou analisados" na página 472

"Exportar documentos a partir de resultados de consulta" na página 475

Referências relacionadas:

Unigens de dados suportadas

Gerar relatórios IBM Cognos BI

Se utilizar o IBM Cognos Business Intelligence (IBM Cognos BI) com IBM Content Analytics with Enterprise Search, os utilizadores do prospecção de análise de conteúdo podem gerar relatórios que incluam valores da faceta, contagens de frequência e valores de correlação. Podem abrir os relatórios no prospecção de análise de conteúdo ou abrir os relatórios com ferramentas de análise IBM Cognos BI.

Os relatórios gerados são guardados no servidor de IBM Cognos BI e podem ser acedidos pela vista Relatórios no prospecção de análise de conteúdo. Os relatórios gerados têm uma ligação de hipertexto que permite aos utilizadores que analisam o relatório regressar a prospecção de análise de conteúdo com as condições de consulta aplicadas quando o relatório foi criado. O utilizador pode continuar a extrair o conteúdo da colecção ou, por exemplo, abrir documentos que correspondam às condições de consulta.

Para integrar o IBM Cognos BI comIBM Content Analytics with Enterprise Search:

 Caso o servidor de IBM Cognos BI utilizado para gerar relatórios seja diferente do servidor principal de IBM Content Analytics with Enterprise Search, instale o servidor de IBM Cognos BI sem o configurar no servidor principal de IBM Content Analytics with Enterprise Search. Consulte o Centro de Informações do IBM Cognos Business Intelligence para obter mais informações acerca dos pré-requisitos e procedimentos de instalação.

- 2. Instale o SDK de IBM Cognos no servidor principal de IBM Content Analytics with Enterprise Search.
- **3**. Criar uma base de dados para armazenar resultados. A base de dados pode estar no servidor principal IBM Content Analytics with Enterprise Search ou num servidor remoto. A base de dados não pode ser uma Cognos Content Database.
- 4. Caso a base de dados não se encontre no servidor principal de IBM Content Analytics with Enterprise Search, copie os módulos do controlador de JDBC para o tipo de base de dados que criou para o servidor principal de IBM Content Analytics with Enterprise Search.

Os exemplos de ficheiros JAR incluem db2jcc.jar e db2jcc_licence_cu.jar para DB2, ojdbc6.jar para Oracle e sqljdbc4.jar para o servidor SQL. Consulte a documentação relativa ao sistema de gestão do sistema para obter detalhes.

- 5. No servidor de IBM Cognos BI, configure um Ligação de Origem de Dados para indicar a base de dados que criou. Utilize a função de teste de IBM Cognos BI para garantir que a Ligação de Origem de Dados consegue estabelecer ligação à base de dados.
- 6. Configure uma colecção para suportar a geração de relatórios:
 - a. Na vista Colecções na consola de administração, expanda uma colecção de análise de conteúdo. No painel de Análise, faça clique no botão para configurar as definições de relatório de IBM Cognos BI.
 - b. Faça clique no botão para iniciar o assistente de configuração e, de seguida, siga as instruções online.
 - **c.** Abra a vista Segurança para autorizar os utilizadores a gerarem relatórios. Por predefinição, nenhum utilizador tem privilégios de geração de relatórios. O modo como autoriza utilizadores depende de os utilizadores terem de iniciar sessão na aplicação.
 - Se a segurança de início de sessão do servidor da aplicação estiver activada, consulte a área de segurança do nível do sistema, faça clique em Acções > Adicionar um utilizador de aplicação ou grupo, especifique um ID de utilizador ou de grupo e seleccione a caixa de confirmação para criar relatórios de IBM Cognos BI.
 - Se não for necessário que os utilizadores iniciem sessão na aplicação, ou se pretender conceder este privilégio a todos os utilizadores, faça clique em Acções > Especificar privilégios predefinidos do utilizador da aplicação e seleccione a caixa de confirmação para criar relatórios de IBM Cognos BI.
- 7. Gere relatórios a partir de prospecção de análise de conteúdo.
 - a. Abra a aplicação e explore uma colecção activada para gerar relatórios.
 - b. Faça clique no ícone para criar um relatório de IBM Cognos BI, disponível nas vistas Facetas, Série cronológica, Desvios, Tendências e Pares de facetas. Caso o ícone não esteja visível, actualize o navegador.
 - c. Quando surgirem as opções para gerar um relatório de IBM Cognos BI, especifique as suas preferências de relatório e faça clique em **Submeter**.
 - d. Após a geração do relatório, abra a vista Relatórios. Faça clique no nome do relatório que pretende ver com as ferramentas de análise de IBM Cognos BI.

Dependendo de como foi configurada a autenticação quando as opções de geração de relatórios de IBM Cognos BI foram configuradas na consola de administração, poderá ter de especificar um ID de utilizador autorizado a aceder ao servidor de IBM Cognos BI antes de ver relatórios. Neste caso, edite

Preferências, abra o separador Relatórios e especifique um ID e palavra-passe de início de sessão válidos de IBM Cognos BI.

Referências relacionadas:

E Centro de Informações do IBM Cognos Business Intelligence

Exportar documentos para IBM Cognos BI

Os utilizadores podem exportar documentos pesquisados, analisados ou procurados de IBM Content Analytics with Enterprise Search para tabelas de esquema em estrela numa base de dados relacional de IBM Cognos Business Intelligence (IBM Cognos BI).

Para além disso, pode criar-se um modelo de IBM Cognos BI com base no esquema em estrela e pode publicar-se um pacote para o modelo. Ao utilizar o pacote, os utilizadores podem efectuar consultas de processamento analítico online (OLAP) em relação aos dados exportados.

Também se pode gerar um ficheiro XML de registo de acções de Cognos para o modelo. Os utilizadores de IBM Cognos BI podem carregar o ficheiro utilizando o IBM Cognos Framework Manager e, de seguida, modificando o modelo consoante for necessário.

Para configurar o IBM Content Analytics with Enterprise Search para exportar documentos para uma base de dados relacional de IBM Cognos BI:

- 1. Configure a base de dados a utilizar para documentos exportados:
 - a. Crie uma base de dados para exportar documentos processados por IBM Content Analytics with Enterprise Search. A base de dados pode situar-se no servidor principal de IBM Content Analytics with Enterprise Search ou num servidor remoto.
 - b. Caso a base de dados não se encontre no servidor principal de IBM Content Analytics with Enterprise Search, copie os módulos do controlador de JDBC para o tipo de base de dados que criou para o servidor principal de IBM Content Analytics with Enterprise Search.

Os exemplos de ficheiros JAR incluem db2jcc.jar e db2jcc_licence_cu.jar para DB2, ojdbc6.jar para Oracle e sqljdbc4.jar para o servidor SQL. Consulte a documentação relativa ao sistema de gestão do sistema para obter detalhes.

- 2. Efectue os seguintes passos caso pretenda criar um modelo de IBM Cognos BI e publicar o pacote para uma análise mais profunda em relação a documentos exportados utilizando o IBM Cognos BI.:
 - a. Caso o servidor de IBM Cognos BI seja diferente do servidor principal de IBM Content Analytics with Enterprise Search, instale o servidor de IBM Cognos BI no servidor principal de IBM Content Analytics with Enterprise Search sem o configurar. Consulte o Centro de Informações do IBM Cognos Business Intelligence para obter mais informações acerca dos pré-requisitos e procedimentos de instalação.
 - b. Instale o SDK de IBM Cognos no servidor principal de IBM Content Analytics with Enterprise Search.
 - c. No servidor de IBM Cognos BI, configure um Ligação de Origem de Dados para indicar a base de dados que criou. Utilize a função de teste de IBM Cognos BI para garantir que a Ligação de Origem de Dados consegue estabelecer ligação à base de dados.
- 3. Configure uma colecção para suportar a exportação de documentos:

- a. Na vista Colecções, na consola de administração, expanda uma colecção de análise de conteúdo e faça clique no ícone Exportar para o tipo de dados que pretende exportar. É possível configurar opções separadas para exportar documentos pesquisados, documentos analisados e interpretados e documentos dos resultados da procura.
- b. Faça clique na opção para exportar documentos para uma base de dados relacional e iniciar o assistente de configuração. Siga as instruções online para concluir o assistente.
- c. Abra a vista Segurança para autorizar os utilizadores a exportar documentos a partir de uma aplicação. Por predefinição, nenhum utilizador tem privilégios de exportação de documentos. O modo como autoriza utilizadores depende de os utilizadores terem de iniciar sessão no servidor da aplicação.
 - Se a segurança de início de sessão do servidor da aplicação estiver activada, consulte a área de segurança do nível do sistema, faça clique em Acções > Adicionar um utilizador ou grupo da aplicação, especifique um ID de utilizador ou ID de grupo e seleccione a caixa de verificação para exportação de documentos.
 - Se não for necessário que os utilizadores iniciem sessão na aplicação, ou se pretender conceder este privilégio a todos os utilizadores, faça clique em Acções > Especificar privilégios predefinidos do utilizador da aplicação e seleccione a caixa de confirmação para exportar documentos.

Conceitos relacionados:

"Exportar documentos para utilização noutras aplicações" na página 469

Tarefas relacionadas:

"Exportar documentos pesquisados ou analisados" na página 472 "Exportar documentos a partir de resultados de consulta" na página 475 "Exportar documentos como ficheiros CSV" na página 492

Referências relacionadas:

Centro de Informações do IBM Cognos Business Intelligence

Exportar documentos para utilização noutras aplicações

Para utilizar informações do IBM Content Analytics with Enterprise Search com outros objectivos, tais como armazém de dados, informação empresarial, classificação e eDiscovery ou aplicações em conformidade, pode exportar documentos de colecções e depois importar os dados exportados para as suas aplicações.

Quando configurar opções para exportar documentos, especifique se os documentos são exportados como ficheiros XML ou CSV para um sistema de ficheiros, exportado para uma base de dados relacional, ou exportados segundo um formato e localização especificados por um suplemento personalizado. O IBM Content Analytics with Enterprise Search não faculta utilitários para importar os documentos exportados para outras aplicações.

Restrição: Para tirar partido de facetas de metadados automáticas, detecção de documentos duplicados e suporte de sinalização de documentos que o IBM Content Analytics with Enterprise Search faculta para auxiliar com a avaliação de conteúdo, o ambiente deve ser executado no mesmo sistema operativo. Por exemplo, se exportar dados do IBM Content Analytics with Enterprise Search e, de seguida, utilizar o IBM Content Collector para adicionar os dados exportados do IBM FileNet P8, os três produtos devem ser executados no mesmo sistema operativo.

Tipos de documentos que pode exportar.

Documentos que podem ser exportados quando são pesquisados, depois de serem interpretados e analisados, ou como resultado de pesquisar ou prospectar uma colecção.

Documentos pesquisados

Pode exportar documentos que foram pesquisados pelos pesquisadores do IBM Content Analytics with Enterprise Search da cache do documento antes de serem interpretados ou analisados. Pode seleccionar esta abordagem, por exemplo, se as aplicações necessitarem de recolher de várias origens.

Neste cenário, presume-se que o sistema de gestão de conteúdos suporta a importação de documentos de um sistema de ficheiros, mas também pretende importar documentos que estão armazenados numa coluna BLOB numa base de dados relacional. Pode configurar um pesquisador da base de dados para pesquisar a coluna BLOB e depois exportar os dados pesquisados para ficheiros XML num sistema de ficheiros. Depois pode importar os ficheiros exportados para o sistema de gestão de conteúdos.

Documentos analisados

Pode exportar documentos analisados da ligação em série de processamento de documentos. Poderá seleccionar esta abordagem, por exemplo, se quiser utilizar as informações não estruturadas numa aplicação de informação empresarial para dados estruturados. Os dados exportados incluem todos os tokens, metadados e facetas gerados de anotações que foram adicionadas aos documentos por analisadores e pelos anotadores do UIMA que a colecção está configurada para utilizar. Neste cenário, presuma que pretende analisar relatórios sobre defeitos em automóveis. Os relatórios contêm dados estruturados, como um código de problema ou a data do relatório. Cada relatório também inclui uma descrição das queixas dos clientes e de como o problema foi resolvido em formato de texto livre. Por exemplo, o relatório de problemas pode ser "Cliente sentiu cheiro a queimado dentro do capô" e a solução do problema pode ser "Foi substituído tubo ferrugento à bomba de combustível."

Pode criar um anotador para extrair palavras-chave específicas da indústria ou padrões de texto não estruturado, como sintomas de um problema, os nomes de peças de substituição e assim por diante. Se configurar uma colecção para utilizar o anotador, podem ser analisados, extraídos e anotados dados desestruturados relevantes, quando documentos pesquisados inserirem a ligação directa de processamento de documentos. Pode exportar estes dados analisados para ficheiros XML, CSV ou para uma base de dados relacional e, de seguida, importar os dados exportados para a aplicação de informação empresarial do utilizador.

Caso utilize o IBM Cognos Business Intelligence (IBM Cognos BI), pode configurar o IBM Content Analytics with Enterprise Search para exportar documentos directamente para uma base de dados relacional. Pode executar consultas de processamento analítico online (OLAP) em relação aos relatórios para efectuar uma análise mais profunda de dados estruturados e não estruturados.

Documentos procurados

Pode consultar uma colecção, limitar os resultados seleccionando facetas e especificando critérios de procura adicionais e, depois, exportar os documentos que correspondam aos resultados. Pode seleccionar esta abordagem, por exemplo, se quiser criar um subconjunto de documentos que necessite de mais investigação.

Neste cenário, presuma que lhe é pedido pelo departamento de conformidade legal que recolha documentos como resposta a um pedido de descoberta em relação a infringir alguma patente. Os documentos de interesse são pesquisados, analisados e armazenados no índice. No entanto, a colecção também inclui documentos que não têm valor para a investigação actual. Quando procurar a colecção, especifique critérios para limitar os resultados para documentos que são relevantes para o pedido de descoberta. Pode exportar os documentos para ficheiros XML e, depois, importar os ficheiros exportados para o sistema eDiscovery, como o IBM eDiscovery Manager.

Noutro cenário, presuma que necessita de treinar uma base de conhecimento do IBM Content Classification. Quando consultar a colecção, pode exportar documentos que correspondam às suas condições de procura como ficheiros XML. Quando os documentos são exportados, um ficheiro catalog.xml que contém informações sobre os campos nos documentos também é exportado. Se importar os ficheiros XML dos documentos e o ficheiro catalog.xml no Classification Workbench, pode utilizar os dados para treinar bases de conhecimento e planos de decisão. Ao pesquisar repetidamente colecções e exportar documentos, pode melhorar a forma como o conteúdo é classificado com o tempo.

Opções de exportação

Para exportar documentos, tem de activar a cache do documento para a colecção.

Quando configurar as opções de exportação para documentos pesquisados ou analisados, pode especificar se os documentos podem ser adicionados ao índice. Por exemplo, se utilizar primeiramente IBM Content Analytics with Enterprise Search como um meio de recolher documentos ou recolher dados de análise acerca do conteúdo, já pode querer exportar os documentos sem os adicionar ao índice.

Quando configurar opções de exportação para documentos pesquisados, pode configurar agendamentos para controlar quando os documentos serão exportados da cache do documento. Poderá criar uma agenda geral para todos os pedidos de exportação e configurar agendas personalizadas para pedidos individuais. Também pode pedir a execução de uma base incremental. Neste caso, apenas são exportados documentos que foram adicionados ao índice após a última vez em que foi executado o programa de exportação.

Quando configurar opções de exportação para documentos pesquisados, analisados ou pesquisados:

- Pode activar ou desactivar a capacidade de exportar documentos da colecção.
- Especifique se pretende exportar documentos como ficheiros XML ou CSV para uma base de dados relacional, ou segundo a lógica num suplemento personalizado.
- Dependendo do formato de exportação, os assistentes podem auxiliar o utilizador a especificar opções de exportação. Por exemplo, se exportar documentos para uma base de dados relacional, um assistente auxilia o utilizador a especificar informações acerca da base de dados destino e os campos e facetas que pretende exportar. Caso utilize o IBM Cognos BI, o assistente também ajuda a especificar opções para explorar directamente para uma base de dados de IBM Cognos BI.
- Os dados de exportação reflectem os documentos na altura em que ocorre a exportação e não inclui qualquer histórico dos documentos. Por exemplo:
 - Se os documentos pesquisados primeiramente forem removidos do espaço de pesquisa, o documento exportado não é removido do sistema de ficheiros de saída.
 - Para uma exportação de documentos pesquisados, os documentos adicionados ao índice após ter início o processo de exportação não são exportados.

Conceitos relacionados:

"Integração com o IBM Cognos BI" na página 463

Criar e implementar um suplemento para exportar documentos ou resultados de inspecções profundas

Tarefas relacionadas:

"Exportar documentos para IBM Cognos BI" na página 466

"Exportar documentos pesquisados ou analisados"

"Exportar documentos a partir de resultados de consulta" na página 475

"Exportar para um directório Windows partilhado" na página 479

"Exportar documentos como ficheiros CSV" na página 492

Referências relacionadas:

"Caminhos de saída para documentos exportados" na página 480

"Formato de ficheiro XML para documentos exportados" na página 481

"Correlações de bases de dados relacionais para documentos exportados de colecções de procura empresarial" na página 485

Exportar documentos pesquisados ou analisados

Para utilizar informações oriundas do IBM Content Analytics with Enterprise Search para outros fins como, por exemplo, armazéns de dados, informação empresarial, ou aplicações de classificação, poderá exportar documentos que tenham sido pesquisados, ou documentos pesquisados e analisados, e depois importar os dados exportados nas suas aplicações.

Como e quando os documentos são exportados depende de como o conteúdo é acrescentado à colecção e se a colecção usa a memória cache do documento ou não. Os documentos são exportados quando os serviços de análise e indexação são executados. Os serviços de análise e indexação iniciam quando é adicionado conteúdo à colecção ou quando o índice é reconstruído.

Documentos pesquisados

A configuração de execução do pesquisador também controla a exportação de documentos novos, alterados e eliminados. Da primeira vez que o pesquisador pesquisa uma origem de dados, são pesquisados todos os documentos. Em pesquisas subsequentes:

- Se o pesquisador estiver configurado para pesquisar todas as actualizações, irá verificar documentos novos, alterados e eliminados. O programa de exportação exporta os documentos novos e alterados. Poderá configurar uma opção para exportar informações relativas a documentos eliminados. Neste caso, quando exportar documentos em formato XML, será criado um ficheiro XML para cada documento eliminado. Na saída de dados XML, o valor do elemento /Document@Type é DELETED.
- Se o pesquisador estiver configurado para pesquisar documentos novos e modificados, não irá verificar documentos eliminados, nem será gerada informação sobre os documentos eliminados.

Se seleccionar a opção de pesquisar apenas documentos novos e modificados, o pesquisador procura documentos com datas de modificação que são posteriores à hora da pesquisa anterior. Se copiar ficheiros para um recurso, a data de modificação pode não se alterar, o que significa que o pesquisador pode não detectar que os ficheiros foram adicionados ao recurso. Por exemplo, se copiar ficheiros para uma pasta Windows, o Windows não muda automaticamente a data de modificação dos ficheiros. Para assegurar que esses ficheiros são pesquisados, seleccione a opção para pesquisar todas as actualizações ou uma pesquisa completa.

 Se o pesquisador estiver configurado para fazer uma pesquisa integral, então todo o espaço de pesquisa será pesquisado e serão exportados todos os documentos que coincidam com os critérios de exportação do utilizador, independentemente de serem documentos actualizados ou eliminados desde a pesquisa anterior.

Documentos importados

Todos os documentos importados para uma colecção passam pelos serviços de análise e indexação. Se configurar a colecção para exportar documentos, todos os documentos importados serão exportados quando forem processados pelos serviços de análise e indexação.

Reconstruir o índice

Se a memória cache estiver activada para uma colecção, os documentos pesquisados e importados ficam guardados na memória cache. Quando reconstruir o índice, os documentos na memória cache são passados pelos serviços de análise e indexação. Deste modo, os documentos podem ser exportados ao reiniciar a construção do índice.

Se alterar as opções de exportação, tais como activar a exportação de documentos analisados enquanto ficheiros XML, tem de reiniciar os serviços de análise e indexação para reflectir essa alteração. Reiniciar os serviços de análise e indexação inicia igualmente algumas acções de exportação. Por exemplo, se a colecção está configurada para exportar documentos enquanto ficheiros CSV, então o processo de exportação cria um directório e ficheiros CSV para guardar os documentos exportado.

Exportar para IBM Cognos BI: Se utilizar o IBM Cognos Business Intelligence, o assistente ajuda-o a especificar informação para a exportação de documentos para uma base de dados relacional ou como ficheiros de valor separado por vírgula (CSV). Verifique também tópicos relacionados com a configuração da integração entre o IBM Content Analytics with Enterprise Search e o IBM Cognos BI.

Exportar para o IBM DB2: Se pretender exportar documentos para uma base de dados do IBM DB2, tem de instalar o Cliente DB2 no servidor IBM Content Analytics with Enterprise Search. Numa instalação distribuída, instale o Cliente DB2 no servidor do pesquisador. Para obter informações acerca das versões suportadas do IBM DB2, consulte *Origens de dados suportadas*.

Procedimento:

Para exportar documentos pesquisados ou analisados:

- Na vista Colecções, expanda a colecção que pretende configurar. No painel Análise e Indexação, certifique-se de que o processo de análise e indexação está em execução.
- 2. Faça clique no ícone para exportar documentos e, de seguida, faça clique em **Configurar opções de exportação de documentos pesquisados e analisados**.
- Na página Opções para Exportar Documentos Pesquisados ou Analisados, especifique opções para os documentos que pretende exportar e como pretende exportá-los.

Exportar como ficheiros XML

Especificar opções para exportar documentos como ficheiros XML:

- Para documentos pesquisados, especifique se pretende exportar metadados de documentos, conteúdo de documentos ou ambos. Terá de especificar os caminhos de saída de dados onde serão gravados os dados exportados. Os directórios de saída de dados devem existir previamente e permitir acesso de escrita/gravação.
- Para documentos analisados, pode exportar os metadados e o conteúdo analisados. Também pode exportar informações que estejam armazenadas na estrutura de análise comum (CAS). Terá de especificar os caminhos de saída de dados onde serão gravados os dados exportados. Os directórios de saída de dados devem existir previamente e permitir acesso de escrita/gravação.

Quando exportar documentos analisados como XML e permitir a CAS como exportação de formato XMI, as informações armazenadas na CAS são convertidas para o formato XML Metadata Interchange (XMI) e exportadas como ficheiros XMI. O objectivo desta opção de exportação é ajudá-lo a detectar e a corrigir um anotador personalizado vendo o conteúdo da CAS. O formato XMI que é utilizado na CAS baseia-se na norma UIMA. Se, de futuro, a norma UIMA mudar, o formato XMI também pode mudar.

- Especificar padrões de URI para identificar os documentos que pretende exportar. O padrão URI é uma expressão regular e pode introduzir um padrão por linha. A expressão regular tem de ser interpretável por Java. Por exemplo, para exportar documentos com uma URI que termine em .pdf, especifique .*.pdf. Para exportar todos os documentos, especifique .*. Os padrões são avaliados em termos de correspondências, pela ordem em que os listar aqui.
- Para documentos pesquisados, especifique se pretende exportar documentos sem os analisar ou adicionar ao índice. Pode, por exemplo, seleccionar esta caixa de verificação, se estiver a recolher documentos para importar para outra aplicação.
- Para documentos analisados, especifique se pretende exportar documentos analisados sem os adicionar ao índice. Pode, por exemplo, seleccionar esta caixa de verificação, se estiver a recolher documentos ou dados de análise para importar para outra aplicação.
- Especifique se pretende exportar informações sobre documentos que tenham sido eliminados do espaço de pesquisa, desde a última vez que o pesquisador verificou documentos novos, alterados e eliminados.
- Especifique se pretende utilizar o nome do campo ou o caminho da faceta para correlacionar metadados quando os documentos são exportados. Pode seleccionar esta opção, por exemplo, se tencionar importar os documentos exportados para o IBM Content Collector. Esta opção preserva as extensões originais do nome do ficheiro quando o conteúdo é exportado e permite que os nomes dos campos e os caminhos das facetas sejam representados como elementos nos ficheiros XML exportados. Se não activar esta opção, os nomes dos campos e os caminhos das facetas são representados como atributos nos ficheiros XML exportados.

Exportar para uma base de dados relacional

As opções são semelhantes às que especificou para exportar documentos como ficheiros XML, tal como especificar padrões de URI para identificar os documentos que pretende exportar. Existem diferenças para a exportação de documentos a partir de colecções de análise de conteúdos e colecções de procura empresarial:

- Se está a configurar opções para uma colecção de análise de conteúdos, faça clique em **Configurar** e execute um assistente para especificar informações sobre a base de dados de destino. Serão exportados documentos para tabelas em esquema de estrela. A base de dados destino tem de existir.
- Se estiver a configurar opções para uma colecção de procura empresarial, especifique o caminho para o ficheiro de correlação da base de dados que controla onde os dados exportados devem ser escritos. Consulte o ficheiro ES-INSTALL_ROOT/default_config/ export_rdb_mapping.xml para ver um exemplo de ficheiro de correlação.

Exportar como ficheiros CSV

Faça clique em **Configurar** e execute um assistente para especificar informações sobre os campos que pretende exportar e onde pretende criar os ficheiros CSV. Também tem de especificar padrões URI para identificar os documentos que pretende exportar.

Exportar com um suplemento personalizado

Especificar o caminho da classe, o nome de classe e as propriedades que pretende passar para o suplemento personalizado.

- 4. Pare e reinicie os serviços de análise e índice para a colecção.
- 5. Para pesquisar os documentos que pretende exportar e iniciar o pesquisador, deve ou importar documentos para a colecção ou configurar, pelo menos, um pesquisador.
- Opcional: Na área Exportar, verifique o estado do pedido de exportação. Por exemplo, poderá ver o número de documentos exportados até ao momento, se o pedido foi concluído e se ocorreram erros.

Conceitos relacionados:

"Exportar documentos para utilização noutras aplicações" na página 469

"Integração com o IBM Cognos BI" na página 463

Criar e implementar um suplemento para exportar documentos ou resultados de inspecções profundas

Tarefas relacionadas:

"Exportar documentos para IBM Cognos BI" na página 466

"Exportar documentos como ficheiros CSV" na página 492

"Exportar documentos a partir de resultados de consulta"

Referências relacionadas:

"Correlações de bases de dados relacionais para documentos exportados de colecções de procura empresarial" na página 485

litro Origens de dados suportadas

Exportar documentos a partir de resultados de consulta

Para utilizar resultados de consulta do IBM Content Analytics with Enterprise Search para outras finalidades, tais como classificação, conformidade ou eDiscovery, pode exportar documentos que correspondam às suas condições de consulta e, de seguida, importar os dados exportados para as suas aplicações. Os utilizadores autorizados podem exportar documentos da aplicação de procura empresarial ou de uma prospecção de análise de conteúdo. As opções que especificar na consola do administrador controlam o modo como os documentos são exportados. A capacidade de exportar documentos não está disponível quando os utilizadores consultam mais do que uma colecção de cada vez.

Se configurar a colecção para utilizar a memória cache do documento, os utilizadores podem exportar conteúdo e metadados de documentos pesquisados, exportar conteúdo analisado com os resultados da análise de texto ou exportar o conteúdo do documento pesquisado e o conteúdo analisado com os resultados da análise de texto. Se a colecção não utilizar a memória cache de documentos, os utilizadores só podem exportar conteúdo analisado com os resultados da análise de texto.

Quando a procura conceptual com base no IBM Content Classification está activada para uma colecção, a API de sequência de SIAPI devolve todos os documentos numa colecção independentemente do valor de limiar do âmbito que foi configurado para pesquisa conceptual na consola de administração. Como resultado, se a pesquisa conceptual estiver activa, todos os documentos são exportados quando um utilizador exportar documentos pesquisados.

A pós-filtragem de resultados da pesquisa não é suportada pela função de exportação. Se um utilizador exportar documentos após procurar uma colecção segura, os documentos que são excluídos dos resultados da pesquisa através da pós-filtragem serão incluídos no conjunto de documentos a exportar. Os documentos que são excluídos dos resultados da pesquisa através de pré-filtragem não serão exportados.

Se alterar as opções de exportação, tal como activar a exportação de documentos como ficheiros XML em vez de ficheiros CSV, tem de reiniciar os serviços de análise e índice para reflectir a alteração.

Exportar para IBM Cognos BI: Se utilizar o IBM Cognos Business Intelligence, o assistente ajuda-o a especificar informação para a exportação de documentos para uma base de dados relacional ou como ficheiros de valor separado por vírgula (CSV). Consulte tópicos relacionados com a configuração da integração entre o IBM Content Analytics with Enterprise Search e o IBM Cognos BI.

Exportar para o IBM DB2: Se pretender exportar documentos para uma base de dados do IBM DB2, tem de instalar o Cliente DB2 no servidor IBM Content Analytics with Enterprise Search. Numa instalação distribuída, instale o Cliente DB2 no servidor do pesquisador. Para obter informações acerca das versões suportadas do IBM DB2, consulte *Origens de dados suportadas*.

Procedimento:

Para permitir aos utilizadores exportar documentos de resultados da consulta:

- Na vista Colecções, expanda uma colecção, faça clique no ícone para exportar documentos procurados. Na vista Colecções, expanda a colecção que pretende configurar. No painel Procura para uma colecção de procura empresarial ou no painel Análise para uma colecção de análise de conteúdos, faça clique no ícone para configurar opções para exportar documentos.
- 2. Na página Opções para exportar documentos procurados, especifique opções para os documentos que pretende exportar e como pretende exportá-los.

Exportar como ficheiros XML

Especifique os caminhos de saída no servidor do controlador no qual os

metadados do documento e o conteúdo do documento devem ser escritos. Se exportar documentos como ficheiros XML para o IBM Content Classification, será automaticamente criado um ficheiro catalog.xml que descreve os campos nos documentos, e exportado junto com eles.

Também tem de especificar padrões URI para identificar os documentos que pretende exportar. O padrão URI é uma expressão regular e pode introduzir um padrão por linha. A expressão regular tem de ser interpretável por Java. Por exemplo, para exportar documentos com uma URI que termine em .pdf, especifique .*.pdf. Para exportar todos os documentos, especifique .*. Os padrões são avaliados em termos de correspondências, pela ordem em que os listar aqui.

Exportar para uma base de dados relacional

Existem diferenças para a exportação de documentos a partir de colecções de análise de conteúdos e colecções de procura empresarial:

- Se está a configurar opções para uma colecção de análise de conteúdos, faça clique em **Configurar** e execute um assistente para especificar informações sobre a base de dados de destino. Serão exportados documentos para tabelas em esquema de estrela. A base de dados relacional de destino tem de existir.
- Se estiver a configurar opções para uma colecção de procura empresarial, especifique o caminho para o ficheiro de correlação da base de dados que controla onde os dados exportados devem ser escritos. Consulte o ficheiro ES-INSTALL_ROOT/default_config/ export_rdb_mapping.xml para ver um exemplo de ficheiro de correlação.

Semelhante a exportar documentos como ficheiros XML, também tem de especificar padrões URI para identificar os documentos que pretende exportar.

Exportar como ficheiros CSV

Faça clique em **Configurar** e execute um assistente para especificar informações sobre os campos que pretende exportar e onde pretende criar os ficheiros CSV. Semelhante a exportar documentos como ficheiros XML, também tem de especificar padrões URI para identificar os documentos que pretende exportar.

Exportar com um suplemento personalizado

Especificar o caminho da classe, o nome de classe e as propriedades que pretende passar para o suplemento personalizado.

- **3**. Opcional: Para agendar quando o programa de exportação é executado, expanda a área Exportar e, de seguida, faça clique em **Configurar um agendamento para exportar documentos pesquisados**.
 - Poderá especificar uma agenda geral que controle a frequência com que o programa de exportação é executado como, por exemplo, em dias da semana ou do mês específicos, ou a horas do dia específicas. Todos os pedidos agendados que utilizem a agenda geral e estejam na fila são processados à hora especificada.
 - Também pode desactivar e permitir o agendamento para pedidos individuais, configurar agendamentos personalizadas para pedidos individuais ou eliminar pedidos de exportação da fila de pedidos de exportação. Se agendar o pedido para ser executado incrementalmente, só serão exportados os documentos que tiverem sido adicionados ao índice após a última execução do programa de exportação.

- Importe os documentos para a colecção ou configure pelo menos um pesquisador para pesquisar os documentos que pretende exportar e inicie o pesquisador.
- 5. Pare e reinicie os serviços de análise e indexação.
- 6. Recolha os documentos que pretende exportar:
 - a. Na vista Segurança, certifique-se de que o privilégio para exportar documentos está atribuído aos utilizadores da aplicação adequados.
 - b. Abra uma aplicação de procura empresarial ou o prospecção de análise de conteúdo e consulte uma colecção. Seleccione facetas e especifique as condições de consulta para restringir os resultados para os documentos que quer exportar.
 - c. Faça clique no ícone para exportar os resultados da consulta actual e especifique opções para a exportação dos documentos. Poderá proceder do seguinte modo:
 - Exporte conteúdo e metadados pesquisados da memória cache de documentos antes que os processos de análise de texto sejam executados. Neste caso, o conteúdo do documento é exportado como saída binária, separado dos metadados do documento. Esta opção está disponível apenas se a colecção estiver configurada para utilizar a cache de documento.
 - Exporte conteúdo analisado com os resultados da análise de texto como, por exemplo, as anotações para análise linguística, classificação e anotadores. Neste caso, o conteúdo do documento é analisado como texto no mesmo ficheiro com os metadados, não sendo exportado como um ficheiro separado. Esta opção está disponível mesmo que a colecção não utilize a memória cache de documentos.
 - Exporte tanto o conteúdo pesquisado, onde o conteúdo do documento é exportado da memória cache de documentos como dados de saída binários, como o conteúdo analisado com os resultados da análise de texto. Esta opção está disponível apenas se a colecção estiver configurada para utilizar a cache de documento.

O pedido de exportação é enviado ao servidor de indexação para fins de processamento, e é apresentado um ID de pedido. Utilize este ID de pedido para supervisionar o pedido. Por exemplo, poderá ver se o pedido foi concluído ou cancelar o pedido.

- 7. Opcional: Se especificou que este pedido de exportação pode ser agendado e pretende especificar um agendamento personalizado para o pedido, edite a área Exportar na consola de administração e faça clique no ícone para o agendamento de pedidos de exportação. Localize o pedido que pretende configurar, certifique-se de que o agendamento está activado para este pedido, seleccione Personalizar agendamento como o tipo de agendamento, faça clique no ícone Editar Agendamento Personalizado e especifique quando pretende que o pedido seja processado.
- 8. Opcional: Na consola de administração, supervisione a área Export para verificar o estado do pedido de exportação. Por exemplo, poderá ver se o pedido foi concluído e se ocorreram erros, cancelar o pedido, ou poderá ver um histórico de todos os pedidos de exportação de documentos procurados.

Conceitos relacionados:

"Exportar documentos para utilização noutras aplicações" na página 469 "Integração com o IBM Cognos BI" na página 463

Criar e implementar um suplemento para exportar documentos ou resultados de inspecções profundas

Tarefas relacionadas:

"Exportar documentos para IBM Cognos BI" na página 466

"Exportar documentos como ficheiros CSV" na página 492

"Exportar documentos pesquisados ou analisados" na página 472

Referências relacionadas:

"Correlações de bases de dados relacionais para documentos exportados de colecções de procura empresarial" na página 485

🕒 Origens de dados suportadas

Exportar para um directório Windows partilhado

Para exportar documentos como ficheiros XML ou ficheiros CSV a partir de um servidor Windows para um directório partilhado noutro servidor Windows, o utilizador administrativo do IBM Content Analytics with Enterprise Search tem de pode aceder ao directório partilhado pelo caminho UNC e ter permissão para ler e escrever no directório.

Para activar o sistema para exportar documentos para um directório Windows partilhado:

- 1. Inicie sessão como administrador do IBM Content Analytics with Enterprise Search predefinido no servidor de controlador principal.
- 2. Aceda ao directório partilhado no servidor Windows remoto ao especificar o caminho UNC no Explorador do Windows. Por exemplo, para exportar documentos para uma exportação designada do directório partilhado num servidor com o nome do sistema central remoteserver, introduza \\remoteserver\export na barra de endereço do Explorador do Windows. Se não conseguir abrir o directório, inicie sessão como o administrador do Windows e conceda permissão de acesso ao directório no ID de administrador do IBM Content Analytics with Enterprise Search.
- 3. Crie um ficheiro no directório partilhado para verificar se o administrador do IBM Content Analytics with Enterprise Search tem permissão para ler e escrever no directório. Se não consegue criar um ficheiro, inicie sessão como o administrador do Windows e conceda permissões de acesso de leitura e escrita ao ID de administrador do IBM Content Analytics with Enterprise Search.
- 4. Após verificar se estão definidas as permissões de acesso correctas, abra a consola de administração do IBM Content Analytics with Enterprise Search e configure as definições para exportar documentos. Quando especifica a pasta de saída, especifique o directório partilhado no formato de caminho UNC, tal como \\remoteserver\export.

Conceitos relacionados:

"Exportar documentos para utilização noutras aplicações" na página 469

Caminhos de saída para documentos exportados

Quando exportar documentos como XML, o programa de exportação cria um directório no caminho de ficheiro de saída que especificar quando configurar opções de exportação.

Tipos de ficheiro exportado

Dependendo dos tipos de documentos que exportar, a saída pode incluir ficheiros .xml, ficheiros .xmi, ficheiros .dat ou todos os tipos de ficheiros.

Documentos pesquisados

Quando exportar documentos pesquisados, os metadados dos documentos são guardados num ficheiro XML com a extensão .xml. O conteúdo do documento é armazenado num ficheiro de dados separado com a extensão .dat. Se o conteúdo for um ficheiro PDF, pode abrir o ficheiro .dat com um leitor de PDF.

Se seleccionar a opção para utilizar nomes de campos ou caminhos de facetas como elementos XML, então a extensão do ficheiro corresponde à extensão do ficheiro original, se for possível determiná-la, em vez de .dat.

Os nomes de campos de metadados para documentos pesquisados exportados corresponde ao campo de metadados original como é definido na origem de dados, como por exemplo o nome da coluna de uma tabela numa base de dados relacional.

Documentos analisados

Quando exportar documentos analisados, os metadados de documentos e o conteúdo pode ser armazenado num ficheiro XML com a extensão .xml. Como o conteúdo é convertido em texto simples quando os documentos são interpretados e analisados, os metadados e o conteúdo podem ser intercalados no mesmo ficheiro XML.

Quando activar CAS como opção de formato XMI, as informações armazenadas na estrutura de análise comum (CAS) são convertidas para o formato XML Metadata Interchange (XMI) e exportadas como ficheiros XMI com a extensão .xmi.

Os nomes dos campos de metadados para documentos analisados exportados é um nome do campo de índice correlacionado. Apenas campos configurados para serem Devolvidos na definição do campo de índice são exportados.

Os documentos analisados também têm notificações que foram adicionadas aos documentos pelos anotadores e outros processos linguísticos e de análise. Apenas as anotações configuradas para serem indexadas como facetas ou campos de índice são incluídas no ficheiro de saída.

Documentos procurados

Quando exporta resultados de uma aplicação de procura empresarial ou de prospecção de análise de conteúdo, pode seleccionar se pretende exportar os resultados do documento pesquisado, os resultados do documento analisado, ou ambos os tipos de resultados.

Estruturas de directório para documentos exportados

Quando os documentos são exportados como XML, o nome do directório de saída é baseado na altura em que ocorre a exportação. Por exemplo, se a exportação começar a 2009/06/11 às 13:00, o nome do directório é 200906111300.

O programa de exportação guarda até 1.000 ficheiros de exportação no directório. Se existem mais de 1.000 ficheiros para exportação, o programa de exportação cria subdirectórios e guarda até 1.000 ficheiros em cada directório. Estes directórios são denominados em sequência, a começar com o número 0.

Os ficheiros de saída também se denominam em sequência, a começar pelo número 0. Sequências diferentes identificam a saída de dados (.dat) e XML (.xml).

Por exemplo:

Caminho de saída exemplo para documentos pesquisados	Caminho de saída exemplo para documentos analisados
Pesquisados	Analisados
content	CAS
200906111300	200906111200
0	exported typesystem.xml
000.dat	0
001.dat	000.xmi
•••	001.xml
999.dat	
1	999.xmi
000.dat	content
001.dat	200906111300
	0
999.dat	001.xml
metadata	002.xm]
200906111300	
0	999.xm]
000.xml	1
001.xml	- 001.xm]
	002.xm]
999.xml	
1	999 xm]
- 000 xml	555 • All 1
001 xm]	
001.700	
 999 xml	
555 • All 1	

Tabela 23. Estruturas de directório exemplo para documentos pesquisados ou analisados

Conceitos relacionados:

"Exportar documentos para utilização noutras aplicações" na página 469

Referências relacionadas:

"Formato de ficheiro XML para documentos exportados"

Formato de ficheiro XML para documentos exportados

Quando exportar documentos para XML, cada documento é exportado como um ficheiro XML separado.

Elementos de XML

O formato do ficheiro XML tem três secções. A primeira secção, Documento, identifica o documento. A segunda secção, Conteúdo, contém informações sobre o

conteúdo do documento. A terceira secção, Metadados, contém informações sobre os metadados do documento, incluindo os nomes dos campos e facetas.

Se seleccionar a opção para utilizar o nome do campo ou caminho da faceta para correlacionar metadados ao exportar documentos, os nomes dos campos e os caminhos das facetas são representados como elementos XML nos ficheiros exportados, uma abordagem que poderá ser útil caso utilize IBM Content Collector. Caso contrário, os nomes dos campos e os caminhos das facetas são representados como atributos na saída XML.

	Obrigatório	
Nome do elemento ou atributo	ou Opcional	Observações
/Document	Obrigatório	Elemento raiz.
/Document@Id Required	Obrigatório	O ID do documento, como o URL.
/Document@Type	Obrigatório	O tipo de documento. Os valores possíveis são NORMAL (para documentos normais), DELETED (para documentos que foram eliminados do espaço de pesquisa) e ERROR (para documentos que devolvem erros).
/Document/Content	Obrigatório	O elemento de segundo nível contém informações sobre o conteúdo do documento.
/Document/ Content@Truncated	Obrigatório	Indica se o conteúdo é truncado pelo pesquisador.
/Document/Content@path	Opcional	O caminho de ficheiro do conteúdo, caso também seja exportado.
/Document/Metadata	Obrigatório	Elemento de segundo nível que contém informações sobre os metadados do documento.
/Document/Metadata/Fields	Opcional	Elemento de terceiro nível que contém informações sobre campos.
/Document/Metadata/ Fields/Field	Opcional	O valor do campo de metadados.
/Document/Metadata/ Fields/Field@Name	Opcional	O nome do campo de metadados.
/Document/Metadata/Facets	Opcional	Elemento de terceiro-nível que contém informações sobre facetas.
/Document/Metadata/ Facets/Facet	Opcional	O nome da faceta, se facultado.
/Document/Metadata/ Facets/Facet/Path	Opcional	O caminho da faceta.
/Document/Metadata/ Facets/Facet@Begin	Opcional	Se esta faceta provém de uma anotação, a posição de carácter que marca o início da anotação.
/Document/Metadata/ Facets/Facet/@End	Opcional	Se esta faceta provém de uma anotação, a posição de carácter que marca o fim da anotação.
/Document/Metadata/ Facets/Facet/Keyword	Opcional	Um valor associado a esta faceta.

Tabela 24. Elementos e atributos na saída de XML para documentos exportados

Exemplo de uma exportação de documentos pesquisados para XML

Este exemplo mostra a saída de XML criada para um documento que foi pesquisado pelo pesquisador de Base de dados JDBC e depois exportado. Todas as colunas pesquisadas são incluídas na saída de XML.

Para ajudar a ler estas informações, alguns elementos são divididos por várias linhas:

```
<?xml version="1.0" encoding="UTF-8"?>
<Document Id="jdbc://jdbc%3Adb2%3A%2F%2Flocalhost%3A50000%2Fexport</pre>
/ADMINISTRATOR.COMPANY/ID/1" Type="NORMAL">
  <Content Truncated="false" Path="D:&#x5C;export&#x5C;crawled&#x5C;20090728161558
   \0\0670.dat" Encoded="false"></Content>
  <Metadata>
    <Fields>
      <Field Name="__$DatabaseName$__">jdbc:db2://localhost:50000/export</Field>
<Field Name="__$TableName$__">ADMINISTRATOR.COMPANY</Field>
<Field Name="__$ContentColumnName$__">DESCRIPTION</Field>
      <Field Name="ID">1</Field>
      <Field Name="NAME">International Business Machines</Field>
      <Field Name="NAME2">IBM</Field>
      <Field Name="CITY">ARMONK</Field>
      <Field Name="STATE">NY</Field>
      <Field Name="COUNTRY">US</Field>
      <Field Name="PHONE">914-499-1900</Field>
    </Fields>
    <Facets></Facets>
  </Metadata>
</Document>
```

Segue-se o ficheiro de conteúdos correspondente do documento exportado. Este ficheiro foi guardado como D:/export/crawled/20090728161558/0670.dat:

A IBM ajudou a tornar pioneiras a tecnologia de informações com os anos, e permanece hoje na dianteira de uma indústria mundial que está a revolucionar a forma em que as empresas, organizações e as pessoas interagem e evoluem.

Exemplo de uma exportação de documentos analisados para XML

Este exemplo mostra a saída de XML criada para documentos que foi pesquisada pelo pesquisador de Base de dados JDBC, analisada na ligação directa de processamento de documentos, e depois exportada. Apenas os campos que foram configurados para serem devolvidos são incluídos na saída de XML.

Para ajudar a ler estas informações, alguns elementos são divididos por várias linhas:

```
<?xml version="1.0" encoding="UTF-8"?>
<Document Id="jdbc://jdbc%3Adb2%3A%2F%2Flocalhost%3A50000%2Fexport
/ADMINISTRATOR.COMPANY/ID/1" Type="NORMAL">
        <Content Truncated="false" Path="D:&#x5C;export&#x5C;analyzed&#x5C;20090728161558
        &#x5C;0&#x5C;0670.dat" Encoded="false">A IBM ajudou a tornar pioneiras a
        tecnologia de informações com os anos, e permanece hoje na dianteira de uma
        indústria mundial que está a revolucionar a forma em que as empresas,
        organizações e as pessoas interagem e evoluem.</Content>
        <Metadata>
        <Fields>
        <Field Name="date">1248765056000</Field>
        <Field Name="$language">en</Field>
        </Field>
        </Field>
        </Field>
        <//right="false">false">field Name="$language">field Name="$language"
```

```
<Field Name="$source">database</Field>
     <Field Name="city">ARMONK</Field>
     <Field Name="name">International Business Machines</Field>
     <Field Name="name2">IBM</Field>
     <Field Name="phone">914-499-1900</Field>
     <Field Name="state">NY</Field>
    </Fields>
    <Facets>
     <Facet>
        <Path>date</Path>
        <Path>2009</Path>
        <Path>7</Path>
        <Path>28</Path>
        <Path>16</Path>
     </Facet>
     <Facet>
        <Path>LOCATION</Path>
        <Path>STATE</Path>
        <Path>CITY</Path>
        <Keyword>ARMONK</Keyword>
     </Facet>
     <Facet>
        <Path>LOCATION</Path>
        <Path>STATE</Path>
        <Keyword>NY</Keyword>
     </Facet>
     <Facet Begin="0" End="10">
        <Path>tkm base word</Path>
        <Path>noun</Path>
        <Path>unk</Path>
        <Keyword>IBM</Keyword>
     </Facet>
      . . .
     <Facet Begin="198" End="209">
        <Path>tkm base word</Path>
        <Path>verb</Path>
        <Keyword>operate</Keyword>
     </Facet>
     <Facet Begin="206" End="216">
        <Path>tkm base word</Path>
        <Path>conj</Path>
        <Keyword>and</Keyword>
     </Facet>
     <Facet Begin="210" End="217">
        <Path>tkm base word</Path>
        <Path>verb</Path>
        <Keyword>thrive</Keyword>
     </Facet>
    </Facets>
  </Metadata>
</Document>
```

Conceitos relacionados:

"Exportar documentos para utilização noutras aplicações" na página 469 **Referências relacionadas**:

"Caminhos de saída para documentos exportados" na página 480

Correlações de bases de dados relacionais para documentos exportados de colecções de procura empresarial

Para exportar documentos para uma base de dados relacional de uma colecção da procura empresarial, tem de criar um ficheiro de correlação de base de dados. O ficheiro de correlação descreve como os metadados nos ficheiros XML, que são criados para os documentos exportados, devem ser correlacionados para tabelas na base de dados.

Restrição: É possível exportar documentos para bases de dados IBM DB2. Antes de exportar documentos, tem de instalar o DB2 Client no servidor IBM Content Analytics with Enterprise Search. Numa instalação distribuída, instale o Cliente DB2 no servidor do pesquisador. Para obter informações acerca das versões suportadas do IBM DB2, consulte *Origens de dados suportadas*.

Antes de começar

Antes de criar um ficheiro de correlação, poderá querer exportar documentos como ficheiros XML para ver como os metadados são correlacionados. Os metadados podem ser correlacionados como campos (metadados ao nível de documentos) ou facetas (metadados ao nível do texto). Devido a diferenças nas correlações de campos, nomes de campos de metadados para os mesmos metadados podem ser diferentes entre exportações de documentos pesquisados e exportações de documentos analisados. Verifique, também, se o valor que deseja inserir numa tabela é representado como um caminho de faceta ou um valor de faceta.

Para exportações de documentos analisados, apenas são exportados os campos configurados para serem devolvidos.

É facultado um ficheiro de correlação exemplo ES_INSTELL_ROOT/default_config/ export_rdb_mapping.xml. Pode modificar este ficheiro e especificá-lo quando configurar opções de exportação na consola de administração. Depois de associar um ficheiro de correlação a uma colecção, o sistema copia o ficheiro para uma pasta do sistema. Se, posteriormente, fizer modificações ao ficheiro, estas só se reflectem se voltar a especificar o ficheiro na consola de administração.

Tabelas de metadados e esquema em estrela

Quando exportar documentos para uma base de dados relacional, o programa de exportação modula a representação XML do documento e insere os dados modulados em tabelas de esquema em estrela.

Nos ficheiros XML para os documentos exportados, os metadados são categorizados em dois tipos:

Metadados ao nível de documentos

Os metadados ao nível de documentos são metadados únicos no documento. Os exemplos deste tipo de metadados são campos e facetas

que são criadas de campos, como a faceta da data do documento. O relacionamento entre um documento e os metadados ao nível de documentos é 1:1.

Metadados ao nível de texto

Os metadados ao nível de texto são metadados que podem ocorrer mais do que uma vez num documento. Um exemplo deste tipo de metadados é uma faceta criada de anotações, como a faceta Part-of-Speech. O relacionamento entre um documento e os metadados ao nível de documentos é *n:n*.

Os dados divididos são inseridos em três tipos de tabelas:

Tabela de factos de documentos

O ID do documento (URI) e as chaves externas para outras tabelas são inseridos nesta tabela. O conteúdo de documentos (binário e analisado por texto) também é inserido numa tabela de factos de documentos.

Tabelas de metadados ao nível de documentos

Os metadados ao nível de documentos são inseridos nestas tabelas. Por exemplo, poderá ter uma tabela para data, localização, modelo e assim por diante.

Tabelas de metadados ao nível de texto

São inseridos metadados ao nível de texto nestas tabelas. Por exemplo, poderá ter tabelas para palavras, frases e assim por diante. Para manter o relacionamento *n:n*, outra tabela faz a ponte para a tabela de factos do documento e a tabela de metadados ao nível de texto.

Para inserir documentos numa base de dados relacional, tem de configurar correlações entre os metadados nos ficheiros XML e nas tabelas de bases de dados. O ficheiro de correlação de bases de dados é um ficheiro que tem três secções:

- Configuração de exportação. Esta secção é necessária.
- Configuração da base de dados.
- Configuração da tabela. Esta secção é diferente para cada tipo de tabela (tabela de factos de documentos, tabela de metadados ao nível de documentos ou tabela de metadados ao nível de texto).

Se uma tabela que definir no ficheiro de correlação não existe na base de dados, o programa de exportação cria-o automaticamente. Se não configurar uma tabela de ponte entre a tabela de factos de documentos e a tabela de metadados ao nível de texto, o programa de exportação cria-o automaticamente.

Secção Configurar Exportação

Todos os elementos, atributos e valores são dependentes de maiúsculas e minúsculas.

	Obrigatório ou		
Nome do elemento ou atributo	Opcional	Observações	
/export	Obrigatóri	oElemento raiz.	
/export/property	Opcional	Utilizado para transmitir propriedades.	
/export/property@name	Opcional	O nome da propriedade de exportação.	
/export/property@value	Opcional	O valor da propriedade de exportação.	

Tabela 25. Exportar elementos de configuração no ficheiro de correlação de bases de dados

As propriedades seguintes são suportadas:

Nome da propriedade	Valor da propriedade
modo	Se substituir for especificado como valor, as tabelas no ficheiro de correlação serão largadas e recriadas quando iniciar a sessão de exportação.
política	Se truncar for especificado como o valor, e o comprimento dos dados for superior ao comprimento da coluna de destino, os dados serão truncados. Caso contrário, o documento não é inserido na tabela.
módulos	O número de módulos a utilizar para inserir ou eliminar dados nas tabelas de metadados ao nível de documentos e tabelas de metadados ao nível de texto. Se este parâmetro for grande, a exportação poderá ser executada mais rapidamente. Este número não pode ultrapassar o número de tabelas. Caso existam duas tabelas de metadados ao nível de documentos e três tabelas de metadados ao nível ext, não serão utilizados mais de dois módulos.
numberOfUnflushedDocs	Depois do número de documentos aqui especificado ser processado, é emitida uma consolidação para a base de dados. Um número baixo pode afectar o rendimento.

Tabela 26. Exportar propriedades de configuração no ficheiro de correlação de bases de dados

Secção de configuração da base de dados

O programa de exportação necessita de criar tabelas na base de dados. O nome de utilizador especificado nesta secção do ficheiro de correlação tem de ter as autorizações seguintes para a base de dados:

- Ligar base de dados
- Criar tabela
- Criar índice

Todos os elementos, atributos e valores são dependentes de maiúsculas e minúsculas.

Tabela 27. Elementos de configuração de bases de dados no ficheiro de correlação de bases de dados

Nome do elemento ou atributo	Obrigatório ou Opcional	Observações
/export/database	Obrigatório	Não aplicável.
/export/database@dburl	Obrigatório	O URL da base de dados.
/export/database@user	Obrigatório	O nome de utilizador a ligar à base de dados.
/export/database@password	Obrigatório	A palavra-passe a ligar à base de dados.
/export/database@driver	Obrigatório	O nome da classe de controladores de JDBC a utilizar para ligar à base de dados.
/export/database@classpath	Obrigatório	O caminho de classes de controladores JDBC a utilizar para ligar à base de dados.

Secção de configuração da tabela de factos de documentos

Todos os elementos, atributos e valores são dependentes de maiúsculas e minúsculas.

Tabela 28. Os elementos de configuração de tabelas de factos de documentos no ficheiro de correlação de bases de dados

Nome do elemento ou atributo	Obrigatório ou Opcional	Observações
/export/database/table	Obrigatório	Não aplicável.
/export/database/table@schema	Obrigatório	Especifique o nome do esquema.
/export/database/table@name	Obrigatório	Especifique o nome da tabela.
/export/database/table@type	Obrigatório	Especifique fact.
/export/database/table/column	Obrigatório	Não aplicável.
/export/database/table/ column@name	Obrigatório	Especifique o nome da coluna.
/export/database/table/ column@key	Obrigatório	Especifique true se a coluna for utilizada para uma chave principal. Caso contrário, especifique false.
/export/database/table/ column@function	Obrigatório	Especifique os atributos seguintes para indicar como é obtido o valor da coluna:
		 id assegura que a base de dados gera e atribui um número que pode ser utilizado como uma chave principal. Utilize este atributo se /export/database/table/column@key for true.
		 docid indica que o ID do documento (URL) é utilizado para o valor. Este atributo é sempre obrigatório.
		 fk indica que esta coluna é utilizada como uma chave externa para manter um relacionamento entre esta tabela e outra tabela. Se especificar este valor, é obrigatório /column/value.
		 text_content indica que o conteúdo analisado por texto é utilizado para o valor. Esta opção é válida para exportações de documentos analisados e pesquisados. Para exportações de documentos pesquisados, é inserido NULL.
		 binary_content indica que o conteúdo em bruto (como pesquisado) é utilizado para o valor. Esta opção é válida para exportações de documentos pesquisados e procurados. Para a exportação de documentos analisados, é inserido NULL. O tipo de coluna tem de ser BLOB.
/export/database/table/ column@type	Obrigatório	Especifique o nome do tipo de dados da coluna como INTEGER ou CHAR. Se o tipo de dados tiver um comprimento, como 10, especifique o comprimento como CHAR(10).

	Obrigatório	
Nome do elemento ou atributo	ou Opcional	Observações
/export/database/table/ column/value	Obrigatório se value@source for column ou a função column@ for fk	Não aplicável.
/export/database/table/ column/value@source	Obrigatório se a função column@ for fk	Especifique column, que indica que o valor surge do valor numa coluna numa tabela. Consulte também @schema, @table e @ name.
/export/database/table/ column/value@schema	Necessário se value@source for column	Especifique o nome do esquema onde tem origem o valor.
/export/database/table/ column/value@table	Necessário se value@source for column	Especifique o nome da tabela onde tem origem o valor.
/export/database/table/ column/value@name	Necessário se value@source for column	Especifique o nome da coluna onde tem origem o valor. Este valor tem de ser uma chave principal de uma tabela especificada com value@schema e value@table.

Tabela 28. Os elementos de configuração de tabelas de factos de documentos no ficheiro de correlação de bases de dados (continuação)

Secção de configuração da tabela de metadados de documentos

Todos os elementos, atributos e valores são dependentes de maiúsculas e minúsculas.

Tabela 29. Elementos de configuração da tabela de metadados de documentos no ficheiro de correlação de bases de dados

Nome do elemento ou atributo	Obrigatório ou Opcional	Observações
/export/database/table	Obrigatório	Não aplicável.
/export/database/table@schema	Obrigatório	Especifique o nome do esquema.
/export/database/table@name	Obrigatório	Especifique o nome da tabela.
/export/database/table@type	Obrigatório	Especifique document.
/export/database/table/column	Obrigatório	Não aplicável.
/export/database/table/ column@name	Obrigatório	Especifique o nome da coluna.
/export/database/table/ column@key	Obrigatório	Especifique true se a coluna for utilizada para uma chave principal. Caso contrário, especifique false.
/export/database/table/ column@function	Obrigatório apenas se column@ key for true	Especifique id para que a base de dados possa gerar e atribuir um número.

Nome do elemento ou atributo	Obrigatório ou Opcional	Observações
/export/database/table/ column@type	Obrigatório	Especifique o nome do tipo de dados da coluna como INTEGER ou CHAR. Se o tipo de dados tiver um comprimento, como 10, especifique o comprimento como CHAR(10).
/export/database/table/ column/value	Obrigatório	Não aplicável.
/export/database/table/ column/value@source	Obrigatório	Especifique onde tem origem o valor da coluna:
		 field indica que o valor tem origem num valor de campo.
		 facet_path indica que o valor tem origem num caminho da faceta.
		 facet_keyword indica que o valor tem origem num valor de faceta.
/export/database/table/ column/value@path	Obrigatório se value@source for facet_path ou facet_keyword	Especifique o caminho da faceta. Utilize um ponto final (.) para concatenar cada caminho. Também pode utilizar um asterisco (*) para especificar o início dos caminhos da faceta. Por exemplo, se quiser inserir o valor 2009, que faz parte da data do caminho da faceta.2009, e o valor 2010, que faz parte da data do caminho da faceta.2010, para a coluna do mesmo ANO, especifique date.* para o valor do caminho da faceta e especifique 2 como o valor value@index. Utilize o asterisco como o último valor depois do ponto final. Se não especificar um asterisco, é utilizado o caminho da faceta exacto que especificar.
/export/database/table/ column/value@index	Necessário se value@source for facet_path	Especifique o nível do caminho da faceta que é utilizado. O índice começa em 0. Por exemplo, se o valor facet_path for date.2009.08.15 e o valor value@index for 1, os pontos de índice para 2009.

Tabela 29. Elementos de configuração da tabela de metadados de documentos no ficheiro de correlação de bases de dados (continuação)

Secção de configuração da tabela de metadados de texto

Todos os elementos, atributos e valores são dependentes de maiúsculas e minúsculas.

Tabela 30. Elementos de configuração da tabela de metadados de texto no ficheiro de correlação de bases de dados

Nome do elemento ou atributo	Obrigatório ou Opcional	Observações
/export/database/table	Obrigatório	Não aplicável.
/export/database/table@schema	Obrigatório	Especifique o nome do esquema.
/export/database/table@name	Obrigatório	Especifique o nome da tabela.

Nome do elemento ou atributo	Obrigatório ou Opcional	Observações
/export/database/table@type	Obrigatório	Especifique text.
/export/database/ table@root_facet	Obrigatório	Especifique o caminho de nível superior de uma faceta onde será inserido um caminho ou valor. O caminho de nível superior é o primeiro caminho, excepto \$. Por exemplo, se o caminho da faceta for \$word.noun.general, o caminho do nível superior é _word.
/export/database/table/column	Obrigatório	Não aplicável.
/export/database/table/ column@name	Obrigatório	Especifique o nome da coluna.
/export/database/table/ column@key	Obrigatório	Especifique true se a coluna for utilizada para uma chave principal. Caso contrário, especifique false.
/export/database/table/ column@function	Obrigatório se column@key for true	Especifique id para que a base de dados possa gerar e atribuir um número.
/export/database/table/ column@type	Obrigatório	Especifique o nome do tipo de dados da coluna como INTEGER ou CHAR. Se o tipo de dados tiver um comprimento, como 10, especifique o comprimento como CHAR(10).
/export/database/table/ column/value	Obrigatório	Não aplicável.
/export/database/table/ column/value@source	Obrigatório	Especifique onde tem origem o valor da coluna:
		 facet_path indica que o valor tem origem num caminho da faceta.
		 facet_keyword indica que o valor tem origem num valor de faceta.
/export/database/table/ column/value@index	Necessário se value@source for facet_path	Especifique o nível do caminho da faceta que é utilizado. O índice começa em 0. Por exemplo, se o valor facet_path for word.noun.general e o valor value@index for 1, o índice indica o nome.

Tabela 30. Elementos de configuração da tabela de metadados de texto no ficheiro de correlação de bases de dados (continuação)

Conceitos relacionados:

"Exportar documentos para utilização noutras aplicações" na página 469 Tarefas relacionadas:

"Exportar documentos pesquisados ou analisados" na página 472

"Exportar documentos a partir de resultados de consulta" na página 475

Referências relacionadas:

🕩 Origens de dados suportadas

Exportar documentos como ficheiros CSV

Ao configurar opções para exportar documentos na consola de administração, um assistente ajuda o utilizador a especificar informações para exportar documentos como ficheiros de valor separado por vírgula (CSV).

Pode importar os ficheiros CSV exportados para aplicações de folha de cálculo como, por exemplo, o Lotus Symphony e Microsoft Excel, para criar relatórios estatísticos. Para além disso, os documentos são exportados como se fossem tabelas de esquema em estrela, o que significa que pode carregar os ficheiros para uma base de dados relacional e depois utilizar uma aplicação de informação empresarial para analisar os dados juntamente com outros dados estruturados.

Caso utilize o IBM Cognos Business Intelligence e tencione exportar documentos para poder gerar relatórios ou efectuar uma análise adicional através de uma base de dados relacional, tem de configurar a integração entre o IBM Content Analytics with Enterprise Search e IBM Cognos BI antes de executar o assistente.

Para exportar documentos como ficheiros CSV:

- 1. Na vista Colecções, expanda a colecção que pretende configurar.
 - Para exportar documentos pesquisados ou analisados, verifique o painel Análise e Indexação e certifique-se de que o processo de análise e indexação está em execução. Faça clique no ícone para exportar conteúdo do documento, bem como metadados.
 - Para exportar os documentos procurados, expanda o painel Procura para uma colecção de procura empresarial ou o painel Análise para uma colecção de análise de conteúdos. Faça clique no ícone para configurar as definições para a exportação de documentos a partir de resultados de procura.
- 2. Faça clique em **Exportar documentos como ficheiros CSV** e, de seguida, faça clique em **Configurar** para iniciar o assistente de exportação. Caso tenha executado anteriormente o assistente, é apresentado o caminho que especificou para a criação de ficheiros CSV.
- 3. Especifique os campos a exportar e como pretende exportá-los. Por predefinição, o ID, o ID do documento (URI) e a Data do documento são exportados como um ficheiros CSV, com o nome doc_fact.csv. O ID é um número automaticamente atribuído, único para cada exportação de documento. Isto significa que o ID é alterado sempre que o documento é exportado.

No caso de documentos pesquisados, todos os campos nativos que estão correlacionados com os campos do índice nesta colecção podem ser exportados. Por exemplo, se uma coluna de uma tabela de base de dados tiver o nome LSTNAME, e se estiver correlacionado com um campo do índice designado lastname, então, nesse caso, a coluna LSTNAME está listada nesta página. No caso de documentos analisados ou procurados, apenas os campos marcados como **Passível de devolução** podem ser exportados.
Configure as seguintes opções para exportar campos:

Não exportar

Especifica que o campo não será exportado quando os documentos forem exportados.

Uma coluna para tabela de factos de documentos

Exporta o campo como uma coluna adicional do ficheiro doc_fact.csv. Por exemplo, se a opção para __\$FileName\$__ estiver seleccionada, o ficheiro doc_fact.csv irá incluir o ID, o ID do documento, a Data do documento e o nome do ficheiro do documento. Um exemplo do conteúdo no ficheiro doc fact.csv é:

1,file:///C:data/xml/apple.xml,1199186392296,apple.xml
2,file:///C:data/xml/banana.xml,1199272792171,banana.xml
3,file:///C:data/xml/citrus.xml,1199272792218,citrus.xml

Uma tabela para dimensão

Exporta o campo como um ficheiro CSV distinto. Por exemplo, filesize.csv é criado para além do ficheiro doc_fact.csv. Caso altere o nome do ficheiro sugerido pelo assistente, certifique-se de que especifica um nome único. Não pode ter dois ficheiros CSV com o mesmo nome.

O ficheiro CSV gerado inclui o ID do documento associado e o valor do campo. Por exemplo, o ficheiro filesize.csv poderá ser: 1,100 2,125 3,98

Os IDs num ficheiro CSV são únicos, mas não são ordenados. Neste exemplo, isso significa que o documento com o ID 1 no ficheiro doc_fact.csv possui o campo __\$FileSize\$__ e o valor do campo é 100.

Caso seleccione a caixa de verificação **Um documento pode ter mais do que um valor para este campo**, é gerado outro ficheiro CSV com o sufixo _brg. Seleccione esta caixa de confirmação caso exista mais do que um campo com o mesmo nome no documento. Por exemplo, existem duas instâncias de campo componente neste documento:

```
URI: file:///C:data/carrepair/001.xml
data:1199186392296
titulo:repair history
cliente:00001
componente:shaft
componente:tire
```

Se exportar o campo componente para o ficheiro component.csv e seleccionar esta caixa de verificação, o ficheiro component_brg.csv também é criado e contém relações entre o ficheiro doc_fact.csv e o ficheiro component.csv. Os exemplos do conteúdo destes ficheiros são:

```
doc_fact.csv
1,file:///C:data/carrepair/001.xml,1199186392296
```

```
component_brg.csv
1,1
1,2
component.csv
1,shaft
2,tire
```

4. Caso esteja a configurar opções de exportação para documentos analisados ou documentos procurados, especifique as facetas que pretende explorar.

Configure as seguintes opções para exportar campos:

Não exportar

Especifica que a faceta não será exportada quando os documentos forem exportados.

Uma tabela para uma dimensão

Exporta todas as subfacetas da faceta para um ficheiro CSV. Caso altere o nome do ficheiro sugerido pelo assistente, certifique-se de que especifica um nome único. Não pode ter dois ficheiros CSV com o mesmo nome.

Um ficheiro CSV que armazena a relação entre este ficheiro e doc_fact.csv é também criado. Por exemplo, se Componente de expressão for seleccionado para ser exportado para um ficheiro cujo nome é phrase.csv, é também criado um ficheiro designado phrase_brg.csv. Os exemplos do conteúdo destes ficheiros são:

doc fact.csv

1,file:///C:/samples/firststep/data/xml/00000000.xml,1199186392296 2,file:///C:/samples/firststep/data/xml/00000001.xml,1199272792171

phrase brg.csv 1,64 1,94 1,101 1,31 2,182 2,170 2,185 2,176 2,134 2,138 phrase.csv 64, noun phrase, nouns, container straw lemon tea 94, noun phrase, adp noun, from ... juice pack 101, noun phrase, nouns, juice pack 31, noun_phrase, mod_noun, lemon ... tea 182, noun phrase, adp noun, of ... thread 170, pred phrase, verb noun, be ... something 185, noun phrase, adp noun, inside ... cup 176, noun_phrase, adp_noun, like ... piece 134, noun phrase, mod noun, tampering ... thread 138, pred phrase, verb noun, tamper ... thread

- 5. Especifique o directório onde os ficheiros CSV devem ser criados e faça clique em **Terminar** para concluir o assistente. O directório tem de existir e permitir acesso para escrita.
- 6. No painel Análise e Indexação, pare e reinicie a análise e a indexação. Caso já existam documentos indexados, tem de reconstruir o índice a partir da memória de cache do documento.
- 7. Opcional: Após a exportação dos documentos, expanda a área Exportar para visualizar o estado dos pedidos de exportação. Por exemplo, poderá ver o número de documentos exportados até ao momento, se o pedido foi concluído e se ocorreram erros.

Conceitos relacionados:

"Exportar documentos para utilização noutras aplicações" na página 469 Tarefas relacionadas:

"Exportar documentos para IBM Cognos BI" na página 466

"Exportar documentos pesquisados ou analisados" na página 472

"Exportar documentos a partir de resultados de consulta" na página 475

Referências relacionadas:

"Formato de ficheiro CSV para documentos exportados"

Formato de ficheiro CSV para documentos exportados

Para cada tarefa de exportação são criados ficheiros CSV como subdirectórios do directório especificado quando o assistente de exportação foi executado.

Esquema do directório

Cada nome de directório tem o formato de uma marca de tempo como, por exemplo, aaaaMMddHHmmss. Por exemplo, se especificou a criação de ficheiros CSV no directório C:\export\csv quando executou o assistente de exportação de CSV, são criados directórios com o formato C:\export\csv\201000712104623 quando se exportarem documentos.

No caso de documentos pesquisados e analisados, é criado um novo directório quando a sessão de análise e indexação é iniciada e os ficheiros CSV são actualizados até a sessão ser parada. No caso de documentos procurados, é criado um directório para cada pedido de exportação.

A codificação de caracteres dos ficheiros CSV gerados é UTF-8.

As entradas são únicas, mas não são ordenadas por ID.

Formato do ficheiro

A codificação de caracteres dos ficheiros CSV gerados é UTF-8.

As entradas são únicas, mas não são ordenadas por ID.

Exemplo

Um exemplo da estrutura do directório e do ficheiro é:

```
CSV
```

```
201000712104623

date_facet.csv

doc_category.csv

doc_fact.csv

doc_subcategory.csv

export_fact.csv

export_metadata.csv

phrase_csv

phrase_brg.csv

word.csv

word_brg.csv

201000712104840

201000712113844
```

Cabeçalhos da coluna

Quando exporta documentos como ficheiros CSV, apenas o ficheiro doc_fact.csv tem cabeçalhos da coluna:

- Os cabeçalhos da coluna para campos exportados num ficheiro CSV são ID, Palavra-Chave. Por exemplo: 4,0range
- Os cabeçalhos da coluna para facetas exportadas num ficheiro CSV são ID, 1.º Caminho da Faceta, 2.º Caminho da Faceta, ... N Caminho da Faceta, Palavra-Chave (o valor da faceta). Por exemplo: 12,word,noun,general,0range
- Os cabeçalhos da coluna que mantêm as relações num ficheiro CSV, como patient_brg_csv, são o ID no ficheiro doc_fact.csv e o ID no ficheiro patient.csv. Por exemplo: 1,12

O ficheiro export_facet.csv apenas temo ID do documento de documentos exportados, como file://C/data/test.txt.

Tarefas relacionadas:

"Exportar documentos como ficheiros CSV" na página 492

Importar ficheiros CSV

Para adicionar conteúdo ao índice sem pesquisar repositórios, pode importar informação sobre os documentos em formato de valores separados por vírgulas (CSV, comma separated values) e tornar a informação pesquisável quando os utilizadores consultarem as colecções.

Para importar ficheiros CSV, execute um assistente na consola de administração e especifique:

- O ficheiro que pretende importar (ou o caminho de directório para vários ficheiros)
- Opções para analisar os dados como, por exemplo, a linha onde começa a leitura, o carácter separador de colunas, quais as linhas a ignorar, etc.
- Como os elementos nos ficheiros CSV devem ser correlacionados a campos e opções de procura para analisar dados de data e dados decimais.
- Se pretende guardar as definições do utilizador para poder executá-las novamente. Ao executar o assistente, o utilizador escolhe se pretende definir novas opções ou reutilizar definições que guardou anteriormente.

Importar ficheiros CSV para uma colecção

Caso pretenda adicionar documentos ao índice sem pesquisar repositórios, pode importar ficheiros em formato de valor separado por vírgula (CSV). Os utilizadores podem consultar este conteúdo da mesma forma que exploram outros conteúdos na colecção.

Para adicionar ficheiros CSV a uma colecção:

- 1. Na vista Colecções, expanda a colecção à qual deseja acrescentar ficheiros CSV.
- 2. Na área Importar, clique em Import CSV files (Importar ficheiros CSV).
- Especifique a localização dos ficheiros que pretende importar. Se especificar um caminho de directório, são importados todos os ficheiros do directório com a extensão .csv. Os ficheiros com mais de 128 colunas ou registos superiores a 512 KB não podem ser importados.

Caminho local

Seleccione esta opção caso o ficheiro CSV se encontre no computador local do utilizador. Faça clique em **Procurar** para seleccionar o ficheiro.

Caminho do Servidor de Controlador

Seleccione esta opção caso o ficheiro, ou os ficheiros, CSV se encontrem no servidor do controlador. Escreva o caminho completo para o ficheiro ou para o directório que contém os ficheiros CSV que pretende adicionar.

4. Especifique se pretende utilizar as predefinições do sistema para importar os ficheiros CSV, reutilizar as definições que guardou quando executou anteriormente o assistente de importação de ficheiros CSV ou reutilizar as definições que estão especificadas num ficheiro de propriedades. Caso reutilize definições guardadas, pode seleccionar as definições que pretende aplicar para importar estes documentos. Caso reutilize definições de um ficheiro de propriedades, faça clique em **Procurar** para seleccionar o ficheiro.

- Especifique como o sistema deve ler os ficheiros CSV. Uma pré-visualização do conteúdo do ficheiro CSV ajuda o utilizador a configurar estas informações. Especifique:
 - O conjunto de codificação de caracteres
 - O carácter ou caracteres separadores de colunas
 - O número da linha onde o analisador deve começar a ler o ficheiro
 - · Se a primeira linha deve ser tratada como informações de cabeçalho
- 6. Correlacione as colunas nos ficheiros CSV com campos de índice, para que os utilizadores possam procurar o conteúdo CSV. Especifique também:
 - Um ID de pedido único (que é utilizado, por exemplo, para supervisionar o pedido de importação e gerar URIs de documento únicos)
 - Se pretende seleccionar a coluna ou colunas que identificam documentos de modo único ou deixar que o sistema gere um identificador único
 - A coluna que pretende utilizar como coluna de data do documento
 - O formato a utilizar para dados de data na coluna de dados ou em colunas correlacionadas a campos de índice de data.Este formato baseia-se na classe Java SimpleDateFormat sensível a maiúsculas. No formato aaaa-MM-dd, aaaa representa o ano, MM representa o mês e dd representa o dia do mês.
 - O formato a utilizar para dados numéricos em colunas que são correlacionados com campos de índice decimais paramétricos. Este formato baseia-se na classe Java DecimalFormat.
 - O fuso horário e locale a utilizar para analisar dados de data e dados decimais
 - O idioma a utilizar para analisar documentos importados. Pode especificar um idioma predefinido para utilizar e, também, especificar que o analisador deve detectar o idioma de origem.
- 7. Especifique se pretende guardar as definições de configuração para reutilizar em outra altura. Para guardar as definições do utilizador, especifique um nome descritivo para que possa seleccioná-lo ao configurar as definições de importação do ficheiro CSV em outra altura. Também pode guardar as definições do utilizador num ficheiro de propriedades e recarregar o ficheiro quando executa novamente o assistente de importação.
- 8. Na área Importar, supervisione o progresso da tarefa de importação. É possível ver o estado da actual tarefa de importação de documentos CSV e das tarefas de importação de documentos CSV à espera de serem processadas.
- 9. Opcional: Abra o histórico de importação de ficheiros CSV para ver informações acerca de tarefas de importação de documentos CSV processadas como, por exemplo, o número de registos lidos e se uma tarefa foi concluída com êxito ou se ocorreram erros.

Se eliminar uma tarefa de importação, todos os documentos que foram adicionados ao índice pela tarefa são eliminados do mesmo.

Integração no Lotus Notes Versão 8

É possível expandir as funções de procura das implementações de cliente do IBM Lotus Notes Versão 8 ao implementar o suplemento de procura do Lotus Notes IBM Content Analytics with Enterprise Search para as colecções de procura empresarial. Este suplemento permite que os utilizadores consultem as colecções a partir da barra de procura do cliente do Lotus Notes.

Para integrar as capacidades de pesquisa do IBM Content Analytics with Enterprise Search com o Lotus Notes, tem de criar o sítio de actualização do suplemento. Os utilizadores podem assim instalar o suplemento nos sistemas de cliente Lotus Notes Versão 8 locais.

Criar o sítio de actualização do suplemento

Para integrar tecnologia de procura do IBM Content Analytics with Enterprise Search com o Lotus Notes Versão 8, tem de implementar o sítio de actualização do Eclipse para pesquisar um servidor da web na sua organização. O sítio de actualização permite aos utilizadores implementar o suplemento de procura para as instalações do cliente locais do Lotus Notes.

Os ficheiros necessários para criar o sítio de actualização são facultados no ficheiro com.ibm.es.notes.search.plugin_9.1.0.zip.

Para criar o sítio de actualização do Eclipse:

- 1. Copie o ficheiro com.ibm.es.notes.search.plugin_9.1.0.zip para um servidor da Web na sua organização.
- 2. Descompacte o conteúdo do ficheiro ZIP para um subdirectório de um directório partilhado no caminho de instalação do servidor da Web.

Por exemplo, para implementar o sítio de actualização no servidor da Web que utilizar para o IBM Content Analytics with Enterprise Search, descompacte o conteúdo do ficheiro zip no directório C:\Program Files\IBM\HTTP Server\htdocs\en_US\search.

Instalar o suplemento de pesquisa no cliente do Lotus Notes versão 8

Para utilizar as funções de pesquisa do IBM Content Analytics with Enterprise Search para o processamento de consultas, pode adicionar o suplemento de pesquisa à barra de pesquisa do Lotus Notes.

O canto superior direito do cliente Lotus Notes inclui uma barra de procura. A barra de procura tem vários suplementos de procura que são fornecidos com a instalação do cliente Lotus Notes básico. Após a instalação do suplemento de procura, pode procurar as colecções de procura IBM Content Analytics with Enterprise Search a partir da barra de procura.

Para instalar o suplemento de procura para o cliente do Lotus Notes:

- 1. Abra o ficheiro NOTES_INSTALL_ROOT\notes.ini, em que NOTES_INSTALL_ROOT é regra geral c:\lotus\notes.
- 2. No fim do ficheiro, adicione uma propriedade denominada OMNIFIND_ENTERPRISE_EDITION_SEARCH_SERVER_URL e especifique o valor

seguinte, em que *hostname* é o nome do sistema central do servidor de pesquisa IBM Content Analytics with Enterprise Search e *port* é a porta do servidor da Web:

http://hostname:port/search/search.do?action=index&q=

- **3**. Guarde e feche o ficheiro.
- 4. Abra a aplicação do cliente Lotus Notes.
- 5. Seleccione Ficheiro (File) > Aplicação (Application) > Instalar (Install).
- 6. Seleccione o botão de opção **Procurar por novas características para instalar** (Search for new features to install).
- 7. Fala clique em Adicionar Localização Remota (Add Remote Location).
- 8. Forneça um nome a esta nova localização, como por exemplo procura empresarial. No campo de URL, insira o URL raiz do sítio de actualização. Por exemplo, se descompactar o ficheiro com.ibm.es.notes.search.plugin_8.4.0.150.zip para o subdirectório search no directório raiz inglês do IBM HTTP Server, o URL que especificar é http://hostname:port/search/.
- 9. Faça clique em Terminar (Finish) para guardar o novo sítio remoto. .
- 10. Seleccione a caixa de verificação junto ao nome do sítio e fala clique em **Seguinte (Next)**.
- Reveja o acordo de licença, seleccione Aceito os termos no acordo de licença (I accept the terms in the license agreement) e faça clique em Seguinte (Next).
- 12. Faça clique em **Terminar (Finish)** para instalar o suplemento. Se receber um aviso de que está a instalar um suplemento não assinado, seleccione o botão de opção **Instalar este suplemento (Install this plug-in)** e faça clique em **OK**.
- **13**. Quando solicitado, faça clique em **Sim (Yes)** para reiniciar o cliente Lotus Notes.

Pode agora fazer clique no texto do menu para a barra de pesquisa do Lotus Notes e seleccione **IBM Content Analytics with Enterprise Search** como uma opção de pesquisa. Se inserir um termo da pesquisa e fizer clique no botão **Pesquisar**, o pedido é direccionado para o servidor de pesquisa e os resultados são apresentados na janela do navegador da Web do Lotus Notes.

Integração com o WebSphere Portal

É possível expandir as capacidades de procura do IBM WebSphere Portal ao integrar a tecnologia IBM Content Analytics with Enterprise Search com WebSphere Portal e através da configuração do WebSphere Portal para utilizar esta tecnologia como o motor de procura predefinido.

Pontos de integração

O programa de instalação do IBM Content Analytics with Enterprise Search faculta scripts de configuração para integração com o WebSphere Portal. O script que executar, e as formas em que o pode integrar, dependem da versão do WebSphere Portal que utilizar.

Depois de executar os scripts de configuração, o IBM Content Analytics with Enterprise Search pode integrar-se com o WebSphere Portal de várias formas:

Portlet de pesquisa

O WebSphere Portal faculta aos utilizadores um único ponto de acesso para interagir com aplicações, conteúdo, processos e pessoas. A estrutura do WebSphere Portal activa novas aplicações, denominadas *portlets*, para serem integradas e implementadas sem afectar outras aplicações no portal.

O IBM Content Analytics with Enterprise Search fornece um portlet de exemplo que pode ser utilizado para procurar colecções de procura empresarial. Não é possível implementar o prospecção de análise de conteúdo como uma portlet.

Depois de implementar a portlet de procura no WebSphere Portal, pode utilizar a interface do WebSphere Portal para consultar as colecções e ver os resultados. Através das definições de configuração do WebSphere Portal, pode configurar o aspecto e o funcionamento da portlet do IBM Content Analytics with Enterprise Search da mesma forma que configura outras portlets.

WebSphere Portal Centro de Procura

O Centro de Procura WebSphere Portal faculta um ponto de partida central para pesquisar todas as origens disponibilizadas para pesquisar através do WebSphere Portal. O Centro de Procura e a portlet de pesquisa Universal permite-lhe pesquisar o conteúdo do WebSphere Portal e quaisquer outras colecções que sejam registadas com o Centro de Procura.

Quando executar os scripts de configuração para integrar o IBM Content Analytics with Enterprise Search with WebSphere Portal, a funcionalidade de procura é integrada como um serviço federado que pode ser utilizado nas colecções de consulta e ouras colecções que estão disponíveis no Centro de Procura.

Barra de Pesquisa do WebSphere Portal

O canto superior de todos os temas da interface do WebSphere Portal inclui uma barra Pesquisa. O comportamento predefinido desta barra é direccionar todos os pedidos de procura para o motor de procura predefinido do Search Center. Para utilizar as funções mais potentes do IBM Content Analytics with Enterprise Search para processamento de consultas, pode alterar este comportamento predefinido de forma a que todos os pedidos de pesquisa sejam redireccionados para a portlet de pesquisa do IBM Content Analytics with Enterprise Search.

Pesquisadores Lista de valores geradores

Configure um pesquisador Lista de valores geradores para recolher documentos de servidores WebSphere Portal, como, por exemplo, sítios do WebSphere Portal, sítios do IBM Web Content Manager e Quickr para documentos WebSphere Portal. depois de criar uma colecção na consola de administração, pode utilizar a portlet Procurar ou uma aplicação de procura empresarial para procurar o conteúdo indexado. Para suportar segurança ao nível de documento, tem de utilizar a portlet Procura no WebSphere Portal e não uma aplicação de procura empresarial autónoma.

O pesquisador Web Content Management e o pesquisador WebSphere Portal são fornecidos para suportar migrações de versões anteriores de IBM Content Analytics with Enterprise Search. Quando cria um pesquisador, crie um pesquisador Lista de valores geradores.

Vantagens da integração

O IBM Content Analytics with Enterprise Search melhora o ambiente de pesquisa do WebSphere Portal facultando suporte para pesquisar um intervalo maior de tipos de origem de dados. Com a portlet facultada, pode procurar sítios da Web, para além de todos os outros tipos de origens de dados que são suportados pelo sistema IBM Content Analytics with Enterprise Search do utilizador.

O IBM Content Analytics with Enterprise Search também faculta benefícios em escalabilidade. O Motor de Procura do Portal é útil para pequenas e médias empresas onde um único servidor é suficiente para suportar o volume de trabalho da procura e da recuperação. Para suportar capacidades ao nível empresarial, o IBM Content Analytics with Enterprise Search pode distribuir o volume de trabalho do processamento de consultas em vários servidores.

Procedimentos

Para obter instruções sobre como configurar a integração, consulte a integração de *WebSphere Portal com IBM Content Analytics with Enterprise Search*. As informações incluem:

- Como executar os scripts de configuração fornecidos para o WebSphere Portal, incluindo o modo como efectuar a integração com um sistema em agrupamentos WebSphere Portal.
- Como configurar o Search Center de WebSphere Portal.
- Como configurar a barra de procura para WebSphere Portal e IBM Lotus Quickr.

Referências relacionadas:

Linegração do WebSphere Portal com IBM Content Analytics with Enterprise Search

Integração com o IBM Content Classification

Se utilizar o IBM Content Classification, pode melhorar a qualidade da pesquisa no IBM Content Analytics with Enterprise Search importando dados do Content Classification. Também pode treinar as bases de conhecimento do Content Classification importando documentos analisados que exportar do IBM Content Analytics with Enterprise Search.

Procura empresarial e exploração de conteúdo

Se configurar uma colecção para utilizar o anotador de Content Classification facultado, os metadados podem ser adicionados a documentos quando estes são processados na ligação directa de processamento de documentos. Os metadados podem melhorar a qualidade da procura devolvendo documentos altamente relevantes nos resultados da consulta. Os metadados também podem ser utilizados como facetas, o que pode ajudar o utilizador a limitar os resultados para especificar documentos de interesse.

Treinar bases de conhecimento

Pode exportar documentos do IBM Content Analytics with Enterprise Search e utilizá-los para treinar uma base de conhecimento do Content Classification. Quando exporta documentos de uma aplicação, o sistema exporta um ficheiro que contém vários documentos. Quando os documentos num ficheiro ultrapassarem um determinado limite, o ficheiro é dividido em vários ficheiros. O sistema também exporta um ficheiro catalog.xml que descreve os campos nos documentos.

Se importar os ficheiros XML dos documentos e o ficheiro catalog.xml no Classification Workbench, pode utilizar os dados para treinar bases de conhecimento e planos de decisão. Ao exportar repetidamente documentos com o tempo, pode treinar de forma interactiva e melhorar a classificação no IBM Content Classification.

Anotador do Content Classification

Utilize o anotador de Content Classification para adicionar metadados com base nos campos do IBM Content Classification e classificações por categorias quando o conteúdo é adicionado a uma colecção.

Quando configura IBM Content Classification para IBM Content Analytics with Enterprise Search:

- Activa o anotador de Content Classification para a colecção.
- Configure o anotador especificando quais os campos do Content Classification que pretende utilizar para analisar e pesquisar documentos. Os campos tornam-se, automaticamente, campos de índice no IBM Content Analytics with Enterprise Search. Tem de correlacionar campos de índice para facetas de modo a permitir aos utilizadores explorar o conteúdo analisado na aplicação de procura empresarial e prospecção de análise de conteúdo.
- É possível melhorar a relevância dos resultados da consulta importando as pontuações por categoria e utilizando as pontuações para ajustar a classificação do documento. Os documentos que pertencem às mesmas categorias que os termos da consulta que um utilizador especificar são de classificação mais

elevada. Quando configura as opções do servidor de procura para a colecção, é possível especificar a forma como a pontuação por categoria influencia a pontuação total do documento.

O anotador do Content Classification suporta os idiomas que são suportados pelo Content Classification Workbench.

Conceitos relacionados:

"Anotadores" na página 243

Tarefas relacionadas:

"Activar anotadores para uma colecção" na página 244

Planos de decisão e pontuações por categoria do Content Classification

Se utilizar o IBM Content Classification, os campos e as pontuações por categoria do Content Classification podem ser utilizados para analisar e classificar documentos que adicionar ao índice do IBM Content Analytics with Enterprise Search.

Para utilizar esta funcionalidade, o IBM Content Classification tem de estar instalado num servidor do IBM Content Analytics with Enterprise Search. Numa configuração de servidor distribuído, instale o servidor Content Classification no servidor de controlador.

Para melhorar o desempenho do processamento de documentos, pode clonar o servidor do Content Classification em cada servidor de processamento de documentos. Especifique localhost como o nome do sistema central no URL do servidor Content Classification quando configurar as opções do Content Classification.

Se activar as opções para importar pontuações por categorias e utilizar pesquisa conceptual, o servidor do Content Classification tem de ser clonado em todos os servidores de pesquisa. Especifique localhost como o nome do sistema central no URL do servidor Content Classification quando configurar as opções do Content Classification.

Também tem de se certificar de que a colecção está configurada para utilizador o anotador de Content Classification. Quando configura anotadores para a colecção, certifique-se de que a caixa de verificação para este anotador está seleccionada.

Planos e campos de decisão

Ao configurar as opções de análise e indexação para uma colecção ou para uma colecção de análise, é possível associar a colecção a um plano de decisão do Content Classification e seleccionar os campos do Content Classification a incluir como campos de índice na colecção. Para utilizar o Content Classification para analisar e classificar o conteúdo, configure os campos de índice a serem analisados.Correlacione os campos de índice com facetas para que os utilizadores possam explorar os resultados das análises quando consultam a colecção.

As regras no plano de decisão especificam condições que podem fazer com que um documento seja sinalizado, como um padrão de palavras específicas ou um campo paramétrico com um valor específico. Quando os documentos pesquisados são processados para o índice, o servidor do Content Classification pode adicionar

campos de metadados a documentos. Os utilizadores podem consultar campos de metadados adicionais da mesma forma que consultam outras facetas ou campos de índice.

Pontuações por categoria

Ao adicionar um plano de decisão a uma colecção, pode especificar que pretende importar as pontuações por categoria das bases de conhecimento do Content Classification que estão associadas ao plano de decisão. As pontuações de categorias podem influenciar a forma como os documentos são classificados nos resultados. Os documentos que pertencem às categorias mais altas do Content Classification têm maior classificação do que outros documentos que correspondam às condições de consulta. Quando configura as opções do servidor de procura para a colecção, é possível especificar a forma como a pontuação por categoria influencia a pontuação total do documento.

Procura conceptual

Se importar as pontuações por categoria, é possível seleccionar activar a procura conceptual quando configura as opções do servidor de procura para a colecção. Ao contrário da procura tradicional por palavra-chave, que procura documentos que contêm a palavra ou frase de procura efectiva, uma procura conceptual localiza documentos que estão relacionados com um conceito.

Ao devolver documentos relacionados com o mesmo conceito, uma pesquisa conceptual pode reduzir o risco de faltarem documentos importantes que estão relacionados com um tópico só porque a consulta não contém as palavras-chave que correspondem ao conteúdo do documento. As categorias do Content Classification agrupam documentos de forma conceptual e podem ajudar a assegurar que os documentos que pertencem às categorias têm uma classificação mais elevada nos resultados da procura.

Quando permitir uma pesquisa conceptual na consola de administração, pode configurar a pontuação do limiar para excluir documentos dos resultados da pesquisa. Esta pontuação é uma pontuação relevante calculada da pontuação do Content Classification para o documento e a pontuação do Content Classification para a consulta.

Com a pesquisa conceptual, as consultas devem estar entre aspas. Por exemplo, "Um dispositivo para analisar texto". Os documentos que contêm muitas ou a maioria das palavras que são especificadas na consulta encontram-se nos primeiros resultados. Além disso, as consultas que incluem frases longas em oposição a poucas palavras podem ser mais eficazes e devolver resultados mais relevantes.

Conceitos relacionados:

"Influenciar o rendimento da consulta e a correcção ortográfica" na página 286 Tarefas relacionadas:

"Configurar os campos de índice e pontuações do Content Classification" "Configurar e implementar conjuntos de documentos" na página 239

Configurar os campos de índice e pontuações do Content Classification

Quando configurar as opções de análise e indexação para uma colecção, pode associar um plano de decisão do IBM Content Classification e utilizar campos do Content Classification e pontuações por categoria e anotar o conteúdo à medida que é adicionado ao índice.

Para configurar os campos de índice e pontuações do Content Classification:

- 1. Expanda a colecção que pretende configurar. No painel Analisar e Indexar, clique em **Configurar** > **IBM Content Classification**.
- Na página Servidor de Classificação de Conteúdo, especifique o URL para o servidor do Content Classification. Para um melhor desempenho numa configuração de vários servidores, instale o servidor do Content Classification no servidor de controladores (localhost).

Pode melhorar o processamento de documentos clonando o servidor do Content Classification em todos os servidores de processamento de documentos. Caso tencione importar pontuações por categorias e utilizar as pontuações para classificação de documentos ou pesquisa conceptual, clone também o servidor do Content Classification em todos os servidores de pesquisa.

- Na página Plano de decisão seleccione o plano de decisão que pretende utilizar e, em seguida, seleccione os campos do Content Classification que pretende adicionar como campos de índice na colecção.
- 4. Se pretender utilizar pontuações por categoria das bases de conhecimento do Content Classification que estão associadas a este plano de decisão para influenciar a forma como os documentos são classificados nos resultados da procura, ou se pretender activar o suporte para procura conceptual, seleccione a opção para importar pontuações por categoria e faça clique em **Terminar**.

Restrição: Se seleccionar uma opção para importar pontuações por categoria, um campo de índice denominado icmscore é criado de forma automática. Não edite nem elimine este campo de índice.

- 5. No painel Analisar e Indexar, faça clique em Configurar > Campos de índice. Especifique atributos para pesquisar documentos e ver documentos nos resultados para cada campo de índice que foi criado de um campo do Content Classification. Se seleccionar o quadrado de confirmação Analisável, os valores do campo são processados como conteúdo de documentos.
- 6. No painel Procurar para uma colecção de procura empresarial, ou no painel Análise para uma colecção de análise de conteúdos, faça clique em Configurar
 > Opções do servidor de procura.
- 7. Se pretender utilizar as pontuações por categoria das bases de conhecimentos do Content Classification para influenciar a forma como os documentos são classificados nos resultados da procura, seleccione a caixa de verificação para activar a pontuação baseada em categorias e especifique um valor para controlar a forma como a pontuação por categoria será aplicada quando for

calculada a classificação do documento. Os processos de procura podem utilizar este factor de classificação para assegurar que os documentos que pertencem às primeiras categorias do Content Classification são mais relevantes nos resultados da procura.

Especifique um valor entre 0 - 100. Por exemplo, se a pontuação predefinida for 60, poderá aumentar o valor deste campo para assegurar que os documentos que pertencem às primeiras categorias do Content Classification têm uma classificação mais elevada nos resultados da procura.

8. Se pretende permitir que os utilizadores localizem documentos que são conceptualmente semelhantes, mesmo que os termos da consulta não correspondam exactamente ao conteúdo do documento, seleccione o quadrado de confirmação para activar a procura conceptual e depois especifique uma classificação de limiar para excluir documentos dos resultados. Os processos de consulta podem utilizar esta classificação para determinar relevância de documentos.

Especifique um valor que seja inferior ou igual a 0 e inferior a 1, como 0.25. Se especificar 0, todos os documentos que correspondam às condições de consulta são devolvidos. Se especificar um valor mais elevado, um número menor de documentos é devolvido nos resultados. O truncamento de resultados afecta a forma como os resultados da análise são informatizados no prospecção de análise de conteúdo. Por exemplo, as contagens da faceta diferem porque os valores de faceta são afectados pelo conjunto de documentos a serem contados. Se o valor do limiar for elevado, os documentos que são puramente relevantes para uma consulta são destinos para contagem de facetas. Ao definir este valor como baixo, mais documentos vêm para o âmbito da contagem de facetas.

9. Para que estas alterações se tornem efectivas, reinicie o analisador de colecções e os servidores de procura.

Conceitos relacionados:

"Influenciar o rendimento da consulta e a correcção ortográfica" na página 286 "Planos de decisão e pontuações por categoria do Content Classification" na página 504

Tarefas relacionadas:

"Configurar e implementar conjuntos de documentos" na página 239

Treinar bases de conhecimento do Content Classification

Pode exportar documentos colecção e utilizar os documentos para treinar as bases de conhecimento do IBM Content Classification.

Para exportar os resultados da procura de uma colecção para o IBM Content Classification:

- 1. Expanda a colecção de onde pretende exportar documentos. No painel Procurar para uma colecção de procura empresarial ou no painel Análise para uma colecção de análise de conteúdo, faça clique no ícone para configurar as definições para exportar documentos a partir de resultados de procura.
- Na página Opções para Exportar Documentos Pesquisados, seleccione a opção para exportar documentos como ficheiros XML para o IBM Content Classification.
- 3. Especifique o caminho de saída para onde serão escritos os ficheiros XML exportados. É criado um ficheiro XML para cada documento que exportar. Um ficheiro catalog.xml que descreve os campos nos documentos é criado automaticamente e exportado com os documentos.

- 4. Opcional: Crie um agendamento para controlar quando é que o programa é executado. Na área Exportar do painel Procurar ou do painel Análise, faça clique no ícone para configurar um agendamento para a exportação de documentos.
 - Pode especificar um agendamento geral que controla a frequência com que o programa de exportação é executado. Por exemplo, o programa pode ser executado em dias específicos da semana ou do mês ou em horas específicas do dia. Todos os pedidos agendados que utilizem a agenda geral e estejam na fila são processados à hora especificada.
 - Também pode desactivar e permitir o agendamento para pedidos individuais, configurar agendamentos personalizadas para pedidos individuais ou eliminar pedidos de exportação da fila de pedidos de exportação. Se agendar o pedido para ser executado incrementalmente, só serão exportados os documentos que tiverem sido adicionados ao índice após a última execução do programa de exportação.
- 5. Recolha os documentos que pretende exportar:
 - a. Abra uma aplicação de procura empresarial ou o prospecção de análise de conteúdo e consulte uma colecção. Seleccione facetas e especifique as condições de consulta para restringir os resultados para os documentos que quer exportar.
 - b. Faça clique no ícone para exportar documentos que correspondam à sua consulta e especifique as opções para exportar os documentos. Poderá proceder do seguinte modo:
 - Exporte documentos e metadados de uma cache de documentos antes dos processos de análise de texto serem executados. Neste caso, o conteúdo do documento é exportado como saída binária, separado dos metadados do documento.
 - Exporte documentos e metadados com os resultados de análise de texto, como anotações de análise linguística, classificação e lógica personalizada. Neste caso, o conteúdo do documento é analisado como texto no mesmo ficheiro com os metadados, não sendo exportado como um ficheiro separado.
 - Exporte documentos com os resultados dos análise de texto, incluindo o conteúdo do documento analisado como texto, e exporte o conteúdo do documento da cache do documento como saída binária.

O pedido de exportação é enviado para o servidor de controlador para processamento, e é apresentado o ID do pedido. Utilize este ID de pedido para supervisionar o pedido. Por exemplo, poderá ver se o pedido foi concluído ou cancelar o pedido.

- 6. No IBM Content Classification, utilize o Classification Workbench para importar os ficheiros XML e o ficheiro catalog.xml, bem como criar uma ou mais bases de conhecimento.
- 7. Configure as bases de conhecimento para tomar conhecimento das categorias nos documentos importados.
- 8. Crie um plano de decisão que remeta para as bases de conhecimento, e defina as regras para adicionar campos de metadados a documentos.
- 9. Adicione as bases de conhecimento e o plano de decisão ao servidor do Content Classification.

Tarefas relacionadas:

"Exportar documentos a partir de resultados de consulta" na página 475 **Referências relacionadas**:

"Caminhos de saída para documentos exportados" na página 480

"Formato de ficheiro XML para documentos exportados" na página 481

Melhorar pontuação com base em categorias e classificação de pesquisa conceptual

Para activar a procura conceptual ou classificação baseada em pontuação do IBM Content Classification, pelo menos uma base de dados de conhecimento treinada deve estar contida num plano de decisão. Esta base de dados de conhecimento deve ser configurada para pontuar categorias ao referir-se ao conteúdo de documento que é enviado do IBM Content Analytics with Enterprise Search.

Para obter resultados de procura mais relevantes, tem de ser treinada uma base de conhecimento para tópicos que são relevantes para documentos numa colecção e que tenham um número suficiente de categorias. Cem ou mais categorias é o apropriado para a pesquisa conceptual.

Para criar um grande número de categorias que seja adequado para a colecção:

- Crie uma colecção para reunir documentos exemplo para treinar uma base de conhecimento de Content Classification. Ao escolher documentos exemplo dos mesmos destinos de pesquisa como uma colecção para procura conceptual ou classificação de base de pontuação, poderá ser educada uma base de conhecimento e tornar-se melhor adaptada ao destino.
- 2. Pesquisar, analisar e indexar os documentos exemplo e exportá-los no formato de ficheiros XML para Content Classification.
- **3.** Importe os ficheiros XML com a ferramenta Content Classification Taxonomy Proposer.
- 4. Agrupe os documentos pelo número de categorias especificado, 100 ou mais.
- 5. Exporte os documentos categorizados para o Classification Workbench e treine uma base de conhecimento para a colecção.
- 6. Implemente a base de conhecimento treinada e um plano de decisão que remeta para a base de conhecimento do servidor do Content Classification no sistema IBM Content Analytics with Enterprise Search.
- Crie uma colecção para utilizar o IBM Content Classification e o plano de decisão implementado. Configure a colecção para importar as pontuações de categoria a serem utilizadas na pontuação baseada em categoria e na procura conceptual.

Comandos, códigos de retorno e IDs da sessão

Pode utilizar comandos para diagnosticar problemas, determinar o estado de diferentes partes do sistema, iniciar e parar sessões ou iniciar e parar o sistema.

Numa instalação de múltiplos servidores, pode executar os comandos a partir de qualquer servidor no sistema. Porém, deve executar os comandos a partir do servidor de controlador. O servidor de controlador pode aceder a informações a partir de todos os outros servidores no sistema.

A maioria dos comandos têm os seguintes formatos:

esadmin command_name arguments
esadmin session_ID action -option

Para obter mais informações sobre todos os comandos, introduza esadmin help. Para obter mais informações sobre um comando específico, introduza esadmin *action* help.

Restrição: Inserir mais do que pedido **esadmin** ao mesmo tempo não suportado. O sistema não impede as submissões múltiplas simultâneas do comando, mas não existe qualquer mecanismo para impedir a inconsistência caso colidam operações simultâneas. Se forem executados vários pedidos **esadmin** ao mesmo tempo, uma operação poderá ser bem-sucedida, mas é provável que as outras operações falhem.

As descrições de comandos estão divididas nas seguintes secções:

- "Comandos esadmin que abrangem todo o sistema" na página 512
- "Comandos esadmin específicos do componente" na página 513.

Alguns comandos de estado específicos dos componentes podem devolver informações extensas. Estas secções seguintes descrevem a informação detalhada que pode ser devolvida de vários comandos de estado:

- "Estado do pesquisador da Web" na página 518
- "Estado de espaço de pesquisa para o pesquisador da Web" na página 519
- "Estado detalhado de espaço de pesquisa para o pesquisador da Web" na página 521
- "Estado detalhado para servidores pesquisados pelo pesquisador Web" na página 523
- "Estado do pesquisador que não pesquisa a Web" na página 524
- "Estado de espaço de pesquisa para pesquisadores que não são da Web" na página 527
- "Estado detalhado de espaço de pesquisa para pesquisadores que não são da Web" na página 533
- "Estado do analisador" na página 537
- "Estado de servidor de procura" na página 538

Comandos esadmin que abrangem todo o sistema

Introduza os seguintes comandos numa linha.

Tabela 31. Comandos esadmin que abrangem todo o sistema

Comando	Descrição	
esadmin system startall	Inicia os componentes do sistema em todos os servidores. Inicia o common communication layer (CCL) apenas no servidor local. Para reciclar o CCL, tem de parar e reiniciar manualmente o CCL em cada servidor remoto.	
	Comando de exemplo:	
	esadmin system startall	
esadmin system stopall	Pára os componentes do sistema em todos os servidores. Pára o CCL apenas no servidor local. Para reciclar o CCL, tem de parar e reiniciar manualmente o CCL em cada servidor remoto.	
	Comando de exemplo:	
	esadmin system stopall	
esadmin system checkall	Verifique o estado de todos os componentes em todos os servidores.	
	Comando de exemplo:	
	esadmin system checkall	
esadmin database.nodel export [-intervalo [de_horas_antes],[a_horas_antes]] -fnome csv_file_in_absolute_path	Exporta o histórico de estatísticas da consulta para um ficheiro em formato CSV. Se omitir a opção de intervalo, todos os registos mantidos do histórico são introduzidos no ficheiro. Com a opção de intervalo, os registos do valor <i>de_horas_antes</i> para o valor <i>para_horas_antes</i> são introduzidos no ficheiro.	
	Comando de exemplo: Exportar todos os registos:	
	esadmin database.nodel export -fname /home/esadmin/query_all.csv	
	Comando de exemplo: Exportar registos das últimas 24 horas (desde 24*60 minutos até à hora actual):	
	esadmin database.nodel export -range 24, -fname /home/esadmin/query_last_24_hours.csvesadmin database.nodel export -range 24, -fname /home/esadmin/query_last_24_hours.csv	
	Comando de exemplo: Exportar registos até há 24 horas:	
	esadmin database.nodel export -range ,24 -fname /home/esadmin/query_before_24_hours.csv	
	Mensagens de exemplo e códigos de retorno:	
	FFQC5303I Database session (node1) (sid: database.node1) CCL session exists. PID: 8357 FFQC5314I Ocorreu o seguinte resultado: 0	

Comandos esadmin específicos do componente

Introduza os seguintes comandos numa linha.

Tabela 32. Comandos esadmin específicos do componente

Comando	Descrição	
esadmin <i>id_sessão_pesquisador</i> start	Inicia uma sessão do pesquisador. Este comando não inicia qualquer actividade de pesquisa.	
	Comando de exemplo:	
	esadmin coll.WEB1.esadmin start	
	Mensagens de exemplo e códigos de retorno:	
	FFQC5310I WEBCrawler1 (sid: col1.WEB1.esadmin) não está em execução. FFQC5314I Result: 0	
esadmin <i>ID_sessão_pesquisador</i> startCrawl -options <i>opção</i>	Aplicável apenas para pesquisadores de origem de dados (não Web). Inicia o pesquisador para pesquisar documentos novos, modificados e eliminados. Utilize -options 2 para iniciar uma pesquisa completa ou -options 3 para iniciar a pesquisa apenas de documentos novos e modificados.	
	Comandos de exemplo:	
	esadmin col3.DB21.esadmin startCrawl esadmin col2.DB21.esadmin startCrawl -options 3	
	Mensagens de exemplo e códigos de retorno:	
	FFQC5303I DB2Crawler1 (sid: col3.DB21.esadmin) já está em execução. PID: 23650 FFQC5314I Result: 0	
esadmin id_sessão_pesquisador pause	Interrompe a pesquisa.	
	Comando de exemplo:	
	esadmin col3.DB21.esadmin pause	
	Mensagens de exemplo e códigos de retorno:	
	FFQC5303I DB2Crawler1 (sid: col3.DB21.esadmin) já está em execução. PID: 23650 FFQC5314I Result: 0	
esadmin <i>ID_sessão_pesquisador</i> resume	Retoma a pesquisa.	
	Comando de exemplo:	
	esadmin col3.DB21.esadmin resume	
	Mensagens de exemplo e códigos de retorno:	
	FFQC5303I DB2Crawler1 (sid: col3.DB21.esadmin) já está em execução. PID: 23650 FFQC5314I Result: 0	
esadmin <i>ID_sessão_pesquisador</i> stopCrawl	Pára a pesquisa.	
	Comando de exemplo:	
	esadmin col3.DB21.esadmin stopCrawl	
	Mensagens de exemplo e códigos de retorno:	
	FFQC5303I DB2Crawler1 (sid: col3.DB21.esadmin) já está em execução. PID: 23650 FFQC5314I Result: 0	

Comando	Descrição
esadmin <i>ID_sessão_pesquisador</i> stop	Pára uma sessão do pesquisador.
	Comando de exemplo:
	esadmin col3.DB21.esadmin stop
	Mansagans de exemplo e códigos de retorno:
	FFQC5303I DB2Crawler1 (sid: col3.DB21.esadmin) já está em execução. PID: 23650 FFQC5314I Result: 0
esadmin <i>id_sessão_pesquisador</i> getCrawlerStatus	Cus Obtém o estado do pesquisador. As informações devolvidas dependem do pesquisador ser um pesquisador da Web ou um pesquisador para todas as outras origens de dados.
	Exemplo de uma ferramenta de um pesquisador da Web: esadmin coll.WEB1.esadmin getCrawlerStatus
	Possíveis códigos de retorno e mensagens de um pesquisador da Web:
	FFQC5303I WebCrawler1 (sid: col1.WEB1.esadmin) já está em execução. PID: 23650
	Exemplo para um pesquisador que não pesquisa a Web:
	esadmin col3.DB21.esadmin getCrawlerStatus
	Possíveis códigos de retorno e mensagens de um pesquisador que não pesquisa a Web:
	FFQC5303I db2crawler (sid: db2col.DB2_96945) já está em execução. PID: 5936
	Para mais informações sobre as mensagens de estado devolvido, consulte "Estado do pesquisador da Web" na página 518 e "Estado do pesquisador que não pesquisa a Web" na página 524.
esadmin <i>id_sessão_pesquisador_ds</i> getCrawlSpaceStatus	Obtém um estado do espaço de pesquisa para qualquer pesquisador que não seja o pesquisador da Web.
	Comando de exemplo:
	esadmin col3.DB21.esadmin getCrawlSpaceStatus
	Mensagens de exemplo e códigos de retorno:
	FFQC5303I DB2Crawler1 (sid: col3.DB21.esadmin) já está em execução. PID: 23650
	Para obter mais informações acerca de mensagens de estado, consulte "Estado do pesquisador que não pesquisa a Web" na página 524.

Comando	Descrição	
esadmin <i>id_sessão_pesquisador_web</i> getCrawlStatus -selections <i>valor</i>	Obtém o estado do espaço de pesquisa geral para o pesquisador da Web.	
	Comando de exemplo:	
	esadmin coll.WEB1.esadmin getCrawlStatus	
	Para obter mais informações sobre mensagens de estado devolvidas, consulte o tópico "Estado do pesquisador da Web" na página 518. Para obter descrições de valores que pode especificar para o parâmetro selections, consulte "Estado de espaço de pesquisa para o pesquisador da Web" na página 519.	
esadmin id_sessão_pesquisador_ds getCrawlSpaceStatusDetail -ts id_servidor_destino	Obtém um estado do espaço de pesquisa para qualquer pesquisador que não seja um pesquisador da Web. Se não especificar a opção de servidor de destino, são devolvidos dados de todos os servidores destino. Por exemplo, se o pesquisador de DB2 pesquisar as bases de dados FOUNTAIN e SAMPLE e não especificar a opção de servidor de destino, é devolvido o estado de todas as tabelas nas bases de dados FOUNTAIN e SAMPLE.	
	Comando de exemplo:	
	esadmin col3.DB21.esadmin getCrawlSpaceStatusDetail -ts FOUNTAIN	
	Mensagens de exemplo e códigos de retorno:	
	FFQC5303I DB2Crawler1 (sid: col3.DB21.esadmin) já está em execução. PID: 23650	
	Para obter mais informações sobre mensagens de estado devolvidas, consulte o tópico "Estado do pesquisador que não pesquisa a Web" na página 524.	
esadmin <i>id_sessão_pesquisador_web</i> getCrawlDetailsPerSite _url_stringselections_num	Obtém o estado detalhado do espaço de pesquisa para o pesquisador da Web.	
-threshold num	Comando de exemplo:	
	esadmin coll.WEB1.esadmin getCrawlDetailsPerSite	
	Para obter mais informações sobre mensagens de estado devolvidas, consulte o tópico "Estado do pesquisador da Web" na página 518. Para obter descrições dos valores que pode especificar para os parâmetros selections e threshold, consulte "Estado detalhado de espaço de pesquisa para o pesquisador da Web" na página 521.	
esadmin monitor	Obtém o estado do analisador.	
<pre>_cid id_colecção</pre>	Comando de exemplo:	
	esadmin monitor getCollectionParserMonitorStatus -cid coll	
	Mensagens de exemplo e códigos de retorno:	
	FFQC5303I Monitor (nodel) (sid: monitor) já está em execução. PID: 12543	
esadmin controller startIndexBuild -cid <i>id_colecção</i>	Inicie uma construção de índice.	
	Comando de exemplo:	
	esadmin controller startIndexBuild -cid col_1	

Comando	Descrição
esadmin controller stopIndexBuild -cid <i>id colecção</i>	Parar uma construção de índice.
	Comando de exemplo:
	esadmin controller stopIndexBuild -cid col_1
esadmin startSearch -cid <i>id_colecção</i>	Inicia os processos de servidor de procura.
	Comando de exemplo:
	esadmin startSearch -cid coll
	Mensagens de exemplo e códigos de retorno:
	FFQC5303I Controller (nodel) (sid: controller) já está em execução. PID: 25917 FFQC5314I Result: 0
esadmin stopSearch -cid <i>id_colecção</i>	Pára os processos de servidor de procura.
	Comando de exemplo:
	esadmin stopSearch -cid coll
	Mensagens de exemplo e códigos de retorno:
	FFQC5303I Controller (node1) (sid: controller) já está em execução. PID: 15292 FFQC5314I Result: 0

Comando	Descrição
esadmin monitor getCollectionSearchMonitorStatus	Obtém o estado do servidor de procura.
-cid id_colecção	Comando de exemplo:
esadmin <i>collection_id</i> . <i>searcher_session_id</i> getSearchStatusDetail	esadmin monitor getCollectionSearchMonitorStatus -cid col_1
esadmin <i>id_sessão_gestor_procura</i> getStatus -cid <i>id_colecção</i>	Mensagens de exemplo e códigos de retorno:
	FFQC5303I Monitor (node1) (sid: monitor) já está em execução. PID: 12649
	Devolve informações detalhadas do estado do índice para uma colecção num determinado servidor de procura. Existe um gestor de procura por cada servidor de procura. Cada sessão de gestor de procura é responsável por supervisionar e operar os índices de procura num servidor de procura específico.
	Comando de exemplo:
	<pre>esadmin col_1.searcher.node1 getSearchStatusDetail</pre>
	Mensagens de exemplo e códigos de retorno:
	FFQC5303I Serviço de procura do índice - Collection col_1 (nodel) (sid: col_1.searcher.nodel) CCL session exists. PID: 5977 FFQC5314I ocorreu o seguinte resultado: <searchstatusdetail> <numberofdocuments>24</numberofdocuments> <lastrefreshedtime>1331618922665</lastrefreshedtime> <indexversion>82</indexversion> <numberofqueryexecutions>40</numberofqueryexecutions> <numberofcachehits> <recentaverageresponsetime>0</recentaverageresponsetime> <recentqueryrate></recentqueryrate></numberofcachehits></searchstatusdetail>
	Comando de exemplo:
	esadmin searchmanager.node1 getStatus -cid col_1
	Mensagens de exemplo e códigos de retorno:
	<pre>FFQC5303I Search Manager (node1) (sid: searchmanager.node1) CCL session exists. PID: 12614 FFQC5314I Ocorreu o seguinte resultado: PID=5977 CacheHits=0 RefreshedTime=1331618922665 IndexVersion=82 QueryRate=0 Port=59855 SessionId=col_1.searcher.node1 CacheHitRate=0.0 Status=1 ResponseTime=0 SessionName=col_1.searcher.node1.1 Para obter mais informações sobre mensagens de estado</pre>
	devolvidas, consulte o tópico "Estado de servidor de procura" na página 538.

Estado do pesquisador da Web

Quando o utilizador executa o comando para obter o estado do pesquisador Web, o comando devolve informação num formato de documento XML. As seguintes informações podem ser devolvidas pelo comando do estado do pesquisador da Web:

```
FFQC5314I Result: <?xml version='1.0' encoding='UTF-8'?>
<CrawlerStatus>
<CrawlerRunLevel Value="Running"/>
<CrawlerThreadStateDist Count="4" Total="200">
<CrawlerThreadState State="FETCHING" Count="100"/>
</CrawlerThreadState State="FETCHING" Count=100>
<ActiveBucketList Count="500">
<ActiveBucket URL="http://w3.ibm.com/"</pre>
                 NumActURLs="355"
                 NumProcURLs="350"
                 TimeRem="5" Duration="1195"/>
</ActiveBucketList>
<CrawlRate Value="75"/>
<RecentlyCrawledURLList Count="40">
<RecentlyCrawledURL URL="http://example.server.com/example.html"/>
<RecentlyCrawledURL URL="http://example.server.com/example.html"/>
<NumURLsThisSession Value="160000"/>
</CrawlerStatus>
```

A tabela seguinte descreve cada elemento de XML e os respectivos atributos possíveis que são devolvidos pelo comando do estado do pesquisador da Web:

Tabela 33. Informações do estado do pesquisador da Web

Elemento	Atributos	Descrição
CrawlerStatus	 CrawlerThreadStateDist ActiveBucketList CrawlRate RecentlyCrawledURLList NumURLsThisSession 	Estado do pesquisador.
CrawlerRunLevel Value	 Cadeia (inglês) "Não iniciada": a sessão do pesquisador existe, mas ainda não recebeu a mensagem de início para processar documentos. "Iniciada": o pesquisador está a iniciar. "Em execução": o pesquisador terminou a inicialização e arranque e está a pesquisar activamente. "Em pausa": o pesquisador foi instruído para suspender a pesquisa activa, mas não para sair. "Interromper": o pesquisador recebeu o sinal para interromper e vai interromper. "Erro": o pesquisador está num estado não recuperável e tem de ser parada e reiniciada para continuar a pesquisar. 	Informações sobre o que o pesquisador está a efectuar.
Estado CrawlerThreadState	Cadeia (inglês)	Actividade do módulo do pesquisador. Este campo mostra o que o módulo ou módulos estão a efectuar.

Tabela 33. Informações do estado do pesquisador da Web (continuação)

Elemento	Atributos	Descrição
ActiveBucket	• URL: String (URL spec)	A actividade actual de um sítio
	O protocolo, o sistema central e a porta cujos URLs estão a ser pesquisados.	
	NumActURLs: Integer (positive)	
	O número de URLs num conjunto quando este foi disponibilizado para pesquisa (activado).	
	NumProcURLs: Integer (nonnegative)	
	O número de URLs de conjunto processados até ao momento, pesquisados ou rejeitados.	
	TimeRem: Integer	
	O número de segundos restantes antes de o conjunto esgotar o tempo de espera.	
	• Duration: Integer (nonnegative)	
	O número de segundos desde que o conjunto foi activado.	
CrawlRate	Value: Integer (nonnegative)	Medida do débito do
	Páginas por segundo a serem pesquisadas (todos os conjuntos combinados).	pesquisador.
RecentlyCrawledURL	URL: String (URL spec)	Uma página pesquisada
	Cadeia que especifica um protocolo, um sistema central, uma porta e um ficheiro pesquisado.	recentemente.
NumURLsThisSession	Value: Integer (nonnegative)	O número de URLs que foram pesquisados desde que esta instância do pesquisador (processo) iniciou a pesquisa.

Estado de espaço de pesquisa para o pesquisador da Web

Quando o utilizador executa o comando para obter o estado do espaço de pesquisa para um pesquisador Web, o comando devolve informação num formato de documento XML. As seguintes informações podem ser devolvidas por um comando de estado do espaço de pesquisa da Web. Os valores do parâmetro selections são máscaras. Por exemplo, para ver as informações do estado da máscara 1 (Número de páginas no armazenamento de dados não processados) e as informações de estado da máscara 2 (Número de sítios descobertos), adicione os valores de máscara e especifique -selections=3.

Tabela 34. Valores de máscara de selecção para um comando de estado do espaço de pesquisa do pesquisador da Web

Bit de máscara	Selecciona
1	Número de páginas na memória de dados não processados.
2	Número de sítios descobertos.
4	Número de sítios com DNS.
8	Número de sítios sem DNS.
16	Número de URLs descobertos.
32	Número de páginas exclusivas guardadas.

Tabela 34. Valores de máscara de selecção para um comando de estado do espaço de pesquisa do pesquisador da Web (continuação)

Bit de máscara	Selecciona
64	Número de URLs pesquisados.
128	Número de URLs que não estão pesquisados.
256	Número de URLs com atraso.
512	Distribuição de códigos de estado de HTTP.

Todos os valores representam totais acumulados para todas as sessões que utilizem a base de dados interna actual:

```
<CrawlStatus>
 <NumPagesInRDS Value="5422386"/>
  <NumSitesDiscovered Value="15332"/>
 <NumSitesWithDNS Value="14832"/>
 <NumSitesWithoutDNS Value="500"/>
 <NumURLsDiscovered Value="15222999"/>
 <NumUniquePagesSaved Value="6234789"/>
 <NumURLsCrawled Value="7800422"/>
 <NumURLsUncrawled Value="7422577"/>
 <NumURLsOverdue Value="14000"/>
 <HTTPCodeDist Count="4" Total="1031000"/>
    <HTTPCode Code="200" Count ="1000000"/>
    <hrp><HTTPCode Code="301" Count ="1000"/>
    <HTTPCode Code="404" Count ="10000"/>
    <http://www.endocodeservice.com/
  </HTTPCode Code="780" Count="20000">
<?CrawlStatus>
```

Os dados devolvidos contêm qualquer um ou todos (possivelmente nenhum) dos seguintes elementos:

Elemento	Atributo	Descrição
CrawlerStatus	 NumPagesInRDS NumSitesDiscovered NumSitesWithDNS NumSitesWithoutDNS NumURLsDiscovered NumUniquePagesSaved NumURLsCrawled NumURLsOverdue 	Informações que podem ser rapidamente obtidas sobre o estado acumulado da pesquisa (todas as sessões).
NumPagesInRDS	 HTTPCodeDist Valor: número inteiro não negativo Quantas páginas estão actualmente na área de publicação (de dados) da memória de dados não processados (RDS) (deste pesquisador apenas). 	Quão cheia a memória de dados não processados (RDS) se está a tornar (das contribuições deste pesquisador apenas).
NumSitesDiscovered	Valor: número inteiro não negativo Quantos sistemas centrais foram descobertos pela pesquisa (ou de valores geradores).	Uma medida da cobertura do pesquisador do domínio a ser pesquisado (contagem de sistema central).

Tabela 35. Informações do espaço do pesquisador para o pesquisador da Web

Elemento	Atributo	Descrição
NumSitesWithDNS	Valor: número inteiro não negativo Quantos sistemas centrais têm endereços de	Uma medida da eficiência possível do pesquisador em obter endereços de IP que são descobertos por nomes de DNS em
	IP associados (processado pelo pesquisador em segundo plano).	URLs.
NumSitesWithoutDNS	Valor: número inteiro não negativo	Uma medida da eficiência possível do pesquieador em obter enderecos de IP que
	Quantos sistemas centrais não têm endereços de IP associados (processado pelo pesquisador em segundo plano).	são descobertos por nomes de DNS em URLs.
NumURLsDiscovered	Valor: número inteiro não negativo	Uma medida da cobertura do pesquisador
	Quantos URLs exclusivos foram visitados pelo pesquisador.	URL).
NumUniquePagesSaved	Valor: número inteiro não negativo	A contribuição deste pesquisador para o tamanho do índice
	Quantas páginas exclusivas foram escritas para o RDS para processamento posterior por outros componentes.	
NumURLsCrawled	Valor: número inteiro não negativo	Uma medida da capacidade do pesquisador
	Quantos URLs exclusivos foram pesquisados pelo pesquisador.	é diferente do número de páginas escritas para o RDS, porque nem todas as páginas pesquisadas são escritas para o RDS.
NumURLsOverdue	Valor: número inteiro não negativo	Uma medida da capacidade do pesquisador de atravessar o espaço da Web
	Quantos URLs exclusivos são elegíveis para serem pesquisados novamente.	ac adavessar o espaço da mes.

Tabela 35. Informações do espaço do pesquisador para o pesquisador da Web (continuação)

Estado detalhado de espaço de pesquisa para o pesquisador da Web

Quando o utilizador executa o comando para obter o estado de espaço da pesquisa para o pesquisador Web, o comando devolve a informação num formato de documento XML. As seguintes informações podem ser devolvidas pelo comando de estado de espaço de pesquisa detalhada. Os valores do parâmetro selections são máscaras. Por exemplo, para ver as informações do estado da máscara 1 (Número de páginas no armazenamento de dados não processados) e as informações de estado da máscara 2 (Número de sítios descobertos), adicione os valores de máscara e especifique -selections=3.

Tabela 36.	Valores de	e máscara	de selecção	para c	o comando	de estado	detalhado	do espaço	de pesq	uisa do
pesquisado	or da Web									

Bit de máscara	Selecciona
1	Número de páginas na memória de dados não processados.
2	Número de sítios descobertos.
4	Número de sítios com DNS.
8	Número de sítios sem DNS.
16	Número de URLs descobertos.
32	Número de páginas exclusivas guardadas.
64	Número de URLs pesquisados.

Bit de máscara	Selecciona
128	Número de URLs que não estão pesquisados.
256	Número de URLs com atraso.
512	Distribuição de códigos de estado de HTTP.

Tabela 36. Valores de máscara de selecção para o comando de estado detalhado do espaço de pesquisa do pesquisador da Web (continuação)

Informações devolvidas de exemplo:

```
<CrawlDetailsPerSite>
             <Site URL=http://w3.ibm.com/">
            <NumURLsDiscovered Value="5422386"/>
            <NumURLsOverdue Value="15332"/>
            <NumURLsCrawled Value="15332"/>
            <NumURLsUncrawled Value="15332"/>
            <NumURLsOverdueBy Threshold="604800" Value="14832"/>
            <NumURLsActivated Value="2200"/>
            <LastActivationTime Value="1076227340"/>
            <LastActivationDuration Value="4300"/>
            <IPAddressList Count="1"/>
                       <IPAddress Value="9.205.41.33"/>
            </IPAddressList>
            <RobotsContent>
                robots content. . .
            </RobotsContent>
            <htps://www.secondensity.com/</htps://www.secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/secondensity.com/s
                         <HTTPCode Code="200" Count ="1000000"/>
                         <HTTPCode Code="301" Count ="1000"/>
<HTTPCode Code="404" Count ="10000"/>
                          <http://www.endocodesection.com/section/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections/sections
            </HTTPCodeDisT>
</CrawlDetailsPerSite>
```

A tabela seguinte descreve cada campo que é devolvido para o estado detalhado do espaço de pesquisa do pesquisador da Web:

Elemento	Atributos	Descrição
CrawlDetailsPerSite	 LastActivationTime: LastActivationDuration: IPAddressList: RobotsContent: HTTPCodeDist: 	As informações podem ser obtidas de forma rápida sobre o estado detalhado de um sítio.
Sítio	URL	URL do sítio da página de raiz
NumURLsDiscovered	Valor	O número de URLs que pode ser descoberto de um sítio.
NumURLsOverdue	Valor	O número de URLs que são elegíveis para voltarem a ser pesquisados a partir do sítio.
NumURLsCrawled	Valor	O número de URLs que são pesquisados para o sítio.
NumURLsUncrawled	Valor	O número de URLs que ainda não foram pesquisados para o sítio.

Elemento	Atributos	Descrição
NumURLsOverdueBy	Limiar, valor: número inteiro (positivo ou negativo) O valor representa o número de URLs que são elegíveis para nova pesquisa. O limiar especifica o tempo durante o qual os URLs estiveram a aguardar nova pesquisa. O limiar é medido como o número de segundos de deslocamento da hora actual. Se o limiar for negativo, significa que uma nova pesquisa dos URLs venceu. Se o limiar for positivo, significa que deverá ocorrer uma nova pesquisa dos URLs.	O número de URLs que se tornaram elegíveis para serem pesquisados de novo há, pelo menos, alguns segundos ou que se tornarão elegíveis dentro do mesmo número de segundos.
NumURLsActivated	Valor	Número de URLs transferidos para a memória durante o último exame deste sítio da Web e disponibilizados para módulos do pesquisador.
LastActivationTime	Valor	O número de segundos desde a época em que os URLs deste sítio foram transferidos para a memória pela última vez.
LastActivationDuration	Valor	O número de segundos que os URLs deste sítio estiveram pela última vez na memória e disponíveis para módulos do pesquisador.
IPAddressList	IPAddress	Todos os endereços de IP conhecidos para o sistema central do servidor deste sítio.
IPAddress	Valor	Endereço de anotação precedida de ponto IPv4 para o sistema central do servidor do sítio da Web.
RobotsContent	Texto	Texto do ficheiro robots, se existir texto.
HTTPCodeDist	HTTPCode	Distribuição de códigos de HTTP das tentativas de transferência deste sítio.
HTTPCode	Código: número inteiro Um código de estado de HTTP ou outro código interno.	Quantas vezes um determinado código de estado de HTTP ocorreu durante o pesquisador deste sítio.

Tabela 37. Informações de estado do espaço de pesquisa para o pesquisador da Web (continuação)

Estado detalhado para servidores pesquisados pelo pesquisador Web

Quando o utilizador executa o comando para obter o estado de espaço da pesquisa para o pesquisador Web, o comando devolve a informação num formato de documento XML. As seguintes informações sobre servidores Web específicos que podem ser pesquisados podem ser devolvidas pelo comando de estado detalhado do espaço de pesquisa.

Informações devolvidas de exemplo:

```
<ServerStatus>
         <Server Name ="www.example.com">
                        <Status>1</Status>
                        <StatusMessage>Running</StatusMessage>
<NumberOfURLs></NumberOfURLs>
                        <NumberOfCrawledURLs>109771</NumberOfCrawledURLs>
                        <NumberOfInsertedURLs>1030</NumberOfInsertedURLs>
                        <NumberOfUpdatedURLs>2053</NumberOfUpdatedURLs>
                        <NumberOfDeletedURLs>2045</NumberOfDeletedURLs>
                        <hr>tatus</hr>
                        <hr/>

                            . . .
                        <StartTime>1100497759852</StartTime>
                        <EndTime></EndTime>
                        <TotalTime>1395055</TotalTime>
        </Server>
</ServerStatus>
```

A tabela seguinte descreve cada campo que é devolvido para o estado detalhado do espaço de pesquisa do pesquisador da Web:

Tabela 38. Informações de estado do espaço de pesquisa para o pesquisador da Web

Elemento	Atributos	Descrição
ServerStatus	Nome	O nome do sistema central do servidor pesquisado.
Estado		O estado do pesquisador para este servidor. Os valores são 0: Inactivo; 1: Em execução; -1: Erro.
StatusMessage		A representação de cadeia para o elemento Status (estado). Idle: Inactivo; Running: Em execução; Error: Erro.
NumberOfURLs		O número de URLs conhecidos para este servidor.
NumberOfCrawledURLs		O número de documentos pesquisados durante a sessão neste servidor.
NumberOfInsertedURLs		O número de documentos inseridos durante esta sessão neste servidor.
NumberOfDeletedURLs		O número de documentos eliminados durante esta sessão neste servidor.
HTTPStatus	Valor (código de retorno HTTP)	O número de documentos com este código de estado HTTP.
StartTime		A hora a que este pesquisador começou a pesquisar este servidor.
EndTime		A hora a que este pesquisador parou de pesquisar este servidor.
TotalTime		O tempo total gasto a pesquisar este servidor, em milissegundos.

Estado do pesquisador que não pesquisa a Web

Quando o utilizador executa o comando para obter o estado do pesquisador para um pesquisador não Web, o comando devolve a informação num formato de documento XML. As seguintes informações podem ser devolvidas pelo comando **getCrawlerStatus** para pesquisadores que não são da Web:

```
FFQC5314I Result: <?xml version='1.0' encoding='UTF-8'?>
<GeneralStatus>
<Status>0</Status>
<StatusMessage>Idle</StatusMessage>
<NumberOfServers>
<NumberOfCompletedServers>1</NumberOfCompletedServers>
<NumberOfTargets>3</NumberOfTargets>
<NumberOfCompletedTargets>3</NumberOfCompletedTargets>
<NumberOfCrawledRecords>115</NumberOfCrawledRecords>
<RunningThreads>0</RunningThreads>
</GeneralStatus>
```

As tabelas seguintes descrevem os elementos XML e os atributos para cada pesquisador, excepto para o pesquisador da Web. Estas informações são devolvidas com o comando do estado do pesquisador.

Tabela 39. Informações de estado do pesquisador para os pesquisadores de NNTP, DB2, base de dados de JDBC e do Notes

Nome do elemento e do atributo	Pesquisador de NNTP	Pesquisadores de bases de dados DB2 e JDBC	Pesquisadores do Notes
Estado	Estado (0, 1, 2, -1)	Estado (0, 1, 2, -1)	Estado (0, 1, 2, -1)
StatusMessage	Estado: 0 - Inactivo, 1 - Em execução, 2 - Em pausa, -1 - Erro	Estado: 0 - Inactivo, 1 - Em execução, 2 - Em pausa, -1 - Erro	Estado: 0 - Inactivo, 1 - Em execução, 2 - Em pausa, -1 - Erro
NumberOfServers	O número de servidores de NNTP no espaço de pesquisa.	O número de bases de dados no espaço de pesquisa.	O número de bases de dados no espaço de pesquisa.
NumberOfCompletedServers	O número de servidores NNTP pesquisados.	O número de bases de dados pesquisadas.	O número de bases de dados pesquisadas.
NumberOfTargets	O número de newsgroups no espaço de pesquisa.	O número de bases de dados no espaço de pesquisa.	O número de vistas e pasta no espaço de pesquisa.
NumberOfCompletedTargets	O número de newsgroups pesquisados.	O número de tabelas pesquisadas.	O número de vistas e pastas pesquisadas.
NumberOfCompletedRecords	O número de artigos pesquisados.	O número de registos pesquisados.	O número de documentos pesquisados.
RunningThreads	O número de módulos do pesquisador.	O número de módulos do pesquisador.	O número de módulos do pesquisador.

Tabela 40. A informação	do estado do	pesquisador p	oara pesquisadores	do Exchange	Server, Co	ontent Manager	e do
Content Integrator							

Nome do elemento e do	Pesquisador do Exchange	Pesquisador do Content	Pesquisador do Content
atributo	Server	Manager	Integrator
Estado	Estado (0, 1, 2, -1)	Estado (0, 1, 2, -1)	Estado (0, 1, 2, -1)
StatusMessage	Estado: 0 - Inactivo, 1 -	Estado: 0 - Inactivo, 1 -	Estado: 0 - Inactivo, 1 -
	Em execução, 2 - Em	Em execução, 2 - Em	Em execução, 2 - Em
	pausa, -1 - Erro	pausa, -1 - Erro	pausa, -1 - Erro
NumberOfServers	O número de servidores do Exchange Server no espaço de pesquisa.	O número de servidores do Content Manager EE no espaço de pesquisa.	O número de repositórios no espaço de pesquisa.

Tabela 40. A informação do estado do pesquisador para pesquisadores do Exchange Server, Content Manager e do Content Integrator (continuação)

Nome do elemento e do atributo	Pesquisador do Exchange Server	Pesquisador do Content Manager	Pesquisador do Content Integrator
NumberOfCompletedServers	O número de servidores do Exchange Server pesquisados.	O número de servidores do Content Manager EE pesquisados.	O número de repositórios pesquisados.
NumberOfTargets	O número de subpastas no espaço de pesquisa.	O número de tipos de item no espaço de pesquisa.	O número de classes no espaço de pesquisa.
NumberOfCompletedTargets	O número de subpastas pesquisadas.	O número de tipos de item pesquisados.	O número de classes de item pesquisadas.
NumberOfCompletedRecords	O número de documentos pesquisados.	O número de documentos pesquisados.	O número de documentos pesquisados.
RunningThreads	O número de módulos do pesquisador.	O número de módulos do pesquisador.	O número de módulos do pesquisador.

Tabela 41. Informação do estado do pesquisador para o Quickr for Domino, Domino Document Manager e para pesquisadores para o sistema de ficheiros UNIX e para o sistema de ficheiros do Windows

Nome do elemento e do atributo	Pesquisador do Quickr for Domino	Pesquisador do Domino Document Manager	Pesquisador do sistema de ficheiros UNIX e Windows
Estado	Estado (0, 1, 2, -1)	Estado (0, 1, 2, -1)	Estado (0, 1, 2, -1)
StatusMessage	Estado: 0 - Inactivo, 1 - Em execução, 2 - Em pausa, -1 - Erro	Estado: 0 - Inactivo, 1 - Em execução, 2 - Em pausa, -1 - Erro	Estado: 0 - Inactivo, 1 - Em execução, 2 - Em pausa, -1 - Erro
NumberOfServers	O número de locais no espaço de pesquisa.	O número de bibliotecas no espaço de pesquisa.	Valor fixo de 1.
NumberOfCompletedServers	O número de locais pesquisados.	O número de bibliotecas pesquisadas.	0 ou 1 se todos os subdirectórios forem pesquisados.
NumberOfTargets	O número de bases de dados de local e bases de dados de sala no espaço de pesquisa.	O número de arquivos no espaço de pesquisa.	O número de subdirectórios no espaço de pesquisa.
NumberOfCompletedTargets	O número de bases de dados de local e de bases de dados de sala.	O número de arquivos pesquisados.	O número de subdirectórios pesquisados.
NumberOfCompletedRecords	O número de documentos pesquisados.	O número de documentos pesquisados.	O número de ficheiros pesquisados.
RunningThreads	O número de módulos do pesquisador.	O número de módulos do pesquisador.	O número de módulos do pesquisador.

Tabela 42. Informações de estado do pesquisador para pesquisadores do WebSphere Portal e Web Content Management

Nome do elemento e do atributo	Pesquisador do WebSphere Portal	Pesquisador do Web Content Management
Estado	Estado (0, 1, 2, -1)	Estado (0, 1, 2, -1)
StatusMessage	Estado: 0 - Inactivo, 1 - Em execução, 2 - Em pausa, -1 - Erro	Estado: 0 - Inactivo, 1 - Em execução, 2 - Em pausa, -1 - Erro

Nome do elemento e do atributo	Pesquisador do WebSphere Portal	Pesquisador do Web Content Management
NumberOfServers	O número de servidores no espaço do pesquisador.	O número de sítios no espaço de pesquisa.
NumberOfCompletedServers	O número de servidores pesquisados.	O número de sítios pesquisados.
NumberOfTargets	O número de servidores no espaço do pesquisador.	O número de sítios no espaço de pesquisa.
NumberOfCompletedTargets	O número de servidores pesquisados.	O número de sítios pesquisados.
NumberOfCompletedRecords	O número de documentos pesquisados.	O número de documentos pesquisados.
RunningThreads	O número de módulos do pesquisador.	O número de módulos do pesquisador.

Tabela 42. Informações de estado do pesquisador para pesquisadores do WebSphere Portal e Web Content Management (continuação)

Estado de espaço de pesquisa para pesquisadores que não são da Web

Quando o utilizador executa o comando para obter o estado de espaço de pesquisa para um pesquisador não Web, o comando regressa a informação num formato de documento XML. As seguintes informações podem ser devolvidas pelo comando getCrawlSpaceStatus para pesquisadores sem ser da Web:

```
FFQC5314I Result: <?xml version='1.0' encoding='UTF-8'?>
<ServerStatus>
        <Server Name ="FOUNTAIN">
            <Status>5</Status>
            <Status>5</Status>
            <StatusP5</Status>
            <NumberOfTargets>1</NumberOfTargets>
            <NumberOfCompletedTargets>1</NumberOfCompletedTargets>
            <NumberOfErrors>0</NumberOfErrors>
            <StattIme>1118354510512</StartTime>
            <ScheduleConfigured>2</ScheduleConfigured>
            <ScheduleTime>1118393377000</ScheduleTime>
            <TotalTime>3874</TotalTime>
            </Server>
</ServerStatus>
```

As tabelas seguintes descrevem os elementos XML e os atributos para cada pesquisador, excepto para o pesquisador da Web. Estas informações são devolvidas com o comando do estado do espaço de pesquisa. Para pesquisadores do Notes, quando o nível de agregação é 0, Server@Name é nome do servidor + nome da base de dados. Quando o nível de agregação é 1, Server@Name é nome do servidor + nome do directório.

Tabela 43. Informações de estado do pesquisador para os pesquisadores de NNTP, DB2, base de dados de JDBC e do Notes

Nome do elemento e do atributo	Pesquisador de NNTP	Pesquisadores de bases de dados DB2 e JDBC	Pesquisadores do Notes
Server@Name	Nome do servidor de newsgroups	Nome da base de dados	Nome da base de dados ou nome do directório

Nome do elemento e do		Pesquisadores de bases	
atributo	Pesquisador de NNTP	de dados DB2 e JDBC	Pesquisadores do Notes
Server/Status	Estado: (0, 1, 2, 3, 4, 5, -1)	Estado (0, 1, 2, 3, 4, 5, -1)	Estado (0, 1, 2, 3, 4, 5, -1)
	 0: Não pesquisado 	 0: Não pesquisado 	 0: Não pesquisado
	• 1: Pesquisa	• 1: Pesquisa	• 1: Pesquisa
	 2: Concluído (não agendado) 	 2: Concluído (não agendado) 	 2: Concluído (não agendado)
	• 3: Em espera	• 3: Em espera	• 3: Em espera
	• 4: Em pausa	• 4: Em pausa	• 4: Em pausa
	• 5: Agendado	 5: Agendado 	 5: Agendado
	• -1: Erro	• -1: Erro	• -1: Erro
Server/StatusMessage	 0: Não pesquisado 1: Pesquisa 2: Concluído (não agendado) 	 0: Não pesquisado 1: Pesquisa 2: Concluído (não agendado) 	 0: Não pesquisado 1: Pesquisa 2: Concluído (não agendado)
	• 3: Em espera	• 3: Em espera	• 3: Em espera
	• 4: Em pausa	• 4: Em pausa	• 4: Em pausa
	 5: Agendado 1: Errer 	 5: Agendado 1: Erres 	 5: Agendado 1: Farra
	• -1: Erro	• -1: Erro	• -1: Erro
Server/NumberOfTargets	O numero de newsgroups no espaço de pesquisa.	O número de bases de dados no espaço de pesquisa.	O numero de vistas e pastas ou directório no espaço de pesquisa.
Server/NumberOf CompletedTargets	O número de newsgroups pesquisados.	O número de tabelas pesquisadas.	O número de vistas e pastas ou directórios pesquisados.
Server/NumberOfErrors	O número de erros.	O número de erros.	O número de erros
Server/StartTime	A hora de início, se aplicável.	A hora de início, se aplicável.	A hora de início, se aplicável.
Server/EndTime	A hora de término, se aplicável.	A hora de término, se aplicável.	A hora de término, se aplicável.
Server/ScheduleConfigured	0, 1, 2	0, 1, 2	0, 1, 2
	 0: O pesquisador não está configurado para agendamento de acordo com os ficheiros de configuração do pesquisador. 1: O pesquieador está 	 0: O pesquisador não está configurado para agendamento de acordo com os ficheiros de configuração do pesquisador. 1: O pesquisador está 	 0: O pesquisador não está configurado para agendamento de acordo com os ficheiros de configuração do pesquisador. 1: O pesquisador está
	configurado para agendamento, mas o agendamento foi desactivado para a sessão	configurado para agendamento, mas o agendamento foi desactivado para a sessão	configurado para agendamento, mas o agendamento foi desactivado para a sessão
	 2: O pesquisador está configurado para agendamento e o agendamento está activado para a sessão 	 2: O pesquisador está configurado para agendamento e o agendamento está activado para a sessão 	 2: O pesquisador está configurado para agendamento e o agendamento está activado para a sessão

Tabela 43. Informações de estado do pesquisador para os pesquisadores de NNTP, DB2, base de dados de JDBC e do Notes (continuação)
Tabela 43	. Informações	de estado	do pesquisad	or para o	s pesquisadores	de NNTF	, DB2,	base d	e dados	de J	IDBC e
do Notes	(continuação)	ł									

Nome do elemento e do atributo	Pesquisador de NNTP	Pesquisadores de bases de dados DB2 e JDBC	Pesquisadores do Notes
Server/ScheduleTime	Agendar hora, se	Agendar hora, se	Agendar hora, se
	aplicável.	aplicável.	aplicável.
Server/TotalTime	O tempo total, se	O tempo total, se	O tempo total, se
	aplicável.	aplicável.	aplicável.
Server/AggregationLevel	0: O pesquisador pesquisa documentos com o modo normal.	0: O pesquisador pesquisa documentos com o modo normal.	 0, 1: 0: O pesquisador do Notes pesquisa documentos com modo normal. (Exceptuando os pesquisador do Note, os outros pesquisadores devolvem sempre 0.) 1: O pesquisador do Notes pesquisa documentos com o modo de directório.

Tabela 44. Pesquisar informação do estado do espaço para pesquisadores do Exchange Server, Content Manager e do Content Integrator

Nome do elemento e do atributo	Pesquisador do Exchange Server	Pesquisador do Content Manager	Pesquisador do Content Integrator
Server@Name	Nome do servidor de Exchange Server.	Servidores do Content Manager EE.	Nome de repositório.
Server/Status	Estado (0, 1, 2, 3, 4, 5, -1)	Estado (0, 1, 2, 3, 4, 5, -1)	Estado (0, 1, 2, 3, 4, 5, -1)
	• 0: Não pesquisado	 0: Não pesquisado 	• 0: Não pesquisado
	• 1: Pesquisa	• 1: Pesquisa	• 1: Pesquisa
	 2: Concluído (não agendado) 	 2: Concluído (não agendado) 	 2: Concluído (não agendado)
	• 3: Em espera	• 3: Em espera	• 3: Em espera
	• 4: Em pausa	• 4: Em pausa	• 4: Em pausa
	• 5: Agendado	• 5: Agendado	• 5: Agendado
	• -1: Erro	• -1: Erro	• -1: Erro
Server/StatusMessage	• 0: Não pesquisado	• 0: Não pesquisado	• 0: Não pesquisado
	• 1: Pesquisa	• 1: Pesquisa	• 1: Pesquisa
	 2: Concluído (não agendado) 	 2: Concluído (não agendado) 	 2: Concluído (não agendado)
	• 3: Em espera	• 3: Em espera	• 3: Em espera
	• 4: Em pausa	• 4: Em pausa	• 4: Em pausa
	• 5: Agendado	• 5: Agendado	• 5: Agendado
	• -1: Erro	• -1: Erro	• -1: Erro
Server/NumberOfTargets	O número de subpastas no espaço de pesquisa.	O número de tipos de item no espaço de pesquisa.	O número de classes de item no espaço de pesquisa.
Server/NumberOf CompletedTargets	O número de subpastas pesquisadas.	O número de tipos de item pesquisados.	O número de classes de item pesquisadas.
Server/NumberOfErrors	O número de erros.	O número de erros.	O número de erros.

Nome do elemento e do atributoPesquisador do Exchang Server		Pesquisador do Content Manager	Pesquisador do Content Integrator	
Server/StartTime	A hora de início, se	A hora de início, se	A hora de início, se	
	aplicável.	aplicável.	aplicável.	
Server/EndTime	A hora de término, se	A hora de término, se	A hora de término, se	
	aplicável.	aplicável.	aplicável.	
Server/ScheduleConfigured	 0, 1, 2 0: O pesquisador não	 0, 1, 2 0: O pesquisador não	 0, 1, 2 0: O pesquisador não	
	está configurado para	está configurado para	está configurado para	
	agendamento de acordo	agendamento de acordo	agendamento de acordo	
	com os ficheiros de	com os ficheiros de	com os ficheiros de	
	configuração do	configuração do	configuração do	
	pesquisador.	pesquisador.	pesquisador.	
	 1: O pesquisador está	 1: O pesquisador está	 1: O pesquisador está	
	configurado para	configurado para	configurado para	
	agendamento, mas o	agendamento, mas o	agendamento, mas o	
	agendamento foi	agendamento foi	agendamento foi	
	desactivado para a	desactivado para a	desactivado para a	
	sessão	sessão	sessão	
	 2: O pesquisador está	 2: O pesquisador está	 2: O pesquisador está	
	configurado para	configurado para	configurado para	
	agendamento e o	agendamento e o	agendamento e o	
	agendamento está	agendamento está	agendamento está	
	activado para a sessão	activado para a sessão	activado para a sessão	
Server/ScheduleTime	Agendar hora, se aplicável.	Agendar hora, se aplicável.	Agendar hora, se aplicável.	
Server/TotalTime	O tempo total, se aplicável.	O tempo total, se aplicável.	O tempo total, se aplicável.	
Server/AggregationLevel	0: O pesquisador pesquisa	0: O pesquisador pesquisa	0: O pesquisador pesquisa	
	documentos com o modo	documentos com o modo	documentos com o modo	
	normal.	normal.	normal.	

Tabela 44. Pesquisar informação do estado do espaço para pesquisadores do Exchange Server, Content Manager e do Content Integrator (continuação)

Tabela 45. Informação do estado do espaço de pesquisa para o Quickr for Domino, Domino Document Manager e para pesquisadores para o sistema de ficheiros UNIX e para o sistema de ficheiros do Windows

Nome do elemento e do atributo	Pesquisador do Quickr for Domino	Pesquisador do Domino Document Manager	Pesquisador do sistema de ficheiros UNIX e Windows
Server@Name	Directório de local	Base de dados da biblioteca	Um valor fixo de sistema central local.
Server/Status	Estado (0, 1, 2, 3, 4, 5, -1) 0: Não pesquisado 1: Pesquisa 2: Concluído (não agendado) 3: Em espera 4: Em pausa 5: Agendado 	Estado (0, 1, 2, 3, 4, 5, -1) • 0: Não pesquisado • 1: Pesquisa • 2: Concluído (não agendado) • 3: Em espera • 4: Em pausa • 5: Agendado	Estado (0, 1, 2, 3, 4, 5, -1) 0: Não pesquisado 1: Pesquisa 2: Concluído (não agendado) 3: Em espera 4: Em pausa 5: Agendado
	• -1: Erro	• -1: Erro	• -1: Erro

Nome do elemento e do atributo	Pesquisador do Quickr for Domino	Pesquisador do Domino Document Manager	Pesquisador do sistema de ficheiros UNIX e Windows
Server/StatusMessage	 0: Não pesquisado 1: Pesquisa 2: Concluído (não agendado) 3: Em espera 4: Em pausa 5: Agendado -1: Erro 	 0: Não pesquisado 1: Pesquisa 2: Concluído (não agendado) 3: Em espera 4: Em pausa 5: Agendado -1: Erro 	 0: Não pesquisado 1: Pesquisa 2: Concluído (não agendado) 3: Em espera 4: Em pausa 5: Agendado -1: Erro
Server/NumberOfTargets	O número de bases de dados de local e bases de dados de sala no espaço de pesquisa.	O número de arquivos no espaço de pesquisa.	O número de subdirectórios no espaço de pesquisa.
Server/NumberOf CompletedTargets	O número de bases de dados de local e de bases de dados de sala.	O número de arquivos pesquisados.	O número de subdirectórios no espaço de pesquisa.
Server/NumberOfErrors	O número de erros.	O número de erros.	O número de erros.
Server/StartTime	A hora de início, se aplicável.	A hora de início, se aplicável.	A hora de início, se aplicável.
Server/EndTime	A hora de término, se aplicável.	A hora de término, se aplicável.	A hora de término, se aplicável.
Server/ScheduleConfigured	 0, 1, 2 0: O pesquisador não está configurado para agendamento de acordo com os ficheiros de configuração do pesquisador. 1: O pesquisador está configurado para agendamento, mas o agendamento foi desactivado para a sessão 2: O pesquisador está configurado para agendamento e o agendamento e o agendamento está activado para a sessão 	 0, 1, 2 0: O pesquisador não está configurado para agendamento de acordo com os ficheiros de configuração do pesquisador. 1: O pesquisador está configurado para agendamento, mas o agendamento foi desactivado para a sessão 2: O pesquisador está configurado para agendamento e o agendamento e o agendamento está activado para a sessão 	 0, 1, 2 0: O pesquisador não está configurado para agendamento de acordo com os ficheiros de configuração do pesquisador. 1: O pesquisador está configurado para agendamento, mas o agendamento foi desactivado para a sessão 2: O pesquisador está configurado para agendamento e o agendamento e o agendamento está activado para a sessão
Server/ScheduleTime	Agendar hora, se aplicável.	Agendar hora, se aplicável.	Agendar hora, se aplicável.
Server/TotalTime	O tempo total, se aplicável.	O tempo total, se aplicável.	O tempo total, se aplicável.
Server/AggregationLevel	0: O pesquisador pesquisa documentos com o modo normal.	0: O pesquisador pesquisa documentos com o modo normal.	0: O pesquisador pesquisa documentos com o modo normal.

Tabela 45. Informação do estado do espaço de pesquisa para o Quickr for Domino, Domino Document Manager e para pesquisadores para o sistema de ficheiros UNIX e para o sistema de ficheiros do Windows (continuação)

Nome do elemento e do atributo	Pesquisador do WebSphere Portal	Pesquisador do Web Content Management
Server@Name	Servidor do WebSphere Portal	URL do valor gerador de procura do Web Content Management
Server/Status	Estado (0, 1, 2, 3, 4, 5, -1)	Estado (0, 1, 2, 3, 4, 5, -1)
	• 0: Não pesquisado	• 0: Não pesquisado
	• 1: Pesquisa	• 1: Pesquisa
	• 2: Concluído (não agendado)	• 2: Concluído (não agendado)
	• 3: Em espera	• 3: Em espera
	• 4: Em pausa	• 4: Em pausa
	• 5: Agendado	• 5: Agendado
	• -1: Erro	• -1: Erro
Server/StatusMessage	• 0: Não pesquisado	• 0: Não pesquisado
	• 1: Pesquisa	• 1: Pesquisa
	• 2: Concluído (não agendado)	• 2: Concluído (não agendado)
	• 3: Em espera	• 3: Em espera
	• 4: Em pausa	• 4: Em pausa
	• 5: Agendado	• 5: Agendado
	• -1: Erro	• -1: Erro
Server/NumberOfTargets	O número de servidores no espaço do pesquisador.	O número de sítios no espaço de pesquisa.
Server/NumberOf CompletedTargets	O número de servidores pesquisados.	O número de sítios pesquisados.
Server/NumberOfErrors	O número de erros.	O número de erros.
Server/StartTime	A hora de início, se aplicável.	A hora de início, se aplicável.
Server/EndTime	A hora de término, se aplicável.	A hora de término, se aplicável.
Server/ScheduleConfigured	0, 1, 2	0, 1, 2
	 0: O pesquisador não está configurado para agendamento de acordo com os ficheiros de configuração do pesquisador. 	 0: O pesquisador não está configurado para agendamento de acordo com os ficheiros de configuração do pesquisador.
	 1: O pesquisador está configurado para agendamento, mas o agendamento foi desactivado para a sessão 	 1: O pesquisador está configurado para agendamento, mas o agendamento foi desactivado para a sessão
	 2: O pesquisador está configurado para agendamento e o agendamento está activado para a sessão 	 2: O pesquisador está configurado para agendamento e o agendamento está activado para a sessão
Server/ScheduleTime	Agendar hora, se aplicável.	Agendar hora, se aplicável.
Server/TotalTime	O tempo total, se aplicável.	O tempo total, se aplicável.
Server/AggregationLevel	0: O pesquisador pesquisa documentos com o modo normal.	0: O pesquisador pesquisa documentos com o modo normal.

Tabela 46. Informações de estado do espaço de pesquisa para os pesquisadores do WebSphere Portal e Web Content Management

Estado detalhado de espaço de pesquisa para pesquisadores que não são da Web

Quando o utilizador executar o comando para obter estado de espaço de pesquisa para pesquisadores não Web, o comando devolve a informação num formato de documento XML. As seguintes informações podem ser devolvidas pelo comando getCrawlSpaceStatusDetail para pesquisadores sem ser da Web:

Tabela 47. Informações detalhadas de estado do espaço de pesquisa para os pesquisadores de NNTP, DB2, base de dados de JDBC e do Notes

Nome do elemento e do atributo	Pesquisador de NNTP	Pesquisadores de bases de dados DB2 e JDBC	Pesquisadores do Notes
Target@Name	Nome do newsgroups	Nome da tabela	Nome da vista ou da pasta
Target@CrawlType	Não aplicável.	0,1 (DB2); 0 (base de dados de JDBC)	0
		• 0: Pesquisa activa (normal)	
		• 1: Pesquisa passiva (DB2 Event Publishing)	
Destino/estado	Estado: (0, 1, 2, 3, 4, -1)	Estado: (0, 1, 2, 3, 4, -1)	Estado: (0, 1, 2, 3, 4, -1)
	• 0: Não pesquisado	• 0: Não pesquisado	 0: Não pesquisado
	• 1: Pesquisa	• 1: Pesquisa	• 1: Pesquisa
	 2: Concluído (não agendado) 	 2: Concluído (não agendado) 	 2: Concluído (não agendado)
	• 3: Em espera	• 3: Em espera	• 3: Em espera
	• 4: Em pausa	• 4: Em pausa	• 4: Em pausa
	• -1: Erro	• -1: Erro	• -1: Erro
Target/StatusMessage	• 0: Não pesquisado	• 0: Não pesquisado	• 0: Não pesquisado
	• 1: Pesquisa	• 1: Pesquisa	• 1: Pesquisa
	 2: Concluído (não agendado) 	 2: Concluído (não agendado) 	 2: Concluído (não agendado)
	• 3: Em espera	• 3: Em espera	• 3: Em espera
	• 4: Em pausa	• 4: Em pausa	• 4: Em pausa
	• -1: Erro	• -1: Erro	• -1: Erro
Target/NumberOf Records	O último artigo no servidor.	O número de registos pesquisados.	O número de documentos pesquisados.
Target/NumberOf CompletedRecords	O número de artigos pesquisados.	O número de registos pesquisados.	O número de documentos pesquisados.
Target/NumberOf InsertedRecords	O número de artigos publicados recentemente.	O número de registos inseridos.	O número de registos inseridos.

Nome do elemento e do atributo	Pesquisador de NNTP	Pesquisadores de bases de dados DB2 e JDBC	Pesquisadores do Notes
Target/NumberOf UpdatedRecords	Não aplicável.	O número de registos actualizados.	O número de registos actualizados.
Target/NumberOf DeletedRecords	Não aplicável.	O número de registos eliminados.	O número de registos eliminados.
Target/StartTime	A data e hora em que o pesquisador foi iniciado pela última vez.	A data e hora em que o pesquisador foi iniciado pela última vez.	A data e hora em que o pesquisador foi iniciado pela última vez.
Target/EndTime	A data e hora em que a pesquisa foi concluída.	A data e hora em que a pesquisa foi concluída.	A data e hora em que a pesquisa foi concluída.
Target/TotalTime	O tempo gasto pelo pesquisador a pesquisar.	O tempo gasto pelo pesquisador a pesquisar.	O tempo gasto pelo pesquisador a pesquisar.
Target/AggregationLevel	0: O pesquisador pesquisa documentos com o modo normal.	0: O pesquisador pesquisa documentos com o modo normal.	 0, 1: 0: O pesquisador pesquisa documentos com o modo normal. O pesquisador pesquisa documentos com o modo normal
Target/LastUpdatedTime	Não aplicável.	Última hora actualizada: • 0: Pesquisa activa (normal) • 1: Pesquisa passiva (DB2 Event Publishing)	Não aplicável.
Target/LastResetTime	Não aplicável.	Estatísticas reinicializadas pela última vez: • 0: Pesquisa activa (normal) • 1: Pesquisa passiva (DB2 Event Publishing)	Não aplicável.

Tabela 47. Informações detalhadas de estado do espaço de pesquisa para os pesquisadores de NNTP, DB2, base de dados de JDBC e do Notes (continuação)

Tabela 48. Informação detalhada do estado do espaço de pesquisa para pesquisadores do Exchange Server, Content Manager e do Content Integrator

Nome do elemento e do atributo	Pesquisador do Exchange Server	Pesquisador do Content Manager	Pesquisador do Content Integrator
Target@Name	Nome da sub-pasta	Nome do tipo de item	Nome da classe de item
Target@CrawlType	0	0	0
Destino/estado	Estado (0, 1, 2, 3, 4, -1) • 0: Não pesquisado • 1: Pesquisa • 2: Concluído (não agendado)	Estado (0, 1, 2, 3, 4, -1) 0: Não pesquisado 1: Pesquisa 2: Concluído (não agendado) 	Estado (0, 1, 2, 3, 4, -1) 0: Não pesquisado 1: Pesquisa 2: Concluído (não agendado)
	 3: Em espera 4: Em pausa -1: Erro 	 3: Em espera 4: Em pausa -1: Erro 	 3: Em espera 4: Em pausa -1: Erro

Nome do elemento e do atributo	Pesquisador do Exchange Server	Pesquisador do Content Manager	Pesquisador do Content Integrator
Target/StatusMessage	• 0: Não pesquisado	• 0: Não pesquisado	• 0: Não pesquisado
	• 1: Pesquisa	• 1: Pesquisa	• 1: Pesquisa
	 2: Concluído (não agendado) 	 2: Concluído (não agendado) 	 2: Concluído (não agendado)
	• 3: Em espera	• 3: Em espera	• 3: Em espera
	• 4: Em pausa	• 4: Em pausa	• 4: Em pausa
	• -1: Erro	• -1: Erro	• -1: Erro
Target/NumberOf Records	Não aplicável.	Não aplicável.	Não aplicável.
Target/NumberOf CompletedRecords	O número de documentos pesquisados.	O número de documentos pesquisados.	O número de documentos pesquisados.
Target/NumberOf InsertedRecords	O número de registos inseridos.	O número de registos inseridos.	O número de registos inseridos.
Target/NumberOf UpdatedRecords	Não aplicável.	O número de registos actualizados.	O número de registos actualizados.
Target/NumberOf DeletedRecords	Não aplicável.	O número de registos eliminados.	O número de registos eliminados.
Target/StartTime	A data e hora em que o pesquisador foi iniciado pela última vez.	A data e hora em que o pesquisador foi iniciado pela última vez.	A data e hora em que o pesquisador foi iniciado pela última vez.
Target/EndTime	A data e hora em que a pesquisa foi concluída.	A data e hora em que a pesquisa foi concluída.	A data e hora em que a pesquisa foi concluída.
Target/TotalTime	O tempo gasto pelo pesquisador a pesquisar.	O tempo gasto pelo pesquisador a pesquisar.	O tempo gasto pelo pesquisador a pesquisar.
Target/AggregationLevel	0: O pesquisador pesquisa documentos com o modo normal.	0: O pesquisador pesquisa documentos com o modo normal.	0: O pesquisador pesquisa documentos com o modo normal.
Target/LastUpdatedTime	Não aplicável.	Não aplicável.	Não aplicável.
Target/LastResetTime	Não aplicável.	Não aplicável.	Não aplicável.

Tabela 48. Informação detalhada do estado do espaço de pesquisa para pesquisadores do Exchange Server, Content Manager e do Content Integrator (continuação)

Tabela 49. Informação detalhada do estado do espaço de pesquisa para o Quickr for Domino, Domino Document Manager e pesquisadores para o sistema de ficheiros UNIX e para sistemas de ficheiros do Windows

Nome do elemento e do atributo	Pesquisador do Quickr for Domino	Pesquisador do Domino Document Manager	Pesquisador do sistema de ficheiros UNIX e Windows
Target@Name	Nome da base de dados PPlace ou nome da base de dados da sala	Nome da base de dados de arquivo	Nome do subdirectório
Target@CrawlType	0	0	0
Destino/estado	Estado (0, 1, 2, 3, 4, -1)	Estado (0, 1, 2, 3, 4, -1)	Estado (0, 1, 2, 3, 4, -1)
	• 0: Não pesquisado	• 0: Não pesquisado	• 0: Não pesquisado
	• 1: Pesquisa	• 1: Pesquisa	• 1: Pesquisa
	 2: Concluído (não agendado) 	 2: Concluído (não agendado) 	 2: Concluído (não agendado)
	• 3: Em espera	• 3: Em espera	• 3: Em espera
	• 4: Em pausa	• 4: Em pausa	• 4: Em pausa
	• -1: Erro	• -1: Erro	• -1: Erro

Nome do elemento e do atributo	Pesquisador do Quickr for Domino	Pesquisador do Domino Document Manager	Pesquisador do sistema de ficheiros UNIX e Windows
Target/StatusMessage	0: Não pesquisado	• 0: Não pesquisado	0: Não pesquisado
	 1: Pesquisa 2: Concluído (não agendado) 	 1: Pesquisa 2: Concluído (não agendado) 	 1: Pesquisa 2: Concluído (não agendado)
	• 3: Em espera	• 3: Em espera	• 3: Em espera
	• 4: Em pausa	• 4: Em pausa	• 4: Em pausa
	• -1: Erro	• -1: Erro	• -1: Erro
Target/NumberOf Records	Não aplicável.	Não aplicável.	Não aplicável.
Target/NumberOf CompletedRecords	O número de documentos pesquisados.	O número de documentos pesquisados.	O número de ficheiros pesquisados.
Target/NumberOf InsertedRecords	O número de registos inseridos.	O número de registos inseridos.	O número de registos inseridos.
Target/NumberOf UpdatedRecords	O número de registos actualizados.	O número de registos actualizados.	O número de registos actualizados.
Target/NumberOf DeletedRecords	O número de registos eliminados.	O número de registos eliminados.	O número de registos eliminados.
Target/StartTime	A data e hora em que o pesquisador foi iniciado pela última vez.	A data e hora em que o pesquisador foi iniciado pela última vez.	A data e hora em que o pesquisador foi iniciado pela última vez.
Target/EndTime	A data e hora em que a pesquisa foi concluída.	A data e hora em que a pesquisa foi concluída.	A data e hora em que a pesquisa foi concluída.
Target/TotalTime	O tempo gasto pelo pesquisador a pesquisar.	O tempo gasto pelo pesquisador a pesquisar.	O tempo gasto pelo pesquisador a pesquisar.
Target/AggregationLevel	0: O pesquisador pesquisa documentos com o modo normal.	0: O pesquisador pesquisa documentos com o modo normal.	0: O pesquisador pesquisa documentos com o modo normal.
Target/LastUpdatedTime	Não aplicável.	Não aplicável.	Não aplicável.
Target/LastResetTime	Não aplicável.	Não aplicável.	Não aplicável.

Tabela 49. Informação detalhada do estado do espaço de pesquisa para o Quickr for Domino, Domino Document Manager e pesquisadores para o sistema de ficheiros UNIX e para sistemas de ficheiros do Windows (continuação)

Tabela 50. Informações detalhadas de estado do espaço de pesquisa para os pesquisadores do WebSphere Portal e Web Content Management

Pesquisador do WebSphere Portal	Pesquisador do Web Content Management
Nome do servidor do WebSphere Portal	O URL do valor gerador de procura que representa o sítio
0	0
Estado: (0, 1, 2, 3, 4, -1)	Estado: (0, 1, 2, 3, 4, -1)
• 0: Não pesquisado	• 0: Não pesquisado
• 1: Pesquisa	• 1: Pesquisa
• 2: Concluído (não agendado)	• 2: Concluído (não agendado)
• 3: Em espera	• 3: Em espera
• 4: Em pausa	• 4: Em pausa
• -1: Erro	• -1: Erro
	Pesquisador do WebSphere Portal Nome do servidor do WebSphere Portal 0 Estado: (0, 1, 2, 3, 4, -1) • 0: Não pesquisado • 1: Pesquisa • 2: Concluído (não agendado) • 3: Em espera • 4: Em pausa • -1: Erro

Nome do elemento e do atributo	Pesquisador do WebSphere Portal	Pesquisador do Web Content Management
Target/StatusMessage	• 0: Não pesquisado	• 0: Não pesquisado
	• 1: Pesquisa	• 1: Pesquisa
	• 2: Concluído (não agendado)	• 2: Concluído (não agendado)
	• 3: Em espera	• 3: Em espera
	• 4: Em pausa	• 4: Em pausa
	• -1: Erro	• -1: Erro
Target/NumberOf Records	Não aplicável.	Não aplicável.
Target/NumberOf CompletedRecords	O número total de registos pesquisados.	O número total de registos pesquisados.
Target/NumberOf InsertedRecords	O número de registos inseridos.	O número de registos inseridos.
Target/NumberOf UpdatedRecords	O número de registos actualizados.	O número de registos actualizados.
Target/NumberOf DeletedRecords	O número de registos eliminados.	O número de registos eliminados.
Target/StartTime	A data e hora em que o pesquisador foi iniciado pela última vez.	A data e hora em que o pesquisador foi iniciado pela última vez.
Target/EndTime	A data e hora em que a pesquisa foi concluída.	A data e hora em que a pesquisa foi concluída.
Target/TotalTime	O tempo gasto pelo pesquisador a pesquisar.	O tempo gasto pelo pesquisador a pesquisar.
Target/AggregationLevel	0: O pesquisador pesquisa documentos com o modo normal.	0: O pesquisador pesquisa documentos com o modo normal.
Target/LastUpdatedTime	Não aplicável.	Não aplicável.
Target/LastResetTime	Não aplicável.	Não aplicável.

Tabela 50. Informações detalhadas de estado do espaço de pesquisa para os pesquisadores do WebSphere Portal e Web Content Management (continuação)

Estado do analisador

Quando o utilizar executa o comando para obter o estado do analisador, o comando devolve a informação num formato de documento XML. As seguintes informações podem ser devolvidas pelo comando de estado do analisador:

FFQC5314I ocorreu o seguinte resultado: <Monitor Type="Parser"> <ParserStatus> <Status>1</Status> </ParserStatus> </Monitor>

A tabela seguinte descreve os elementos de XML para informações que são devolvidas pelo comando de estado do analisador:

Tabela 51. Elementos para o comando de estado do analisador

Elemento	Descrição
Estado	 0: A sessão do analisador para esta colecção está parada. 1: A sessão do analisador para esta colecção está em execução.

Estado de servidor de procura

Quando o utilizador executa o comando para obter o estado do servidor de procura, o comando devolve a informação num formato de documento XML. As seguintes informações podem ser devolvidas pelo comando de estado do servidor de procura:

```
FFQC5314I Result: <?xml version="1.0"?>
<Monitor Type="Search" Count="1">
<SearchStatus Name="Search Manager (nodel)" SearchID=
"searchmanager.nodel" HostName="myComputer.svl.ibm.com">
<Status>1</Status>
</SearchStatus>
</Monitor>
```

A tabela seguinte descreve os elementos de XML das informações devolvidas pelo comando de estado do servidor de procura:

Tabela 52. Elementos do comando de estado do servidor de procura

Elemento	Descrição	
SearchStatusName	O nome e ID da sessão do gestor de procura que está a supervisionar e actualizar o índice de procura desta colecção.	
HostName	O nome do servidor onde está a ser executado o índice de procura.	
Estado • 0 se o índice de procura desta colecção não estiver em execu • 1 se o índice de procura desta colecção estiver em execu		

Está disponível o estado detalhado do servidor de procura. O comando para devolver o estado do servidor de procura pode devolver a seguinte informação:

```
FFQC5303I Search Manager (nodel) (sid: searchmanager.nodel)
já está em execução. PID: 15711
FFQC5314I Result: PID=18390
CacheHits=3
QueryRate=1
Port=44008
SessionId=coll.runtime.node1
CacheHitRate=0.333
ResponseTime=70
Status=1
SessionName=coll.runtime.node1.1
```

A tabela seguinte descreve os itens nas informações devolvidas do comando de estado detalhado do servidor de procura:

Tabela 53. Itens do comando detalhado de estado do servidor de pro	cura
--	------

Item	Descrição
CacheHits	O número de resultados obtido da cache de procura.
QueryRate	O número de consultas recebidas no último intervalo de tempo. Por predefinição, o intervalo de tempo é cinco minutos.
Port	O número de porta utilizado pelo índice de procura para escutar por ou receber consultas.
SessionId	O ID da sessão para o índice de procura desta colecção.
CacheHitRate	O número de resultados obtido da cache de procura como uma percentagem de todos os resultados da procura.
ResponseTime	O tempo de resposta médio em milissegundos para o intervalo de tempo especificado. (A predefinição é cinco minutos.)

Item	Descrição	
Estado	 0 se o índice de procura desta colecção não estiver em execução. 	
	• 1 se o índice de procura desta colecção estiver em execução.	
SessionName	nName O nome da sessão para o índice de procura desta colecção.	

Tabela 53. Itens do comando detalhado de estado do servidor de procura (continuação)

Códigos de retorno de comandos esadmin

Os seguintes códigos podem ser devolvidos para comandos esadmin:

Tabela 54. Códigos de retorno de comandos esadmin

Código	Nome	Descrição
0	CODE_ERROR_NONE	O comando terminou com êxito.
102	CODE_ERROR_INSTANTIATION_EXCEPTION	Ocorreu um erro ao inicializar uma rotina de tratamento de comandos.
103	CODE_ERROR_ACCESS_EXCEPTION	Ocorreu um erro de acesso ilegal ao inicializar uma rotina de tratamento de comandos.
104	CODE_ERROR_EXECUTE_EXCEPTION	
105	CODE_ERROR_THROWABLE	
106	CODE_ERROR_NO_SUCH_METHOD	
107	CODE_ERROR_INVALID_SESSION	
108	CODE_ERROR_INVALID_PARAMETER	
109	CODE_ERROR_SESSION_NOT_RUNNING	

Obtenção de IDs de sessões

Utilize o comando esadmin check para apresentar uma lista de componentes e os respectivos IDs de sessão. A tabela seguinte apresenta uma lista de sessões comuns, os respectivos IDs, o servidor em que se encontram e o estado da sessão.

Tabela 55. Exemplos de nomes de sessão, servidores de origem, IDs de sessão e estados de sessão

	Servidor onde a sessão est	tá	
Sessão	a ser executada	ID de sessão	Estado de sessão
configmanager	servidor de controlador	10433	Iniciado
controlador	servidor de controlador	10464	Iniciado
customcommunication	servidor de controlador	Não aplicável	Não aplicável
identificação	servidor de controlador	10649	Iniciado
supervisor	servidor de controlador	10682	Iniciado
parserservice	servidor de controlador	10718	Iniciado
resource.node1	servidor de controlador	10759	Iniciado
samplecpp	servidor de controlador	10827	Iniciado
sampletest	servidor de controlador	10857	Iniciado
scheduler	servidor de controlador	10889	Iniciado
searchmanager.node1	servidor de controlador	10927	Iniciado
utilities.node1	servidor de controlador	10384	Iniciado

Conceitos relacionados:

"Supervisionar o sistema" na página 423

"Efectuar cópia de segurança e restaurar o sistema" na página 455

E Recolher informações para a análise de problemas

Demonstration Mensagens para o IBM Content Analytics with Enterprise Search

Tarefas relacionadas:

"Supervisionar pesquisadores" na página 423

"Iniciar o sistema" na página 417

"Parar os servidores de colecção ou todo o sistema" na página 419

"Administrar os servidores de procura no modo autónomo" na página 421

Dependência de maiúsculas e minúsculas

Componentes do sistema diferentes, tais como sintaxe da consulta, ligações rápidas, nomes de campos e assim por diante, processam maiúsculas e minúsculas de forma diferente.

Sintaxe de consultas

A procura é sensível a maiúsculas e minúsculas, excepto nos seguintes casos:

Nomes de elementos e de atributos de XML

Sensível a maiúsculas e minúsculas. Os termos e valores de atributos são sensíveis a maiúsculas e minúsculas mesmo em consultas de XML. Por exemplo, neste documento:

```
<book>
<Author>
<Name>Fernando</Name>
<Contact Type="eMail">fer@nando.org</Contact Type>
<Contact Type="Telefone">+1 408 876 4242</Contact Type>
</Author>
<book>
```

As consultas seguintes não devolverão o documento:

- @xmlxp::'autor[Nome ftcontains ("Fernando")]'
- @xmlxp::'//contact[@type="eMail"]'
- @xmlf2::'<autor><nome>Fernando</nome></autor>
- @xmlf2::'<CONTACT TYPE="email">nando</contacto>

Mas consultas seguintes devolverão o documento:

- @xmlxp::'Autor[Nome ftcontains ("fernando")]'
- @xmlxp:://Contact[@Type="email"]'
- @xmlf2::'<Autor><Nome>fernando</Autor><Nome>
- @xmlf2::'<Contact Type="email">nando</Contacto>

Listas de controlo de acessos (ACLs, access control lists) Sensível a maiúsculas e minúsculas.

URLs nos termos docid: e samegroupas:

Sensível a maiúsculas e minúsculas. No entanto, as partes de URL nos termos site: ou url: não são sensíveis a maiúsculas e minúsculas. Por exemplo, num documento com o URL http://www.here.com/HR/:

- A consulta docid:http://www.here.com/hr não devolverá o documento.
- As consultas url:hr e url:HERE devolverão o documento.
- A consulta site:HERE.com devolverá o documento.

Âmbitos

Não sensíveis a maiúsculas e minúsculas. Por exemplo, tanto o âmbito Scope:RESEARCH como scope::research devolverão documentos do âmbito designado Research.

Termos de caracteres globais

Não sensíveis a maiúsculas e minúsculas. Por exemplo, o termo Fe*na*do é equivalente a fe*na*do.

Nomes de campos

Não sensíveis a maiúsculas e minúsculas. Por exemplo, as consultas Title:Expenses, TITLE:expenses, e title:expenses são todas equivalentes. Todos os nomes de campos são insensíveis a maiúsculas e minúsculas se forem originários de um ficheiro de correlação de XML.

Ligações rápidas

Não sensíveis a maiúsculas e minúsculas.

URIs resumidos e definições hierárquicas baseadas em padrões de URI

Os URIs são sensíveis a maiúsculas e minúsculas, mas os nomes de grupos de URI resumidos não são sensíveis a maiúsculas e minúsculas. Não é possível definir dois nomes de grupos de URI resumidos que difiram apenas no tipo de letra.

Nomes de campos

Não sensíveis a maiúsculas e minúsculas. Esta regra aplica-se aos nomes de campos em consultas, correlações de XML, definições de classes hierárquicas e a todas as outras interfaces administrativas onde sejam especificados nomes de campos. Se configurar um nome de campo em maiúsculas ou maiúsculas e minúsculas na interface, este pode ser mudado para minúsculas pelo sistema de modo a aparecer em minúsculas da próxima vez que visualizar a configuração.

Dicionários

Esta regra depende do tipo de dicionário:

- Não sensíveis a maiúsculas e minúsculas. Esta regra aplica-se a dicionários de sugestão de ortografia.
- Sensível a maiúsculas e minúsculas. Esta regra aplica-se a dicionários de sinónimos, a dicionários de palavras de paragem e a dicionários de palavras hierárquicas.

Nomes e IDs de colecção

Sensível a maiúsculas e minúsculas. Se especificar um nome ou ID de colecção na API de administração ou de procura, este terá de corresponder exactamente ao tipo de letra da colecção com esse nome. No entanto, embora os IDs de colecção sejam sensíveis a maiúsculas e minúsculas, não é possível especificar dois IDs de colecção que difiram apenas no tipo de letra. Esta restrição também se aplica a pesquisadores, nomes de origens de dados e IDs.

Aplicação de procura empresarial e prospecção de análise de conteúdo

Sensível a maiúsculas e minúsculas. Não é possível especificar dois nomes de aplicações ou IDs que difiram apenas no tipo de letra.

Detecção e resolução de problemas da administração

A administração inclui processos do pesquisador, analisador, indexação, segurança, análise de texto e procura.

Os problemas com a administração podem incluir scripts de cópia de segurança que não funcionam, problemas de sockets, problemas de portas, sessões que terminam de forma inesperada, etc.

Ferramentas para solucionar problemas

Estão disponíveis ferramentas para solucionar problemas no sistema IBM Content Analytics with Enterprise Search.

Recolher informações para a análise de problemas

É possível utilizar um utilitário para arquivar e enviar ficheiros de registo, dados de configuração e outras informações sobre resolução de problemas para o IBM Support Software para análise e resolução de problemas.

Para reunir ficheiros de registo e outras informações de resolução de problemas, execute o comando **esservice**. Não existe suporte para recolher este tipo de informação na consola de administração.

Executar o utilitário de serviço

Se pretende reunir informações de todos os servidores do sistema, tem de executar o comando **esservice** em cada servidor.

- 1. Inicie sessão como administrador. Numa configuração de servidores múltiplos, inicie sessão no servidor para o qual pretende recolher a informação.
- 2. Introduza o seguinte comando:

AIX ou Linux

\$ES_INSTALL_ROOT/bin/esservice.sh -options

Windows

%ES_INSTALL_ROOT%\bin\esservice.bat -options

Opções:

-help

Fornece informações de ajuda para o comando.

-file zip_FileName

Especifica o nome do ficheiro de arquivo que contém os dados a enviar para o IBM Support Software. Se não especificar esta opção, o utilitário utilizar a seguinte convenção para designar o ficheiro ZIP:

service_component_yyyyMMdd_HHmmssz.zip

em que:

component

Numa configuração de servidores múltiplos, identifica o servidor onde o utilitário de serviço é executado: controlador, pesquisador ou procura.

aaaaaMMdd

É o ano, mês e dia actuais em que o utilitário de serviço é executado.

HHmmss

É a hora, minuto e segundo actuais em que o utilitário de serviço é executado.

z É a distância em fuso horário do GMT quando o utilitário de serviço é executado. Por exemplo, a Hora Local do Pacífico é GMT -0800.

Se especificar o nome de um ficheiro existente, o utilitário de serviço substitui o ficheiro existente com a nova informação de serviço e imprime uma mensagem de aviso sobre o ficheiro substituído.

Sugestão: Se especificar um nome de ficheiro em vez de utilizar a convenção de nomenclatura predefinida, certifique-se de que inclui um identificador de componente no nome do ficheiro. Numa configuração de vários servidores, o identificador de componentes ajuda o IBM Software Support a identificar o ficheiro de arquivo que contém informações sobre um servidor específico. Por exemplo, poderá especificar index.zip, crawler.zip, search1.zip ou search2.zip.

-maxzip N

Faz com que o ficheiro de arquivo de saída seja dividido quando o tamanho dos dados de entrada é maior do que o número especificado (*N*) de GB. O valor predefinido é de 20 GB e o valor mínimo é de 5 GB. Quando o ficheiro de arquivo é dividido, o utilitário insere um número antes da extensão do ficheiro para aumentar o nome do ficheiro. Por exemplo

service_crawler_20070123_101312-0800.zip service_crawler_20070123_101312-0800_2.zip service_crawler_20070123_101312-0800_3.zip

-cid collection_ID_1,collection_ID_2,collection_ID_3

Especifica uma lista separada por vírgulas de IDs de colecção para as colecções que desactivaram ficheiros de registo de documentos que pretende incluir no ficheiro de arquivo.

Sugestão: Para determinar o ID de uma colecção, pode utilizar a consola de administração ou utilizar o comando **esadmin report collections**.

-nocore

Afecta apenas os ficheiros no directório ES_NODE_ROOT/logs. Especifica que não serão incluídos mais ficheiros essenciais C++ no ficheiro de arquivo.

-noheapdump

Afecta apenas os ficheiros no directório ES_NODE_ROOT/logs. Especifica que não serão incluídos ficheiros essenciais de Java no ficheiro de arquivo.

-pattern pattern

Afecta apenas os ficheiros no directório ES_NODE_ROOT/logs. Especifica os critérios de padrão para os ficheiros a incluir no ficheiro de arquivo. Por exemplo:

-pattern command.*.log

Corresponde a todos os ficheiros iniciados por command e terminados por .log.

pattern -[a-z]*_[0-9].log

Corresponde a todos os ficheiros que contêm caracteres alfabéticos

em minúsculas seguidos de um carácter de traço sublinhado e uma sequência de números que terminam com .log.

Se o padrão contém caracteres especiais como um asterisco (*), ponto de interrogação (?)e assim por diante tem de incluir o padrão entre aspas duplas. Para obter mais informações sobre a sintaxe do padrão, consulte http://java.sun.com/j2se/1.4.2/docs/api/java/util/regex/Pattern.html.

-pwd

Faz com que todas as palavras-passe codificadas sejam deixadas como foi especificado no ficheiro ES_INSTALL_ROOT/nodeinfo/es.cfg. Especifique esta opção se o IBM Software Support precisar das palavras-passe para solucionar um problema.

Se não especificar esta opção, o utilitário cria um ficheiro tmp/es.cfg para incluir no arquivo. Se um campo de palavra-passe no ficheiro tmp/es.cfg especificar uma palavra-passe, a palavra-passe é substituída por uma cadeia da letra x (por exemplo, xxxx). Se o campo da palavra-passe estiver em branco, o campo permanece em branco no ficheiro tmp/es.cfg.

Directrizes de utilização

- Pode executar o utilitário de serviço quando o sistema IBM Content Analytics with Enterprise Search está em execução ou parado. Também pode executar o utilitário de serviço num sistema que não esteja totalmente instalado.
- O utilitário reúne a informação sobre o sistema operativo como o número de processadores, a quantidade de memória e espaço livre em disco num ficheiro designado os_system.out, e inclui este ficheiro no arquivo.
- O utilitário do serviço não inclui dados da aplicação de procura empresarial (ES_INSTALL_ROOT/installedApps) nem dados de colecção do arquivo. Por exemplo, o arquivo não inclui metadados do pesquisador, dados de credenciais do utilizador, dados do analisador, ficheiros de índice, dicionários, motores de análise de texto e assim por diante.

Referências relacionadas:

"Comandos, códigos de retorno e IDs da sessão" na página 511

Obter informação sobre erros de CCL

É possível utilizar um utilitário para gerar um relatório que mostre o histórico de todos os pedidos pendentes recebidos pelo servidor de Common Communication Layer (CCL). Este relatório pode ajudar a identificar onde ocorreu um problema, por exemplo, o motivo por que um processo parou de responder.

Também é possível especificar opções para assegurar que um administrador recebe uma notificação por correio electrónico quando ocorre um erro de CCL.

Para gerar um relatório sobre a actividade de CCL, executa um comando **reportccl**. Não existe suporte para gerar este relatório na consola de administração. O IBM Support Software poderá pedir-lhe que crie este relatório para análise e resolução de problemas.

Sugestão: Se executar o utilitário **esservice**, o relatório de CCL é automaticamente incluído no arquivo do ficheiro que é criado quando executar esse utilitário.

Executar o utilitário reportccl

Para gerar um relatório que mostre a actividade de CCL:

- 1. Inicie sessão como administrador. Numa configuração de vários servidores, inicie sessão em qualquer servidor do sistema.
- 2. Certifique-se de que o servidor de CCL está em execução.
- 3. Introduza o seguinte comando:

AIX ou Linux

\$ES_INSTALL_ROOT/bin/reportccl.sh

Windows

%ES_INSTALL_ROOT%\bin\reportccl.bat

O relatório é criado num ficheiro chamado reportccl.log no directório ES_NODE_ROOT/logs.

Receber correio electrónico sobre erros de CCL

Se pretender ser notificado quando ocorrerem erros de CCL, é necessário configurar as opções de correio electrónico para as mensagens do sistema. Para o fazer, seleccione a página Registos quando estiver na vista Sistema da consola de administração. Na página onde configurar as opções de correio electrónico para as mensagens do sistema:

- Identifique o servidor de SMPT e o ID de utilizador ou IDs de utilizadores para receber o correio electrónico.
- Especifique se pretende receber o correio electrónico sobre todas as mensagens ou especifique os IDs de mensagem para mensagens específicas sobre as quais pretende receber correio electrónico.

Problemas ao iniciar e parar o sistema

Estão disponíveis sugestões sobre a resolução de problemas para ajudar o utilizador a solucionar problemas que podem ocorrer quando o sistema é iniciado ou parado.

Um servidor do sistema pára

Se instalar o IBM Content Analytics with Enterprise Search numa configuração de servidores múltiplos e um problema fizer com que um dos servidores pare, tem de reiniciar o servidor falhado, em seguida, tem de parar e reiniciar todo o sistema.

Sintomas

Se tentar iniciar o sistema quando um servidor estiver parado, o sistema reconhece que o servidor não está disponível e não tenta sincronizá-lo ou comunicar com ele.

Resolver o problema

Reinicie o servidor que parou e, em seguida, insira os seguintes comandos para parar e reiniciar o sistema:

esadmin system stopall esadmin system startall

O comando esadmin stop tem de preceder o comando stopccl

Não insira o comando **esadmin stop** depois de inserir o comando **stopccl** ou enquanto o comando **stopccl** está a ser executado.

Sintomas

O comando esadmin stop poderá parar de responder.

Causas

O comando **stopccl** não é síncrono. Poderá continuar a limpar sessões enquanto o comando **esadmin stop** tenta contactar as sessões que estão a ser limpas.

Resolver o problema

Para evitar este problema, insira sempre o comando **esadmin stop** antes de inserir o comando **stopccl**.

Se o comando **esadmin stop** não responde, pode editá-lo premindo as teclas Ctrl-C.

Problemas com pesquisa de documentos

Estão disponíveis sugestões para solucionar problemas que possam ocorrer com os pesquisadores. Por exemplo, o pesquisador da Web poderá não ser capaz de aceder a um sítio protegido ou um pesquisador poderá parar inesperadamente.

Pesquisador da Web pára inesperadamente

O pesquisador da Web pára de responder.

Sintomas

Quando supervisionar o pesquisador Web na consola de administração, o ícone de estado para um componente do pesquisador Web é um ícone X vermelho.

Resolver o problema

Para solucionar este problema, tem de parar o pesquisador Web numa linha de comandos, antes de poder reiniciar o pesquisador Web na consola de administração.

- 1. No servidor do controlador, inicie sessão como administrador.
- 2. Na linha de comandos, insira o seguinte comando para obter o ID de sessão do pesquisador da Web:

esadmin report sessions -type crawler -format full

É apresentada uma lista de informação da configuração da sessão para todos os pesquisadores no sistema. Para localizar as informações de configuração da sessão para o pesquisador da Web que está a ter problemas, veja o campo Nome de visualização para a configuração de cada pesquisador. Depois de localizar o nome de visualização do pesquisador, anote o valor de ID da sessão correspondente, por exemplo:

```
>esadmin report sessions -type crawler -format full FFQC5323I Relatório de sessões
```

ID de sessão: col_12345.WEB_54321 Nome de visualização: WEBCrawler1 Descrição: ID de colecção: col_12345 ID de nó: node1 Tipo: crawler Sub-tipo: WEB Utilizador:

```
Palavra-passe:
Domínio:
Sinalizadores:
Dir configuração:
col_12345.WEB_54321
null
Dir dados: null
Dir registo: null
Destino: null
Propriedades: {init_heap=16, max_heap=100}
    _____
ID de sessão: col 56789.NNTP 98765
Nome de visualização: NNTPCrawler1
Descrição:
ID de colecção: col_56789
ID de nó: nodel
Tipo:
                       crawler
Sub-tipo:
                      NNTP
Utilizador:
Palavra-passe:
Domínio:
Dominio:.Sinalizadores:0Dir configuração:col_56789.NNTP_98765Dir dados:nullDir registo:nullDestino:null
Propriedades: {init_heap=16, max_heap=100}
FFQC5324I ------ Fim de relatório. Total: 2 ------
```

- Insira o comando seguinte para parar o pesquisador da Web, onde session_ID corresponde ao valor de ID de sessão no relatório de sessões: esadmin session_ID stop
- 4. Na consola de administração, inicie o pesquisador Web.

Visualizar valor de início de sessão do pesquisador Web no navegador Firefox

Se o pesquisador Web não conseguir aceder a um sítio seguro, pode utilizar o suplemento LiveHTTPHeaders de Firefox para ver os valores de início de sessão que o pesquisador está a tentar utilizar. Utilize esta informação para solucionar problemas de definições de autenticação baseadas em formulários na configuração do pesquisador Web.

Sintomas

O pesquisador Web não pode aceder às páginas a pesquisar. Os erros a indicar que não foi possível encontrar as páginas são devolvidos (como HTTP 404 Não encontrado).

Causas

Os valores de autenticação baseados em formulário que o pesquisador Web está configurado para utilizar podem estar incorrectos.

Resolver o problema

Para ver quais os valores que são passados pelo navegador Firefox quando o pesquisador Web tenta iniciar sessão num sítio seguro:

1. Instale o suplemento LiveHTTPHeaders no navegador Firefox. Aceda a http://livehttpheaders.mozdev.org/. Em seguida, faça clique na ligação de descarregamento para LiveHTTPHeaders para instalar automaticamente o

suplemento no Firefox. Para obter informações adicionais, consulte http://livehttpheaders.mozdev.org/installation.html.

- 2. Reinicie o navegador Firefox.
- 3. Limpe todos os cookies armazenados. Abra o menu Ferramentas, seleccione **Opções**, faça clique no separador **Privacidade** e, em seguida, faça clique em **Limpar agora**.
- 4. Inicie o suplemento LiveHTTPHeaders. Abra o menu Ferramentas e seleccione LiveHTTPHeaders.
- 5. Quando a interface Live HTTP Headers é apresentada, especifique o URL para o sítio protegido que pretende pesquisar.
- 6. Quando a página de autenticação é apresentada, insira o ID de utilizador, palavra-passe e outros dados de autenticação que sejam necessário para o formulário.
- 7. Quando a página superior do sítio a pesquisar for apresentada, faça clique em **Guardar todas** para guardar os cabeçalhos de HTTP.

Controlar redirecção de HTTP do pesquisador da Web

O comportamento predefinido do pesquisador da Web é anexar a porta HTTP predefinida, 80, ao nome do sistema central no cabeçalho do pedido. Como resultado, a pesquisa pode ser redireccionada para um sítio que é diferente do que tencionava adicionar ao espaço de pesquisa.

Sintomas

O pesquisador da Web pesquisa sítios que não pretendia incluir no espaço da pesquisa. Por exemplo, se pesquisar my.internal.net, o nome do sistema central no cabeçalho do pedido de HTTP fica my.internal.net:80. Quando o pesquisador se ligar a esta porta, o endereço pode ser processado para my.external.net.

Resolver o problema

Para evitar que o pesquisador da Web anexe automaticamente a porta predefinida de HTTP nomes de sistema central num pedido de GET:

- 1. Edite o ficheiro crawl.properties no directório de configuração do pesquisador da Web, como ES_NODE_ROOT/ *col_52628.WEB_22581*/crawl.properties.
- 2. Adicione a propriedade seguinte:

add_default_port_in_host_field=false

3. Reinicie o pesquisador da Web.

O botão Seguinte não está disponível durante a configuração do pesquisador

Quando configura um pesquisador, o botão **Seguinte** não fica disponível depois de seleccionar o tipo de pesquisador.

Sintomas

Não é apresentada qualquer mensagem e não é possível avançar para a página seguinte no assistente do pesquisador.

Causas

A causa provável é que a execução activa de scripts não está activada no navegador. A consola de administração utiliza scripts. Se o navegador da Web não permitir a execução de scripts, este problema pode ocorrer.

Resolver o problema

Se utilizar o Internet Explorer, siga estes passos para solucionar o problema:

- 1. Abra o Internet Explorer e faça clique em Ferramentas > Opções da Internet.
- 2. Faça clique no separador Segurança.
- 3. Seleccione a zona Internet e faça clique em Nível de personalização.
- 4. Desloque-se até **Execução de scripts** e faça clique na opção **Activar** para **Execução activa de scripts**.
- 5. Encerre todas as sessões abertas do navegador.
- 6. Inicie a consola de administração na nova sessão do navegador e configure o pesquisador.

Ocorre o erro COEN1001E ao ligar a Documentum

Depois de configurar um conector de Documentum num IBM Content Integrator, o erro COEN1001E ocorre ao testar a ligação.

Sintomas

Ocorre um erro COEN1001E:

Não é possível utilizar as classes de Documentum porque os DLLs suportados (Windows) ou as bibliotecas partilhadas (UNIX) não podem ser carregadas.

Poderá querer ver outros erros, como a classe Java não é encontrada ou as mensagens relativas ao caminho de classes.

Causas

Em Windows, o Documentum DFC está instalado no directório C:\Program Files\Documentum\share. Se definir a variável de ambiente PATH como PATH='C:\Program Files\Documentum\Shared';%PATH%", o problema ocorre porque o caminho de directório para o DC contém espaço em branco ou caracteres em branco, como os Ficheiros de Programa.

O problema ocorre independentemente de utilizar aspas ou aspas duplas no nome do caminho no ficheiro config.bat (variável VBR_CLASSPATH) ou no ficheiro RMIBridge.bat (variável PATH).

Resolver o problema

Para evitar este problema, utilize o nome abreviado do caminho:

 Para determinar o nome do caminho abreviado, insira o comando dir /x. Neste exemplo, o nome do caminho abreviado é PROGRA~1: C:\>dir /x

05/22/2009 06:14 PM <DIR> DOCUME~1 Documents and Settings 05/27/2009 01:44 PM <DIR> PROGRA~1 Program Files

 Defina o valor do caminho para VBR_CLASSPATH no ficheiro config.bat e o valor para PATH no ficheiro RMIBridge.bat sem colocar o caminho entre aspas simples ou aspas duplas. Por exemplo:

Problemas com a análise e indexação de documentos

As sugestões de resolução de problemas estão disponíveis para ajudar o utilizador a solucionar problemas que possam ocorrer quando os documentos são analisados ou indexados. Por exemplo, o analisador poderá não iniciar.

Não é possível iniciar o analisador

O analisador parece parar de responder quando tenta iniciá-lo.

Sintomas

O analisador pára de responder e não é iniciado. Quando executa o comando **esadmin check**, irá ver mensagens que indicam que o extractor de texto parou de responder quando se tentou inicializar.

Causas

Uma memória de cache de ficheiro de extractor de texto danificada em HOME/.oit poderá impedir que o analisador seja iniciado.

Resolver o problema

Se a mensagem do estado devolvido dos comandos **esadmin check** indicar que o extractor de texto está a tentar iniciar:

- 1. Insira o comando seguinte para parar o sistema: esadmin stop
- Mude para o directório principal do administrador e localize o ficheiro \$HOME/.oit.
- 3. Elimine o directório .oit/ e todos os ficheiros nesse directório.
- 4. Insira o comando seguinte para reiniciar o sistema: esadmin start

Analisador permanece no estado de espera

Depois de estar em execução algum tempo, o analisador entra em estado de espera.

Sintomas

Ao supervisionar o analisador, as mensagens indicam que o analisador aguarda reinício.

Causas

Esta situação pode ocorrer se forem carregados ficheiros de correlação incorrectos a nível semânticos no sistema. Os ficheiros inválidos fazem com que o analisador não consiga iniciar ou entrar neste estado de espera.

Outro possível problema que pode fazer com que o analisador entre no estado de espera é quando não é possível iniciar uma sessão no extractor de texto.

Resolver o problema

Reveja o ficheiro de registo de auditoria, para que o analisador determine se a causa do problema são ficheiros de correlação incorrectos. Reveja os ficheiros de registo de erros da colecção e do sistema para obter informações adicionais sobre a causa do problema.

Analisador falha ao iniciar depois de associar um motor de análise de texto a uma colecção

Se o analisador falhar ao iniciar depois de associar um motor de análise de texto com uma colecção, poderá ser necessário aumentar o tamanho de pilha de JVM.

Sintomas

O analisador não é iniciado.

Resolver o problema

O motor de análise de texto (ficheiro pear), que inclui os anotadores personalizados que estão associados a uma colecção, é executado numa caixa delimitada específica da colecção. A caixa delimitada é uma sessão separada chamada sessão de processador de documentos. Para alterar o tamanho de pilha de JVM dessa sessão, tem de modificar o seguinte ficheiro de configuração, onde *ID_colecção* é o ID da colecção a que o ficheiro pear está associado: ES NODE ROOT/master config/*ID_colecção* config.ini

No ficheiro, procure uma expressão como session*N*.type=indexservice e session*N*.subtype=docproc para obter o número da sessão do processador de documentos.

Depois de obter o número de sessão, altere o tamanho da pilha na definição seguinte:

sessionN.max_heap=size_in_MB

É necessário ter cuidado ao aumentar o tamanho da pilha. Para obter recomendações sobre a memória, consulte os requisitos para o IBM Content Analytics with Enterprise Search.

Após ter guardado as respectivas alterações, reinicie o sistema do IBM Content Analytics with Enterprise Search.

Referências relacionadas

Requisitos para o IBM Content Analytics with Enterprise Search

Sincronização falha quando ficheiros maiores de 8 MB se encontram no directório master_config

Se um ficheiro alheio ao sistema, maior do que 8 MB, estiver colocado no directório ES_NODE_ROOT/master_config, poderá causar a falha das operações que requerem a sincronização dos dados de configuração.

Sintomas

Quando administrar o sistema (adicionar uma colecção, remover uma colecção, adicionar um pesquisador, remover um pesquisador, e assim por diante), os dados

de configuração são sincronizados em todo o sistema. Se um ficheiro for maior do que 8 MB for adicionado ao directório ES_NODE_ROOT/master_config podem ocorrer problemas:

FFQC4814E ES_ERR_FILE_SIZE_LIMIT_EXCEEDED

Os erros apresentados são semelhantes aos apresentados aqui:

FFQC1027E Ocorreu um erro ao sincronizar os ficheiros de configuração do sistema para o servidor my_server.www.ibm.com [node1].
FFQC1029E Ocorreu um erro ao sincronizar o ficheiro local /data/nodes/node_774/master_config/core.13906 para ficheiro remoto /data/nodes/node_774/config/core.13906 no servidor my_server.www.ibm.com.
FFQT0008E Ocorreu um erro ao ler ficheiro /data/nodes/node_774/master_config/core.13906.
FFQC4814E O ficheiro /data/nodes/node_774/master_config/core.13906 excede o limite de tamanho de 8388608.
O tamanho do ficheiro é de 588369920.

Neste caso, um ficheiro essencial foi criado no directório ES_NODE_ROOT/ master_config. O sistema tenta copiar (ou sincronizar) este ficheiro no directório ES_NODE_ROOT/config, mas como foi aplicado um limite de tamanho para o ficheiro de configuração, de 8 MB, a operação de sincronização falha.

Resolver o problema

Para solucionar este problema, remova todos os ficheiros alheios ao sistema e maiores do que 8 MB do directório ES_NODE_ROOT/master_config e, em seguida, insira os seguintes comandos para reiniciar o sistema:

esadmin stop esadmin start

O extractor de texto falha quando detecta documentos corrompidos

Quando o extractor de texto detecta um documento corrompido, rejeita o documento e, em seguida, tenta reiniciar a sessão do extractor de texto. Em alguns casos, o extractor de texto termina de forma anómala e cria uma cópia da memória principal.

Sintomas

O extractor de texto cria uma cópia da memória principal quando tenta reiniciar depois de rejeitar um documento corrompido.

Causas

Se um documento estiver corrompido e a sessão do extractor de texto rejeitar o documento, é feita uma tentativa para reiniciar a sessão do extractor de texto dado que um documento corrompido pode corromper documentos em boas condições que cheguem posteriormente. Por vezes, se o extractor de texto não parar e reiniciar de forma normal, pode ocorrer uma cópia da memória principal.

Resolver o problema

Determine se o ficheiro rejeitado está corrompido. Por exemplo, verifique se consegue abrir o documento na aplicação original. Se o ficheiro não estiver corrompido, verifique se é um tipo de documento que seja suportado pelo extractor de texto.

Se o tipo de documento for suportado, poderá ser necessário abrir um relatório do problema no Oracle para determinar se existe um defeito na tecnologia Oracle Outside In Content Access que é utilizada pelo extractor de texto. Contacte o IBM Software Support para obter assistência.

Referências relacionadas

"Regras de análise predefinidas para o extractor de texto" na página 272

Os dados de saída de rastreio de pilha aparecem depois de o extractor de texto parar o processamento

Se o extractor de texto terminar anormalmente num sistema Linux, é escrito um rastreio de pilha na consola onde IBM Content Analytics with Enterprise Search foi iniciado. Pode ignorar estes dados de saída.

Sintomas

O extractor de texto produz um grande volume de dados de saída semelhante aos seguintes dados de rastreio de pilha:

```
*** glibc detected *** estask: double free or corruption (out): 0x59f03d00 ***
====== Backtrace: ========
/lib/libc.so.6[0x835595]
/lib/libc.so.6(cfree+0x59)[0x8359d9]
/usr/lib/libstdc++.so.6(_ZdlPv+0x21)[0xb09581]
/opt/IBM/es/lib/libffq.oss.so(_ZN2es3oss14threadFunctionEPv+0x5e)[0x5571516e]
/lib/libpthread.so.0[0x56b4a5ab]
/lib/libc.so.6(clone+0x5e)[0x89dcfe]
====== Memory map: =======
007ae000-007c8000 r-xp 00000000 fd:00 62096345 /lib/ld-2.5.so
007c8000-007c9000 r--p 00019000 fd:00 62096345 /lib/ld-2.5.so
007c9000-007ca000 rw-p 0001a000 fd:00 62096345 /lib/ld-2.5.so
007cc000-0090b000 r-xp 00000000 fd:00 62095377 /lib/libc-2.5.so
0090b000-0090c000 ---p 0013f000 fd:00 62095377 /lib/libc-2.5.so
0090c000-0090e000 r--p 0013f000 fd:00 62095377 /lib/libc-2.5.so
0090e000-0090f000 rw-p 00141000 fd:00 62095377 /lib/libc-2.5.so
0090f000-00912000 rw-p 0090f000 00:00 0
. . .
fffb5000-fffc8000 rwxp 7ffffffea000 00:00 0 [stack]
fffc8000-fffca000 rw-p 7ffffffd000 00:00 0
ffffe000-fffff000 r-xp ffffe000 00:00 0
```

[esadmin@discover1 △]\$

Causas

Os dados de saída de rastreio de pilha poderão ser escritos na consola se a sessão to extractor de texto terminar anormalmente.

Resolver o problema

Pode ignorar estes dados de saída.

Problemas com aplicações e servidores de procura

Estão disponíveis sugestões de resolução de problemas para ajudar o utilizador a solucionar problemas que possam ocorrer durante a consulta às colecções. Por exemplo, poderá ter problemas ao iniciar sessão na aplicação, ou poderá ter tempos de resposta à consulta lento.

Nenhum processo de procura configurado (erro FFQC4033E)

Se visualizar uma mensagem de erro FFQC4033E, os ficheiros de configuração do sistema poderão estar corrompidos ou poderá não ter espaço em disco suficiente.

Sintomas

O erro seguinte é apresentado na consola de administração: FFQC4033E Nenhum processo de procura configurado para partição *partition_ID* da colecção *collection ID* no servidor *server name*.

Diagnosticar o problema

Certifique-se de que o computador não ficou sem espaço em disco. Se tiver, liberte algum espaço em disco.

Resolver o problema

Se tiver espaço em disco livre suficiente, execute os passos para solucionar o problema:

1. Insira o comando seguinte: esadmin configmanager sync

Tente aceder a colecção utilizando a consola de administração. Se conseguir visualizar a colecção, pare aqui. Se não conseguir procurar a informação da colecção na consola de administração, avance para o passo seguinte.

- Remova a colecção do sistema. A cópia principal da informação de configuração da colecção poderá estar corrompida. Para remover a colecção:
 - a. Pare o sistema. Insira: esadmin stop
 - b. Se existir apenas uma colecção no sistema, remova o ficheiro ES_NODE_ROOT/master_config/collections.ini. Se existir mais do que uma ligação, remova todas as propriedades da colecção corrompida no ficheiro collections.ini. Ordene o ficheiro collections.ini dando-lhe um novo nome como collections.ini.sort. Em seguida, remova as propriedades que estão associadas à colecção. Altere o nome do ficheiro collections.ini.sort novamente para collections.ini. Este passo remove todas as ligações com a colecção e impede-a de ser reconhecida no sistema. Os passos detalhados deste procedimento são os seguintes:
 - 1) Insira os seguintes comandos. Pode utilizar um editor de texto da sua escolha. Estes exemplos mostram o vi editor.

sort collections.ini > collections.ini.sort
vi collections.ini

2) Remova todas as linhas do ficheiro que começam pelo nome da colecção corrompida. Por exemplo, se o ficheiro collections.ini tem o seguinte conteúdo, e a mensagem de erro indicar que a colecção com o ID col2 é a colecção corrompida, remova cada linha que começa por collection2:

```
collection1.configfile=col1_config.ini
collection1.datadir=/home/ofadmin/node/data/col1
collection1.description=
collection1.displayname=Collection1
collection1.flags=0
collection1.id=col1
collection1.sectiontype=collection
collection1.type=1
collection2.configfile=col2_config.ini
collection2.datadir=/home/ofadmin/node/data/col2
collection2.description=
collection2.displayname=Collection2
```

collection2.flags=0
collection2.id=col2
collection2.sectiontype=collection
collection2.type=2

Depois de remover estas linhas, o ficheiro tem o seguinte conteúdo:

```
collection1.configfile=coll_config.ini
collection1.datadir=/home/ofadmin/node/data/col1
collection1.description=
collection1.displayname=Collection1
collection1.flags=0
collection1.id=col1
collection1.sectiontype=collection
collection1.type=1
```

- 3) Insira o seguinte comando: mv collections.ini.sort collections.ini
- c. Remova os artigos seguintes. O sistema poderá ter todos estes artigos. Para este exemplo, *collection_ID* é o ID de colecção da colecção corrompida.
 - Todos os directórios de configuração da sessão da colecção associada. Isto significa qualquer directório no ES_NODE_ROOT/master_config/ and ES_NODE_ROOT/config/ que contém collection_ID como prefixo (por exemplo, collection_ID.indexer).
 - Qualquer ficheiro de configuração em ES_NODE_ROOT/master_config/ que tenha o prefixo collection_ID (por exemplo, collection_ID_log.prp, collection_ID_docTracking.prp).
 - No directório ES_NODE_ROOT/master_config/datalistener, remova o ficheiro de configuração de escuta de dados da colecção (por exemplo, rds collection ID.DL.ini).
- 3. Reinicie o sistema. Insira: esadmin start

Se os passos anteriores não solucionarem o problema, contacte o IBM Software Support e forneça as seguintes informações:

- O nome, marca de tempo e tamanho de todos os ficheiros e directórios sob os directórios ES_NODE_ROOT/master_config/ e ES_NODE_ROOT/config/.
- Os ficheiros collections.ini e collection_ID_config.ini nos directórios ES_NODE_ROOT/master_config/ and ES_NODE_ROOT/config/.
- O ficheiro de registo de erro do sistema no directório ES_NODE_ROOT/logs/.
- Os ficheiros configmanager_audit_*timestamp*.log e utilities_audit_*timestamp*.log no directório ES_NODE_ROOT/logs/audit.

Iniciar sessão numa aplicação demora muito tempo

Se demorar muito tempo (tal como 30 segundos ou mais) para que os utilizadores finais iniciem sessão na aplicação de procura empresarial ou na prospecção de análise de conteúdo, existem passos que pode executar para melhorar o desempenho.

Sintomas

Quando a segurança está activada no IBM Content Analytics with Enterprise Search, os utilizadores finais têm de esperar algum tempo depois de especificarem as respectivas credenciais de início de sessão antes da aplicação de procura ser apresentada.

Resolver o problema

Existem várias formas para configurar as origens de dados, a aplicação ou o componente de gestão de identidade (ICM) para melhorar o desempenho. As sugestões seguintes utilizam servidores Domino como um exemplo, mas as sugestões aplicam-se a outros tipos de repositórios que pode configurar para a procura segura.

- Crie um índice de texto completo no Notes Address Book (NAB) em cada servidor Domino que seja pesquisado por um pesquisador. A existência de um índice NAB melhora significativamente o desempenho das tarefas de procura de grupo.
- Execute a validação de utilizador apenas quando ocorrer a extracção de grupo. Esta abordagem não melhora o desempenho, mas melhora a experiência do utilizador em tentativas subsequentes de início de sessão na aplicação de procura empresarial.
- Personalize o código de IMC para que a validação e a extracção de grupo seja efectuada para um dos servidores Domino, em seguida, parta do princípio que os mesmos nomes de utilizador, palavras-passe e grupos são válidos para todos os servidores Domino. Esta abordagem pode melhorar muito o desempenho.
- Personalize o código de IMC para que utilize o Java Naming and Directory Interface (JNDI) para procurar os nomes de utilizadores e de grupos canónicos directamente do Domino através do Lightweight Directory Access Protocol (LDAP). É possível conseguir mais melhorias do desempenho alterando o código de IMC para que procure directamente no servidor de LDAP, em seguida, utilize a mesma informação para todos os objectos Identity.

Por exemplo, se utilizar a portlet Procura para procurar colecções no WebSphere Portal, pode comentar a validação do utilizador e o código de extracção de grupo nas partes do IMC da portlet Procura. Se tiver um directório de LDPA empresarial que seja partilhado por todos os servidores Domino do utilizador, pode recorrer às APIs PUMA de WebSphere Portal para obter as informações de nomes e grupos de utilizadores e, em seguida, utilizar essas informações para todos os objectos Identity (Identidade).

Tempos de resposta de procura lentos no Linux

Poderá ser necessário alterar a política de permuta de memória em sistemas Linux que utilizam kernel 2.4 para conseguir tempos de resposta melhores.

Sintomas

Irá ter tempos de resposta longos para pedidos de procura e tempos de resposta longos quando utiliza a consola de administração do WebSphere Application Server.

Causas

A política de permuta de ficheiros predefinida do Linux no kernel 2.4 ou é agressiva. Isto significa que o kernel permuta dados da aplicação em preferência para os dados de memória cache do ficheiro embora os dados da aplicação sejam acedidos mais vezes. Isto pode causar problemas para o sistema que é executado com um carregamento elevado.

Normalmente, o tamanho de um índice de texto é maior do que a memória física disponível, o que significa que não encaixa na memória principal de imediato. O sistema operativo, o servidor de aplicações Web, e o IBM Content Analytics with Enterprise Search também requerem memória física (aproximadamente 1.5 GB).

Durante a actividade de procura, o componente de procura precisa de aceder aos dados indexados. Como a memória física não consegue manter todos os dados da aplicação e indexados, o sistema operativo inicia a permuta de página que está actualmente em memória. A política de permuta do Linux 2.4 permuta as páginas de memória das aplicações actualmente em execução. Esta permuta faz com que essas aplicações respondam de forma mais lenta às actividades do utilizador porque têm de efectuar a permuta das páginas de memória novamente para a memória física.

Resolver o problema

A política de permuta do Linux kernel 2.4 baseia-se no ficheiro /proc/sys/vm/pagecache. Os valores predefinidos ("1 15 100") permitem ao Linux utilizar toda a memória física para colocação em memória cache dos ficheiros. Pode alterar esses valores com o seguinte comando como utilizador raiz (root): echo '1 15 75' > /proc/sys/vm/pagecache

ou pode utilizar o seguinte comando: sysctl -w vm.pagecache="1 15 75"

Também pode adicionar a seguinte linha ao ficheiro /etc/sysctl.confuse: vm.pagecache="1 15 75"

O último valor na cadeia de comando (neste caso: 75) especifica a percentagem de memória física livre que permanece disponível depois de todas as aplicações terem sido carregadas. Por exemplo:

- Memória física principal: 6 GB
- Memória da aplicação e do sistema operativo: 1.5 GB
- Memória física livre restante: 4.5 GB
- Valor de memória cache da página: 75%

Neste exemplo, o sistema operativo não utiliza mais memória para a memória cache do ficheiro que a que está disponível quando todas as aplicações estão em execução.

Nota: Uma alteração em parâmetros relacionados com o sistema poderá influenciar o desempenho total do sistema.

Com o Linux kernel 2.6, a política de permuta alterou-se. O valor predefinido é agora de 60% e encontra-se no ficheiro /proc/vm/swappiness.

Não é possível registar o ID da aplicação

Inconsistências de palavra-passe entre várias cópias do ficheiro ES_INSTALL_ROOT/nodeinfo/es.cfg podem impedir que uma aplicação seja registada em IBM Content Analytics with Enterprise Search.

Sintomas

Falharam as tentativas de registar um ID da aplicação com o servidor IBM Content Analytics with Enterprise Search. Ocorre o erro seguinte:

FFQM5007E A instância da aplicação [*instance_name*] não é válida. Verifique o nome da aplicação e a palavra-passe associada.

Causas

O ficheiro es.cfg que é utilizado pela aplicação de procura empresarial é diferente da es.cfg que é utilizada pelo servidor IBM Content Analytics with Enterprise Search. O es.cfg no IBM Content Analytics with Enterprise Search é actual, mas o ficheiro na máquina cliente pode incluir palavras-passe descodificadas e outras diferenças.

Resolver o problema

Para resolver este problema, substitua o es.cfg na máquina cliente pelo ficheiro es.cfg que está no servidor IBM Content Analytics with Enterprise Search antes de registar um ID da aplicação. As palavras-passe e os valores de CK em ambos os ficheiros devem corresponder.

Conceitos relacionados

Registar IDs de aplicações

Clonar a portlet Pesquisa para WebSphere Portal

Se uma portlet for clonada na consola de administração do WebSphere Portal, poderá ver duas instâncias da mesma aplicação da portlet. Se quiser ter duas portlets ESSearchPortlet para colecções à parte, tem de instalar a portlet ESSearchPortlet como uma aplicação à parte, não cloná-la.

Sintomas

Quando a portlet Procura para as colecções de procura empresarial é copiada ou clonada, as definições globais predefinidas são partilhadas entre todas as portlets do IBM Content Analytics with Enterprise Search. Pode ser confuso ver as definições que se aplicam em todas as instâncias da portlet. Além disso, a capacidade de associar a aplicação de procura empresarial IBM Content Analytics with Enterprise Search a uma determinada instância de portlet não funciona correctamente. O sistema utiliza uma abordagem primeiro a entrar primeiro a sair, o que significa que as definições de configuração na primeira instância da portlet são utilizadas por todas as instâncias da portlet subsequentes.

Causas

A portlet Pesquisa de IBM Content Analytics with Enterprise Search utiliza a sessão HTTP para armazenar atributos transientes. Várias instâncias da portlet partilham a mesma HttpSession. Uma vez que todas as instâncias de ESSearchPortlet clonadas estarem na mesma sessão de HTTP, os atributos armazenados na sessão de HTTP são partilhados entre todas as instâncias. Por exemplo:

```
request.getSession().setAttribute("inputQueryString", queryString);
request.getSession().setAttribute("builtQueryString", builtQueryString);
request.getSession().setAttribute("userFeedBack", userFeedBack.toString());
request.getSession().setAttribute("userFeedBackConcepts", userFeedBackConcepts.toString());
request.getSession().setAttribute("queryLanguage",(String) parameters.get("queryLanguage"));
request.getSession().setAttribute("selectedCollectionIDs", selectedCollectionIDs);
```

Resolver o problema

Para implementar várias instâncias de ESSearchPortlet no WebSphere Portal:

- 1. Crie uma nova portlet ESSearchPortlet denominada, por exemplo ESSearchPortlet2.war.
- 2. Actualize o ficheiro portlet.xml, para especificar o novo nome da portlet: <name>applicationName</name> <value>ESSearchPortlet2</value>

- 3. Actualize o ficheiro InstallESSearchPortlet6.xml com a nova informação da portlet. Este ficheiro está no ficheiro es.wp61.install.jar.
- 4. Insira o comando seguinte para instalar ESSearchPortlet2: xmlaccess.bat -user username -password password -in InstallESSearchPortlet2.xml -url "http://portal61_hostname/wps/config/" Estão agora implementadas duas portlets ESSearchPortlet (ESSearchPortlet e ESSearchPortlet2).
- 5. Utilize a consola de administração do IBM Content Analytics with Enterprise Search para associar cada portlet a colecções que pretende pesquisar.

Pode implementar portlets adicionais repetindo estes passos.

Problemas operacionais

Estão disponíveis sugestões de resolução de problemas para ajudar o utilizador a solucionar problemas operacionais que possam ocorrer. Por exemplo, poderá encontrar conflitos de portas ou uma condição de disco cheio, ou poderá ter problemas ao iniciar sessão no sistema ou ao executar os utilitários do sistema.

Conflitos de portas causam erros de socket

Se visualizar mensagens de erro sobre problemas de socket, poderá ter um conflito de portas.

Sintomas

Poderá ver a seguinte mensagem:

FFQ00273W Ocorreu um aviso interno - Mensagem de Excepção: {0} at java.net.PlainSocketImpl.socketBind(Native Method) at java.net.PlainSocketImpl.bind(PlainSocketImpl.java:357) at java.net.ServerSocket.bind(ServerSocket.java:341)

```
at java.net.ServerSocket.<init>(ServerSocket.java:208)
```

```
at java.net.ServerSocket.<init>(ServerSocket.java:120)
```

Resolver o problema

Para solucionar o conflito de portas, mude os números de portas que são utilizados para o IBM Content Analytics with Enterprise Search para as portas disponíveis.

Tarefas relacionadas

"Alterar o número de portas do sistema" na página 32

Problemas de atribuição de sockets causam falha de operações com erro FFQO0143E

No sistema operativo Windows, as operações como, por exemplo, análise, construção de um índice, procura e assim por diante, falham com a mensagem de erro FFQ00143E (ES_ERR_SOCKET_ADDRESS_IN_USE) porque não é possível atribuir um endereço de socket.

Sintomas

Este erro pode ocorrer em qualquer sessão ou componente do sistema. No exemplo seguinte, o erro ocorreu durante uma construção de índice porque não estavam disponíveis sockets de cliente suficientes no servidor onde o IBM Content Analytics with Enterprise Search está instalado:

5/29/07 14:30:03.721 JST [Error] [ES_ERR_SOCKET_ADDRESS_IN_USE] [col_2000] [col_2000.indexer.delta] EXAMPLE1:6612:580:oss:src\trevi\oss\src\Socket.cpp:oss::Socket::connect:581 FFQ00143E O endereço e número de porta especificado já estão a ser utilizados. Nome do servidor: 192.168.100.101. Número de porta: 6002.

5/29/07 14:30:03.846 JST [Error] [ES_ERR_CCLSRV_SOCKET_COMM_FAILED] [col_2000] [col_2000.indexer.delta] EXAMPLE1:6612:580:oss::ccl:src\trevi\oss\ccl\src\Session.cpp:oss::ccl:: Session::openSession:102 FFQ00253E A comunicação de socket falhou.

5/29/07 14:30:04.003 JST [Error] [ES_ERR_CTRL_ACQUIRE_SESSION_READ_LOCK] [col_2000] [col_2000.indexer.delta] EXAMPLE1:6612:580:control::configmanager:src\trevi\control\configmanager\src \CtrlConfigManager.cpp:control::configmanager::CtrlConfigManager::getSession ReadLock:1207 FFQC0017E Não foi possível adquirir o bloqueio de leitura col_2000.indexer de configuração da sessão.

5/29/07 14:30:04.128 JST [Error] [ES_ERR_CTRL_INDEXER_BUILD_THREAD_FAILED] [col_2000] [col_2000.indexer.delta] EXAMPLE1:6612:580:control::indexer:src\trevi\control\indexer\src\ BuilderThread.cpp:control::indexer::BuilderThread::run:291 FFQC1417E 0 módulo de construção de índice encontrou um erro.

Causas

Por predefinição, o Windows não permite a todos os utilizadores configurar ligações de cliente em portas acima de 5000. Depois de fechar um socket, a ligação de socket mantém no estado TIME_WAIT durante aproximadamente mais dois minutos (a quantidade de tempo depende da configuração do sistema). Depois de terminar o período de espera, o socket é libertado e o endereço pode ser reutilizado.

Por vezes, será aberto um número grande de portas aleatórias no computador, porque as chamadas de API que utilizam o cliente do Common Communication Layer (CCL) implicam operações frequentes de cada sessão.

Se forem efectuadas mais de 4000 ligações (portas 1024 através de 5000) antes de as portas serem libertadas (depois de o estado TIME_WAIT terminar), em seguida, as tentativas de abrir um socket de cliente numa porta acima de 5000 serão rejeitadas pelo sistema operativo. Poderá ser devolvida uma mensagem que indica que o endereço já está em utilização.

Resolver o problema

É possível evitar este problema modificando os parâmetros de TCP/IP no registo do Windows:

- 1. Efectue a cópia de segurança do registo do Windows.
- 2. Inicie o editor de registos do Windows: No menu Windows Iniciar, seleccione **Executar** e depois insira regedit.
- Localize a seguinte chave: HKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet\ Services\Tcpip\Parameters
- Faça clique com o botão direito do rato na pasta Parâmetros e seleccione Novo > Valor de DWORD.
- 5. Dê o seguinte nome à chave MaxUserPort.
- 6. Faça duplo clique na chave MaxUserPort, altere o valor para 65534 e seleccione Decimal como base.

- Faça clique com o botão direito do rato na pasta Parâmetros e seleccione Novo > Valor de DWORD.
- 8. Dê o seguinte nome à chave **TcpTimedWaitDelay**.
- Faça duplo clique sobre a chave TcpTimedWaitDelay, altere o valor para 30 e seleccione Decimal como base.
- 10. Reinicie o computador.

Para obter informações adicionais sobre este problema, consulte:

- O artigo sobre a biblioteca de MSDN, *Avoiding TCP/IP Port Exhaustion*, em http://msdn2.microsoft.com/en-us/library/aa560610.aspx.
- Microsoft Knowledge Base Artigo 169271 em http://support.microsoft.com/ default.aspx?scid=kb;[LN];196271.

Recuperar de uma falha de disco cheio

É possível recuperar o sistema depois de uma condição de disco cheio causar uma falha num componente.

Sintomas

Espaço em disco disponível insuficiente para executar processos do IBM Content Analytics with Enterprise Search.

Resolver o problema

Para recuperar de uma falha da disco cheio:

- 1. Corrija a causa da falha no servidor onde ocorreu o problema. Por exemplo, pode adicionar mais espaço livre ou adiciona espaço ao disco.
- 2. Inicie sessão como administrador. Para uma configuração de vários servidores, inicie sessão no servidor do controlador.
- 3. Inicie o sistema inserindo o comando esadmin start.
- 4. Verifique se quaisquer componentes anteriormente em execução foram parados ou se encontram no estado de erro (como pesquisadores, analisadores ou processos de procura). Se assim for, inicie sessão na consola de administração, altere para o modo de supervisão e reinicie os componentes parados.
 - Se o servidor que falhou for o servidor do pesquisador, é provável que a maioria dos pesquisadores esteja no estado de erro. Neste caso, tem de parar e reiniciar os pesquisadores.
 - Se o servidor que falhou for um servidor de procura, é provável que a maioria dos processos de procura esteja parada. Neste caso, reinicie os servidores de procura.
 - Se o servidor que falhou for o servidor do analisador ou do controlador, é provável que a maioria dos analisadores esteja parada. Neste caso, tem de reiniciar os analisadores.

Não é possível efectuar cópia de segurança ou restaurar a informação da base de dados depois de executar a cópia de segurança e restaurar scripts

Se receber mensagens de erro durante a cópia de segurança e restauro, certifique-se de que os servidores do sistema estão em execução.

Sintomas

Quando efectuar a cópia de segurança ou o restauro do IBM Content Analytics with Enterprise Search, poderá receber as seguintes mensagens:

FFQF0111E esbackup.sh: Não foi possível efectuar cópia de segurança da informação da base de dados, rc=3.

FFQF0032E esrestore.sh: Não foi possível restaurar a informação da base de dados, rc=3.

Resolver o problema

Para solucionar o problema, certifique-se primeiro de que o sistema está em conformidade com as directrizes de utilização para efectuar cópia de segurança e restauro do sistema. Também tem de certificar-se de que o sistema está iniciado.

Para iniciar o sistema:

- 1. Inicie sessão como administrador. Numa configuração de vários servidores, inicie sessão no servidor do controlador.
- 2. Insira o comando seguinte para iniciar o sistema: esadmin start
- 3. Execute os scripts esbackup e esrestore.

Conceitos relacionados

"Efectuar cópia de segurança e restaurar o sistema" na página 455

Tarefas relacionadas

"Efectuar a cópia de segurança do sistema" na página 456

"Restaurar o sistema" na página 457

Reiniciar registo centralizado

Para reiniciar o registo centralizado depois de o servidor do controlador falhar, tem de reiniciar o sistema.

Sintomas

Se o serviço CCL (common communication layer server) no servidor do controlador falhar, as mensagens de erro geradas a partir de outros servidores não podem ser escritas nos ficheiros de registo de erros centralizados no servidor de indexação. Apenas os ficheiros de registo central no servidor do controlador podem ser visualizados através da consola administrativa.

Causas

Numa instalação de vários servidores, os seguintes ficheiros de registo de erros centralizados são mantidos no servidor do controlador, onde *collectionID* é o ID de colecção e *date* é a data em que o ficheiro de registo é criado:

ES_NODE_ROOT/logs/system_date.log
ES_NODE_ROOT/logs/collectionID_date.log

Quando ocorrem erros num dos outros servidores, os erros são enviados para o servidor do controlador para que possam ser guardados no ficheiro de registo de erros centralizado correcto e para que possam ser vistos posteriormente através da consola administrativa. Se o serviço CCL no servidor do controlador falhar, os outros servidores detectam a condição e começam a escrever erros nos ficheiros de registo locais, que também são designados ES_NODE_ROOT/logs/system_date.log e ES_NODE_ROOT/logs/collectionID_date.log.

Estes servidores não conseguem detectar quando o serviço CCL no servidor do controlador está novamente disponível, e assim continuam a escrever erros nos ficheiros de registo de erros locais. OS ficheiros de registo de erros locais que não se encontram no servidor do controlador não podem ser visualizados através da consola administrativa. Apenas os ficheiros de registo central no servidor do controlador podem ser visualizados através da consola administrativa.

Resolver o problema

Se o serviço CCL no servidor do controlador falhar, e reparar que estão a ser escritos erros nos ficheiros de registo de erro locais noutros servidores, insira os seguintes comandos para reiniciar o sistema e activar o registo centralizado para ser restabelecido correctamente:

```
esadmin system stopall
esadmin system startall
```

Definir a palavra-passe de administrador codificada para ser a mesma em todos os servidores

A codificação da palavra-passe do administrador do IBM Content Analytics with Enterprise Search tem de ser idêntica em todos os servidores do sistema.

Sintomas

Ao utilizar os scripts **eschangepw.sh** ou **eschangepw.bat**, a mesma palavra-passe é codificada de uma forma no servidor de procura e de outra forma no servidor do pesquisador e no servidor do controlador.

Causas

O valor de CK no ficheiro es.cfg é utilizado como uma chave de codificação. Este valor é definido quando o produto é instalado, e baseia-se no ID e palavra-passe de administrador. Para uma instalação de servidores múltiplos, o valor de CK é o mesmo em todos os servidores apenas se especificar o mesmo administrador e valores de palavra-passe quando executa o programa de instalação em cada servidor. Se a chave de codificação difere, as palavras-passe codificadas não coincidem.

Resolver o problema

Se os valores da chave de codificação no ficheiro es.cfg nos servidores não coincidirem:

- 1. Copie o valor de um servidor para outro.
- Execute o script eschangepw.sh (em AIX ou Linux) ou o script eschangepw.bat (em Windows) para definir a mesma palavra-passe codificada em cada servidor.

Tarefas relacionadas

"Alterar a palavra-passe de administrador na configuração de múltiplos servidores" na página 27
Documentação do produto

Pode ler a documentação sobre o IBM Content Analytics with Enterprise Search no formato de PDF ou HTML.

O programa de instalação do IBM Content Analytics with Enterprise Search instala automaticamente o Centro de Informações, que inclui versões de HTML da documentação. Para uma instalação de múltiplos servidores, o Centro de Informações é instalado em todos os servidores de procura. Se não instalar o Centro de Informações, o Centro de Informações é aberto num sítio da Web da IBM. O Centro de Informações no sítio da IBM poderá conter informação mais actual do que a informação contida nos Centros de Informações instalados.

Para ver versões instaladas dos documentos em PDF, avance para o directório ES_INSTALL_ROOT/docs/*locale*. Por exemplo, para localizar documentos em inglês, avance para ES_INSTALL_ROOT/docs/en_US.

Para aceder às versões PDF da documentação em todos os idiomas disponíveis, consulte a documentação do IBM Content Analytics with Enterprise Search Versão 3.0.

Para aceder a transferência de produtos, pacotes de correcções, notas técnicas e outras informações de resolução de problemas, consulte o suporte do IBM Content Analytics with Enterprise Search.

A seguinte tabela mostra a documentação disponível, nomes de ficheiros e localizações.

Título	Nome do ficheiro	Localização
Centro de Informações		http://pic.dhe.ibm.com/infocenter/ analytic/v3r0m0/
Manual de Iniciação Rápida (Este documento também está disponível em cópia impressa em inglês, francês e japonês.)	ContentAnalytics300 _qsg_ <i>locale de duas letras</i> .pdf	ES_INSTALL_ROOT/docs/ <i>locale</i>
Manual de Instalação	iiysi.pdf	ES_INSTALL_ROOT/docs/locale
Manual de Administração	iiysa.pdf	ES_INSTALL_ROOT/docs/ <i>locale</i>
Manual de Programação	iiysp.pdf	ES_INSTALL_ROOT/docs/en_US/
Integração de Análise de Texto	iiyst.pdf	ES_INSTALL_ROOT/docs/locale
Guia de análise de conteúdo	iiysc.pdf	ES_INSTALL_ROOT/docs/locale
Referência de Mensagens	iiysm.pdf	<pre>ES_INSTALL_ROOT/docs/locale</pre>

Tabela 56. Documentação do produto

Sugestão: Para obter mais informações sobre actualizações à documentação que ainda não estão incorporadas no Centro de Informações ou nas publicações do produto, consulte Actualizações de documentação para o IBM Content Analytics with Enterprise Search Versão 3.0.

Funções de acessibilidade

As funções de acessibilidade ajudam os utilizadores com qualquer incapacidade física, como mobilidade restrita ou visão limitada, a utilizar com êxito os produtos da tecnologia de informação.

A IBM procura fornecer produtos com acesso facilitado para todos, independentemente da idade ou capacidade.

Funções de acessibilidade

A lista que se segue inclui as principais funções de acessibilidade incluídas no IBM Content Analytics with Enterprise Search:

- · Operação apenas com o teclado
- Interfaces que são frequentemente utilizadas por leitores de ecrã

O Centro de Informações do IBM Content Analytics with Enterprise Search e as publicações relacionadas com o mesmo, suportam a acessibilidade. As funções de acessibilidade do Centro de Informações estão descritas em http://pic.dhe.ibm.com/infocenter/analytic/v3r0m0/topic/com.ibm.discovery.es.nav.doc/accessibility_info.htm.

Navegação com o teclado

Este produto utiliza as teclas de navegação padrão do Microsoft Windows.

Também pode utilizar os seguintes atalhos do teclado para navegar e avançar no programa de instalação do IBM Content Analytics with Enterprise Search.

Acção	Atalho
Realçar um botão de opção	Tecla de seta
Seleccionar um botão de opção	Tecla Tab
Realçar um botão de acção	Tecla Tab
Seleccionar um botão de acção	Tecla Enter
Ir para a janela seguinte, anterior ou cancelar	Realçar um botão de acção ao premir a tecla Tab e, em seguida, premir Enter
Desactivar a janela activa	Ctrl + Alt + Esc

Tabela 57. Teclas de atalho para o programa de instalação

Informações sobre interfaces

As interfaces do utilizador para a consola de administração, a aplicação de procura empresarial, o Content Analytics Miner e as aplicações de personalizador são interfaces baseadas no navegador que pode visualizar no Microsoft Internet Explorer ou Mozilla Firefox. Consulte a ajuda online para o Internet Explorer ou Firefox para obter uma lista de teclas de atalhos e outras funcionalidades de acessibilidade para o navegador.

Informações relacionadas com acessibilidade

Pode visualizar as publicações sobre o IBM Content Analytics with Enterprise Search em Adobe Portable Document Format (PDF) através do Adobe Reader. Os PDFs são fornecidos num CD que é enviado juntamente com o produto, mas também podem ser acedidos em http://www.ibm.com/support/ docview.wss?uid=swg27023678.

IBM e a acessibilidade

Para obter mais informações sobre o compromisso da IBM com a acessibilidade, consulte o IBM Human Ability and Accessibility Center.

Glossário

Este glossário define termos utilizados nas interfaces e na documentação do produto IBM Content Analytics with Enterprise Search.

lista de controlo de acessos (ACL)

Em segurança informática, é uma lista associada a um objecto que identifica todos os indivíduos que podem aceder ao objecto e os respectivos direitos de acesso.

função administrativa

Uma classificação de um utilizador que concede acesso a outro utilizador.

motor de análise

Ver motor de análise de texto.

resultados da análise

As informações que são produzidas por anotadores. Os resultados de análise são escritos numa estrutura de dados denominada estrutura de análise comum. Os resultados de análise produzidos pelos motores de análise de texto personalizados (anotadores) podem ser disponibilizados para procura através da respectiva inclusão no índice.

anotação

Informações sobre um grupo de recursos de rede de texto. Por exemplo, uma anotação pode indicar que um grupo de recursos de rede de texto representa um nome de empresa. Na Unstructured Information Management Architecture (UIMA), uma anotação é uma estrutura funcional especial.

anotador

Um componente de software que executa tarefas de análise linguística específicas e produz e regista anotações. Um anotador é o componente lógico de análise num motor de análise.

anotadores base

Um conjunto de motores de análise de texto personalizada utilizado para o processamento de análise de documentos predefinido.

procura Booleana

Uma procura na qual um ou mais termos da procura são combinados utilizando operadores como AND, NOT e OR.

classe hierárquica

Um objecto que contém especificações que podem influenciar a classificação relativa de um documento nos resultados da procura.

palavra hierárquica

Uma palavra que pode influenciar a classificação relativa de um documento nos resultados de procura. Durante o processamento da consulta, a importância de um documento que contenha uma palavra hierárquica pode ser aumentada ou diminuída, dependendo de uma classificação predefinida para a palavra.

árvore de categorias

Uma hierarquia de categorias.

certificado

Em segurança informática, um documento digital que associa uma chave

pública à identidade do proprietário do certificado, permitindo, assim, que o proprietário do certificado seja autenticado. Um certificado é emitido por uma entidade certificadora e é digitalmente assinado por essa autoridade.

entidade certificadora

Uma organização ou empresa fidedigna de terceiros que emite certificados digitais utilizados para criar assinaturas digitais e pares de chaves públicas-privadas. A autoridade certificadora garante a identidade dos indivíduos a quem é concedido o certificado exclusivo.

normalização de caracteres

Um processo no qual as formas variantes de um carácter, tais como maiúsculas e marcas diacríticas, são reduzidas a um formato comum.

clítico Uma palavra que funciona sintacticamente em separado, mas que está ligada foneticamente a outra palavra. Um clítico pode ser escrito como ligado ou separado da palavra à qual está associado. Exemplos comuns de clíticos incluem a parte final de uma contracção em inglês (*wouldn't* ou *you're*).

colecção

Um conjunto de dados e opções para pesquisar, analisar, indexar e procurar as origens de dados.

estrutura de análise comum (CAS)

Uma estrutura que armazena o conteúdo e os metadados de um documento, e todos os resultados de análise produzidos por um motor de análise de texto. Durante a análise de documentos, todas as trocas de dados são processadas utilizando a estrutura de análise comum.

consumidor de estrutura de análise comum (consumidor CAS)

Um consumidor que executa o processamento final nos resultados da procura que são armazenados na estrutura de análise comum. Por exemplo, um consumidor indexa o conteúdo da estrutura de análise comum num motor de procura ou preenche uma base de dados relacional com resultados de análise específicos.

Nível de comunicação comum (Common communication layer, CCL)

As infra-estrutura de comunicações que une os vários componentes (controlador, analisador, pesquisador, servidor de índice) do IBM Content Analytics with Enterprise Search.

extracção de conceitos

Função de análise de texto que identifica itens de vocabulário significativos (tais como pessoas, locais ou produtos) em documentos de texto e produz uma lista desses itens. Ver, também, extracção de temas.

correlação

Uma indicação sobre quão relevante é um valor da faceta em documentos que correspondem às condições da consulta. A classificação da correlação mede o nível de exclusividade e elevada frequência em comparação a outros documentos que correspondem à consulta. Um valor da correlação que seja superior a 1.0 representa uma anomalia que pode necessitar de mais investigação.

espaço de pesquisa

Um conjunto de origens que correspondem a padrões especificados (tais como Uniform Resource Locators (URLs), nomes de bases de dados, caminhos de sistemas de ficheiros, nomes de domínios, e endereços de IP) que um pesquisador lê para obter itens para indexação.

pesquisador

Um programa de software que obtém documentos de origens de dados e recolhe informações que podem ser utilizadas para criar índices de procura.

credencial

Informações detalhadas, adquiridas durante a autenticação, que descrevem o utilizador, quaisquer associações de grupo e outros atributos de identidade relacionados com a segurança. As credenciais podem ser utilizadas para executarem um grande número de serviços, tais como autorização, examinação e delegação. Por exemplo, as informações de início de sessão (ID de utilizador e palavra-passe) de um utilizador são credenciais que permitem ao utilizador aceder a uma conta.

motor de análise de texto personalizada

Um motor de análise de texto que é criado utilizando o kit de desenvolvimento de software (SKD) Unstructured Information Management Architecture (UIMA) e pode ser adicionado ao conjunto de motores de análise de texto de padrão (também conhecidos como anotadores base). Ver, também, motor de análise de texto.

origem de dados

Qualquer repositório de dados do qual seja possível obter documentos, tal como a web, bases de dados relacionais e não relacionais e sistemas de gestão de conteúdos.

tipo de origem de dados

Um agrupamento de origens de dados de acordo com o protocolo que está a ser utilizado para aceder aos dados.

arquivo de dados

Uma estrutura de dados em que os documentos são mantidos na sua forma analisada.

desenfileirar

Remover os itens de uma fila.

diacrítico

Uma marca que indica uma alteração no valor fonético de um carácter ou combinação de caracteres.

descobridor

Uma função de um pesquisador que determina quais as origens de dados que estão disponíveis para o pesquisador obter informações.

nome distinto

O nome que identifica de modo único uma entrada num dicionário. Um nome distinto é constituído por atributo:pares de valores, separado por vírgulas. Para além disso, também é um conjunto de pares de valores de nome (tais como NP=nome da pessoa e P=País ou região) que identificam de modo único uma entidade num certificado digital.

Document Object Model

Um sistema no qual um documento estruturado, tal como um ficheiro XML, é visualizado como uma árvore de objectos que podem ser acedidos e actualizados de forma programada.

Arquivo Domino Document Manager

Uma base de dados Domino Document Manager que é utilizada para organizar documentos. Os arquivos contêm bases de dados Domino.

Biblioteca Domino Document Manager

Uma base de dados Domino Document Manager que é o ponto de entrada para o Domino Document Manager.

Domino Internet Inter-ORB Protocol (DIIOP)

Uma tarefa de servidor que é executada no servidor e que trabalha com o Solicitador de Pedido Objecto do Domino para permitir a comunicação entre applets Java criadas com as classe de Java do Notes e o servidor Domino. Os utilizadores de navegadores e servidores Domino utilizam o DIIOP para comunicarem e trocarem dados de objectos.

classificação dinâmica

Um tipo de classificação no qual os termos na consulta são analisados em relação aos documentos que estão a ser procurados para determinar a classificação de resultados. Ver, também, classificação baseada em texto. Compare com classificação estática.

resumo dinâmico

Um tipo de resumo no qual os termos da procura são destacados e os resultados da procura contêm frases que representam da melhor forma os conceitos do documento que o utilizador procura. Compare com resumo estático.

colocar em fila

Colocar uma mensagem ou item numa fila.

carácter de alteração de controlo

Um carácter que suprime ou selecciona um significado especial para um ou mais caracteres que venham depois.

faceta Uma propriedade claramente definida de um assunto. As facetas de um determinado assunto são mutuamente exclusivas e colectivamente exaustivas. Os esquemas de classificação por facetas distinguem-se dos esquemas de categorização hierárquica, uma vez que pode ser usada mais do que uma faceta para encontrar itens interessantes.

valor da faceta

A combinação de uma faceta com uma cadeia de caracteres específica, tal como uma faceta denominada Cidade combinada com uma cadeia Nova Iorque.

procura por facetas

Um processo de procura de informações mediante a filtragem de um conjunto de tópicos que vai seleccionando progressivamente apenas os valores válidos de um sistema de classificação por facetas, que é uma colecção predefinida de facetas.

caminho das funcionalidades

Caminho utilizado para aceder ao valor de uma funcionalidade numa estrutura funcional de Unstructured Information Management Architecture (UIMA).

estrutura das funcionalidades

A estrutura de dados subjacente que representa o resultado da análise de texto. Uma estrutura funcional é uma estrutura atributo-valor. Cada estrutura funcional é de um certo tipo, e cada tipo tem um conjunto especial de funcionalidades válidas ou atributos, muito semelhante a uma classe de Java.

procura federada

Uma capacidade de procura que permite procuras através de vários serviços de procura e devolve uma lista consolidada de resultados da procura.

federação

O processo de combinar sistemas de atribuição de nomes de forma a que o sistema agregado possa processar nomes compostos de qualquer sistema de atribuição de nomes.

campo Uma área em que é introduzida uma categoria específica de dados ou informações de controlo.

procura por campo

Uma consulta restrita a um determinado campo.

texto de forma livre

Texto não estruturado que consiste em palavras ou frases.

procura de texto livre

Uma procura na qual o termo da procura é exprimido como texto de forma livre.

frequência

Uma indicação sobre quantos documentos no conjunto de documentos consultados contêm um determinado valor da faceta.

índice de texto completo

Uma estrutura de dados que faz referência a itens de dados para permitir uma procura de documentos que contenham termos da consulta.

procura aproximada

Uma procura que devolve palavras com ortografia semelhante à do termo da consulta.

procura híbrida

Uma combinação de procura Booleana e procura de texto livre.

gestão de identidade

Um conjunto de APIs que controlam o acesso a dados protegidos e que permitem aos utilizadores procurar uma colecção sem que lhes seja exigido que especifiquem um ID de utilizador e palavra-passe para cada repositório da colecção.

índice Consulte índice de texto completo.

memória cache do índice

Uma memória tampão que retém dados permitindo a reconstrução do índice sem implicar nova pesquisa dos documentos.

campo de índice

Um campo que existe apenas no índice para representar dados que sejam comuns entre múltiplas origens de entrada. Os campos de entrada podem ajudar os utilizadores a obter documentos sem terem de saber os nomes precisos dos campos.

extracção de informações

Um tipo de extracção de conceitos que reconhece automaticamente itens de vocabulário relevantes, tais como nomes, termos e expressões, em documentos de texto.

endereço de IP

Um endereço exclusivo para um dispositivo ou unidade lógica que utiliza o padrão IP.

Java Database Connectivity (JDBC)

Uma norma industrial para conectividade independente da base de dados entre a plataforma Java e um vasto leque de bases de dados. A interface JDBC fornece uma API ao nível de chamada para acesso de base de dados com base em SQL.

JavaScript

Uma linguagem de scripts da web utilizada em navegadores e servidores da web.

JavaServer Pages (JSP)

Uma tecnologia de escrita de servidor que permite que código Java seja incorporado dinamicamente dentro de páginas da web (ficheiros HTML) e que seja executado quando a página é apresentada, de forma a devolver conteúdo dinâmico a um cliente.

Java virtual machine (JVM)

Uma implementação de software que executa código Java compilado (applets e aplicações).

Katakana

Um conjunto de caracteres constituído por símbolos utilizados em um dos dois alfabetos fonéticos japoneses comuns, que é utilizado principalmente para escrever palavras estrangeiras de forma fonética.

ficheiros de base de dados de chaves

Ver conjunto de chaves.

conjunto de chaves

Em segurança informática, um ficheiro que contém chaves públicas, chaves privadas, raízes fidedignas e certificados. Ver também ficheiro de arquivo de chaves.

ficheiro de arquivo de chaves

Um conjunto de chaves que contém chaves públicas que são armazenadas como certificados de assinante e chaves privadas que são armazenadas em certificados pessoais.

identificação de idioma

Uma função de procura que determina o idioma de um documento.

lema A forma base de uma palavra, bem como as formas flexionadas da mesma categoria gramatical.

formação de lemas

Um processo que determina o lema para cada forma de uma palavra que ocorre num texto. O lema de uma palavra inclui a sua forma base, bem como as formas flexionadas da mesma categoria gramatical. Por exemplo, o lema de ir inclui todas as formas verbais do verbo. Os lemas dos nomes englobam as formas singular e plural (como chapéu e chapéus). Os lemas dos adjectivos englobam as formas comparativas e superlativas (como bom, melhor e boníssimo). Os lemas dos pronomes englobam as diversas categorias gramaticais do mesmo pronome (como eu, me, mim e migo).

afinidade lexical

Relação entre palavras de procura num documento que estão próximas umas das outras ao nível do significado. A afinidade lexical é utilizada para calcular a pertinência de um resultado.

biblioteca

Um objecto de sistema que serve como um directório para outros objectos. Ver, também, Biblioteca do Domino Document Manager.

ditongo

Dois ou mais caracteres que se ligam de forma a aparecerem como um só carácter. Por exemplo, ai e eu são caracteres que podem ser apresentados como ditongos.

Lightweight Directory Access Protocol (LDAP)

Um protocolo aberto que utiliza o TCP/IP para fornecer acesso a directórios que suportam um modelo de X.500 que não incorre nos requisitos de recursos do protocolo X.500 Directory Access Protocol (DAP), mais complexo. Por exemplo, o LDAP pode ser utilizado para localizar pessoas, organizações e outros recursos num directório da Internet ou intranet.

procura linguística

Um tipo de procura que procura, obtém e indexa um documento com termos que são reduzidos à forma base (por exemplo, para que *caracteres* seja indexado como *carácter*) ou expandidos com a forma base (tal como com palavras compostas).

análise de ligações

Um método baseado na análise de hiperligações entre documentos e utilizado para determinar quais as páginas na colecção que são importantes para os utilizadores.

federador local

Um objecto cliente criado pelas APIs de procura e índice que permitem aos utilizadores pesquisar um conjunto de colecções heterogéneas e obter um conjunto unificado de resultados da procura.

Local do Lotus Quickr

Um local da Web que é fornecido pelo Lotus Quickr permite a participantes dispersados geograficamente colaborarem em projectos e comunicarem online num espaço de trabalho estruturado e seguro.

Sala do Lotus Quickr

Uma área particionada de um local do Lotus Quickr que está restrita a membros autorizados que partilham um interesse comum e uma necessidade de trabalhar colectivamente.

carácter de máscara

Um carácter que é utilizado para representar caracteres opcionais antes, no meio e depois de um termo da procura. Os caracteres de máscara são utilizados normalmente para encontrar variações de um termo num índice. Ver, também, carácter global.

administrador principal

Uma função administrativa que permite a um utilizador gerir todo o sistema IBM Content Analytics with Enterprise Search.

tipo MIME

Um padrão da Internet para identificar o tipo de objecto que está a ser transferido através da Internet.

supervisor

Um utilizador que tem autoridade para observar processos de nível de colecção.

carácter de mudança de linha

Um carácter de controlo faz com que a posição de impressão ou de apresentação se mova uma linha para baixo.

segmentação n-gram

Um método de segmentação que considera a sobreposição de sequências de um número específico de caracteres como uma única palavra. Consulte também segmentação. Contraste com Segmentação de espaços em branco baseada em Unicode.

directiva de não seguimento

Uma directiva numa página da web que fornece instruções aos robôs (como o pesquisador da Web) para não seguirem as ligações que se encontram nessa página.

directiva sem índice

Uma directiva numa página da web que fornece instruções aos robôs (como o pesquisador da Web) para não incluírem o conteúdo dessa página no índice.

Chamada de procedimento remoto do Notes (NRPC)

Mecanismo de comunicação do Lotus Notes utilizado para todas as comunicações de Notes para Notes.

termo opaco

Um termo de consulta que não é analisado pelo analisador de consultas linguísticas. Em vez disso, os termos opacos são identificados pela sua sintaxe específica de implementação, como sintaxe específica para pesquisar documentos XML com um idioma de consulta XML. Os termos de consulta opacos começam com o carácter @ e o identificador do idioma de consulta. Por exemplo, @xmlf2 especifica que a consulta deve ser processada pelo idioma de consulta do fragmento XML e @xmlp especifica que a consulta deve ser processada pelo idioma de consulta XPath.

operador

Um utilizador que tem a autoridade para observar, iniciar e parar processos de nível de colecção.

procura paramétrica

Um tipo de procura que procura objectos que contenham um atributo ou valor numérico, tal como datas, números inteiros ou outros tipos de dados numéricos dentro de um intervalo especificado.

analisador

Um programa que interpreta documentos adicionados ao arquivo de dados. O analisador extrai informações dos documentos e prepara-os para indexar, procurar e obter.

controlador de analisador

Um serviço que alimenta o serviço do analisador com documentos. Existe um controlador de analisador por cada colecção. O serviço de controlador de analisador de uma colecção corresponde ao analisador da colecção na consola de administração.

serviço analisador

O serviço que processa a totalidade da análise de documentos e do processamento da análise de texto em colecções de documentos. Existe sempre um serviço analisador em execução, no mínimo.

local Uma localização virtual visível no portal onde indivíduos e grupos se encontram para colaborarem. Num portal, cada utilizador tem um local

pessoal para trabalho privado, e os indivíduos e grupos têm acesso a uma variedade de locais partilhados, que tanto podem ser locais públicos como restritos. Consulte também Local do Lotus Quickr.

classificação popular

Um tipo de classificação que sobe a classificação existente de um documento com base na popularidade do documento.

arquivo de motor de processamento

Um ficheiro de arquivo zip .pear que inclui um motor de análise Unstructured Information Management Architecture (UIMA) e todos os recursos requeridos para o utilizar em análises personalizada.

procura de proximidade

Uma procura de texto que devolve um resultado quando dois padrões de procura ocorrem dentro de uma distância específica um do outro.

servidor proxy

Um servidor que faz de intermediário para pedidos HTTP da web que sejam alojados por uma aplicação ou um servidor da web. Um servidor proxy faz de substituto para os servidores de conteúdo na empresa.

ligação rápida

Associação entre um Uniform Resource Identifier (URI) e palavras-chave ou expressões.

classificação

A atribuição de um valor inteiro a cada documento nos resultados da procura a partir de uma consulta. A ordem dos documentos nos resultados da procura é baseada na pertinência em relação à consulta. Uma classificação mais alta significa uma correspondência mais aproximada. Ver, também, classificação dinâmica e classificação estática.

arquivo de dados não processados

Estrutura de dados onde os documentos pesquisados são armazenados antes de serem enviados ao analisador. Os pesquisadores escrevem no arquivo de dados não processados e o analisador lê a partir desse arquivo de dados não processados. Uma vez analisados os documentos, estes são removidos do arquivo de dados não processados. Não confundir com arquivo de dados.

anotador de expressões globais

Um componente de software que detecta entidades ou unidades de informação num documento de texto, como números de produtos, com base em expressões regulares que descrevem os padrões exactos que são pesquisados no documento de texto. Se uma das expressões globais corresponder a partes do texto do documento, o anotador de expressões globais cria as anotações correspondentes que incluem a correspondência, ou parte da mesma. Estas expressões anotadas são posteriormente armazenadas no índice, através de um ficheiro de correlação de índice, ou numa base de dados compatível com JDBC, através de um ficheiro de correlação de base de dados.

federador remoto

Um federador de servidor que associa um conjunto de objectos passíveis de serem procurados.

Protocolo Robots Exclusion

Um protocolo que permite aos administradores de sítios da Web indicarem aos robôs visitantes quais as partes do sítio que não devem ser visitadas pelo robô. sala Um programa que permite aos utilizadores criarem documentos para outros lerem, responderem a comentários de outros utilizadores e reverem o estado do projecto e as datas de conclusão. Os utilizadores também podem conversar com outros utilizadores que se encontrem na mesma sala. Consulte também Sala do Lotus Quickr.

categoria baseada em regras

Categorias criadas por regras que especificam quais os documentos que estão associados a quais categorias. Por exemplo, pode definir regras para associação de documentos que contenham ou excluam determinadas palavras ou que correspondam a um padrão de Uniform Resource Identifier (URI), com categorias específicas.

aplicação de procura

Um programa que processa consultas, pesquisa o índice, devolve os resultados da procura e obtém os documentos de origem.

cache de procura

Uma memória tampão que mantém os dados e os resultados de pedidos de procura anteriores.

motor de procura

Um programa que aceita um pedido de procura e devolve uma lista de documentos ao utilizador.

resultados da procura

Uma lista de documentos que correspondem ao pedido da procura.

Secure Sockets Layer (SSL)

Um protocolo de segurança que fornece privacidade de comunicações. Com SSL, as aplicações de cliente/servidor podem comunicar de uma forma que se destina a impedir a leitura não autorizada, a adulteração e a falsificação de mensagens.

token de segurança

Informações sobre a identidade e a segurança que são utilizadas para autorizar o acesso a documentos numa colecção. Diferentes tipos de origem de dados suportam diferentes tipos de tokens de segurança. Os exemplos incluem funções de utilizador, IDs de utilizador, IDs de grupo e outras informações que podem ser utilizadas para controlar o acesso a conteúdos.

página da lista de valores geradores

No WebSphere Portal, uma página XML que contém ligações às páginas disponíveis num portal. Os pesquisadores podem utilizar a lista de valores geradores para identificar os documentos a pesquisar. A página da lista de valores geradores contém também metadados armazenados em conjunto com os documentos pesquisados no índice.

Uniform Resource Locator (URL) inicial

O ponto de partida para uma pesquisa.

segmentação

A divisão de texto em unidades lexicais distintas, tais como palavras, orações, frases, parágrafos ou lemas. Consulte também segmentação n-gram e Segmentação de espaços em branco baseada em Unicode.

procura semântica

Um tipo de procura por palavra-chave que incorpora a análise linguística e contextual. Ver também análise de texto.

servlet

Um programa de Java que é executado num servidor da web e estende a

funcionalidade do servidor gerando conteúdos dinâmicos como resposta aos pedidos de clientes da web. Os servlets são utilizados frequentemente para ligarem bases de dados à web.

shingle

Uma cadeia de tokens (palavras) consecutivos retirados de uma frase. Por exemplo, de "Esta é uma frase muito curta.", os shingles de 3 palavras (ou trigramas) são:

Esta é uma é uma frase uma frase muito frase muito curta

Os shingles podem ser utilizados em linguística estatística. Por exemplo, se dois textos diferentes tiverem muitos shingles em comum, os textos estão provavelmente relacionados de alguma forma.

página de erros esporádicos

Um tipo de página da web que fornece informações sobre o motivo pelo qual não é possível devolver a página da web solicitada. Por exemplo, em vez de devolver um simples código de estado, o servidor de HTTP pode devolver uma página que explica detalhadamente o código de estado.

classificação estática

Um tipo de classificação no qual os factores sobre os documentos que estão a ser classificados, tais como a data, o número de ligações que apontam para o documento e etc., aumentam a classificação. Contraste com classificação dinâmica.

resumo estático

Um tipo de resumo no qual os resultados da procura contêm um resumo específico, armazenado do documento. Contraste com resumo dinâmico.

stemming

Consultar stemming de palavras.

palavra de paragem

Uma palavra que é utilizada frequentemente, tal como *o*, *um* ou *e*, que é ignorada pela aplicação de procura.

remoção de palavras de paragem

O processo de remover palavras de paragem da consulta para ignorar palavras comuns e devolver resultados mais relevantes.

resumo

O processo de incluir instruções não redundantes nos resultados da procura para descrever resumidamente o conteúdo de um documento. Ver, também, resumo dinâmico e resumo estático.

dicionário de sinónimos

Um dicionário que permite ao utilizador procurar sinónimos dos termos da consulta quando procuram uma colecção.

taxonomia

Uma classificação de objectos em grupos baseada em semelhanças. Uma taxonomia organiza os dados em categorias e subcategorias. Ver também árvore de categorias.

análise de texto

O processo de extrair semântica e outras informações do texto para melhorar a possibilidade de obtenção de dados numa colecção. Ver também procura semântica.

análise de texto

Uma forma de processamento de idioma natural que inclui técnicas de aprendizagem linguística, estatística e de máquina para analisar texto e extrair informação chave para integração empresarial.

motor de análise de texto

Um componente de software que é responsável por encontrar e representar conteúdos semânticos e de contexto em textos.

classificação baseada em texto

O processo de atribuir um número inteiro a um documento que signifique a importância do documento em relação aos termos numa consulta. Um valor inteiro mais elevado significa uma correspondência mais aproximada com a consulta. Ver, também, classificação dinâmica.

extractor de texto

Um componente que utiliza a tecnologia de filtragem de documentos baseada em Oracle Outside In Content Access para identificar formatos de documentos.

segmentação de texto

Ver segmentação.

extracção de temas

Um tipo de extracção de conceitos que reconhece automaticamente itens de vocabulário relevantes em documentos de texto para extrair o tema ou tópico de um documento. Ver, também, extracção de conceitos.

token As unidades textuais básicas que são indexadas. Os tokens podem ser as palavras num idioma ou outras unidades de texto adequadas para indexação.

segmentação

O processo de analisar a entrada em tokens.

segmentador

Um programa de segmentação que digitaliza texto e determina se, e quando, uma série de caracteres pode ser reconhecida como um token.

carácter de seguimento

Um carácter que tem a última posição numa palavra.

sistema de tipos

O sistema de tipos define os tipos de objectos (estruturas funcionais) que podem ser identificados num documento por um motor de análise de texto. O sistema de tipos define todas as estruturas funcionais possíveis em termos de tipos e funcionalidades. Pode definir qualquer número de tipos diferentes num sistema de tipos. Um sistema de tipos é específico de domínio e aplicação.

Segmentação de espaços em branco baseada em Unicode

Um método de segmentação que utiliza propriedades de carácter Unicode para distinguir entre token e caracteres separadores. Consulte também segmentação. Contraste com segmentação n-gram.

Uniform Resource Identifier (URI)

Uma cadeia de caracteres compacta que identifica um recurso abstracto ou físico.

Uniform Resource Locator (URL)

O endereço exclusivo de um recurso informativo que está acessível numa rede, como a Internet. O URL inclui o nome abreviado do protocolo utilizado para aceder ao recurso informativo e as informações utilizadas pelo protocolo para localizar o recurso informativo.

Unstructured Information Management Architecture (UIMA)

Uma arquitectura da IBM que define um contexto para implementar sistemas para a análise de dados não estruturados.

agente do utilizador

Uma aplicação que procura a web e deixa informações próprias nos sítios que visita. Por exemplo, um pesquisador da Web é um agente do utilizador.

pesquisador da Web

Um tipo de pesquisador que explora a web obtendo um documento da web e seguindo as ligações nesse documento.

procura de termo ponderado

Uma consulta em que é dada mais importância a determinados termos.

carácter global

Um carácter que é utilizado para representar caracteres opcionais antes, no meio ou depois de um termo da procura.

stemming de palavras

Um processo de normalização linguística no qual as formas variantes de uma palavra são reduzidas a um formato comum. Por exemplo, palavras como *programação*, *programado*, e *programável* são reduzidas a *programa*.

XML Path Language (XPath)

Uma linguagem destinada a identificar ou endereçar de forma exclusiva partes de dados de XML de origem, para utilização com tecnologias relacionadas com XML, como XSLT, XQuery e analisadores de XML. XPath é um padrão do World Wide Web Consortium.

Avisos

Estas informações foram desenvolvidas para produtos e serviços disponibilizados nos E.U.A.

Os produtos, serviços ou funcionalidades descritos neste documento poderão não ser disponibilizados pela IBM noutros países. Consulte o seu representante IBM para obter informações sobre os produtos e serviços actualmente disponíveis na sua área. Quaisquer referências, nesta publicação, a produtos, programas ou serviços IBM não significam que apenas esses produtos, programas ou serviços IBM possam ser utilizados. Qualquer outro produto, programa ou serviço, funcionalmente equivalente, poderá ser utilizado em substituição daqueles, desde que não infrinja nenhum direito de propriedade intelectual da IBM. No entanto, é da inteira responsabilidade do utilizador avaliar e verificar o funcionamento de qualquer produto, programa ou serviço não IBM.

Neste documento, podem ser feitas referências a patentes ou a pedidos de patente pendentes. O facto de este documento lhe ser fornecido não lhe confere nenhum direito sobre essas patentes. Caso solicite pedidos de informação sobre licenças, tais pedidos deverão ser endereçados, por escrito, para:

IBM Director of Licensing IBM Corporation North Castle Drive Armonk, NY 10504-1785 E.U.A.

Para endereçar os seus pedidos de informação sobre licenças relacionados com informações de conjunto de caracteres de duplo byte (DBCS, Double Byte Character Set), contacte o Departamento de Propriedade Intelectual do seu país ou envie-os, por escrito, para:

Intellectual Property Licensing Legal and Intellectual Property Law IBM Japan Ltd. 1623-14, Shimotsuruma, Yamato-shi Kanagawa 242-8502 Japão

O parágrafo seguinte não se aplica ao Reino Unido nem a nenhum outro país onde estas cláusulas sejam inconsistentes com a lei local: A INTERNATIONAL BUSINESS MACHINES CORPORATION FORNECE ESTA PUBLICAÇÃO "TAL COMO ESTÁ", SEM GARANTIA DE QUALQUER ESPÉCIE, EXPLÍCITA OU IMPLÍCITA, INCLUINDO, MAS NÃO SE LIMITANDO ÀS GARANTIAS IMPLÍCITAS DE NÃO INFRACÇÃO, COMERCIALIZAÇÃO OU ADEQUAÇÃO A UM DETERMINADO FIM. Alguns Estados não permitem a exclusão de garantias, explícitas ou implícitas, em determinadas transacções; esta declaração pode, portanto, não se aplicar ao seu caso.

Esta publicação pode conter imprecisões técnicas ou erros de tipografia. A IBM permite-se fazer alterações periódicas às informações aqui contidas; essas alterações serão incluídas nas posteriores edições desta publicação. A IBM pode introduzir melhorias e/ou alterações ao(s) produto(s) e/ou programa(s) descrito(s) nesta publicação em qualquer momento, sem aviso prévio.

Quaisquer referências, nesta publicação, a sítios da Web não IBM são fornecidas apenas para conveniência e não constituem, em caso algum, aprovação desses sítios da Web. Os materiais existentes nesses sítios da Web não fazem parte dos materiais destinados a este produto IBM e a utilização desses sítios da Web será da exclusiva responsabilidade do utilizador.

A IBM pode usar ou distribuir quaisquer informações que lhe forneça, da forma que julgue apropriada, sem incorrer em nenhuma obrigação para consigo.

Os licenciados deste programa que pretendam obter informações sobre o mesmo com o objectivo de permitir: (i) a troca de informações entre programas criados independentemente e outros programas (incluindo este) e (ii) a utilização recíproca das informações que tenham sido trocadas, deverão contactar a IBM através do seguinte endereço:

IBM Corporation J46A/G4 555 Bailey Avenue San Jose, CA 95141-1003 E.U.A.

Tais informações poderão estar disponíveis, sujeitas aos termos e às condições adequados, incluindo, em alguns casos, o pagamento de um encargo.

O programa licenciado descrito neste documento e todo o material licenciado disponível para o programa são fornecidos pela IBM nos termos das Condições Gerais IBM (IBM Customer Agreement), do Acordo de Licença Internacional para Programas IBM (IPLA, IBM International Program License Agreement) ou de qualquer acordo equivalente entre ambas as partes.

Estas informações contêm exemplos de dados e relatórios utilizados em operações comerciais diárias. Para ilustrá-los o melhor possível, os exemplos incluem nomes de indivíduos, firmas, marcas e produtos. Todos estes nomes são fictícios e qualquer semelhança com nomes e moradas reais é mera coincidência.

Quaisquer dados de desempenho aqui contidos foram determinados num ambiente controlado. Assim sendo, os resultados obtidos noutros ambientes operativos podem variar significativamente. Algumas medições podem ter sido efectuadas em sistemas ao nível do desenvolvimento, pelo que não existem garantias de que estas medições sejam iguais nos sistemas disponíveis habitualmente. Para além disso, algumas medições podem ter sido calculadas por extrapolação. Os resultados reais podem variar. Os utilizadores deste documento devem verificar os dados aplicáveis ao seu ambiente específico.

As informações relativas a produtos não IBM foram obtidas junto dos fornecedores desses produtos, dos seus anúncios publicados ou de outras fontes de divulgação ao público. A IBM não testou esses produtos e não pode confirmar a exactidão do desempenho, da compatibilidade ou de quaisquer outras afirmações relacionadas com produtos não IBM. Todas as questões sobre as capacidades dos produtos não IBM deverão ser endereçadas aos fornecedores desses produtos.

Todas as afirmações relativas às directivas ou tendências futuras da IBM estão sujeitas a alterações ou descontinuação sem aviso prévio, representando apenas metas e objectivos.

LICENÇA DE COPYRIGHT:

Esta publicação contém programas de aplicação exemplo na linguagem de origem, que ilustra técnicas de programação em várias plataformas operativas. Pode copiar, modificar e distribuir estes programas exemplo de qualquer forma, sem encargos para com a IBM, com a finalidade de desenvolver, utilizar, comercializar ou distribuir programas de aplicação conformes à interface de programação de aplicações e destinados à plataforma operativa para a qual os programas exemplo são escritos. Estes exemplos não foram testados exaustivamente sob todas as condições. Deste modo, a IBM não garante nem se responsabiliza pela fiabilidade, assistência ou funcionamento implícito destes programas.

Cada cópia, ou qualquer parte destes programas exemplo, ou qualquer trabalho derivado dos mesmos, tem de incluir um aviso de direitos de autor, do seguinte modo:

© (nome da empresa) (ano). Algumas partes deste código são derivadas de Programas Exemplo da © Copyright IBM Corp. _introduza o ano ou anos_.

Algumas partes deste produto são:

- Oracle[®] Outside In Content Access, Copyright © 1992, 2012, Oracle.
- Materiais Licenciados IBM XSLT Processor Propriedade da IBM © Copyright IBM Corp., 1999-2012.

Este produto utiliza fornecedores criptográficos aprovados pelo FIPS 140-2; IBMJCEFIPS (certificado 376) e/ou IBMJSSEFIPS (certificado 409) e/ou IBM Crypto for C (ICC (certificado 384) para criptografia. Os certificados estão listados no sítio NIST da Web em http://csrc.nist.gov/cryptval/140-1/1401val2004.htm.

Marcas comerciais

IBM, o logótipo IBM e ibm.com são marcas comerciais ou marcas comercias registadas da International Business Machines Corp., registadas em várias jurisdições a nível mundial. Outros nomes de produtos e serviços poderão ser marcas comerciais da IBM ou de outras empresas. Está disponível uma lista actual de marcas comerciais da IBM na Web em Informação sobre copyright e marcas comerciais em www.ibm.com/legal/copytrade.shtml.

Os seguintes termos são marcas comerciais ou marcas comerciais registadas de outras empresas:

Adobe, o logótipo Adobe, PostScript e o logótipo PostScript são marcas comerciais registadas ou marcas comerciais da Adobe Systems Incorporated nos Estados Unidos e/ou noutros países.

Intel e Pentium são marcas comerciais ou marcas comerciais registadas da Intel Corporation ou das respectivas subsidiárias nos E.U.A e noutros países.

Java e todas as marcas comerciais e logótipos baseados em Java são marcas comerciais ou marcas comerciais registadas da Oracle e/ou respectivas filiais.

Linux é uma marca comercial registada da Linus Torvalds nos E.U.A. e/ou noutros países.

Microsoft, Windows, Windows NT e o logótipo do Windows são marcas comerciais da Microsoft Corporation nos Estados Unidos e/ou noutros países.

UNIX é uma marca comercial registada de The Open Group nos Estados Unidos e noutros países.

Outros nomes de empresas, produtos ou serviços podem ser marcas comerciais ou marcas de serviço de terceiros.

Índice Remissivo

Α

adicionar servidores 29 administração do sistema 25 administrador alterar a palavra-passe em servidores múltiplos 27 alterar a palavra-passe num único servidor 26 agendamento pesquisadores 57, 60 agentes do utilizador 135 agrupamentos de documentos 240 AIX criação de miniaturas 231 AIX, Linux suporte do protocolo IPv6 41 alertas configuração de servidor SMTP 449 descrição 445 documentos indexados 446 documentos pesquisados 446 espaço livre nos servidores 447 nível de colecção 446 nível de sistema 447 opções de correio electrónico 446, 447 receber correio electrónico para 450 tempos de resposta da procura 446 alertas de espaço livre 447 âmbitos criar 298 descrição 297 Formatos URI 164 padrões URI 297, 298 procurar 297 analisador de ASCII 269 analisadores analisador de ASCII 269 analisar tipos de documentos 267 análise de ligações 227 análise de termos compostos 178 caracteres numéricos 283 caracteres SBCS 283 criação de miniaturas 229 descrição 17, 177 detecção de documentos de chegada 178, 434 detecção de idioma 280 detecção de páginas de códigos 281 detecção do formato do documento 265 estado de espera 551 falha ao iniciar 552 ficheiros de correlacionamento incorrectos 551 ficheiros sem extensões 269 geração de miniaturas em AIX 231 idiomas suportados 279 opções de analisadores 178 opções de serviços de indexação 178 processamento linguístico 279

analisadores (continuação) procura XML nativa 179 segmentação híbrida 282 segmentação n-gram 282, 283 selecção do tipo de analisador 265 sem resposta 551 sessões múltiplas para o extractor de texto 275 supervisionar 434 tarefas de análise de dados 17 textos bidireccional em ficheiros PDF 275 tipos de documentos 268 tipos de documentos desconhecidos 269 tipos de documentos do extractor de texto suportados 272 tipos de documentos para o extractor de texto 269 tipos de documentos predefinidos 266 tipos MIME 268 análise de conteúdos análise de frequência 11 análises altamente escaláveis 12 análises periódicas 12 correlações 11 descrição geral 9 encadeamento de análise 10 exportar facetas 12 relatórios de facetas 12 análise de ligações descrição 227 análise de postura configuração 262 análise de texto correlacionar elementos de XML 218 dicionários lexicais 190, 191 estruturas de análise comum 219, 220 ficheiros de regras 260 índice de facetas 261 inspecção profunda 307 motores de análise de texto 216, 217 reconstruir o índice de facetas 441 reimplementar recursos 441 termos de identificação de interesse 262 análise de texto âncora análise global 395 descrição 363 indexar de documentos 396 segurança de colecção 395 análise de texto personalizada correlacionar a estrutura de análise comum com o índice 219 correlacionar a estrutura de análise comum com uma base de dados relacional 220 correlacionar elementos de XML 218 análise de texto personalizada (continuação) correlacionar resultados da análise com o índice 219 correlacionar resultados da análise com uma base de dados relacional 220 descrição 214 dicionários lexicais 190, 191 motores de análise de texto 216, 217 análise global análise de texto âncora 363, 395 descrição 18 detecção de documentos duplicados 228, 363, 382 personalizadas descrição 233 análise global personalizada descrição 233 anotador de Voz do cliente 245 Anotador do Content Classification 503 anotadores 214 activar 244 administração 243 Content Classification 503 descrição geral 10 Voz do cliente 245 APIs descrição 22 aplicação ESAdmin iniciar sessão 25 aplicações ficheiro config.properties 346 aplicações de análise IDs da aplicação 382 segurança ao nível da colecção 382 aplicações de procura empresarial aceder 324 alterar o número da porta 34 associar a colecções 381 configuração de expansão de grupo 360 descrição 23 deslocar resultados 361 ficheiro config.properties 326 fornecidas 323 ID de aplicação predefinido 323 IDs da aplicação 382 iniciar 324 modo de análise 347 personalizadas 323 personalizar 325, 347 personalizar o cabeçalho de apresentação 355 propriedade searchable.sessional 361 segurança ao nível da colecção 382 árvores de categorias descrição 237 autenticação descrição 375 Autenticação base de HTTP 146, 147

autenticação baseada em formulário 146, 148
autenticação de início de sessão único servidor de aplicação da Web incorporado 368
WebSphere Application Server 371
autorização, descrição 375

В

bases de conhecimento, Content Classification 507 bases de dados da Oracle acesso com os pesquisadores da DB2 74 acesso com os pesquisadores de Base de dados IDBC 98 Bases de dados de DB2 acesso com os pesquisadores da DB2 74 acesso com os pesquisadores de Base de dados IDBC 98 bases de dados do SQL Server acesso com os pesquisadores da DB2 74 acesso com os pesquisadores de Base de dados JDBC 98 **BigInsights** análise global personalizada 233

С

cache de procura configurar 287 descrição 287 cadeia do contexto de segurança do utilizador 388 cadeia USC 388 Caminhos de saída de XML, documentos exportados 480 campo de data apresentar 223 hora epoch 223 personalizadas 223 predeterminado 223 campos filtrar 221, 222 campos analisáveis, descrição 195 campos de índice analisável 195 atributos 195 campos ordenáveis 195 correlacionamento com facetas nas colecções de procura empresarial 182, 184, 186 correlacionamento para elementos HTML 210 correlacionar com coleccões de análise de conteúdos 252, 255 correlacionar com elementos XML 206, 208 correlacionar com metadados de HTML 209 correlacionar metadados extraídos do conteúdo binário 212, 213 correspondência exacta 195

campos de índice (continuação) criar 194 descrição 192 devolvido 195 exportar 205 filtrar 221, 222 formato XML 205 importar 205 ligar a facetas 192 ligar a origens de entrada 192 nomes reservados 199 ocultar origens de entrada e facetas 206 procura de texto livre 195 procura facetada 195 procura paramétrica 195 procura por campo 195 campos de índice XML correlacionar elementos com 218 descrição 218 campos ordenáveis ordenação de cadeias 195 ordenação numérica 195 campos paramétricos descrição 195 formatos de data 223 ordenação numérica 195 caracteres globais em consultas 296 expansão da consulta 296, 297 expansão do índice 296 categorias agrupamentos 234 agrupamentos de documentos 240 árvores de categorias 237 baseadas em regras 234, 235 criar 237 descrição 234 imbricar subcategorias 237 categorias baseadas em regras configurar 237 descrição 235 centro de informações alterar a raiz de contexto do URL 35 alterar o número da porta 35 Chinês caracteres SBCS 283 expansão de sinónimos 283 classificação baseada em texto 313 classificação dinâmica 313 classificação estática descrição 314 classificar resultados da procura agregação de resultados 320 classificação baseada em texto 313 descrição 313 dicionários de palavras hierárquicas 318 dinâmicos 313 estático 314 filtros de classificação de documentos 320 importância do documento 314 padrões URI 318, 319 regras de expansão de consulta 320 clonagem colecções 53

clonagem (continuação) pesquisadores 57 codificação de byte múltiplo 281 codificação de byte único 281 codificação unicode 281 códigos de estado de HTTP 430 recebido por pesquisadores Web 429 relatório do pesquisador Web 429 códigos de retorno 511 cofre de credenciais 380 colecções activar cache de documento 51 activar miniaturas 51 administração 45 analisar 177 associar a aplicações 381 clonagem 53 criar 51 descrição 15 detecção de documentos duplicados 228 determinar o ID 55 edicão 53 eliminar 54 exportar dados de configuração 458 exportar documentos 469 federação 45 ignorar controlos de acesso de nível de documento 414 importar dados de configuração 458 indexação 177 para servidores BigInsights 51 processamento de documentos 177 segurança de documentos duplicados 382 segurança de texto âncora 395 segurança do ID da aplicação 382 servidores de procura 285 colecções de análise de conteúdo análise de postura 247 anotadores 243 configuração de análise de postura 262 configurar 247 criar através da exportação de documentos 304 Definições de colecção 247 Definições de escala de tempo 247 definições de escala de tempo cíclica 247 dicionários de utilizador 257 dicionários de utilizador de ficheiros CSV 259 facetas de campo de data 247 facetas de metadados de documentos 247 ficheiros de regras 260 índice de facetas 261 inspecção profunda 307 procura facetada 249 reconstruir o índice de facetas 441 reimplementar recursos 441 resultados de inspecção profunda 308 resumo de todas as funções 45 sinalizadores de documentos 303 supervisionar 441

colecções de análise de conteúdo (continuação) termos de identificação de interesse 262 vistas contextuais 263 colecções de procura empresarial anotadores 243 dicionários lexicais 189, 190, 191 exportar documentos para colecções de análise de conteúdos 304 resumo de todas as funções 45 sinalizadores de documentos 303 colecções dinâmicas 469 colecções federadas procura facetada 180 colecções federadas da procura empresarial 45 Comando Add-SPSolution 132 comando de colecções do relatório esadmin 55 comando de exportação esadmin 458 comando de importação esadmin 458 comando esadmin 511 comando esadmin startSearch 421 comando esadmin stopSearch 421 comando esadmin system startall 511 comando esadmin system stopall 511 comando esadmin verifySecureCrawler 402 comando eschangewaspw configuração de servidor distribuído 374 configuração de um único servidor 372 comando esviewlog 453 comando startLocal 422 comando stopLocal 422 Comando stsadm 132 comandos 511 componentes do sistema iniciar 420 parar 420 Conector Java para Content Manager EE 70, 72 conectores de Documentum 550 configuração da Assistência de Directório 408 configuração da tabela de factos de documentos 485 configuração da tabela de metadados de documentos 485 configuração da tabela de metadados de texto 485 configuração de adaptadores de circuito fechado 37 configuração de expansão de grupo 360 Configuração de Servidor Fidedigno 405 configuração de servidor LDAP Pesquisadores do Notes 410 configuração de servidor SMTP 449 Configuração de utilizadores do Domino, pesquisadoresQuickr for Domino 407 Configuração do Exchange Web Service 87 configuração em WebSphere Application Server 128

consola de administração descrição 20 iniciar sessão 25 consultas populares, supervisionar 439 consultas recentes, supervisionar 439 Content Analytics Miner descrição 23 Content Classification classificação com base em categorias 509 configuração 506 exportar para 507 integração 503, 504 planos de decisão 504, 506 pontuações por categoria 504, 506 procura conceptual 504, 506, 509 treinar bases de conhecimento 507 conteúdo de documento, descrição 195 controladores JDBC pesquisadores de Base de dados JDBC 98 controlos de acesso autenticação de início de sessão 365 desactivar para uma colecção 414 descrição 375 Ficheiro de chave LTPA 392 gestão de identidade 388, 391 requisitos para o Lotus Domino 405 requisitos para sistemas de ficheiros do Windows 411 segurança ao nível do documento 386 segurança de início de sessão único 392 tokens LTPA 392 validação do utilizador actual 387 cookies para pesquisa da Web configurar 151 descrição 150 formato 150 cópia de segurança do sistema 455, 456 Coreano análise de termos compostos 178 caracteres SBCS 283 expansão de sinónimos 283 correlacionar a estrutura de análise comum com bases de dados relacionais 220 elementos de XML com a estrutura de análise comum 218 elementos HTML para campos do índice 210 elementos XML com campos de índice 208 estruturas de análise comuns com o índice 219 resultados da análise com bases de dados relacionais 220 credenciais de ligação 380 criação de miniaturas em AIX 231 tipos de documentos suportados 229 criar âmbitos 298 facetas para colecções de análise de conteúdo 251 filtros de campos 222

criar *(continuação)* ligações rápidas 295 relatórios de pesquisadores Web 429

D

Data Event Publisher, configuração do pesquisador de DB2 77 Data Listener configurar 160 reiniciar 160, 442 supervisionar 442 detalhes do módulo, supervisionar 426 detalhes do URI supervisionar 436 detecção automática idiomas 280 páginas de códigos 281 detecção de documentos duplicados activar a segurança 382 análise global 228, 382 descrição 228, 363 diagrama de fluxo de dados 23 dicionários de palavras de paragem adicionar ao sistema 293 descrição 292 utilizar com uma colecção 294 dicionários de palavras hierárquicas adicionar ao sistema 317 descrição 316 utilizar com uma colecção 318 dicionários de sinónimos adicionar ao sistema 291 descrição 289 segmentação n-gram 283 utilizar com uma colecção 291 dicionários de utilizador configurar 257 importar ficheiros CSV 259 dicionários lexicais adicionar ao sistema 190 associar a colecções 191 descrição 189 directivas para não indexar configurar 154 descrição 153 directivas para não seguir configurar 154 descrição 153 directório master_config falhas de sincronização 552 limitações de tamanho de ficheiro 552 documentação encontrar 565 HTML 565 PDF 565 documentação HTML 565 documentação PDF 565 documentos abandonados rastrear 437 documentos binários correlacionar metadados com campos de índice 212, 213 extracção de metadados 212 extrair metadados 213

documentos de XML procura XML nativa 179 procurar 206 documentos docm indexação 276 documentos HTML procurar 209 documentos pesquisados para Content Classification 507 domínios do Lotus Domino 405 domínios do Windows 411

Ε

edição aplicações de Data Listener 160 coleccões 53 propriedades da aplicação 326, 346 efectuar a cópia de segurança do sistema 455 Elementos de metadados de HTML correlacionamento com campos de índice 209 elementos de XML correlacionamento com campos de índice 206, 208 correlacionar com a estrutura de análise comum 218 procurar 218 elementos Dublin Core 210 elementos HTML correlacionamento com campos de índice 210 elementos Dublin Core 210 eliminar colecções 54 pesquisadores 60 encriptação de segurança 363 endereços de IP adaptador de circuito fechado 37 Suporte IPv6 em AIX, Linux 41 suporte IPv6 em Windows 38 suporte para duplos 37 endereços de IP, alterar 35 erros CCL, utilitário reportcel 543, 545 espaço de pesquisa alertas sobre 446 configuração do pesquisador da Web 139 descrição 16 edição 59 espaço global do pesquisador da Web 151 ESSearchPortlet clonagem 559 resolução de problemas 559 sessão HTTP 559 ESSPSolution.wsp 132 estado do sistema pesquisadores 423 Pesquisadores da Web 426 estatísticas de consultas exportar 439 supervisionar 439 estruturas de análise comum correlacionar com bases de dados relacionais 220

estruturas de análise comum (continuação) correlacionar com o índice 219 correlacionar elementos de XML com 218 descrição 214 Exchange Server pesquisadores do 2000 e do 2003 formatos URI 164 expansão da consulta descrição 296 expansão do índice descrição 296 exportações de documentos analisados 469 exportações de documentos pesquisados 469 exportar 507 caminhos de saída de XML 480 campos de índice 205 como ficheiros CSV 492 descrição 469 Directório partilhado Windows 479 documentos 481 documentos analisados 472 documentos pesquisados 472 documentos procurados 476 estatísticas de consultas 439 ficheiros CSV formato de ficheiro exportado 495 ficheiros de correlação da base de dados relacional 485 formato de ficheiro CSV 495 formato XML, campos de índice 205 formato XML, documentos exportados 481 para bases de dados de IBM Cognos BI 466 resultados de inspecção profunda 308 extensões do nome do ficheiro excluir de espaços de pesquisa da Web 139 suportadas pelo extractor de texto 269 suportou o extractor de texto 272 extractor de texto analisar tipos de documentos 267 associar tipos de documentos 269 descrição 265 resolução de problemas em documento corrompido 553 saída de rastreio de pilha 554 sem resposta 551 sessões múltiplas 275 tipos de documentos predefinidos 266, 272

F

facetas campos de data 247 colecções de análise de conteúdo 249 colecções de procura empresarial 180 configuração de dados 182 configuração hierárquica 182 configuração simples 182

facetas (continuação) Criar para colecções de análise de conteúdos 251 descrição geral 11 documento de metadados 247 exportar 12 indexação 261 intervalos de data para colecções de análise de conteúdos 255 intervalos de data para colecções de procura empresarial 186 intervalos decimais para colecções de análise de conteúdos 252 intervalos decimais para colecções de procura empresarial 184 procura de colecções federadas 180 raiz 251 relatórios 12 subfacetas 251 termos de interesse 262 ver correlações de campo de índice 192 facetas de data configuração 182 descrição 180 facetas de intervalo de data colecções de análise de conteúdo 255 colecções de procura empresarial 186 descrição 180 facetas de intervalo decimal colecções de análise de conteúdo 252 colecções de procura empresarial 184 descrição 180 facetas de metadados automáticos 247 facetas hierárquicas configuração 182 descrição 180 facetas planas configuração 182 descrição 180 facetas raiz para exploração de conteúdo 251 factores hierárquicos para dicionários de palavras hierárquicas 316 para padrões URI 318, 319 FFQC4814E resolução de problemas 552 Ficheiro AutoRunComponents.properties 420 ficheiro CCLServer_date.log 32 ficheiro config.properties 372, 374 descrições da propriedade 326 edição 346 personalizar 347 ficheiro cookies.ini configurar 151 descrição 150 formato 150 ficheiro crawl.rules 151 ficheiro crawler_rdb_plugin.xml 102, 106 ficheiro de chave LTPA exportar 368 gerar 368 importar 368 servidor de aplicação da Web incorporado 368 WebSphere Application Server 371

ficheiro es.cfg 26, 27, 35, 37, 372, 374 ficheiro followindex.rules configurar 154 descrição 153 ficheiro global.rules 151 ficheiro mimetypes.xml 265, 268, 276 ficheiro nodes.ini 35 ficheiro parser_config.xml 211, 265, 268 ficheiro parserTypes.cfg 269 ficheiro stellent.properties 269 ficheiro stellenttypes.cfg 269 ficheiro stellentTypes.cfg 265 ficheiros ccl.properties 35 ficheiros CSV importar para colecções 497 Importar para colecções 497 importar para um dicionário de utilizador 259 ficheiros de arquivo formatos suportados 164 formatos URI 164 pesquisa 164 ficheiros de correlação da base de dados relacional 485 ficheiros de correlação para documentos exportados 485 ficheiros de esquema para vistas-resumo 351 ficheiros de registo caracteres DBCS 453 comando esviewlog 453 configuração de servidor SMTP 449 descrição 21, 445 filtrar 451 localização predefinida 445 níveis de gravidade 448 opções de correio electrónico 450 rotação 448 supervisionar 451 tamanho máximo 448 visualizar 451 ficheiros de regras configurar 260 ficheiros pat 260 ficheiros robots.txt conformidade com o pesquisador da Web 136 identificação do agente de utilizador 135 ficheiros tar formatos URI 164 pesquisa 164 Ficheiros XML pesquisa 209 ficheiros zip formatos URI 164 pesquisa 164 filtrar campos 221, 222 filtros de classificação de documentos 320 filtros de resultados para títulos e URLs 356 filtros de URL de expressão regular 357 firewalls, pesquisa de documentos do Exchange Server 2000 e 2003 96, 404 formato de ficheiros XML campos de índice 205

formato de ficheiros XML (continuação) documentos exportados 481 inspecções profundas 309 formatos de data detectado por predefinição 223 personalizadas 223 fragmentos XML, procura XML nativa 179 função administrativa do personalizador de aplicações 376 função do administrador da colecção 376 função do administrador de análise de texto 376 função do administrador principal 376 função do categorizador com base em regras 376 função do editor de árvore da faceta 376 função do editor do dicionário do utilizador 376 função do operador 376 função do supervisor 376 funções administrativas configuração 376 funções de acessibilidade para este produto 567

G

gestão de identidade cadeia de consulta de XML 388 configuração 394 contexto de segurança do utilizador 388 desactivar 388 descrição 388 extracção de grupo 388 perfis de utilizadores 391 suporte de início de sessão único 392 grupos de agregação de resultados 320

Η

histórico do tempo de resposta, supervisionar 439 hora epoch 223

IBM Cognos BI exportar para 466 geração de relatórios 464 integração 463, 464, 466 **IBM** Connections configuração do pesquisador do Lista de valores geradores 125 correlacionar campos com campos de índice 128 utilizador do administrador de procura 128 IBM Content Analytics with Enterprise Search alterar a palavra-passe em servidores múltiplos 27 alterar a palavra-passe num único servidor 26

IBM Content Analytics with Enterprise Search (continuação) alterar endereços de IP 35 códigos de retorno 511 comandos 511 consola de administração 20 descrição geral dos componentes 15 diagrama de fluxo de dados 23 IDs de sessão 511 integração com o IBM Cognos BI 463 integração com o Lotus Notes 499 servidores de índice 18 servidores de procura 19 servidores do pesquisador 16 Suporte de IP duplo 37 IBM Web Content Manager configuração do pesquisador do Lista de valores geradores 125 identificação 16 idiomas códigos de dois caracteres 279 detecção automática 280 procurar 279 suportados 279 idiomas de texto complexos 279 idiomas de texto simples 279 idiomas ideográficos 279 IDs da aplicação 382 IDs de colecção 55 IDs de sessão 511 imagens de miniatura redimensionar 232 importância do documento descrição 314 dicionários de palavras hierárquicas 318 expansão da consulta 320 padrões URI 318, 319 importar ficheiros CSV para colecções 497 ficheiros CSV para um dicionário de utilizador 259 índice configuração de módulo 178 descrição 177 incremental 177 partições 177 processos globais 227 supervisionar 434 índice de registo da consulta 301 índices alertas sobre 446 âmbitos 297 caracteres globais 296 descrição 18 Formatos URI 164 remover URIs 225 texto âncora 396 URIs contraídos 299, 300 iniciar aplicações de procura empresarial 324 componentes do sistema 420 Data Listener 442 personalizadores de aplicações 347 prospecção de análise de conteúdo 324

iniciar (continuação) prospecção de análise de conteúdo com condições de arranque 324 servidores 417 servidores de pesquisador 423 servidores de procura 438 iniciar URLs para pesquisadores da Web 139, 144 inspecção profunda descrição 307 exportar resultados 308 inspecções profundas formato XML 309 integração WebSphere Portal 501 integração com o Lotus Notes 499 intervalo máximo para nova pesquisa 144 intervalo mínimo para nova pesquisa 144 intervalos da nova pesquisa de pesquisadores da Web 144 IOCP, configuração de pesquisadores 118

J

Japonês caracteres SBCS 283 expansão de sinónimos 283 java.util.regex.Pattern class 357

L

ligações rápidas criar 295 descrição 294 Formatos URI 164 procurar 294 locales analisar 279 procurar 279 Lotus Notes barra de procura 499 instalação de suplementos 499 integração 499 sítio de actualizações de suplementos 499 Lotus Quickr for Domino configuração do pesquisador 122

Μ

memória cache de documentos activar para uma colecção 51 exportar 492 exportar documentos analisados 472 exportar documentos pesquisados 472, 476 ficheiros CSV 492 mensagens de erro configuração de servidor SMTP 449 receber correio electrónico para 448, 450 visualizar ficheiros de registo 451 metadados do extractor de texto correlacionamento com campos de índice 212, 213 correlacionar para conteúdo do documento 213 nomes de campos de metadados predefinidos 212 miniatura activar para uma colecção 51 modo de análise para aplicações de procura empresarial 347 módulo bos.iocp.rte 118 módulos pesquisador da Web 427 serviço de índice 178 Módulos da porta de conclusão E/S, configuração de pesquisadores 118 módulos de pesquisador, configuração do pesquisador do sistema de ficheiros 158 motores de análise de texto adicionar ao sistema 216 associar a coleccões 217 correlacionar a estrutura de análise comum com bases de dados relacionais 220 correlacionar elementos de XML 218 correlacionar resultados da análise com bases de dados relacionais 220 correlacionar resultados da análise com o índice 219 descrição 214

Ν

nível de isolamento para DB2 81 NNTP pesquisadores, configurar 109 notificações de correio electrónico configuração de servidor SMTP 449 para alertas 450 novidades desta edição 1 números de portas alterar a porta da aplicação 34 alterar a porta do Centro de Informações 35 alterar a porta do sistema 32 predefinição 31

0

opções de procura campos de data 223

Ρ

páginas de códigos detecção automática 281 suportadas 281 páginas de erros esporádicos, pesquisadores da Web 145 palavra-passe, administrador 26, 27 palavra-passe de administrador alterar em múltiplos servidores 27 alterar num único servidor 26 palavras-passe codificadas 564

295 parar componentes do sistema 420 servidores 417, 419 servidores de procura 421, 438 servidores do pesquisador 423 partições de índice 177 perfis de utilizadores configuração 394 descrição 391 personalizador de análise ficheiros de esquema 351 filtros de resultados para títulos e URLs 356 iniciar 347 vista-resumo 351 personalizador de procura filtros de resultados para títulos e URLs 356 imagens do cabeçalho de apresentação 355 iniciar 347 personalizar aplicações 346 pesquisa de texto livre, descrição 195 pesquisador do Notes número de documentos pesquisados 425 supervisionar 425 pesquisadores activar segurança ao nível do documento 57 agendamento 57, 60 Agente para sistemas de ficheiros Windows 61, 63 aplicações de Data Listener 160 associar tipos de pesquisadores 57 base de dados de JDBC 100 Base de dados JDBC 98, 102, 106 Case Manager 64 Content Integrator 65, 67 Content Manager 69 criar 59 DB2 74 descrição 16 descrição geral da configuração 57 Domino Document Manager 82 edição 59 eliminar 60 estado do sistema 423 Exchange Server 85 Exchange Server 2000 e 2003 95, 96, 404 ficheiros de arquivo 164 FileNet P8 96 Formatos URI 164 Lista de valores geradores 125 NNTP 109 nomes de campos reservados 199 Notes 121 Quickr for Domino 122 script de configuração de Case Manager 65 script de configuração de FileNet P8 65 segurança ao nível do documento 383

Palavras-chave em ligações rápidas 294,

pesquisadores (continuação) SharePoint 129 Sistema de ficheiros UNIX 133 Sistema de ficheiros Windows 157 supervisionar 423 suplementos 161 suporte para externos 160 valor inicial para 59 valores base para 57 Web 134 Web Content Management 154, 156 WebSphere Portal 155, 156 Pesquisadores da lista de valores geradores integração com o WebSphere Portal 501 pesquisadores da Web códigos de estado de HTTP 430 configuração 134 configuração de cookies 151 detalhes do módulo 426 detalhes do sítio 426 directivas para não indexar 154 directivas para não seguir 154 estado de URL 426 ficheiro followindex.rules 154 sítios activos 428 velocidade da pesquisa 428 Pesquisadores da Web agentes do utilizador 135 comprimento máximo da ligação 139 comprimento máximo do directório 139 controlar redirecção de HTTP 549 cookies 150 criar relatórios sobre 429 detalhes do módulo 427 directivas para não indexar 153 directivas para não seguir 153 espaço global do pesquisador 151 estado do sistema 426 ficheiro followindex.rules 153 ficheiros robots.txt 135, 136 formato de cookie 150 histórico de pesquisadores 426 intervalos da nova pesquisa 144 limitar o espaço de pesquisa 139 páginas de erros esporádicos 145 pesquisar regras 139 servidores proxy 149 sítios activos 426 sítios da Web protegidos por palavra-passe 146, 147, 148 supervisionar 426 suporte JavaScript 138 URLs iniciais 139, 144 URLs pesquisados recentemente 426 visitar URLs logo que possível 144 Pesquisadores de Base de dados JDBC Colunas BLOB 98 colunas com um grande binário ou conteúdo de texto 98 configuração 98 controladores suportados 98 formatos URI 164 pesquisar tabelas múltiplas 102, 106

Pesquisadores de Base de dados JDBC (continuação) suplemento para pesquisar tabelas múltiplas 102, 106 pesquisadores de bases de dados JDBC pesquisar tabelas múltiplas 100 suplemento para pesquisar tabelas múltiplas 100 pesquisadores de DB2 Configuração do Data Event Publisher 77 configuração do WebSphere MQ 79 Pesquisadores do Agente para sistemas de ficheiros Windows comandos do servidor de agente 63 configuração 61 formatos URI 164 Pesquisadores do Case Manager formatos URI 164 requisitos de configuração 65 pesquisadores do Content Integrator scripts de configuração 67 Pesquisadores do Content Integrator configuração 65 configuração do conector 67 Documentum no Linux no System z 68 formatos URI 164 pesquisadores do Content Manager configurar em AIX 70 configurar em Linux 70 instalação em Windows 72 Pesquisadores do Content Manager configuração 69 Formatos URI 164 Pesquisadores do DB2 Colunas BLOB 74 colunas com um grande binário ou conteúdo de texto 74 configuração 74 configuração da publicação de eventos 76 formatos URI 164 instalação de WebSphere MQ 76 nível de isolamento 81 publicação de eventos 74 WebSphere II Classic Federation 81 Pesquisadores do Domino Document Manager configuração 82 configuração de IOCP 118 configuração do protocolo DIIOP 117 formatos URI 164 instalação em Windows 116 instalação no sistema operativo AIX 114 instalação no sistema operativo Linux 114 protocolo NRPC 114, 116 Pesquisadores do Exchange Server configuração 85 configuração de metadados 92 Configuração do Exchange Web Service 87 configuração do servidor da caixa de correio 88

Pesquisadores do Exchange Server (continuação) Configuração do servidor da caixa de correio 90 filtros de caixas de correio 85 formatos URI 164 pastas pessoais 85 permissões de segurança 85 procura segura 93 Serviço da Web do ESExchangeService 88 serviço web ESCommonServices 90 tipos de itens 85 Pesquisadores do Exchange Server 2000 e 2003 configuração 95 documentos protegidos 96, 404 Pesquisadores do FileNet P8 formatos URI 164 requisitos de configuração 65 Pesquisadores do Notes configuração 110 configuração de IOCP 118 configuração de segurança de nível de documento 405 configuração de servidor LDAP 410 configuração do protocolo DIIOP 117 desempenho da segurança de nível de documento 409 expansão de grupos 410 formatos URI 164 instalação em Windows 116 instalação no sistema operativo AIX 114 instalação no sistema operativo Linux 114 pesquisa da base de dados 110 pesquisa do directório 110 pesquisa do valor da coluna 110 pesquisar anexos 121 pesquisar campos 121 pesquisar novamente documentos 120 protocolo NRPC 114, 116 regras de mapeamento de campos 121 segurança ao nível do documento 110 Servidor Fidedigno do Lotus Domino 405 validação de credenciais actuais 405 Pesquisadores do Quickr for Domino configuração 122 configuração da Assistência de Directório 408 configuração de IOCP 118 configuração de utilizadores Domino 407 configuração do protocolo DIIOP 117 formatos URI 164 instalação em Windows 116 instalação no sistema operativo AIX 114 instalação no sistema operativo Linux 114 protocolo NRPC 114, 116 segurança Utilizador Local 407

Pesquisadores do SharePoint formatos URI 164 implementação de serviço da Web 132 suporte de elemento social 129 suporte de procura protegida 129 Pesquisadores do sistema de ficheiros do Windows configuração de segurança de nível de documento 411 Pesquisadores do Sistema de ficheiros UNIX configuração 133 formatos URI 164 módulos de pesquisador múltiplos 158 tamanho máximo do documento 160 Pesquisadores do Sistema de ficheiros Windows configuração 157 formatos URI 164 módulos de pesquisador múltiplos 158 tamanho máximo do documento 160 Pesquisadores do Web Content Management configuração 154 copiar URLs do sítio 156 formatos URI 164 Pesquisadores do WebSphere Portal configuração 155 copiar URLs do sítio 156 formatos URI 164 pesquisadores externos aplicações de Data Listener 160 configurar 160 Pesquisadores Lista de valores geradores configuração 125 configuração de início de sessão único 371 correlacionar campos do IBM Connections para campo de índice 128 formatos URI 164 pesquisadores Notes 110 planos de decisão, Content Classification 504, 506 pontuações por categoria, Content Classification 504, 506 pós-filtragem procura facetada 385 privilégios de utilizadores de aplicações 378 processamento de texto anotadores 214 dicionários lexicais 189 estruturas de análise comum 214 motores de análise de texto 214 processos globais descrição 227 procura conceptual, Content Classification 504, 506 procura empresarial descrição geral 9 interface do utilizador 13 procura facetada colecções federadas 180

procura por campo dependência de maiúsculas e minúsculas 195 descrição 195 expansão de consulta de texto livre 195 factores hierárquicos 195 ordenação de cadeias 195 sugestões ortográficas 195 procura semântica 179, 214, 218 procura XML nativa 179 procurar documentos de XML 218 ligações rápidas 294 profundidade do Caminho URL 139 propostas de agrupamentos 240 propriedade sameGroupDataSources 360 propriedade searchable.sessional 361 propriedade useWorkManagerInIMC 360 propriedades do pesquisador descrição 16 edicão 59 prospecção de análise de conteúdo aceder 324 alterar o número da porta 34 associar a colecções 381 condições de arranque 324 configuração de expansão de grupo 360 deslocar resultados 361 ficheiro config.properties 326 fornecidas 323 ID de aplicação predefinido 323 iniciar 324 personalizadas 323 personalizar 325, 347 personalizar a vista vista-resumo 351 propriedade searchable.sessional 361 URL endereçável 324 prospecção de conteúdo análise de frequência 11 análises altamente escaláveis 12 análises periódicas 12 correlações 11 descrição geral 11 exportar facetas 12 relatórios de facetas 12 protocolo DIIOP, configuração de pesquisadores 117 protocolo IPv6 AIX, Linux 41 Windows 38 protocolo NRPC configuração do pesquisador 116 variável de ambiente NOTESNTSERVICE 116 protocolo NRPC, configuração de pesquisadores 114 protocolo Robots Exclusion conformidade com o pesquisador da Web 136 identificação do agente de utilizador 135 publicação de eventos configuração do pesquisador de DB2 77, 79

publicação de eventos *(continuação)* definir DB2 pesquisadores 76 descrição 74

Q

Quickr for WebSphere Portal configuração do pesquisador do Lista de valores geradores 125

R

rastreio de documentos 437 recolha de conteúdo empresarial 10 reconstruir o índice de facetas 441 recursos de análise descrição 247 RegisterApplicationID, detecção e correcção de problemas 558 regras de categorias configurar 237 conteúdo do documento 235, 237 padrões URI 235, 237 regras de domínio para pesquisadores da Web 139 regras de expansão de consulta 320 regras de prefixo para pesquisadores da Web 139 Regras do endereço de IP para pesquisadores da Web 139 regras do endereço para pesquisadores da Web 139 reimplementar os recursos de análise 441 relatórios de detalhes do sítio criar 429 descrição 426 relatórios do histórico de pesquisadores criar 429 descrição 426 relatório de códigos de estado de HTTP 429 relatório de sítio 429 remover URIs de um índice 225 rendimento da consulta 286 resolução de problemas analisadores sem resposta 551 clonar a portlet Pesquisa 559 clonar ESSearchPortlet 559 comando esservice 543 directório master_config 552 erros de CCL 545 estado de espera do analisador 551 falhas de sincronização 552 falhas do extractor de texto 553 falhas no analisador 552 ligar a Documentum 550 palavras-passe codificadas 564 rastreio de pilha do extractor de texto extractor 554 Redirecção do pesquisador da Web 549 registo do ID da aplicação RegisterApplicationID 558 reunir ficheiros de registo 543 restauração do sistema 455, 457

restaurar o sistema 455 restaurar scripts descrição 455 executar 457 resultados da procura agrupar 299, 300 caracteres globais 296 classificação 319 classificação através de expansão de consulta 320 classificação baseada em texto 313 classificação dinâmica 313 classificação estática 314 configuração de padrões URI 318 contrair 299, 300, 416 descrição 313 expansão do carácter global 297 filtros de expressão global 356 filtros de resultados para títulos e URLs 356 filtros de URL de expressão regular 357 importância do documento 314 personalizar resumos 288 pós-filtragem 384 pré-filtragem 384 resumo dinâmico 288 resumos 288 resultados da procura contraídos configurar 300 descrição 299 restrições de segurança 416 resumo dinâmico 288 resumos dinâmicos 288 personalizar 288 resumos de documentos personalizar 288 revisitar URLs logo que possível 144

S

script esbackup.bat 456 script esbackup.sh 456 script eschangepw 26, 27, 564 script escrcm.sh 70 script escrcm.vbs 72 script escrdb2.sh 76 script escrdb2.vbs 76 script escrfilenet 65 script escrnote.sh 114 script escrnote.vbs 116 script escrvbr.sh 67 script escrvbr.vbs 67 script esrestore.bat 457 script esrestore.sh 457 script startccl 457 scripts esbackup.bat 456 esbackup.sh 456 escrcm.sh 70 escrcm.vbs 72 escrdb2.sh 76 escrdb2.vbs 76 escrnote.sh 114 escrnote.vbs 116 escrvbr.sh 67

scripts (continuação) escrvbr.vbs 67 esrestore.bat 457 esrestore.sh 457 startccl 457 scripts da cópia de segurança descrição 455 executar 456 segurança activar para uma colecção 363 análise de texto âncora 395 autenticação 375 autenticação de início de sessão 365 codificação 363 cofre de credenciais 380 configuração de expansão de grupo 360 configuração de início de sessão único WebSphere Application Server 371 Configuração de início de sessão único servidor de aplicação da Web incorporado 368 configuração de servidor distribuído 374 configuração de um único servidor 372 configuração do servidor de aplicação da Web incorporado 365 controlos de acesso 375 credenciais de ligação 380 descrição 363 Desempenho do Lotus Notes 409, 410 detecção de documentos duplicados 382 Documentos do Lotus Domino 405 domínios do Windows 411 gestão de identidade 388, 394 global, WebSphere Application Server 370 ID de administrador predefinido 376 IDs da aplicação 382 ignorar controlos de acesso de nível de documento 414 início de sessão de aplicação 365 nível de colecção 381, 414 nível de documento 383, 384, 386, 387, 394, 414 nível de sistema 375 perfis de utilizadores 391 pós-filtragem de facetas 385 resultados da procura contraídos 416 servidor do pesquisador indisponível 402 suplementos do pesquisador 161 suporte de início de sessão único 392 segurança ao nível da colecção análise de texto âncora 395 descrição 363 detecção de documentos duplicados 382 IDs da aplicação 382 segurança ao nível do documento configuração do pesquisador 57 controlos de acesso indexados 386

segurança ao nível do documento (continuação) descrição 363, 383 Desempenho do Lotus Notes 409, 410 Documentos do Lotus Domino 405 facetas e pós-filtragem 385 gestão de identidade 388, 394 para documentos do Lotus Domino 405 para sistemas de ficheiros do Windows 411 perfis de utilizadores 391 pós-filtragem de resultados 384 pré-filtragem de resultados 384 restrições de facetas 383 suplementos do pesquisador 161 suporte de início de sessão único 392 tokens de segurança 386 validação de credenciais actuais 387 validação em tempo real 387 segurança de início de sessão único configuração 394 gestão de identidade 392 Segurança Utilizador Local, pesquisadoresQuickr for Domino 407 serviço da Web do ESExchangeService 88 Serviço web ESCommonServices 90 serviços da web para o Exchange Server 88, 90 serviços da Web para SharePoint 132 servidor de agente comandos 63 servidor de aplicação da Web incorporado Configuração de início de sessão único 368 servidores adicionar ao sistema 29 alterar endereços de IP 35 iniciar 417 iniciar serviços locais 422 parar 417, 419 parar serviços locais 422 Suporte de IP duplo 37 servidores de procura cache de procura 287 consultas populares 439 consultas recentes 439 descrição 19, 285 dicionários de palavras de paragem 292 dicionários de palavras hierárquicas 316 dicionários de sinónimos 289, 291 estatísticas de consultas 439 histórico do tempo de resposta 439 iniciar 421, 438 parar 421, 438 rendimento da consulta 286 sugestões ortográficas 286 supervisionar 438 suporte de introdução adiantada 301, 303

servidores de procura (continuação) utilizar dicionários de palavras de paragem 293, 294 utilizar dicionários de palavras hierárquicas 317, 318 utilizar dicionários de sinónimos 291 servidores do pesquisador iniciar 417, 423 parar 423 Servidores Fidedignos do Lotus Domino 405 servidores proxy 149 servidores proxy HTTP 149 sinalizadores configurar 303 sintaxe da consulta XML, nativa 179 sistema operativo AIX configuração da publicação de eventos 76 configuração do pesquisador do Content Integrator 67 configuração do pesquisador do Content Manager 70 Sistema Operativo AIX configuração do pesquisador do Domino Document Manager 114 configuração do pesquisador do Notes 114 configuração do pesquisador do Quickr for Domino 114 sistema operativo Linux configuração da publicação de eventos 76 configuração do pesquisador do Content Integrator 67 configuração do pesquisador do Content Manager 70 configuração do pesquisador do Domino Document Manager 114 configuração do pesquisador do Notes 114 configuração do pesquisador do Quickr for Domino 114 pesquisar origens do Documentum 68 sistema operativo Windows configuração da publicação de eventos 76 configuração do pesquisador 116 configuração do pesquisador do Content Integrator 67 configuração do pesquisador do Content Manager 72 sítios da Web activos, supervisionar 426, 428 sítios da Web protegidos por palavra-passe 146 Autenticação base de HTTP 147 autenticação baseada em formulário 148 subfacetas para exploração de conteúdo 251 sugestões ortográficas 286 supervisionar analisadores 434 códigos de estado de HTTP 430 colecções 423

supervisionar (continuação) configuração do servidor 423 consultas populares 439 consultas recentes 439 Data Listener 442 descrição 20 Detalhes do módulo do pesquisador da Web 427 detalhes do pesquisador do Notes 425 detalhes do URI 436 documentos abandonados 437 estatísticas de consultas 439 ficheiros de registo com caracteres DBCS 453 histórico do tempo de resposta 439 pesquisadores 423 Pesquisadores da Web 426 serviços de indexação 434 servidores de procura 438 sítios activos do pesquisador da Web 428 topologia do sistema 423 velocidade da pesquisa do pesquisador da Web 428 suplemento de tabelas estruturadas múltiplas 100, 102, 106 suplemento no Lotus Notes instalação 499 sítio de actualizações 499 suplementos Pesquisadores de Base de dados JDBC 102, 106 pesquisadores de bases de dados JDBC 100 pesquisar tabelas estruturadas múltiplas 100, 102, 106 suplementos do pesquisador 161 suplementos para pesquisadores 161 suporte de introdução adiantada colecções de n-gram 301 configurar 303 descrição 301 índice de registo da consulta 301 procura federada 301 Suporte JavaScript em pesquisadores da Web 138 suporte linguístico análise de texto personalizada 214 códigos de linguagem 279 detecção de idioma 280 detecção de páginas de códigos 281 dicionários de palavras de paragem 292 dicionários de palavras hierárquicas 316 dicionários de sinónimos 289 dicionários lexicais 189 locales 279 procura semântica 179, 214 procura XML nativa 179 segmentação híbrida 282 segmentação n-gram 282

Т

tabelas de esquema em estrela 485 tamanho da pilha JVM 552 tamanho máximo do documento, configuração do pesquisador do sistema de ficheiros 160 Tecnologia Apache PDFBox 275 tempo de resposta da procura alertas sobre 446 supervisionar 439 termos compostos, análise 178 termos de identificação de interesse 262 texto alternativo indexação 211 textos bidireccional em ficheiros PDF 275 tipo de categorização baseadas em regras 235 tipos de documentos analisar 267 detectar 265 para o extractor de texto 269 para serviços do analisador 268 predefinição 266 suportados pelo extractor de texto 272 tipos de documentos desconhecidos 269 tipos de origens de dados Armazéns de objectos do IBM FileNet P8 96 base de dados de Software AG Adabas 74 bases de dados CA-Datacom 74 bases de dados da Oracle 74, 98 bases de dados da Sybase 74 Bases de dados de DB2 74, 98 bases de dados de IMS 74 bases de dados de JDBC 98, 100, 102, 106 bases de dados de VSAM 74 bases de dados do Domino Document Manager 82 bases de dados do Informix 74 bases de dados do Notes 110, 121 bases de dados do SQL Server 74, 98 bases de dados relacionais 74 Caixas de correio do Exchange Server 85 DB2 para bases de dados do iSeries 74 DB2 para z/OS 74 IBM Connections 125 Lotus Quickr for Domino 122 newsgroups de NNTP 109 Pastas públicas do Exchange Server 85, 95 Quickr for Domino bases de dados 122 Quickr for WebSphere Portal 125 Repositórios do Content Integrator 65, 67 sistemas de ficheiros partilhados Windows 63 sistemas de ficheiros UNIX 133 sistemas de ficheiros Windows 61, 157 sítios da Web 134

tipos de origens de dados (continuação) sítios de Web Content Management 154 sítios SharePoint 129 Soluções de IBM Case Manager 64 suportados 24 Tipos de artigos do Content Manager EE 69 WebSphere Portal 125, 155 tipos de pesquisadores associar numa colecção 57 valores base para 57 tipos MIME para serviços do analisador 268 Tipos MIME, incluindo em espaços de pesquisa da Web 139 títulos de anexos nos resultados 121 tokens de segurança configuração do pesquisador 386 desactivar para uma colecção 414 segurança ao nível do documento 386 topologia do sistema 423

U

UIMA adicionar motores de análise de texto ao sistema 216 associar a colecções 217 correlacionar a estrutura de análise comum com bases de dados relacionais 220 correlacionar a estrutura de análise comum com o índice 219 correlacionar elementos de XML 218 correlacionar resultados da análise com bases de dados relacionais 220 correlacionar resultados da análise com o índice 219 descrição 214 estruturas de análise comum 219, 220 URIs âmbitos 297, 298 contraídos nos resultados da procura 299, 300 formatos no índice 164 influenciar classificações estáticas 318, 319 ligações rápidas 294, 295 regras de categorias 235, 237 remover de um índice 225 visualizar detalhes sobre 436 URIs contraídos configurar 300 descrição 299 restrições de segurança 416 URL enderecável, prospecção de análise de conteúdo 324 URLs recentemente pesquisados, supervisionar 426

utilitário esservice 543 utilitário reportcel 545 utilizador do WebSphere Application Server palavra-passe para a configuração de um único servidor 372 palavra-passe para a configuração do servidor distribuído 374

V

validação de consulta 387 validação de credenciais actuais 387, 405, 411 variável de ambiente NOTESNTSERVICE 116 velocidade da pesquisa, supervisionar 428 visitar URLs logo que possível 144 vistas contextuais configurar 263 visualizar Detalhes do URI 436 ficheiros de registo com caracteres DBCS 453

W

WebSphere Application Server configuração de início de sessão único 371 WebSphere II Classic Federation 81 WebSphere MQ, configuração de pesquisadores de DB2 79 WebSphere MQ, configuração do servidor do pesquisador 76 WebSphere Portal Descrição da barra de procura 501 Descrição do Search Center 501 integração 501 Windows exportar para o directório partilhado 479 suporte do protocolo IPv6 38

X

XPath, procura XML nativa 179



Número do Programa: 5724-Z21

SC17-5499-00

