

**Enterprise Search beheren**

**Opmerking**

Lees eerst de algemene informatie onder "Kennisgevingen en handelsmerken" op pagina 389.

**Derde uitgave (november 2006)**

Dit document bevat eigendomsinformatie van IBM. U vindt deze auteursrechtelijk beschermde informatie in de licentieovereenkomst. De informatie in deze publicatie bevat geen productgaranties en de instructies in deze handleiding kunnen niet als zodanig worden geïnterpreteerd.

U kunt IBM-publicaties online of via uw plaatselijke IBM-vertegenwoordiger bestellen:

- Als u publicaties online wilt bestellen, gaat u naar het IBM Publications Center op [www.ibm.com/shop/publications/order](http://www.ibm.com/shop/publications/order).
- Als u een IBM-vertegenwoordiger bij u in de buurt zoekt, gaat u naar de wereldwijde IBM-adressenlijst op [www.ibm.com/planetwide](http://www.ibm.com/planetwide).

IBM mag informatie die door u wordt verstrekt, gebruiken of distribueren op elke manier haar goedgevondt zonder daarbij verplichtingen jegens u aan te gaan.

© Copyright IBM Corp. 2004, 2006.

---

# Inhoudsopgave

## Wat is enterprise search? . . . . . 1

Typen gegevensbronnen die door enterprise search worden ondersteund. . . . .	2
Overzicht van componenten van enterprise search. . . . .	3
Crawlers van enterprise search . . . . .	4
Parsers van enterprise search . . . . .	5
Indexen van enterprise search . . . . .	7
Zoekservers voor enterprise search . . . . .	9
De beheerconsole van enterprise search . . . . .	10
Een enterprise search-systeem bewaken . . . . .	10
Logbestanden van enterprise search . . . . .	11
Enterprise search aanpassen . . . . .	11
Voorbeeld van een zoekprogramma voor enterprise search . . . . .	12
De gegevensstroom in enterprise search . . . . .	13

## Enterprise search - Systeembeheer . . . 17

Aanmelden bij de beheerconsole . . . . .	21
Het beheerderswachtwoord voor enterprise search wijzigen voor een configuratie met één server . . . . .	22
Het beheerderswachtwoord voor enterprise search wijzigen voor een configuratie met meerdere servers . . . . .	23
TCP-poortnummers voor enterprise search . . . . .	26
Het poortnummer voor het enterprise search-systeem wijzigen . . . . .	27

## Collecties van enterprise search. . . . . 29

Een collectie maken met de Collectiewizard . . . . .	29
Een collectie maken met behulp van de view Collecties . . . . .	31
Een collectie bewerken . . . . .	32
Een collectie wissen . . . . .	34

## Beheer van enterprise search-crawler 35

Een crawler maken . . . . .	37
Crawlereigenschappen bewerken . . . . .	38
Een crawlruimte bewerken . . . . .	38
Een crawler wissen . . . . .	39
Crawlerroosters . . . . .	40
Content Edition-crawlers . . . . .	40
Directe werkstand-toegang tot WebSphere II Content Edition-repository's . . . . .	42
Serverwerkstand-toegang tot WebSphere II Content Edition-repository's . . . . .	43
DB2-crawlers . . . . .	45
De crawlerserver configureren op UNIX voor DB2-crawlers . . . . .	47
De crawlerserver configureren op Windows voor DB2-crawlers . . . . .	48
WebSphere Information Integrator Event Publisher Edition configureren voor DB2-crawlers . . . . .	49
WebSphere MQ configureren voor DB2-crawlers . . . . .	52
DB2-databases op een klassieke gegevensbronserver crawlen . . . . .	53
DB2 Content Manager-crawlers. . . . .	54

De crawlerserver configureren op UNIX for DB2 Content Manager-crawlers . . . . .	55
De crawlerserver configureren op Windows for DB2 Content Manager-crawlers. . . . .	57
Domino Document Manager-crawlers . . . . .	58
Exchange Server-crawlers. . . . .	61
JDBC-databasecrawlers . . . . .	62
NNTP-crawlers . . . . .	64
Notes-crawlers . . . . .	65
Tips voor het crawlen van Lotus Domino-databases . . . . .	67
De crawlerserver configureren op UNIX voor het crawlen van Lotus Domino-bronnen . . . . .	68
De crawlerserver configureren op Windows voor het crawlen van Lotus Domino-bronnen. . . . .	71
Servers configureren die het DIIOP-protocol gebruiken . . . . .	72
De I/O Completion Port op AIX configureren voor het crawlen van Lotus Domino-bronnen . . . . .	73
QuickPlace-crawlers . . . . .	75
UNIX-bestandssysteemcrawlers. . . . .	79
Web-crawlers . . . . .	80
Configuratie van de gebruikersagent . . . . .	80
Gebruik van het Robots Exclusion Protocol in de webcrawler . . . . .	81
Ondersteuning voor JavaScript . . . . .	84
Regels voor het beperken van de webcrawlruimte . . . . .	84
URL-verbindingen met de webcrawler testen . . . . .	88
Intervalinstellingen voor opnieuw crawlen in de webcrawler . . . . .	88
Opties voor het bezoeken van URL's met de webcrawler . . . . .	89
De manier waarop de webcrawler omgaat met soft error-pagina's . . . . .	90
Ondersteuning voor het crawlen van beveiligde websites . . . . .	91
Websites die worden bediend door proxyservers . . . . .	93
Beheer van cookies . . . . .	94
Configuratie van algemene webcrawlruimte . . . . .	96
De instructies no-follow en no-index . . . . .	98
Instructies van het type no-follow en no-index in webpagina's . . . . .	98
Configureren welke datum de webcrawler gebruikt voor gecrawlde documenten . . . . .	99
Web Content Management-crawlers . . . . .	100
WebSphere Portal-crawlers . . . . .	102
De URL kopiëren om te crawlen vanaf WebSphere Portal . . . . .	104
Windows-bestandssysteemcrawler . . . . .	105
Ondersteuning voor Data Listener-programma's configureren. . . . .	107
Aangepaste crawler plug-ins . . . . .	108
Ondersteuning voor het crawlen van gearchiveerde bestanden . . . . .	110
URI-indelingen in een enterprise search-index . . . . .	111

## Beheer van enterprise search-parser 123

Werken met categorieën . . . . .	124
Regelgebaseerde categorieën . . . . .	124
Categoriestructuren . . . . .	126
Een type categorisering selecteren . . . . .	127
Categorieën configureren . . . . .	128
Werken met XML-zoekvelden . . . . .	129
XML-zoekvelden . . . . .	130
XML-elementen aan zoekvelden toewijzen . . . . .	130
Werken met HTML-zoekvelden . . . . .	133
HTML-zoekvelden . . . . .	133
HTML-metagegevens-elementen toewijzen aan zoekvelden . . . . .	134
Aangepaste tekstverwerking . . . . .	136
Tekstanalyseprogramma's aan het systeem toevoegen . . . . .	137
Een tekstanalyse aan een collectie koppelen . . . . .	138
XML-elementen toewijzen aan de Common Analysis Structure . . . . .	139
De Common Analysis Structure toewijzen aan de index . . . . .	140
De Common Analysis Structure toewijzen aan een relationele database . . . . .	141
Threads voor de parserservice configureren . . . . .	142
Geavanceerde analyse voor samengestelde termen inschakelen . . . . .	143
Ondersteuning voor native XML-zoekopdrachten inschakelen . . . . .	144

## Detectie van documentindelingen voor enterprise search . . . . . 145

Documenttypen die standaard worden ondersteund . . . . .	146
Documenttypen die zijn gekoppeld aan collectie- en Stellant-parsers . . . . .	146
Documenttypen koppelen aan een collectieparser . . . . .	147
Standaardserviceregels voor collectieparsers . . . . .	149
Documenttypen koppelen aan een Stellant-parser . . . . .	149

## Taal- en codetabelondersteuning voor enterprise search. . . . . 153

Automatische taaldetectie . . . . .	154
Automatische codetabeldetectie . . . . .	155
Taalkundige analyse van Chinese, Japanse en Koreaanse documenten . . . . .	156
N-gram-segmentering . . . . .	156
Witruimten uit tekst verwijderen . . . . .	156

## Beheer van enterprise search-index 159

Rooster opstellen voor het opbouwen van de index	160
Het rooster voor indexering wijzigen . . . . .	161
Indexroosters in- en uitschakelen . . . . .	161
Indexen gelijktijdig opbouwen . . . . .	162
Opties die van invloed zijn op de doorzoekbare indexweergave . . . . .	163
Geïndexeerde opties voor het doorzoeken van documenten . . . . .	164
Jokertekens in query's . . . . .	168

Scopes . . . . .	172
Scopes configureren . . . . .	173
Samengevouwen URI's . . . . .	173
URI's samenvouwen in de zoekresultaten . . . . .	175
URI's uit de index verwijderen . . . . .	175

## Beheer van zoekserver voor enterprise search . . . . . 177

Zoekcache . . . . .	178
Een zoekcache configureren . . . . .	178
Aangepaste synoniemenwoordenboeken . . . . .	179
Synoniemenwoordenboeken aan het systeem toevoegen . . . . .	180
Een synoniemenwoordenboek aan een collectie koppelen . . . . .	181
Aangepaste stopwoordenboeken . . . . .	181
Stopwoordenboeken aan het systeem toevoegen . . . . .	182
Een stopwoordenboek aan een collectie koppelen . . . . .	183
Dynamisch samenvatten . . . . .	184
Documentsamenvattingen aanpassen in de beheerconsole . . . . .	184
Documentsamenvattingen aanpassen door de eigenschappen te bewerken . . . . .	185
Werken met snellinks . . . . .	185
Snellinks . . . . .	186
Snellinks configureren . . . . .	186

## Ranking van documenten in enterprise search . . . . . 189

Op tekst gebaseerde score . . . . .	189
Statische ranking . . . . .	190
Standaardwaarden voor statische documentranking herstellen . . . . .	190
Aangepaste gewogen woordenboeken . . . . .	192
Gewogen woordenboeken toevoegen aan het systeem . . . . .	193
Een gewogen woordenboek koppelen aan een collectie . . . . .	194
Documentranking op basis van URI-patronen . . . . .	194
De scores beïnvloeden van documenten die aan URI-patronen voldoen . . . . .	195
Documentranking op basis van wegingsklassen	196
Velden toewijzen aan wegingsklassen . . . . .	198
Wegingsfactor voor wegingsklassen configureren . . . . .	199
Standaardwaarden voor wegingsklassen . . . . .	200

## Zoekprogramma's voor enterprise search . . . . . 203

Zoekprogramma's aan collecties koppelen . . . . .	204
Functies van het voorbeeldzoekprogramma . . . . .	204
Eigenschappen van zoekprogramma . . . . .	206
Eigenschappen van het voorbeeldzoekprogramma wijzigen . . . . .	219
Zoekprogramma's aanpassen . . . . .	220
Het voorbeeldzoekprogramma klonen . . . . .	221
Toegang krijgen tot zoekprogramma's . . . . .	223
De zoekservers configureren om uitsluitend veilige (SSL) opdrachten te accepteren . . . . .	224

<b>Externe bronnen voor enterprise search</b>	<b>227</b>
Externe bronnen aan het systeem toevoegen	227
Zoekprogramma's aan externe bronnen koppelen	230
<b>Beveiliging van enterprise search</b>	<b>231</b>
Beveiliging tijdens de installatie	232
Verificatie versus toegangsbesturing	233
Beheerdersrollen	234
Beheerders configureren	235
Beveiliging op collectieniveau	236
Analyse van duplicaatdocumenten	236
Zoekprogramma-ID's	236
Beveiliging op documentniveau	237
Zoekresultaten vooraf en achteraf filteren	238
Validatie door middel van beveiligingstokens	239
Validatie van actuele legitimatiegegevens tijdens queryverwerking	241
Analyse van ankertekst	248
Beveiliging inschakelen voor enterprise search	250
Algemene beveiliging en LDAP-gebruikersregister configureren in WebSphere Application Server	250
Beveiliging inschakelen voor een enterprise search-systeem met één server	252
Beveiliging inschakelen voor een enterprise search-systeem met meerdere servers	253
Vereisten in de crawlerinstellingen voor de ondersteuning van beveiliging	254
Toegang beheren tot beveiligde Exchange Server-documenten	258
Beveiliging op documentniveau afdwingen voor Lotus Domino-documenten	259
Beveiliging op documentniveau afdwingen voor documenten van het Windows-bestandssysteem	263
Beveiliging uitschakelen voor enterprise search	265
Beveiliging uitschakelen voor een bedrijfs-toepassing in WebSphere Application Server	265
Beveiliging op documentniveau uitschakelen	267
Beveiliging uitschakelen voor samengevoegde zoekresultaten	268
<b>Integratie van enterprise search met WebSphere Portal</b>	<b>271</b>
Installatiescripts voor integratie tussen enterprise search en WebSphere Portal	272
Enterprise search installeren in WebSphere Portal versie 5.1	273
De zoekbalk van WebSphere Portal versie 5.1 configureren voor gebruik van enterprise search	276
Enterprise search verwijderen uit WebSphere Portal versie 5.1	278
Enterprise search installeren in WebSphere Portal versie 6	279
Het Zoekcentrum van WebSphere Portal versie 6 configureren voor enterprise search	282
De zoekbalk van WebSphere Portal versie 6 configureren voor gebruik van enterprise search	283
Enterprise search verwijderen uit WebSphere Portal versie 6	285

<b>Migratie vanuit WebSphere Portal naar enterprise search</b>	<b>287</b>
Een collectie migreren vanuit WebSphere Portal	287
Instellingen van gemigreerde collecties	289
Logbestand van de Migratiewizard	290
<b>Een enterprise search-systeem starten en stoppen</b>	<b>291</b>
Een enterprise search-systeem starten	291
Een enterprise search-systeem stoppen	294
Besturen welke componenten worden gestart of gestopt	295
Zoekservers in de stand-alone werkstand beheren	296
<b>Activiteit van enterprise search beheeren</b>	<b>299</b>
Schatting maken van het aantal documenten in een collectie	299
Een collectie bewaken	300
Gegevens over een URI bekijken	301
Crawlers bewaken	302
Gegevens over webcrawleractiviteit bekijken	304
Threadgegevens webcrawler	305
Actieve sites van de webcrawler	305
Crawlsnelheid webcrawler	306
Webcrawlerrapporten maken	307
HTTP-statuscodes die naar de webcrawler worden verzonden	308
De parser bewaken	312
De indexactiviteit voor een collectie bewaken	313
De indexeerwachtrij van enterprise search bewaken	314
De zoekservers bewaken	315
De Data Listener bewaken	317
Documenttracering	317
Logbestanden configureren voor documenttracering	318
Rapporten over gewiste documenten bekijken	318
Logbestanden over gewiste documenten bekijken	320
<b>Logbestanden en alerts van enterprise search</b>	<b>323</b>
Alerts	323
Alertsignalen op collectieniveau configureren	324
Alerts op systeemniveau configureren	325
Logbestanden configureren	326
SMTP-servergegevens configureren	327
E-mail over gelogde berichten ontvangen	328
Logbestanden bekijken	330
<b>Backup en herstel van een enterprise search-systeem</b>	<b>333</b>
Een backup van het enterprise search-systeem maken	334
Het enterprise search-systeem herstellen	335
<b>Opdrachten, retourcodes en sessie-ID's in enterprise search</b>	<b>337</b>

<b>Hoofdlettergevoeligheid in enterprise search . . . . .</b>	<b>367</b>
<b>Documentatie bij enterprise search</b>	<b>369</b>
<b>Toegankelijkheidsfuncties in WebSphere Information Integrator OmniFind Edition. . . . .</b>	<b>371</b>
<b>Woordenlijst met termen voor enterprise search . . . . .</b>	<b>373</b>

<b>Toegang krijgen tot informatie over Content Management en Discovery . .</b>	<b>387</b>
Commentaar op de documentatie . . . . .	387
Contact opnemen met IBM. . . . .	388
<b>Kennisgevingen en handelsmerken</b>	<b>389</b>
Kennisgevingen . . . . .	389
Handelsmerken . . . . .	391
<b>Trefwoordenregister . . . . .</b>	<b>393</b>



---

## Wat is enterprise search?

Een enterprise search-systeem biedt uitgebreide mogelijkheden voor het via een enkele query doorzoeken van een willekeurig aantal gestructureerde en niet-gestructureerde gegevensbronnen. Snelle responstijden voor query's en een volledige resultaatset waarin de resultaten zijn gerangschikt op belang, stellen u niet alleen in staat belangwekkende documenten op te sporen, maar ook om daar de juiste informatie in te vinden.

IBM OmniFind Enterprise Edition is de nieuwe naam voor IBM WebSphere Information Integrator OmniFind Edition, een naam die u nog steeds in het product en de documentatie zult tegenkomen terwijl we deze naamswijziging doorvoeren.

De verschillende componenten van enterprise search, die met OmniFind Enterprise Edition worden geïnstalleerd, verzamelen informatie die in uw hele bedrijf verspreid is. Door in een webbrowser een query op te geven, kunt u lokale en niet-lokale databases, samenwerkingssystemen, contentbeheersystemen, bestandsystemen en interne en externe websites doorzoeken, allemaal tegelijkertijd.

Een enterprise search-systeem is ontworpen om naadloos met uw bestaande systemen te worden geïntegreerd en neemt alle logistiek voor zijn rekening die nodig is om gegevens van uiteenlopende bronnen te verzamelen en die gegevens zodanig te indexeren dat ze snel kunnen worden opgehaald. Door de gegevens te onderwerpen aan taalkundige en andere analyses is enterprise search in staat om alleszins relevante resultaten te produceren. Om in de verschillende typen repository's te zoeken, is het niet nodig om te leren omgaan met verschillende interfaces.

Het is ook mogelijk om ondersteuning toe te voegen voor het doorzoeken van gegevensbronnen waarvan u liever niet hebt dat ze worden opgenomen in een bedrijfszoekindex. Met de mogelijkheden die enterprise search biedt voor federatief zoeken kunt u deze externe bronnen doorzoeken op hetzelfde moment dat u geïndexeerde gegevensbronnen doorzoekt.

### Zoek kwaliteit

Om te garanderen dat de gebruikers de informatie die ze zoeken ook daadwerkelijk vinden, biedt OmniFind Enterprise Edition ondersteuning aan de IBM Unstructured Information Management Architecture (UIMA). UIMA is een open raamwerk dat een algemene, standaardinterface voor tekstanalyse beschrijft. Dankzij uitgebreide semantische analyse is enterprise search in staat begrippen, latente betekenissen, relaties, feiten en andere relevante gegevens te vinden, gegevens die in een niet-gestructureerde tekst vaak verscholen blijven. De informatie die tijdens de analyse wordt verzameld, kan worden gebruikt om de kwaliteit van de zoekresultaten te verhogen en om andere programma's, zoals die voor business intelligence en data mining, op een kwalitatief hoger plan te brengen.

### Beveiliging

Beveiliging is een integraal onderdeel van enterprise search. Alleen gebruikers die gemachtigd zijn om het systeem te beheren, kunnen dat doen. Met de beveiligingsmechanismen die beschikbaar zijn in IBM WebSphere Application Server kunt u beheerdersrollen configureren en bepalen welke gebruikers toegang hebben tot de verschillende beheerfuncties.

Daarnaast kunt u opties gebruiken om tijdens het verzamelen van gegevens beveiligingstokens aan gegevens te koppelen. Als uw zoekprogramma's de beveiliging inschakelen, kunt u die tokens (die in de index bij de documenten worden bewaard) gebruiken om toegangsbesturing af te dwingen en om te garanderen dat alleen gebruikers met de juiste legitimatiegegevens in staat zijn query's op de gegevens uit te voeren en de resultaten te zien.

Voor bepaalde typen gegevensbronnen kunt u opties configureren om de aanmeldingsgegevens van een gebruiker tijdens de verwerking van query's te vergelijken met de actuele toegangsgegevens. Deze extra beveiligingslaag garandeert dat de bevoegdheden van een gebruiker in real-time worden gecontroleerd bij de native gegevensbron. Op die manier kunt u voorkomen dat er problemen ontstaan doordat de legitimatiegegevens van een gebruiker veranderen nadat een document en de bijbehorende beveiligingstokens zijn geïndexeerd.

#### **Verwante onderwerpen**

Beveiliging van enterprise search

“Overzicht van componenten van enterprise search” op pagina 3

De verschillende componenten van enterprise search verzamelen gegevens in uw hele bedrijf, analyseren, ontleden en categoriseren de informatie en stellen een index samen die door de gebruikers kan worden doorzocht.

“De gegevensstroom in enterprise search” op pagina 13

De componenten van enterprise search die u met WebSphere Information Integrator OmniFind Edition installeert, zijn nauw met elkaar vervlochten om te garanderen dat de gegevens ongehinderd door het systeem kunnen stromen.

Integratie van aangepaste tekstanalyse

Basisconcepten gebruikt bij verwerking van tekstanalyse

#### **Verwante verwijzing**

“Typen gegevensbronnen die door enterprise search worden ondersteund”

Er is vooraf gedefinieerde ondersteuning voor een groot aantal typen gegevensbronnen beschikbaar.

---

## **Typen gegevensbronnen die door enterprise search worden ondersteund**

Er is vooraf gedefinieerde ondersteuning voor een groot aantal typen gegevensbronnen beschikbaar.

Nadat u IBM OmniFind Enterprise Edition hebt geïnstalleerd, kunt u beginnen met het verzamelen van gegevens van de volgende typen gegevensbronnen:

- Itemtypen (documenten, resources en items) van IBM DB2 Content Manager
- IBM DB2 Universal Database-databases (DB2 UDB)
- Databases van IBM Domino Document Manager (voorheen Domino.Doc)
- Databases van IBM Lotus Notes
- Databases van IBM Lotus QuickPlace
- IBM WebSphere Information Integrator Content Edition-repository's: Documentum, FileNet Panagon Content Services, FileNet P8 Content Manager, Hummingbird Document Management (DM), Microsoft SharePoint, OpenText Livelink en Portal Document Manager (PDM) itemklassen
- Roepnaamtabellen van IBM WebSphere Information Integrator voor diverse databasesystemen, waaronder IBM DB2 UDB for z/OS, IBM Informix, Microsoft SQL Server, Oracle en Sybase

- IBM WebSphere Portal-sites
- IBM Workplace Web Content Management-sites
- Openbare mappen van Microsoft Exchange Server
- Microsoft SQL Server-databases
- Bestandssystemen van Microsoft Windows
- NNTP-nieuwsgroepen (Network News Transfer Protocol)
- Oracle-databases
- Bestandssystemen van UNIX
- Websites op internet en in uw intranet

U kunt ook ondersteuning toevoegen voor het doorzoeken van de volgende typen externe bronnen zonder documenten van deze bronnen toe te voegen aan de index van enterprise search:

- Databases die het protocol Java database connectivity (JDBC) ondersteunen (alleen DB2 UDB- en Oracle-databasesystemen). Voor elke tabel die u geschikt maakt om te worden doorzocht, wordt er een afzonderlijk externe bron gecreëerd.
- LDAP-servers (Lightweight Directory Access Protocol). Voor elke LDAP-server wordt er één externe bron gemaakt.

De meest recente informatie over ondersteunde gegevensbronnen en ondersteunde productversies vindt u op de pagina met systeemvereisten op de website OmniFind Enterprise Edition Support.

#### **Verwante onderwerpen**

Externe bronnen voor enterprise search

“De gegevensstroom in enterprise search” op pagina 13

De componenten van enterprise search die u met WebSphere Information Integrator OmniFind Edition installeert, zijn nauw met elkaar vervlochten om te garanderen dat de gegevens ongehinderd door het systeem kunnen stromen.

---

## **Overzicht van componenten van enterprise search**

De verschillende componenten van enterprise search verzamelen gegevens in uw hele bedrijf, analyseren, ontleden en categoriseren de informatie en stellen een index samen die door de gebruikers kan worden doorzocht.

Een enterprise search-*collectie* vertegenwoordigt die groep bronnen die door de gebruikers met een enkele query kan worden doorzocht. Als u een collectie maakt, geeft u op welke bronnen u wilt opnemen en configureert u opties voor de manier waarop de gebruikers de geïndexeerde gegevens kunnen doorzoeken.

U kunt meerdere collecties maken en elk van die collecties kan gegevens uit tal van gegevensbronnen bevatten. Zo kunt u een collectie maken die documenten uit IBM DB2 Universal Database-, IBM Lotus Notes- en IBM DB2 Content Manager-databases bevat. Als gebruikers deze collectie doorzoeken, kan het zoekresultaat in principe documenten uit elk van die gegevensbronnen bevatten.

Dankzij ondersteuning voor federatieve zoekopdrachten kunnen gebruikers met één enkele query meerdere collecties doorzoeken. De zoekresultaten kunnen dan documenten uit elk van de collecties en alle externe bronnen binnen uw enterprise search-systeem bevatten.

Bij het maken en beheren van een collectie komen de volgende activiteiten kijken:

### **Gegevens verzamelen**

De *crawler* componenten verzamelen documenten uit gegevensbronnen, hetzij doorlopend, hetzij overeenkomstig een schema dat u opgeeft. Door regelmatig te crawlen, garandeert u dat gebruikers altijd toegang hebben tot de meest recente informatie.

### **Gegevens analyseren**

De *parser* componenten extraheren tekst uit alle door de crawler opgehaalde documenten en voeren daar taalkundige en andere analyses op uit. Deze gedetailleerde analyse van de content verhoogt de kwaliteit van de zoekresultaten.

### **Gegevens indexeren**

De *index* componenten worden op gezette tijden uitgevoerd en voegen aan de index informatie toe over nieuwe en gewijzigde documenten. De indexcomponenten voeren tevens een algemene analyse uit van de documenten in een collectie; ook dit is bedoeld om de kwaliteit van de zoekresultaten te verbeteren.

### **Gegevens doorzoeken**

De *zoek* componenten doorzoeken de index en werken bij het verwerken van zoekopdrachten en het afbeelden van de resultaten samen met uw zoekprogramma's.

Andere componenten van WebSphere Information Integrator OmniFind Edition stellen u in staat om beveiligingsvoorkeuren op te geven, de systemactiviteiten te bewaken en optredende problemen op te lossen. Daarnaast bevat het product een werkend voorbeeld van een zoekprogramma. Dit kunt u gebruiken als een sjabloon voor het maken van uw eigen zoekprogramma's.

#### **Verwante onderwerpen**

"De gegevensstroom in enterprise search" op pagina 13

De componenten van enterprise search die u met WebSphere Information Integrator OmniFind Edition installeert, zijn nauw met elkaar vervlochten om te garanderen dat de gegevens ongehinderd door het systeem kunnen stromen.

Beheer van enterprise search-crawler

Beheer van enterprise search-parser

Beheer van de enterprise search-index

Beheer van zoekserver voor enterprise search

## **Crawlers van enterprise search**

Crawlers verzamelen documenten van gegevensbronnen, zodat die documenten kunnen worden geanalyseerd, geïndexeerd en doorzocht.

De crawlercomponent die met WebSphere Information Integrator OmniFind Edition wordt meegeleverd, heeft de volgende functies:

- Als u een crawler configureert, gaan de *ontdekkingsprocessen* op zoek naar informatie over de bronnen die beschikbaar zijn om te worden gecrawld, zoals de namen van alle views en mappen in een Lotus Notes-database of de namen van alle bestandssystemen op een UNIX-server.
- Nadat u de bronnen die u wilt crawlen hebt geselecteerd en de crawler hebt gestart, beginnen de crawlercomponenten met het verzamelen van gegevens op de bronnen, zodat die gegevens kunnen worden geanalyseerd en geïndexeerd.

Een enkele collectie kan meerdere crawlers hebben en elke crawler is bedoeld om gegevens te verzamelen op een bepaald type gegevensbron. U zou bijvoorbeeld

drie crawlers kunnen maken om in één en dezelfde collectie gegevens samen te brengen van bestandssystemen, Notes-databases en relationele databases. Of u zou meerdere crawlers van hetzelfde type kunnen maken en voor elk van die crawlers een verschillend rooster kunnen instellen, aangepast aan de frequentie waarmee de gegevens die door elk van die crawlers worden verzameld, worden gewijzigd.

Nadat u de webcrawler hebt gestart, blijft deze continu doorlopen. U geeft welke URL's er moeten worden gecrawld; de crawler gaat dan geregeld na of er wellicht nieuwe gegevens zijn bijgekomen of bestaande gegevens zijn gewijzigd. Andere typen crawlers kunt u handmatig starten en stoppen, of uitvoeren aan de hand van een rooster. In dat rooster geeft u op wanneer de crawler voor het eerst moet worden uitgevoerd en hoe vaak er op de gegevensbronnen moet worden gekeken of er nieuwe of gewijzigde documenten zijn.

*Crawlereigenschappen* zijn regels die bepalen hoe een bepaalde crawler zich tijdens het verzamelen van gegevens dient te gedragen. U kunt bijvoorbeeld regels opgeven die bepalen op welke manier de crawler omgaat met systeemresources. De groep bronnen die in aanmerking komt om te worden gecrawld, vormt de *crawlruimte* van een crawler. Nadat u een crawler hebt gemaakt, kunt u de crawlereigenschappen nog altijd wijzigen om verandering te brengen in de manier waarop de crawler gegevens verzamelt. Bovendien kunt u de crawlruimte aanpassen en nieuwe bronnen toevoegen, of juist bronnen verwijderen die niet meer doorzocht hoeven te worden.

#### **Verwante onderwerpen**

Beheer van enterprise search-crawler

#### **Verwante taken**

Crawlers bewaken

## **Parsers van enterprise search**

Een enterprise search-parser analyseert documenten die zijn verzameld door een crawler, en maakt ze gereed voor indexering.

De parsercomponent die bij WebSphere Information Integrator OmniFind Edition wordt geleverd, analyseert de content en de metagegevens van documenten. De resultaten van deze analyse worden opgeslagen in een datastore, waar ze toegankelijk zijn voor de indexeringscomponent. De parser voert de volgende taken uit:

- Tekst uit het document extraheren, ongeacht de indeling van het document. Zo kan de parser de tekst in XML- en HTML-documenten onderscheiden van de tags. Met behulp van zogenaamde Stellent for IBM WebSphere Information Integrator OmniFind Edition Outside In Viewer Technology kan de parser tevens tekst extraheren uit binaire bestandsindelingen, zoals Microsoft Word- en Adobe Acrobat PDF-documenten.
- De tekensetcodering van elk document vaststellen. Voordat er enige taalkundige analyse wordt uitgevoerd, gebruikt de parser deze informatie om alle tekst te converteren naar Unicode.
- De brontaal van elk document vaststellen.
- Tekst extraheren en tokens toevoegen om de vindbaarheid van gegevens te verbeteren. Tijdens deze fase voert de parser de volgende taken uit:
  - De tekens normaliseren. Dit behelst de normalisering van onder meer hoofdlettergebruik en diakritisch tekens, zoals e-trema.

- De structuur van alinea's, zinnen, woorden en spaties analyseren. Via taalkundige analyse haalt de parser samengestelde woorden uit elkaar en wijst de parser tokens toe die het opzoeken van woorden en synoniemen mogelijk maken.
- Ontleedregels toepassen die u voor de collectie hebt opgegeven. Bij het configureren van de parser kunt u de volgende activiteiten configureren:

#### **Regels voor veldkoppeling van XML- en HTML-documenten**

Met deze optie kunnen gebruikers zowel gestructureerde als niet-gestructureerde content in XML- en HTML-documenten doorzoeken. Als u XML-elementen of HTML-metagegevens-elementen koppelt aan zoekvelden in de index van enterprise search, kunnen gebruikers de veldnamen opgeven in query's en op die manier specifieke onderdelen van XML- en HTML-documenten doorzoeken. Query's waarbij specifieke velden worden doorzocht, kunnen nauwkeurigere zoekresultaten opleveren dan query's in de vorm van vrije tekst, waarbij de volledige inhoud van documenten wordt doorzocht.

#### **Categories**

Met deze optie kunnen gebruikers documenten onderzoeken aan de hand van de categorieën waartoe die documenten behoren. De gebruikers kunnen categorieën in de zoekresultaten selecteren en het bladeren vervolgens beperken tot documenten die tot dezelfde categorie behoren.

Als u een collectie maakt, geeft u op of u gebruik wilt maken van categorisering. Als u *regelgebaseerde* categorieën gebruikt, worden documenten aan categorieën gekoppeld aan de hand van regels die u opgeeft. Op regels gebaseerde categorieën kunt u configureren met enterprise search-collecties die u maakt en met collecties die u migreert vanuit IBM WebSphere Portal.

#### **Aangepaste tekstanalyse**

Programma-ontwikkelaars kunnen aangepaste analyseprogramma's schrijven om complexe taalkundige analyse uit te voeren op de gegevens die u moet doorzoeken. Deze programma's kunt u inpluggen in het enterprise search-systeem en gebruiken om aantekeningen te maken bij de content van uw collecties. Door deze aantekeningen te indexeren, maakt u collecties geschikt voor semantisch zoeken (zoeken op betekenissen).

Bijvoorbeeld: gebruikers kunnen zoeken naar termen die bij elkaar in de buurt of in dezelfde zin staan, of ze kunnen zoeken naar relaties tussen verschillende termen. Een gebruiker zou bijvoorbeeld op zoek kunnen zijn naar documenten waarin een IBM-vertegenwoordiger met de naam Jansen ter sprake komt, en niet een IBM-technicus met diezelfde naam.

#### **Ondersteuning van n-gram-segmentering**

Om de vindbaarheid van Chinese, Japanse en Koreaanse documenten te verbeteren, kunt u de n-gram-segmenteringsmethode van lexicale analyse inschakelen. Bij deze vorm van analyse worden woorden niet van elkaar gescheiden aan de hand van spaties. Nadat een collectie eenmaal is gemaakt, is het niet meer mogelijk de segmenteringsmethode te wijzigen.

#### **Ondersteuning voor het doorzoeken van XML-documenten met native XML-zoekopdrachten**

Een native XML-zoekopdracht kan nauwkeurigere zoekresultaten ople-

veren omdat ook de XML-markup wordt doorzocht. Zo kan in de query worden opgegeven dat een bepaald woord moet voorkomen in een bepaald XML-element.

### **Klassen om het relatieve belang van velden te verhogen**

Als u velden toewijst om klassen te ondersteunen, kunt u daarmee invloed uitoefenen op de manier waarop documenten worden gerangschikt in het zoekresultaat. U zou bijvoorbeeld de score van titelvelden een hoger belang kunnen geven; als de term die u zoekt dan voorkomt in de titel van een document, krijgt dat document een hogere plaats in de zoekresultaten dan een document waarin de term niet in de titel staat.

### **Verwante onderwerpen**

Werken met categorieën

XML-zoekvelden

HTML-zoekvelden

Integratie van aangepaste tekstanalyse

Tekstanalyse in enterprise search

### **Verwante taken**

De parser bewaken

## **Indexen van enterprise search**

De indexcomponenten van enterprise search worden op gezette tijden uitgevoerd en voegen informatie over nieuwe en gewijzigde documenten toe aan de index.

Om te garanderen u dat gebruikers altijd toegang hebben tot de meest recente informatie in de bronnen die ze doorzoeken, bestaat het bouwen van een index uit twee stadia:

### **De hoofdindex bouwen**

Tijdens het bouwen van de hoofdindex wordt de index helemaal opnieuw opgebouwd, zodat de structuur optimaal geordend of "georganiseerd" is. De indexeringsprocessen lezen alle gegevens die door de crawlers zijn verzameld en door de parser zijn geanalyseerd.

### **Delta-indexen bouwen**

Bij het bouwen van een delta-index wordt de informatie toegevoegd die is gecrawld sinds de laatste keer dat de hoofdindex is gebouwd.

Bij het configureren van indexopties voor een collectie kunt u roosters voor het bouwen van de hoofd- en de delta-indexen opgeven. De frequentie waarmee u de index het beste kunt bouwen, hangt af van de beschikbare systeemresources en van de vraag of de bronnen die worden geïndexeerd, statische of dynamische content bevatten.

Om te zorgen dat nieuwe informatie snel beschikbaar is, moet u opgeven dat er vaak een index wordt gebouwd. Plan ook af en toe de bouw van een hoofdindex, zodat alle nieuwe informatie wordt samengebracht, nieuwe content wordt geanalyseerd en de snelheid van de index wordt geoptimaliseerd.

U kunt de indexeringsprocessen ook starten zonder ze in een rooster te plaatsen. Als u bijvoorbeeld bepaalde parsingsregels wijzigt en wilt dat die wijzigingen beschikbaar komen voor uw zoekprogramma's, kunt u het bouwen van de hoofd-

index starten meteen nadat de gegevens zijn gecrawld en geanalyseerd. Het is dus niet nodig om te wachten tot de index volgens het rooster de volgende keer wordt gebouwd.

Om het gebruik van resources binnen de perken te houden, kunt u opgeven door hoeveel collecties de indexeringsprocessen kunnen worden gedeeld en kunt u meerdere bouwopdrachten voor indexen tegelijkertijd indienen. Door meerdere indexen gelijktijdig te bouwen, zorgt u dat het bouwen van een zeer grote hoofd-index het bouwen van delta-indexen voor andere collecties niet in de weg staat. Het bouwen van indexen is een resource-intensief proces, en vooral bij grote systemen is het dus van belang dat u de systeembelasting in de gaten houdt en het rooster voor het bouwen van hoofd- en delta-indexen daaraan aanpast.

Bij het bouwen van een index voeren de indexeringsprocessen een globale documentanalyse uit. Tijdens deze fase worden er algoritmen uitgevoerd om duplicaatdocumenten op te sporen en om de linkstructuur van documenten te analyseren. Bovendien worden er in webdocumenten speciale bewerkingen uitgevoerd op ankertekst (de tekst die de doelpagina in een hypertext-link beschrijft).

U kunt opties opgeven voor de volgende indexeringsactiviteiten:

- Om het voor gebruikers mogelijk te maken jokertekens te gebruiken, kunt u ondersteuning inbouwen voor het uitbreiden van zoektermen in de index, of kunt u opgeven dat de zoektermen tijdens de verwerking van de query moeten worden uitgebreid. U moet hier een keuze maken tussen resourcegebruik en responstijd van de query.
- U kunt scopes configureren. Met een *scope* kunt u beperkingen opleggen aan hetgeen de gebruikers in de collectie kunnen zien. U zou bijvoorbeeld één scope kunnen definiëren voor de URI's van documenten van de Technische Dienst en een andere scope voor de URI's van documenten van de afdeling HR. Als het zoekprogramma overweg kan met scopes, is het voor gebruikers mogelijk documenten te zoeken en op te halen van alleen die subsets van de collectie.
- U kunt opties opgeven voor het samenvouwen van de zoekresultaten voor documenten die dezelfde URI-prefix hebben. Bovendien kunt u een groepsnaam opgeven, zodat documenten met verschillende URI-prefixen bij elkaar kunnen worden samengevouwen in de zoekresultaten.
- Nadat de index is gebouwd, is het mogelijk om URI's van documenten die de gebruikers niet mogen doorzoeken, te verwijderen.

#### **Verwante onderwerpen**

Beheer van de enterprise search-index

Jokertekens in query's

Scopes

Samengevouwen URI's

Documentranking op basis van URI-patronen

#### **Verwante taken**

Rooster opstellen voor het opbouwen van de index

Indexen gelijktijdig opbouwen

URI's uit de index verwijderen

De indexactiviteit voor een collectie bewaken

De indexeerwachtrij van enterprise search bewaken



## Zoekservers voor enterprise search

De zoekservers voor enterprise werken bij het verwerken van query's, het doorzoeken van de index en het afbeelden van de zoekresultaten samen met uw zoekprogramma's.

De zoekservers voor enterprise search worden geïnstalleerd op het moment dat u WebSphere Information Integrator OmniFind Edition installeert. Als u de zoekservers configureert voor een collectie, kunt u opties opgeven voor de manier waarop die collectie moet worden doorzocht:

- U kunt een zoekcache configureren waarin zoekresultaten die vaak worden opgevraagd, worden bewaard. Een dergelijke cache kan de zoek- en ophaalsnelheid verhogen.
- U kunt een standaardtaal opgeven voor het doorzoeken van documenten in de collectie.
- Als uw programmaontwikkelaars woordenboeken op maat maken, kunt u die woordenboeken aan collecties koppelen:
  - Als gebruikers zoeken in een collectie die gebruikmaakt van een *synoniemenwoordenboek*, worden documenten die synoniemen van gezochte woorden bevatten, eveneens opgenomen in de zoekresultaten.
  - Als gebruikers zoeken in een collectie die gebruikmaakt van een *stopwoordenboek*, worden de stopwoorden uit de query verwijderd voordat de query wordt verwerkt.
  - Als gebruikers zoeken in een collectie die gebruikmaakt van een *gewogen woordenboek*, wordt het belang van de documenten die de "gewogen woorden" bevatten, verhoogd of verlaagd aan de hand van de wegingsfactor van die woorden in het woordenboek.
- Als u vooraf denkt dat bepaalde documenten relevant zijn voor bepaalde zoekopdrachten, kunt u zogenaamde "snellinks" configureren. Een *snellink* koppelt een bepaalde URI aan specifieke trefwoorden en termen. Als een query een of meer trefwoorden of termen bevat die u hebt opgegeven in de definitie van een snellink, wordt de bijbehorende URI automatisch afgebeeld als onderdeel van de zoekresultaten.

In een configuratie met meerdere servers is er bescherming tegen storingen beschikbaar, niet alleen op serverniveau, maar ook op collectieniveau. Als een collectie op één zoekserver om welke reden dan ook niet meer beschikbaar is, worden de query's voor die collectie automatisch doorgestuurd naar een andere zoekserver.

### Verwante onderwerpen

Zoekprogramma's voor enterprise search

Zoekcache

Aangepaste synoniemenwoordenboeken

Aangepaste stopwoordenboeken

Aangepaste gewogen woordenboeken

Snellinks

### Verwante taken

De zoekservers bewaken

## De beheerconsole van enterprise search

De beheerconsole van enterprise search draait in een webbrowser, wat betekent dat beheerders er altijd en overal toegang toe kunnen krijgen. Dankzij de geavanceerde beveiliging krijgen allen gebruikers die speciaal gemachtigd zijn toegang tot de beheerfuncties.

De beheerconsole voor enterprise search wordt op de zoekservers geïnstalleerd op het moment dat u WebSphere Information Integrator OmniFind Edition installeert.

De beheerconsole bevat wizards die u helpen bij het uitvoeren van een aantal elementaire beheertaken. Zo helpt de Collectiewizard u bij het maken van een collectie en kunt u met behulp van die wizard uw werk opslaan als concept. Crawlerwizards zijn specifiek voor een bepaalde gegevensbron en helpen u bij het selecteren van de bronnen die door de gebruikers doorzocht moeten kunnen worden.

Voor andere beheertaken kunt u een selectie maken uit de afzonderlijke items die u wilt beheren. Als u een collectie wilt wijzigen, kunt u op de pagina Indexeren bijvoorbeeld het rooster voor indexering wijzigen en kunt u op de pagina Parser regels voor het analyseren van XML-documenten aanpassen.

### Verwante onderwerpen

Enterprise search - Systeembeheer

Beheerdersrollen

### Verwante taken

Aanmelden bij de beheerconsole

## Een enterprise search-systeem bewaken

Met de beheerconsole van enterprise search kunt u de systeemactiviteiten bewaken en de werking waar nodig aanpassen.

Nadat u WebSphere Information Integrator OmniFind Edition hebt geïnstalleerd en minimaal één collectie hebt gedefinieerd, kunt u gedetailleerde gegevens oproepen voor elke hoofdactiviteit (crawlen, analyseren of parsen, indexeren en zoeken). De informatie omvat onder meer responstijden en voortgangsgegevens, zoals hoeveel documenten er tijdens een bepaalde crawl- of indexeringsessie zijn gecrawld of geïndexeerd.

De meeste activiteiten kunt u starten en stoppen. U kunt een activiteit bijvoorbeeld ook onderbreken (pauze), de configuratie ervan wijzigen of een probleem oplossen, en de werking hervatten zodat de activiteit kan worden voortgezet.

U kunt ook waarschuwingen of "alerts" configureren. Deze maken het mogelijk dat u een e-mail ontvangt wanneer bij activiteiten een bepaalde "bewaakte" gebeurtenis optreedt. U kunt bijvoorbeeld worden gewaarschuwd als de responstijd voor zoekopdrachten een bepaalde drempelwaarde overschrijdt.

Als een document uit het enterprise search-systeem is gewist, kunt u dat document traceren en bepalen wanneer, waar en waarom dat document gewist is. Het kan bijvoorbeeld zijn dat de parser het document niet kon analyseren of dat een beheerder het document uit de index heeft verwijderd.

### Verwante onderwerpen

Activiteit van enterprise search beheren

Een enterprise search-systeem starten en stoppen

## Logbestanden van enterprise search

Er worden logbestanden gemaakt voor afzonderlijke collecties en voor sessies op systeemniveau.

Als u de opties voor logbestanden voor een enterprise search-collectie of voor het systeem configureert, geeft u op welke typen berichten er moeten worden gelogd, zoals foutberichten en waarschuwingsberichten. U kunt tevens opgeven hoe vaak het systeem oude logbestanden moet overschrijven, om plaats te maken voor nieuwe berichten. Daarnaast kunt u opgeven dat u bij bepaalde soorten berichten (waaronder waarschuwingen of "alerts"), of bij alle berichten, een e-mail wilt ontvangen.

Als u een logbestand wilt bekijken, selecteert u eenvoudig het gewenste logbestand. De bestandsnaam bevat informatie over het tijdstip waarop het bestand is gemaakt en over de component waarop de berichten betrekking hebben. Bovendien kunt u weergavefilters opgeven. U kunt bijvoorbeeld opgeven dat u alleen foutberichten wilt zien, of alleen berichten van een bepaalde enterprise search-sessie.

### Verwante onderwerpen

Logbestanden en alerts van enterprise search

Alerts

Berichten voor enterprise search

### Verwante taken

Logbestanden configureren

SMTP-servergegevens configureren

E-mail over gelogde berichten ontvangen

Logbestanden bekijken

## Enterprise search aanpassen

De API's (Application Programming Interfaces) voor enterprise search maken het mogelijk om niet alleen zoekprogramma's op maat te maken, maar ook programma's op maat voor het bijwerken van de content van collecties, programma's op maat voor tekstanalyse en woordenboeken op maat voor synoniemen, stopwoorden en gewogen woorden.

Na installatie van WebSphere Information Integrator OmniFind Edition, zijn de volgende groepen API's beschikbaar voor het uitbreiden van enterprise search-collecties:

### Search and Index API (SI-API)

Met deze API kunt u zoekprogramma's op maat en een aangepaste beheerinterface maken.

### Data Listener API

Met deze API kunt u gegevens van externe crawlers ontvangen. Die externe crawlers kunnen verbinding maken met de Data Listener van enterprise search en gegevens toevoegen aan, of verwijderen uit, een collectie.

### Crawler-plugins

Met plugin-API's kunt u metagegevens toevoegen aan documenten wanneer deze worden gecrawld, of beveiligingstokens koppelen aan de hand waarvan de bedrijfs- en beveiligingsregels van uw organisatie kunnen worden gehandhaafd.

U kunt de vindbaarheid van informatie verbeteren door aangepaste programma's voor taalkundig analyse te integreren met uw enterprise search-collecties. Nadat u tekstanalyseprogramma's op maat in het systeem hebt opgenomen, kunt u die programma's koppelen aan collecties. Zodra gebruikers een collectie doorzoeken, profiteren ze dan van de woordkoppelingen die door uw programma's in de index zijn ingebouwd. Zo kunnen gebruikers bijvoorbeeld zoeken naar begrippen en relaties tussen termen, en niet alleen naar de termen zelf.

Daarnaast kunt u de vindbaarheid van informatie verbeteren door het integreren van woordenboeken op maat die bijvoorbeeld afkortingen bevatten, of jargon dat specifiek is voor uw bedrijfstak. Nadat u deze woordenboeken op maat in het systeem hebt opgenomen, kunt u ze koppelen aan collecties. Zodra gebruikers een collectie doorzoeken, profiteren ze van de volgende voordelen:

- Als een query woorden bevat waarvoor synoniemen zijn gedefinieerd, worden ook documenten die synoniemen van de zoektermen bevatten, opgenomen in de zoekresultaten.
- Als een query stopwoorden bevat, worden die stopwoorden uit de query verwijderd, zodat er geen irrelevante documenten worden opgenomen in de zoekresultaten.
- Als een query gewogen woorden bevat, worden documenten die die woorden bevatten in de zoekresultaten hoger of lager gerangschikt, afhankelijk van de wegingsfactor van die woorden in het woordenboek.

#### **Verwante onderwerpen**

Zoekprogramma's voor enterprise search

Aangepaste synoniemenwoordenboeken

Aangepaste stopwoordenboeken

Aangepaste gewogen woordenboeken

Overzicht van de Search and Index API

Data listener

Crawler-plugins

#### **Verwante taken**

Ondersteuning voor Data Listener-programma's configureren

## **Voorbeeld van een zoekprogramma voor enterprise search**

Er wordt een voorbeeld van een zoekprogramma voor enterprise search meegeleverd. Dit kunt u ongewijzigd gebruiken, maar u kunt het ook gebruiken als sjabloon voor het ontwikkelen van uw eigen zoekprogramma's op maat.

Het voorbeeldzoekprogramma wordt geïnstalleerd op het moment dat u WebSphere Information Integrator OmniFind Edition installeert. Het programma geeft een demonstratie van de meeste zoek- en ophaalfuncties die in enterprise search beschikbaar zijn. Maar het is ook gewoon een werkend programma, dat u in staat stelt alle actieve collecties en externe bronnen in uw enterprise search-systeem te doorzoeken. U kunt het voorbeeldprogramma gebruiken om nieuwe collecties en externe bronnen te testen voordat u ze beschikbaar stelt aan uw gebruikers.

Het voorbeeldzoekprogramma maakt ook duidelijk dat er ondersteuning is voor federatieve zoekopdrachten; u kunt er immers meerdere collecties en externe bronnen tegelijk mee doorzoeken.

Bij bepaalde typen crawlers kunt u de component Identiteitenbeheer van enterprise search gebruiken om de actuele legitimatiegegevens te controleren op het moment

dat de gebruikers het zoekprogramma openen. Als het te doorzoeken domein wordt beschermd met SSO-beveiliging (single sign-on), kunnen er SSO-mechanismen worden gebruikt om de gebruiker gedurende de hele zoeksessie te valideren. Is dit niet het geval, dan kan de component Identiteitenbeheer de legitimatiegegevens van de gebruiker in versleutelde vorm opslaan in een profiel en die legitimatiegegevens gebruiken om verboden document uit te sluiten uit de zoekresultaten.

Voor het aanpassen van het voorbeeld-zoekprogramma kunt u gebruik maken van het programma Search Application Customizer; dit is een grafische gebruikersinterface waarin u meteen kunt zien wat de gevolgen zijn van de wijzigingen die u aanbrengt. U kunt het zoekprogramma aanpassen door wijzigingen aan te brengen in het configuratiebestand van het programma.

Als u een zoekprogramma op maat wilt maken, kunt u de Search and Index API voor enterprise search gebruiken.

#### **Verwante onderwerpen**

Zoekprogramma's voor enterprise search

Functies van het voorbeeldzoekprogramma

Overzicht van de Search and Index API

#### **Verwante taken**

Toegang krijgen tot zoekprogramma's

Eigenschappen van het voorbeeldzoekprogramma wijzigen

Zoekprogramma's aanpassen

---

## **De gegevensstroom in enterprise search**

De componenten van enterprise search die u met WebSphere Information Integrator OmniFind Edition installeert, zijn nauw met elkaar vervlochten om te garanderen dat de gegevens ongehinderd door het systeem kunnen stromen.

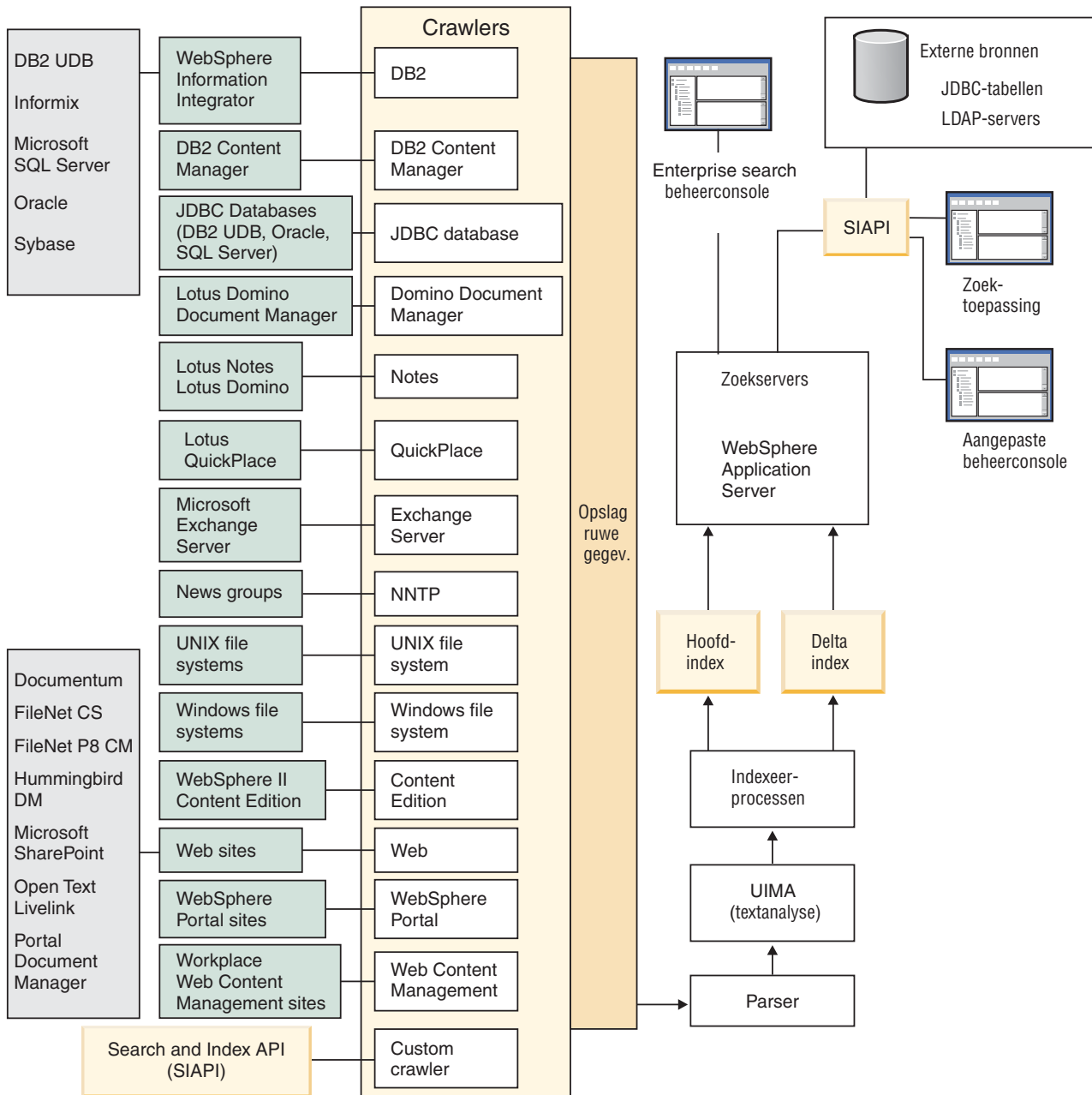
Crawlers verzamelen documenten uit gegevensbronnen in het hele bedrijf. De parser haalt de nuttige informatie uit de gecrawelde documenten en genereert tokens die de documenten bijvoorbeeld koppelen aan categorieën of die kunnen helpen vast te stellen hoe relevant de documenten zijn voor termen in een zoekopdracht. De verzamelde gegevens worden opgeslagen in een index, zodat ze op een efficiënte manier weer kunnen worden opgehaald.

Via een webbrowser en een zoekprogramma kunnen gebruikers de geïndexeerde collecties en externe bronnen doorzoeken. Het zoekprogramma kan in de browser een lijst afbeelden met resultaten waarop de gebruiker kan klikken. Bij meer geavanceerde zoekprogramma's wordt er dynamisch gegenereerde content afgebeeld op basis van informatie in verschillende bronnen.

Bijvoorbeeld: een programma voor het doorzoeken van een catalogus kan de producten die aan een zoekopdracht voldoen, op een speciale manier afbeelden. Met een enkele query kunnen documenten op verschillende soorten gegevensbronnen worden doorzocht, zoals een combinatie van documenten in IBM DB2 Content Manager en Lotus Notes.

De beheerder bepaalt uiteindelijk welke gegevens er worden verzameld en op welke manier deze worden gecrawld, geanalyseerd, geïndexeerd en doorzocht. Door de systeemactiviteiten te bewaken, kan de beheerder aanpassingen maken om de doorvoer van gegevens te optimaliseren.

Het volgende schema geeft aan hoe de gegevensstromen door een enterprise search-systeem lopen.



Figuur 1. Stroomschema van de gegevens in een enterprise search-systeem

### Verwante onderwerpen

“Wat is enterprise search?” op pagina 1

Een enterprise search-systeem biedt uitgebreide mogelijkheden voor het via een enkele query doorzoeken van een willekeurig aantal gestructureerde en niet-gestructureerde gegevensbronnen. Snelle responstijden voor query's en een volledige resultaatset waarin de resultaten zijn gerangschikt op belang, stellen u niet alleen in staat belangwekkende documenten op te sporen, maar ook om daar de juiste informatie in te vinden.

“Overzicht van componenten van enterprise search” op pagina 3

De verschillende componenten van enterprise search verzamelen gegevens in uw hele bedrijf, analyseren, ontleden en categoriseren de informatie en stellen een index samen die door de gebruikers kan worden doorzocht.

**Verwante verwijzing**

“Typen gegevensbronnen die door enterprise search worden ondersteund” op pagina 2

Er is vooraf gedefinieerde ondersteuning voor een groot aantal typen gegevensbronnen beschikbaar.





---

## Enterprise search - Systeembeheer


U kunt de beheerconsole van enterprise search gebruiken om collecties en externe bronnen te maken en beheren, componenten te starten en stoppen, systeem-activiteit en logbestanden te bewaken, beheerders te configureren, zoekprogramma's te koppelen aan collecties en externe bronnen en informatie op te geven om de beveiliging te verbeteren.

### view Collecties

Gebruik de onderstaande richtlijnen om uw eerste collectie te maken en het systeem te beheren.

#### Aanmelden

Meld u aan bij de beheerconsole van enterprise search. De view Collecties vormt het ingangspunt voor het maken en beheren van collecties.

**Tip:** Voor hulp bij de beheerconsole kunt u op elk gewenst moment klikken op **Help** op de werkbalk of op **Help voor deze pagina**. Als voor berichten gedetailleerde uitleg en herstelacties beschikbaar zijn, kunt u klikken op het pictogram  **Meer informatie** aan het eind van het bericht om de details weer te geven.

#### Een collectie maken

Kies een van de volgende methoden om een nieuwe, lege collectie te maken:

- Als u een collectie wilt maken met behulp van de Collectiewizard, klikt u op **Collectiewizard** en volgt u de instructies in de wizard.
- Als u een collectie wilt maken met behulp van de view Collecties, klikt u op **Collectie maken**, voert u gegevens in de velden op de pagina Collectie maken in en klikt u vervolgens op **OK**.

#### De collectie configureren

Open een nieuwe, lege collectie en voeg content toe, geef aan op welke manier gegevens moeten worden gecrawld en maak de gegevens doorzoekbaar.

Klik voor de nieuwe collectie op  **Bewerken** en selecteer een pagina om de opties voor de collectie op te geven.

**Waarschuwing:** Als u wilt teruggaan naar de vorige weergegeven pagina of de informatie in de beheerconsole wilt vernieuwen, klikt u op **Vorige** en **Vernieuwen** in de beheerconsole van enterprise search. Als u in de webbrowser op **Terug** of **Vernieuwen** klikt, kan dat leiden tot inconsistente resultaten en zelfs tot gegevensverlies.

- Op de pagina Algemeen kunt u opties opgeven die van toepassing zijn op de hele collectie:
  - U kunt algemene opties bewerken om de naam of beschrijving van de collectie te wijzigen, of de geschatte grootte van de collectie aan te passen.
  - U kunt informatie over de collectie bekijken die u niet kunt wijzigen, zoals het collectie-ID of de statische-rankingmethode voor de ranking van documenten in de zoekresultaten.

- Als bij het maken van de collectie beveiliging is ingeschakeld, kunt u de beveiligingsopties op documentniveau in- of uitschakelen.
- Configureer op de pagina Crawler ten minste één crawler.  
Een collectie kan gegevens uit verschillende gegevensbronnen bevatten. Voor elk type gegevensbron dat u in de collectie wilt opnemen, moet u ten minste één crawler configureren. Wanneer u een crawler maakt, krijgt u tijdens de configuratie ervan hulp van een wizard die specifiek van toepassing is op het type gegevens voor de crawler.
- Op de pagina Parser kunt u opties configureren voor de manier waarop de gecrawlde gegevens moeten worden geanalyseerd, zodat hiernaar op effectieve wijze kan worden gezocht:
  - U kunt opgeven of XML-documenten moeten worden geanalyseerd, zodat ze kunnen worden doorzocht met behulp van native XML-zoekopdrachten.
  - U kunt categorieën aan documenten koppelen, zodat gebruikers in een subset van de collectie kunnen zoeken of per categorie kunnen zoeken in de zoekresultaten.
  - U kunt XML-elementen en HTML-metagegevens-elementen toewijzen aan zoekvelden in de index, zodat gebruikers de veldnamen kunnen opgeven in query's en specifieke gedeelten van documenten kunnen doorzoeken.
  - Als u aangepaste tekstanalyseprogramma's aan het enterprise search-systeem hebt toegevoegd, kunt u een programma kiezen voor gebruik met de collectie en vervolgens tekstverwerkingsopties opgeven om de vindbaarheid van informatie te verbeteren.
  - U kunt wegingsklassen aan velden koppelen om de ranking van zoektermen in de zoekresultaten te beïnvloeden.
- Op de pagina Index kunt u roosters configureren voor het bouwen van de index. U kunt opgeven dat de index regelmatig moet worden opgebouwd, zodat gebruikers altijd de beschikking hebben over de laatste informatie. Daarnaast kunt u de volgende (optionele) acties uitvoeren:
  - Instellen dat gebruikers jokertekens in zoektermen kunnen opgeven.
  - Scopes opgeven, zodat gebruikers in een bepaald deel van de collectie kunnen zoeken in plaats van alle documenten in de index.
  - Zoekresultaten samenvouwen, zodat de documenten uit dezelfde bron in de zoekresultaten worden samengevouwen.
  - Wegingsfactoren toewijzen om de ranking in de zoekresultaten te beïnvloeden van documenten waarvan het URI-patroon overeenkomt.
  - URI's uit de index verwijderen. Mogelijk wilt u voorkomen dat gebruikers bepaalde documenten kunnen bekijken nadat de collectie is gemaakt.
- Op de pagina Zoeken kunt u opties opgeven voor het zoeken van documenten in de collectie:
  - U kunt cacheruimte reserveren voor de zoekresultaten en de standaardtaal van de collectie wijzigen.
  - Als u aangepaste woordenboeken voor synoniemen, stopwoorden en gewogen woorden aan het enterprise search-systeem hebt toegevoegd, kunt u de woordenboeken selecteren die u wilt gebruiken wanneer gebruikers de collectie doorzoeken.
  - U kunt opgeven hoeveel ruimte in de zoekresultaten moet worden bijgemaakt voor een samenvatting voor elk document.

- Als u specifieke URI's in de zoekresultaten wilt weergeven wanneer bepaalde trefwoorden of zinnen in een query voorkomen, kunt u snel-links configureren.
- Op de pagina Loggen kunt u de volgende acties uitvoeren:
  - Opties opgeven voor de typen berichten die u wilt loggen en opgeven hoe vaak de logbestanden opnieuw moeten worden gebruikt.
  - Opties opgeven voor het ontvangen van alerts (waarschuwingen) voor collectieactiviteiten. Op die manier zou u bijvoorbeeld kunnen worden gewaarschuwd als de gemiddelde responstijd voor zoekopdrachten een bepaalde grenswaarde overschrijdt.
  - Opties opgeven voor het ontvangen van e-mail wanneer bepaalde berichten of typen berichten worden gelogd.
  - Opties opgeven voor het loggen van informatie op basis waarvan u kunt bepalen wanneer, waar en waarom een document uit het enterprise search-systeem is gewist.

## De componenten starten

Als u de gegevensbronnen voor het crawlen en de opties voor het verzamelen en doorzoeken van de gegevens hebt opgegeven, kunt u de processen voor het samenstellen van de collectie starten. De volgorde waarin u de componenten start, is van doorslaggevend belang. Gegevens kunnen pas worden geanalyseerd en geïndexeerd als de gegevens door de crawler zijn gecrawld, en de zoekservers kunnen pas zoekopdrachten verwerken als de hoofdindex is opgebouwd.

## Externe bronnen

Als u gegevensbronnen wilt doorzoeken zonder de gegevens te crawlen of indexeren, klikt u op **Externe bronnen** op de werkbalk om opties op te geven waarmee u de gegevensbronnen doorzoekbaar kunt maken. U moet informatie opgeven om de JDBC-databases (Java Database Connectivity) en de LDAP-servers (Lightweight Directory Access Protocol) te activeren die voor enterprise search worden benaderd. Als u de externe bronnen aan de zoekprogramma's hebt gekoppeld, kunnen gebruikers deze bronnen doorzoeken op hetzelfde moment dat zij collecties doorzoeken met gegevens die zijn gecrawld, geanalyseerd en geïndexeerd.

## De view System

Als u lid bent van de enterprise search-beheerdersrol, kunt u klikken op **System** op de werkbalk om de volgende activiteiten uit te voeren. Beheerders, operators en bewakers van collecties hebben alleen toegang tot deze view als een enterprise search-beheerder deze machtiging heeft ingesteld.

- Data Listener-programma's op de client configureren zodat de collecties op de client kunnen worden bijgewerkt.
- Aangepaste tekstanalyseprogramma's aan het systeem toevoegen.
- Aangepaste woordenboeken voor synoniemen, stopwoorden en gewogen woorden aan het systeem toevoegen.
- Aangeven hoeveel collecties parallel indexen kunnen opbouwen en opgeven of de hoofdindexen voor een collectie gelijktijdig kunnen worden uitgevoerd met opgebouwde delta-indexen.
- Alerts configureren voor events op systeemniveau.
- Opties opgeven voor het loggen van berichten die zijn gegenereerd door sessies op systeemniveau.

- Informatie opgeven over de mailservers zodat u e-mails kunt ontvangen over de enterprise search-activiteiten.

## De view Beveiliging

Als u lid bent van de enterprise search-beheerdersrol, kunt u klikken op **Beveiliging** om de beveiligingsopties op te geven. Beheerders, operators en bewakers van collecties hebben geen toegang tot deze view.

Als u de beveiliging inschakelt in IBM WebSphere Application Server, kunt u de view Beveiliging gebruiken om de beheerdersrollen te configureren. Op deze manier kunt u meer gebruikers de mogelijkheid geven systeembeheer uit te voeren, maar kunt u de toegang van elk van die gebruikers wel beperken tot bepaalde functies en collecties.

U kunt de view Beveiliging ook gebruiken om opties voor identiteitenbeheer te configureren. U kunt bijvoorbeeld opties opgeven waarmee de legitimatiegegevens van gebruikers in profielen worden opgeslagen, die kunnen worden gebruikt om de actuele legitimatiegegevens van gebruikers tijdens de queryverwerking te valideren. Als de te doorzoeken bron wordt beveiligd met SSO-beveiliging (Single Sign-On), kunt u ook opties opgeven voor het gebruik van SSO-verificatiemethoden om de actuele legitimatiegegevens van gebruikers tijdens de queryverwerking te valideren.

Als u nog geen eigen zoekprogramma's hebt gemaakt, kunt u het voorbeeldzoekprogramma gebruiken om alle collecties en externe bronnen te doorzoeken. Als u een aangepast zoekprogramma hebt gemaakt, kunt u de view Beveiliging gebruiken om het programma te koppelen aan de collecties en externe bronnen die ermee kunnen worden doorzocht.

## De view Monitor

U kunt op elk gewenst moment op het  **Monitor** klikken om de systeem- of collectiecomponenten te bewaken. Als uw beheerdersrol dat toelaat, kunt u de componentprocessen ook starten en stoppen terwijl u ze bewaakt.

### Verwante onderwerpen

“Beheer van enterprise search-crawler” op pagina 35

U moet crawlers configureren voor de verschillende gegevenstypen die u in een collectie wilt opnemen. Een collectie kan een willekeurig aantal crawlers bevatten.

“Activiteit van enterprise search beheren” op pagina 299

Tijdens het bewaken van systeem- en collectieactiviteiten kunt u de status van verschillende processen bekijken, mogelijke problemen controleren of de configuratie-instellingen aanpassen om de prestaties te verbeteren.

### Verwante taken

“Een enterprise search-systeem starten” op pagina 291

Om in te stellen dat gebruikers een collectie kunnen doorzoeken, moet u de systeemprocessen starten en vervolgens de servers starten waarmee de collectie wordt gecrawld, geanalyseerd en geïndexeerd (de zoekservers worden automatisch gestart).

“Zoekservers in de stand-alone werkstand beheren” op pagina 296

Om te zorgen dat de zoekservers altijd beschikbaar zijn, kunt u de zoekservers voor afzonderlijke collecties starten zelfs als de indexserver niet actief is.

“sEen enterprise search-systeem stoppen” op pagina 294

Als u wijzigingen hebt aangebracht in een enterprise search-server of als u problemen moet oplossen, moet u de server mogelijk stoppen en opnieuw starten.

“Een collectie maken met de Collectiewizard” op pagina 29

Als u enterprise search voor het eerst gebruikt, kunt u gebruikmaken van een wizard die u kan helpen bij het maken van een collectie. De wizard levert details voor elke stap in het proces en biedt de mogelijkheid de instellingen op te slaan terwijl u de stappen doorloopt.

“Een collectie maken met behulp van de view Collecties” op pagina 31

U kunt de view Collecties gebruiken om een lege collectie te maken. Vervolgens kunt u de collectie bewerken en opties opgeven op basis waarvan gegevens aan de collectie worden toegevoegd en de collectie doorzoekbaar wordt gemaakt.

---

## Aanmelden bij de beheerconsole

Om een enterprise search-systeem te beheren, geeft u een URL in een webbrowser op en meldt u zich vervolgens aan bij de beheerconsole.

### Voordat u begint

U moet zich aanmelden met een gebruikers-ID met toegangsmachtigingen voor de beheerconsole van enterprise search:

- Als u algemene beveiliging niet inschakelt in WebSphere Application Server, kan alleen de enterprise search-beheerder die is opgegeven tijdens de installatie van WebSphere II OmniFind Edition toegang krijgen tot de beheerconsole.
- Als u algemene beveiliging inschakelt in WebSphere Application Server, kunt u de beheerconsole van enterprise search gebruiken om beheerdersrollen te configureren. De gebruikers-ID's die u configureert, moeten aanwezig zijn in het gebruikersregister van WebSphere Application Server. Als u beheerdersrollen configureert, verleent u meer gebruikers toegang tot de beheerconsole terwijl u kunt bepalen met welke functies en collecties elke gebruiker kan werken.

### Procedure

Ga als volgt te werk om u aan te melden bij de beheerconsole van enterprise search:

1. Voer de URL voor de beheerconsole in de webbrowser in. Bijvoorbeeld:  
`http://ZoekServer.com/ESAdmin/`  
*ZoekServer.com* is de hostnaam van de zoekserver voor enterprise search. Afhankelijk van de webserverconfiguratie kan het zijn dat u het poortnummer moet opgeven. Bijvoorbeeld:  
`http://ZoekServer.com:9080/ESAdmin/`
2. Voer uw gebruikers-ID en wachtwoord in op de welkomspagina en klik op **Aanmelden**.

De view Collecties verschijnt. Deze view vormt het startpunt voor het beheren van het systeem en de collecties. Als u beheerdersrollen gebruikt, zijn de acties die u kunt uitvoeren en de collecties die worden weergegeven afhankelijk van uw beheerdersrol.

Als uw sessie gedurende een bepaalde periode inactief is, wordt u automatisch afgemeld. Als u wilt doorgaan met het beheren van het systeem, moet u zich weer aanmelden.

Als u klaar bent met het beheren van de collecties, klikt u op **Afmelden** om u af te melden van de console. Vervolgens kunt u zich aanmelden met een ander ID en wachtwoord of u kunt de webbrowser sluiten om de beheerconsole af te sluiten.

#### **Verwante onderwerpen**

“Beheerdersrollen” op pagina 234

Enterprise search gebruikt het concept van rollen om de toegang tot verschillende functies in de beheerconsole te besturen.

#### **Verwante taken**

“Een enterprise search-systeem starten” op pagina 291

Om in te stellen dat gebruikers een collectie kunnen doorzoeken, moet u de systeemprocessen starten en vervolgens de servers starten waarmee de collectie wordt gecrawld, geanalyseerd en geïndexeerd (de zoekservers worden automatisch gestart).

“sEen enterprise search-systeem stoppen” op pagina 294

Als u wijzigingen hebt aangebracht in een enterprise search-server of als u problemen moet oplossen, moet u de server mogelijk stoppen en opnieuw starten.

---

## **Het beheerderswachtwoord voor enterprise search wijzigen voor een configuratie met één server**

Het wachtwoord voor de enterprise search-beheerder wordt in gecodeerde vorm opgeslagen. U kunt het wachtwoord wijzigen met behulp van het script `eschangepw`.

### **Beperkingen**

Wachtwoorden kunnen de volgende speciale tekens bevatten:

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

Als u op AIX-, Linux- en Solaris-systemen een wachtwoord opgeeft dat speciale tekens bevat, moet u het volledige wachtwoord tussen enkele aanhalingstekens plaatsen.

Bijvoorbeeld: `'mypwd@$%'`

Als u op een Windows-systeem een wachtwoord met speciale tekens opgeeft, moet u het volledige wachtwoord tussen dubbele aanhalingstekens plaatsen.

Bijvoorbeeld: `"my?+!pwd"`

### **Over deze taak**

Het wachtwoord van het oorspronkelijke enterprise search-beheerders-ID is opgegeven tijdens de installatie van WebSphere II OmniFind Edition.

Als u het wachtwoord wilt wijzigen, moet u het script `eschangepw` uitvoeren om de wijziging in het hele enterprise search-systeem door te voeren. Met het installatieprogramma worden twee omgevingsvariabelen gemaakt die u met het script `eschangepw` kunt gebruiken:

#### **ES\_INSTALL\_ROOT**

De enterprise search-installatiedirectory.

## ES\_NODE\_ROOT

De enterprise search-gegevensdirectory. Het wachtwoord voor het enterprise search-beheerders-ID wordt opgeslagen in het bestand es.cfg in deze directory.

### Procedure

Ga als volgt te werk om het enterprise search-beheerderswachtwoord te wijzigen in een configuratie met één server:

1. Meld u aan als de enterprise search-beheerder en stop het enterprise search-systeem:

```
esadmin system stopall
```

**Belangrijk:** Als het systeem is gestopt, kunnen gebruikers geen zoekopdrachten meer indienen.

2. Wijzig het systeemwachtwoord voor het enterprise search-beheerders-ID met behulp van besturingssysteemopdrachten (onder AIX, Linux of Solaris) of met behulp van de functie voor het wijzigen van wachtwoorden (onder Windows).
3. Voer het volgende script uit, waarbij *NieuweWaarde* het wachtwoord is dat u in stap 2 hebt opgegeven:

#### AIX, Linux of Solaris

```
eschangepw.sh NieuweWaarde
```

#### Windows

```
eschangepw NieuweWaarde
```

4. Start het enterprise search-systeem opnieuw:

```
esadmin system startall
```

#### Verwante verwijzing

“Opdrachten, retourcodes en sessie-ID’s in enterprise search” op pagina 337  
U kunt opdrachten gebruiken om een diagnose te maken van problemen, de status vast te stellen van de verschillende delen van het systeem, sessies te starten en te stoppen of het systeem te starten en te stoppen.

---

## Het beheerderswachtwoord voor enterprise search wijzigen voor een configuratie met meerdere servers

Het wachtwoord voor de enterprise search-beheerder wordt in gecodeerde vorm opgeslagen. Als u het wachtwoord wilt wijzigen, moet u het script eschangepw uitvoeren om de wijziging op alle computers in het enterprise search-systeem door te voeren.

### Beperkingen

Wachtwoorden kunnen de volgende speciale tekens bevatten:

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

Als u op AIX-, Linux- en Solaris-systemen een wachtwoord opgeeft dat speciale tekens bevat, moet u het volledige wachtwoord tussen enkele aanhalingstekens plaatsen.

Bijvoorbeeld: 'mypwd@\$%'

Als u op een Windows-systeem een wachtwoord met speciale tekens opgeeft, moet u het volledige wachtwoord tussen dubbele aanhalingstekens plaatsen.

Bijvoorbeeld: "my?+!pwd"

### Over deze taak

Het wachtwoord voor het enterprise search-beheerders-ID (dat is opgegeven tijdens de installatie van WebSphere II OmniFind Edition), moet hetzelfde zijn op alle enterprise search-servers.

Als u het wachtwoord wilt wijzigen en de wijziging wilt doorvoeren in het hele enterprise search-systeem, moet u het script `eschangepw` uitvoeren op elke computer waarop u met enterprise search werkt. In de onderstaande procedure wordt een volgorde voorgesteld voor het wijzigen van het wachtwoord op alle servers. U bent niet verplicht deze volgorde te volgen, maar u moet wel alle stappen uitvoeren die voor elk servertype vereist zijn.

Met het installatieprogramma worden twee omgevingsvariabelen gemaakt die u met het script `eschangepw` kunt gebruiken:

#### ES\_INSTALL\_ROOT

De enterprise search-installatiedirectory.

#### ES\_NODE\_ROOT

De enterprise search-gegevensdirectory. Het wachtwoord voor het enterprise search-beheerders-ID wordt opgeslagen in het bestand `es.cfg` in deze directory.

### Procedure

Ga als volgt te werk om het enterprise search-beheerderswachtwoord te wijzigen in een configuratie met meerdere servers:

1. Meld u op de enterprise search-indexserver aan als de enterprise search-beheerder en stop het enterprise search-systeem:  
`esadmin system stopall`

**Belangrijk:** Als het systeem is gestopt, kunnen gebruikers geen zoekopdrachten meer indienen.

- a. Wijzig het systeemwachtwoord voor het enterprise search-beheerders-ID met behulp van besturingssysteemopdrachten (onder AIX, Linux of Solaris) of met behulp van de functie voor het wijzigen van wachtwoorden (onder Windows).
- b. Voer het volgende script uit, waarbij *NieuweWaarde* het wachtwoord is dat u in stap 1a hebt opgegeven:

#### AIX, Linux of Solaris

```
eschangepw.sh NieuweWaarde
```

#### Windows

```
eschangepw NieuweWaarde
```

2. Voer de volgende stappen uit op de overige computers in het enterprise search-systeem:
  - a. Meld u aan als enterprise search-beheerder.
  - b. Stop de CCL (Common Communication Layer) voor enterprise search:



### AIX, Linux of Solaris

stopccl.sh

### Windows-opdrachtaanwijzing

stopccl

### Systeembeheer van Windows-services

- 1) Start Windows-services: **Start** → **Programma's** → **Systeembeheer** → **Services**.
  - 2) Klik met de rechtermuisknop op **WebSphere Information Integrator OmniFind Edition** en klik op **Stop**.
- c. Wijzig het systeemwachtwoord voor het enterprise search-beheerders-ID met behulp van besturingssysteemopdrachten (onder AIX, Linux of Solaris) of met behulp van de functie voor het wijzigen van wachtwoorden (onder Windows). Dit wachtwoord moet gelijk zijn aan het wachtwoord dat u in stap 1a op pagina 24 hebt opgegeven.
- d. Voer het volgende script uit, waarbij *NieuweWaarde* het wachtwoord is dat u in stap 1a op pagina 24 hebt opgegeven:

### AIX, Linux of Solaris

eschangepw.sh *NieuweWaarde*

### Windows

eschangepw *NieuweWaarde*

- e. Start de CCL opnieuw:

### AIX, Linux of Solaris

startccl.sh -bg

### Windows-opdrachtaanwijzing

startccl

### Systeembeheer van Windows-services

Ga als volgt te werk om de CCL op de achtergrond te starten:

- 1) Start Windows-services: **Start** → **Programma's** → **Systeembeheer** → **Services**.
  - 2) Klik met de rechtermuisknop op **WebSphere Information Integrator OmniFind Edition** en klik op **Properties**.
  - 3) Klik op de tab **Log On**.
  - 4) Wijzig het wachtwoord door een nieuwe waarde voor het wachtwoord op te geven en klik vervolgens op **OK**.
  - 5) Klik met de rechtermuisknop op **WebSphere Information Integrator OmniFind Edition** en klik op **Start**.
3. Meld u op de enterprise search-indexserver aan als de enterprise search-beheerder en start het enterprise search-systeem opnieuw:
- ```
esadmin system startall
```

### Verwante verwijzing

“Opdrachten, retourcodes en sessie-ID's in enterprise search” op pagina 337  
U kunt opdrachten gebruiken om een diagnose te maken van problemen, de status vast te stellen van de verschillende delen van het systeem, sessies te starten en te stoppen of het systeem te starten en te stoppen.

## TCP-poortnummers voor enterprise search

U kunt het beste de standaardpoortnummers controleren die in een enterprise search-systeem worden gebruikt, zodat eventuele poortconflicten kunnen worden voorkomen tijdens het configureren van resources of het toewijzen van poortnummers aan andere programma's.

Als u een firewall configureert, moet u de toegang tot bepaalde poortnummers expliciet inschakelen. Daarnaast moet u ervoor zorgen dat alle enterprise search-servers zich binnen de firewall bevinden.

*Tabel 1. Poortnummers configureren voor enterprise search*

| Poortnaam                                                                                                                                               | Poortnummers                   | Locatie van configuratie                                                                               |
|---------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------|--------------------------------------------------------------------------------------------------------|
| Common Communication Layer (CCL)                                                                                                                        | 6002                           | ES_NODE_ROOT/nodeinfo/es.cfg en ES_NODE_ROOT/master_config/nodes.ini op alle enterprise search-servers |
| HTTP op de zoekservers                                                                                                                                  | 80                             | HTTP_SERVER_ROOT/conf/http.conf op de zoekservers                                                      |
| Beheerconsole van WebSphere Application Server versie 6                                                                                                 | 9060                           | Op de zoekservers                                                                                      |
| Beheerconsole van WebSphere Application Server versie 5.1                                                                                               | 9090                           | Op de zoekservers                                                                                      |
| DB2-crawler                                                                                                                                             | 6000, 6001, 6002, 60003, 50000 | Op de crawlerserver                                                                                    |
| Data Listener                                                                                                                                           | 6668                           | ES_NODE_ROOT/master_config/datalistener/dlConfig.prp op de crawlerserver                               |
| Informatiecentrum                                                                                                                                       | 8888                           | Op de zoekservers                                                                                      |
| Anonieme of dynamische poorten voor CCL, bestandsoverdracht (ESFTP) en indexkopie                                                                       | 49152 t/m 65535                | Op alle enterprise search-servers                                                                      |
| Cloudscape Network Server                                                                                                                               | 1527                           | Op de crawlerserver                                                                                    |
| Aangepaste communicatie                                                                                                                                 | 8890                           | Op de crawlerserver                                                                                    |
| Clientverbindingen op afstand met de DB2-server (alleen gebruikt voor releases eerder dan WebSphere Information Integrator OmniFind Edition versie 8.4) | 50000                          | Op de crawlerserver                                                                                    |
| WebSphere Information Integrator Content Edition FastObjects-database                                                                                   | 6001 (6002 als alternatief)    | Op de beheerconsole van WebSphere Information Integrator Content Edition                               |
| WebSphere Information Integrator Content Edition RMI-proxyconnector (Remote Method Invocation)                                                          | 1251 (RMI-poort)               | Op de beheerconsole van WebSphere Information Integrator Content Edition                               |

---

## Het poortnummer voor het enterprise search-systeem wijzigen

Als het poortnummer dat het enterprise search-systeem voor communicatie gebruikt, conflicteert met een poortnummer dat door een ander product wordt gebruikt, moet u het poortnummer voor enterprise search wijzigen.

### Over deze taak

Het poortnummer voor het enterprise search-systeem wordt opgegeven tijdens de installatie van WebSphere II OmniFind Edition. In een configuratie met meerdere servers wordt voor alle servers hetzelfde poortnummer opgegeven.

Als het poortnummer niet kan worden gebruikt (als het poortnummer bijvoorbeeld is toegewezen aan een ander product op dezelfde server), verschijnt het volgende foutbericht in het bestand `CCLServer_datum.log`, waarbij *datum* de datum is waarop het logbestand is gemaakt):

```
FFQ00273W Er is een interne fout opgetreden - Uitzonderingsbericht: {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)
```

### Procedure

Ga als volgt te werk om het poortnummer te wijzigen dat door enterprise search wordt gebruikt:

1. Meld u op de computer waarop het poortnummer moet worden gewijzigd aan als de enterprise search-beheerder en stop het enterprise search-systeem:  
`esadmin system stopall`

**Belangrijk:** Als het systeem is gestopt, kunnen gebruikers geen zoekopdrachten meer indienen.

2. Open het bestand `ES_NODE_ROOT/nodeinfo/es.cfg` in een editor, zoek de volgende eigenschap, geeft een nieuwe waarde op voor het poortnummer, sla het bestand op en sluit het bestand vervolgens:

```
CCLPort=nieuw_poort_nummer
```

3. Start de CCL (Common Communication Layer) opnieuw voor enterprise search:

**AIX, Linux of Solaris**

```
startccl.sh
```

**Windows-opdrachtaanwijzing**

```
startccl
```

**Systeembeheer van Windows-services**

Ga als volgt te werk om CCL op de achtergrond te starten:

- a. Start Windows-services: **Start** → **Programma's** → **Systeembeheer** → **Services**.
  - b. Klik met de rechtermuisknop op **WebSphere Information Integrator OmniFind Edition** en klik op **Start**.
4. Ga naar de indexserver, meld u aan als de enterprise search-beheerder en stop de CCL:

**AIX, Linux of Solaris**

```
stopccl.sh
```

## Windows-opdrachtaanwijzing

stopccl

### Systeembeheer van Windows-services

- a. Start Windows-services: **Start** → **Programma's** → **Systeembeheer** → **Services**.
  - b. Klik met de rechtermuisknop op **WebSphere Information Integrator OmniFind Edition** en klik op **Stop**.
5. Bewerk het bestand ES\_NODE\_ROOT/nodeinfo/es.cfg :
- a. Zoek de volgende eigenschap, waarbij *computer\_naam* de naam is van de computer waarvoor u het poortnummer hebt gewijzigd in stap 2 op pagina 27. *N* in de eigenschap *nodeN* is een getal waarmee de enterprise search-server wordt aangegeven.  
`nodeN.destination=computer_naam`
  - b. Zoek de volgende subeigenschap, geef hetzelfde poortnummer op als het nummer dat u voor de server in stap 2 op pagina 27 hebt opgegeven, sla het bestand op en sluit het bestand:  
`nodeN.port=nieuw_poort_nummer`
6. Start het enterprise search-systeem opnieuw:  
`esadmin system startall`

### Verwante verwijzing

“Opdrachten, retourcodes en sessie-ID's in enterprise search” op pagina 337  
U kunt opdrachten gebruiken om een diagnose te maken van problemen, de status vast te stellen van de verschillende delen van het systeem, sessies te starten en te stoppen of het systeem te starten en te stoppen.

“TCP-poortnummers voor enterprise search” op pagina 26

U kunt het beste de standaardpoortnummers controleren die in een enterprise search-systeem worden gebruikt, zodat eventuele poortconflicten kunnen worden voorkomen tijdens het configureren van resources of het toewijzen van poortnummers aan andere programma's.

---

## Collecties van enterprise search

Een enterprise search-collectie bevat de volledige set bronnen die door de gebruikers met een enkele query kan worden doorzocht. Met behulp van federatie kunnen gebruikers met een enkele query meerdere collecties doorzoeken.

Wanneer u een collectie maakt, geeft u instellingen op die voor de hele collectie gelden. De collectie is leeg totdat u er content aan toevoegt.

U kunt op twee manieren collecties aan een enterprise search-systeem toevoegen:

- Als u geen ervaring hebt met de beheerconsole van enterprise search of als u nog niet precies weet op welke manier de collectiecomponenten samenwerken, kunt u de Collectiewizard gebruiken om een collectie te maken. De Collectiewizard begeleidt u door de taken die moeten worden uitgevoerd om een collectie te maken en met behulp van de wizard kunt u uw werk opslaan als concept.
- Als u vertrouwd bent met de beheerconsole, wilt u mogelijk een collectie maken door specifieke pagina's te selecteren die u wilt beheren in de view Collecties.

Als u een collectie hebt gemaakt, kunt u de opties in de view Collecties gebruiken om de instellingen voor de collectie, het enterprise search-systeem en de beveiliging te bewerken en bewaken.

### Collectiefederatie

Als uw zoekprogramma ondersteuning biedt voor federatie, kunnen gebruikers tegelijkertijd meerdere collecties doorzoeken. Daarnaast kunt u met behulp van federatie de groottebeperking voor een collectie omzeilen (20.000.000 documenten per collectie). Zo kunnen gebruikers twee collecties doorzoeken die allebei 20.000.000 documenten bevatten.

De zoekkwaliteit is afhankelijk van de scores die door de afzonderlijke collecties worden gegenereerd, die vervolgens worden samengevoegd om zo de eindresultaatset te maken. De resultaten zijn hetzelfde als wanneer u twee afzonderlijke zoekopdrachten uitvoert en de resultaten vervolgens samenvoegt en de ranking bepaalt.

#### Verwante taken

“Een collectie bewaken” op pagina 300

U kunt algemene gegevens bekijken over de status van elke component in de collectie of opties selecteren om gedetailleerde informatie te bekijken over de activiteiten van elke component of URI.

---

## Een collectie maken met de Collectiewizard

Als u enterprise search voor het eerst gebruikt, kunt u gebruikmaken van een wizard die u kan helpen bij het maken van een collectie. De wizard levert details voor elke stap in het proces en biedt de mogelijkheid de instellingen op te slaan terwijl u de stappen doorloopt.

### Voordat u begint

Om een collectie te maken, moet u de rol van beheerder van enterprise search hebben.

Als u content aan een collectie wilt toevoegen of opties wilt opgeven voor de manier waarop de content in de collectie kan worden geanalyseerd, geïndexeerd of doorzocht, moet u een enterprise search-beheerder of beheerder van de collectie zijn.

## Beperkingen

Met de Collectiewizard kunt u de volgende crawlertypen maken:

- Content Edition
- DB2
- DB2 Content Manager
- Exchange Server
- Notes
- UNIX-bestandssysteem
- Web
- Windows-bestandssysteem

## Over deze taak

Tijdens het maken van een collectie kunt u aangeven dat u de collectie wilt opslaan als concept. Als een collectie is opgeslagen als concept, kunnen beheerders die zijn gemachtigd de collectie te beheren, wijzigingen in de collectie aanbrengen. Mogelijk wilt u dat een collectiebeheerder die ervaring heeft met Lotus Notes-bronnen een Notes-crawler configureert. Later kan een collectiebeheerder die ervaring heeft met UNIX-systemen, de conceptcollectie bewerken en een UNIX-bestandssysteem-crawler configureren.


## Procedure

Ga als volgt te werk om een collectie te maken met de Collectiewizard:

1. Klik op **Collecties** om de view Collecties te openen.
2. Klik op **Collectiewizard**.
3. Volg de aanwijzingen in de wizard om een lege collectie te maken en content aan de collectie toe te voegen.

U moet algemene informatie over de collectie configureren en minimaal één crawler maken. Voor de overige configuratieopties kunt u de standaardwaarden accepteren of opties opgeven voor de nieuwe collectie.

4. Als u de collectie wilt opslaan voordat deze gereed is, klikt u op **Opslaan als concept**.

De collectie wordt samen met de overige conceptcollecties weergegeven in de view Collecties. Als u beveiliging voor de collectie hebt ingeschakeld, wordt het pictogram  **Collectiebeveiliging is ingeschakeld** naast de collectienaam weergegeven.

5. Om terug te gaan naar de collectie die u aan het maken bent, klikt u op **Terug naar wizard** in de view Collecties.
6. Klik op **Voltoeien** om de collectie te maken.

De nieuwe collectie wordt samen met de overige collecties weergegeven in de view Collecties.

Nadat u een collectie hebt gemaakt, moet u de processen starten zodat de gegevens kunnen worden gecrawld, geanalyseerd, geïndexeerd en doorzocht. U kunt het voorbeeldzoekprogramma gebruiken om de nieuwe collectie te doorzoeken totdat u de collectie hebt gekoppeld aan de zoekprogramma's die de collectie kunnen doorzoeken.

#### **Verwante onderwerpen**

“Enterprise search - Systeembeheer” op pagina 17

U kunt de beheerconsole van enterprise search gebruiken om collecties en externe bronnen te maken en beheren, componenten te starten en stoppen, systeemactiviteit en logbestanden te bewaken, beheerders te configureren, zoekprogramma's te koppelen aan collecties en externe bronnen en informatie op te geven om de beveiliging te verbeteren.

---

## **Een collectie maken met behulp van de view Collecties**

U kunt de view Collecties gebruiken om een lege collectie te maken. Vervolgens kunt u de collectie bewerken en opties opgeven op basis waarvan gegevens aan de collectie worden toegevoegd en de collectie doorzoekbaar wordt gemaakt.

#### **Voordat u begint**

Om een collectie te maken, moet u de rol van beheerder van enterprise search hebben.

Als u content aan een collectie wilt toevoegen of opties wilt opgeven voor de manier waarop de content in de collectie kan worden geanalyseerd, geïndexeerd of doorzocht, moet u een enterprise search-beheerder of beheerder van de collectie zijn.

#### **Over deze taak**

Voor informatie over de waarden die u voor een nieuwe collectie kunt opgeven, klikt u tijdens het maken van de collectie op **Help**.

#### **Procedure**

Ga als volgt te werk om een collectie te maken via de view Collecties:

1. Klik in de view Collecties op **Collectie maken**.
2. Geef de gewenste informatie op de pagina Een collectie maken op of selecteer de gewenste waarden in de volgende velden:
  - **Naam collectie**. Geef een beschrijvende naam op voor de content of het doel van de collectie.
  - **Collectiebeveiliging**. Geef aan of u beveiliging voor de collectie wilt inschakelen. Als de collectie eenmaal is gemaakt, kunt u deze instelling niet meer wijzigen. Als collectiebeveiliging is ingeschakeld, kunt u later opties opgeven voor het afdwingen van toegangsbesturing op documentniveau.
  - **Belang van document (statisch rankingmodel)**. Geef een strategie op voor het toewijzen van een statische rankingfactor die wordt gebruikt voor de rangschikking van documenten in de zoekresultaten. Als de collectie eenmaal is gemaakt, kunt u deze waarde niet meer wijzigen.
  - **Type categorisering**. Geef aan of gebruikers documenten kunnen zoeken op basis van de categorieën waartoe de documenten behoren.
  - **Te gebruiken taal**. Geef de standaardtaal op voor het zoeken van documenten in de collectie.

3. Accepteer de standaardwaarden voor de volgende velden of geef de opties op die u voor deze collectie wilt gebruiken:
  - **Beschrijving.** Standaard wordt er geen beschrijving gemaakt.
  - **Geschat aantal documenten.** Standaard is de geschatte grootte van de collectie 1.000.000 documenten. Deze waarde wordt in het systeem gebruikt om een schatting te maken van de geheugen- en schijfresources die voor de collectie nodig zijn (en niet om de grootte van de collectie te beperken).
  - **Locatie voor collectiegegevens.** De standaardlocatie voor collectiegerelateerde bestanden bevindt zich op de indexserver. Als de collectie eenmaal is gemaakt, kunt u deze waarde niet meer wijzigen.
  - **Collectie-ID.** Het standaardcollectie-ID is gebaseerd op de collectienaam. Als de collectie eenmaal is gemaakt, kunt u deze waarde niet meer wijzigen. Als u een aangepast collectie-ID opgeeft, kunnen de zoekprogramma's de collectie aanroepen met dit ID, en dus niet met het meestal cryptische ID dat door het systeem wordt gegenereerd.
  - **N-gramsegmentering.** De standaardsegmenteringsmethode is witruimtesegmentering van Unicode. Selecteer de optie voor het gebruik van n-gramsegmentering alleen als uw collectie Chinese, Japanse of Koreaanse documenten bevat en de parser n-gramsegmentering moet gebruiken om woorden van elkaar te scheiden. Als de collectie eenmaal is gemaakt, kunt u deze waarde niet meer wijzigen.
4. Klik op **OK**.

In de view Collecties wordt de nieuwe collectie samen met de andere collecties in het enterprise search-systeem weergegeven. Als u beveiliging voor de collectie hebt ingeschakeld, wordt het pictogram  **Collectiebeveiliging is ingeschakeld** naast de collectienaam weergegeven.

De collectie is leeg totdat u er content aan toevoegt. Om content aan een nieuwe collectie toe te voegen, selecteert u de collectie in de view Collecties. Vervolgens kunt u de collectie bekijken, bewerken, minimaal één crawler maken, en opties opgeven voor de manier waarop de gegevens moeten worden geanalyseerd, geïndexeerd en doorzocht.

Vervolgens moet u de processen starten voor het crawlen, analyseren, indexeren en doorzoeken van de collectie. U kunt het voorbeeldzoekprogramma gebruiken om de nieuwe collectie te doorzoeken totdat uw aangepaste zoekprogramma's gereed zijn.

#### **Verwante onderwerpen**

"Enterprise search - Systeembeheer" op pagina 17

U kunt de beheerconsole van enterprise search gebruiken om collecties en externe bronnen te maken en beheren, componenten te starten en stoppen, systeemactiviteit en logbestanden te bewaken, beheerders te configureren, zoekprogramma's te koppelen aan collecties en externe bronnen en informatie op te geven om de beveiliging te verbeteren.

---

## **Een collectie bewerken**

U kunt collecties bewerken om informatie op te geven over de documenten die u in een collectie wilt opnemen.

### **Voordat u begint**

Om een collectie te bewerken, moet u de rol van beheerder van enterprise search hebben of beheerder van een collectie zijn.




## Over deze taak

Wanneer u een collectie bewerkt, geeft u opties op voor het crawlen van de gegevensbronnen, het analyseren van documenten, het opbouwen van de index, het doorzoeken van de geïndexeerde content en het loggen van foutberichten. Als u een collectie maakt, moet u deze bewerken om content aan de collectie te kunnen toevoegen. Later kunt u de collectie bewerken om de content bij te werken of de manier te wijzigen waarop de gegevens worden gecrawld, geanalyseerd, geïndexeerd, doorzocht of gelogd.

## Procedure

Ga als volgt te werk om een collectie te bewerken:

1. Klik op **Collecties** in de werkbalk om de view Collecties te openen.
2. Zoek de collectie die u wilt bewerken in de lijst van collecties en klik vervolgens op  **Bewerken**.
3. Breng wijzigingen aan op een van de volgende pagina's:

### Crawlen

Geef de gegevensbronnen op die u wilt crawlen en geef opties op voor de manier waarop de content moet worden gecrawld. Elke collectie moet ten minste één crawler bevatten en één collectie kan gegevens uit meerdere typen gegevensbronnen bevatten. Voor elk type gegevens dat u in uw collectie wilt opnemen, moet u minimaal één crawler configureren.

**Parser** Geef opties op voor de manier waarop gecrawlde documenten moeten worden geanalyseerd. U kunt categorieën configureren, zodat gebruikers subsets van de collectie kunnen doorzoeken, en u kunt regels configureren waarmee gebruikers bepaalde gedeelten van XML- en HTML-documenten kunnen doorzoeken. Als u aangepaste tekstanalyseprogramma's aan het enterprise search-systeem toevoegt, kunt u deze selecteren om de content in de collectie te analyseren en annoteren. U kunt wegingsklassen aan velden koppelen om de ranking van zoektermen in de zoekresultaten te beïnvloeden.

### Trefwoordenregister

Geef roosters op voor het opbouwen van de volledige index en het bijwerken van de index met nieuwe en gewijzigde content. U kunt ook opties configureren om het gebruik van jokertekens in query's in te schakelen, de weergave van de collectie te beperken tot een reeks URI's, de zoekresultaten van dezelfde website samen te vouwen en URI's uit de index te verwijderen.

### Zoeken

Geef opties op voor de zoekprocessen. Zo kunt u een zoekcache configureren en een zoektaal selecteren. Daarnaast kunt u snellinks configureren, om de weergave van vooraf bepaalde URI's in de zoekresultaten in te schakelen wanneer een gebruiker bepaalde woorden of woordcombinaties in een query opgeeft. Als u aangepaste woordenboeken aan het enterprise search-systeem hebt toegevoegd, kunt u de woordenboeken selecteren die u bij het doorzoeken van deze collectie wilt gebruiken.

### Loggen

Geef de typen berichten op die u wilt loggen en geef opties op voor het maken en overschrijven van logbestanden. U kunt alertsignalen configureren zodat u op de hoogte wordt gesteld als zich bepaalde gebeur-

tenissen voordoen. Daarnaast kunt u opties opgeven voor de ontvangst van e-mail wanneer bepaalde berichten of bepaalde typen berichten worden gelogd. U kunt opties selecteren voor het loggen van informatie over documenten die uit het enterprise search-systeem zijn gewist.

### **Algemeen**

Geef algemene informatie op over de collectie en bekijk instellingen die niet kunnen worden gewijzigd. Als bij het maken van de collectie de beveiliging is ingeschakeld, kunt u beveiligingsopties op document-niveau configureren.

---

## **Een collectie wissen**

Als u een collectie volledig wist, wordt alle informatie over de collectie van het enterprise search-systeem verwijderd.

### **Voordat u begint**

Om een collectie te wissen, moet u de rol van beheerder van enterprise search hebben.

U moet alle processen die aan de collectie zijn gekoppeld, stoppen voordat u de collectie wist.

### **Over deze taak**

Het wissen van een collectie kan veel tijd in beslag nemen. Nadat u hebt bevestigd dat u een collectie wilt wissen, wist het systeem alle collectiegegevens.

**Tip:** Mogelijk ziet u een bericht over een timeout van de aangevraagde bewerking hoewel het proces nog op de achtergrond wordt uitgevoerd. Om te bepalen of de taak is voltooid, klikt u op **Vernieuwen** in de beheerconsole (klik niet op **Vernieuwen** in de webbrowser). Het wissen is voltooid wanneer de naam van de collectie niet langer in de lijst van collecties verschijnt.

### **Procedure**

U kunt als volgt een collectie wissen:

1. Klik op **Collecties** om de view Collecties te openen.
2. Zoek in de lijst van collecties de collectie die u wilt wissen en klik op  **Wissen**.

---

## Beheer van enterprise search-crawler

U moet crawlers configureren voor de verschillende gegevenstypen die u in een collectie wilt opnemen. Een collectie kan een willekeurig aantal crawlers bevatten.

### Crawlers configureren

U kunt crawlers maken, bewerken en wissen met behulp van de beheerconsole van enterprise search. Meestal zal een deskundige op het gebied van het te crawlen gegevenstype de crawler configureren. Om bijvoorbeeld een crawler in te stellen voor het crawlen van Lotus Notes-gegevensbronnen, moet de collectiebeheerder een Notes-beheerder zijn of nauw samenwerken met iemand die verstand heeft van de te crawlen databases.

Als u een crawler maakt, wordt u bij het opgeven van de eigenschappen waarmee wordt bepaald hoe de crawler de systeemresources gebruikt, geholpen door een wizard die specifiek van toepassing is op het gegevenstype dat wordt gecrawld. Daarnaast helpt de wizard u bij het selecteren van de bronnen die u wilt doorzoeken.

U kunt op elk gewenst moment wijzigingen aanbrengen in bestaande crawlers. Indien gewenst kunt u de crawlereigenschappen of gedeelten van de crawlruimte bewerken. Ook voor deze taken kunt u de crawlerwizards gebruiken.

### Nieuwe crawlers voorzien van basiswaarden

U kunt een crawler maken door de standaardwaarden van het systeem te gebruiken of door de waarden te kopiëren die voor een andere crawler van hetzelfde type zijn opgegeven. Als u een bestaande crawler als basis voor een nieuwe crawler gebruikt, kunt u snel meerdere crawlers maken die over dezelfde eigenschappen beschikken en deze bijvoorbeeld configureren voor het crawlen van verschillende bronnen of de crawlers met behulp van verschillende roosters uitvoeren.

Door een crawler te kopiëren, kunt u de werkbelasting voor het crawlen verdelen tussen verschillende crawlers die dezelfde regels gebruiken. Zo kunt u een Notes-crawler kopiëren als u dezelfde eigenschappen en crawlregels voor velden op een andere Lotus Notes-server wilt gebruiken. De enige verschillen kunnen de databases zijn die door elke crawler worden gecrawld, en de beveiligingsinstellingen op documentniveau.

### Crawlertypen combineren in een collectie

Enterprise search-crawlers zijn ontworpen om informatie te verzamelen van specifieke typen gegevensbronnen. Bij het configureren van crawlers voor een collectie moet u bepalen hoe deze verschillende typen gegevensbronnen moeten worden gecombineerd, zodat gebruikers op een eenvoudige manier de enterprisegegevens kunnen doorzoeken. Als u wilt dat gebruikers met één enkele query Microsoft Windows-bestandssystemen en openbare mappen van Microsoft Exchange Server kunnen doorzoeken, maakt u een collectie met Windows-bestandssysteem-crawlers en Exchange Server-crawlers.

Wanneer u meerdere crawlertypen in één enkele collectie combineert, moet u ervoor zorgen dat alle crawlers dezelfde methode voor statische ranking kunnen

gebruiken. (Deze methode kunt u opgeven tijdens het maken van de collectie.) Als u bijvoorbeeld een combinatie gebruikt van webbronnen (die documentlinks als rankingfactor gebruiken) en NNTP-bronnen (die doorgaans de documentdatum als rankingfactor gebruiken), kan dit ten koste gaan van de kwaliteit van de zoekresultaten.

## Beveiliging op documentniveau inschakelen

Als u tijdens het maken van een collectie beveiliging voor de collectie inschakelt, kunt u opties voor de beveiliging op documentniveau opgeven. Elke crawler kan beveiligingstokens koppelen aan de documenten die ermee worden gecrawld. Als u tijdens het configureren van de crawler aangeeft dat u beveiliging op documentniveau wilt gebruiken, koppelt de crawler de beveiligingstokens die u bij elk document opgeeft, waarna deze tokens worden toegevoegd aan de index voor de documenten.

Als u beveiliging inschakelt in uw aangepaste zoekprogramma's, kunnen de programma's gebruikmaken van de beveiligingstokens die de crawlers aan de documenten koppelen voor de verificatie van gebruikers. Op deze manier kunt u de toegang tot bepaalde documenten in een collectie beperken en instellen dat andere documenten door alle gebruikers kunnen worden doorzocht. U kunt voor een collectie bijvoorbeeld instellen dat alle gebruikers toegang hebben tot de documenten in de openbare mappen van Microsoft Exchange Server, terwijl alleen de gebruikers met bepaalde gebruikers-ID's toegang hebben tot de documenten in de Lotus Notes-databases.

U kunt aangepaste bedrijfsregels toepassen om de waarde van de beveiligingstokens te bepalen door de regels in een Java-klasse te coderen. Wanneer u de crawlereigenschappen configureert, geeft u de naam op van de plugin die de crawler moet gebruiken bij het crawlen van documenten. De beveiligingstokens die door de plugin worden toegevoegd, worden in de index opgeslagen en kunnen worden gebruikt om de toegang tot documenten te besturen.

Wanneer u bepaalde typen crawlers configureert, kunt u extra beveiligingsopties instellen. Zo kunt u opgeven dat gebruikers tijdens de verwerking van query's moeten worden gevalideerd. Als u deze optie inschakelt, worden de legitimatiegegevens van gebruikers vergeleken met de actuele toegangsgegevens die worden bijgehouden door de gegevensbronnen die worden doorzocht. De validatie van actuele legitimatiegegevens kan worden uitgevoerd in plaats van of als aanvulling op de validatie die is gebaseerd op beveiligingstokens in de enterprise search-index.

### Verwante onderwerpen

*"Enterprise search - Systeembeheer"* op pagina 17

U kunt de beheerconsole van enterprise search gebruiken om collecties en externe bronnen te maken en beheren, componenten te starten en stoppen, systeemactiviteit en logbestanden te bewaken, beheerders te configureren, zoekprogramma's te koppelen aan collecties en externe bronnen en informatie op te geven om de beveiliging te verbeteren.

*"Beveiliging op documentniveau"* op pagina 237

Als bij het maken van de collectie de beveiliging is ingeschakeld, kunt u beveiligingsopties op documentniveau configureren. Als u beveiligingsopties op documentniveau configureert, zorgt u dat in de zoekresultaten alleen de documenten worden afgebeeld waarvoor de gebruiker die de zoekopdracht heeft ingediend, weergavemachtigingen heeft.

### Verwante taken

“Crawlers bewaken” op pagina 302

U kunt algemene gegevens bekijken over de status van elke crawler in een collectie of opties selecteren om gedetailleerde informatie te bekijken over de crawleractiviteiten.

### Verwante verwijzing

“Vereisten in de crawlerinstellingen voor de ondersteuning van beveiliging” op pagina 254

Om informatie te verzamelen waarmee beveiliging op documentniveau kan worden afgedwongen, moeten de crawlers gemachtigd zijn voor toegang tot de native beveiligingsgegevens. Voor bepaalde gegevenssoorten moeten aanvullende stappen worden uitgevoerd om een beveiligde omgeving te kunnen configureren.

---

## Een crawler maken

Tijdens het maken van een crawler geeft u het type crawler op dat u wilt maken. Een wizard helpt u bij het opgeven van informatie over de gegevens die u in de collectie wilt opnemen.

### Voordat u begint

Om een crawler te maken, moet u de rol van beheerder van enterprise search hebben of beheerder zijn van de collectie die eigenaar is van de crawler.

### Over deze taak

U moet ten minste één crawler voor een collectie maken. Het type crawler dat u maakt, is afhankelijk van het type gegevens dat u in de collectie wilt opnemen. Met behulp van een wizard die specifiek van toepassing is op het type crawler dat u maakt, kunt u de opties voor de crawler opgeven. De wizard helpt u bijvoorbeeld bij het opgeven van opties voor de manier waarop de systeemresources in de crawler moeten worden gebruikt. Daarnaast helpt de wizard u bij het selecteren van de gegevensbronnen die u in de collectie wilt opnemen.

### Procedure

Ga als volgt te werk om een crawler te maken:

1. Open een collectie, selecteer de pagina Crawler en klik op **Crawler maken**.
2. Selecteer het crawlertype en de basiswaarden voor de crawler:
  - a. Selecteer het type crawler dat ondersteuning biedt voor het type gegevens dat u wilt crawlen (bijvoorbeeld websites, Lotus Notes-databases of UNIX-bestandssystemen).

Nadat u een crawlertype hebt geselecteerd, worden er opties weergegeven voor de manier waarop u de crawler wilt maken.
  - b. Selecteer de basiswaarden voor de crawler:

#### **Standaardwaarden van het systeem gebruiken voor de nieuwe crawler**

Hiermee stelt u de crawler in met de standaardwaarden.

Als u deze optie selecteert, klikt u op **Volgende** om te beginnen met de configuratie van de nieuwe crawler.

### Waarden van een bestaande crawler klonen voor de nieuwe crawler

Hiermee neemt u de instellingen over die voor een andere crawler van dit type zijn opgegeven.

Als u deze optie selecteert, wordt een lijst weergegeven van alle crawlers die overeenkomen met het afgebeelde crawlertype. Selecteer de crawler die u voor de nieuwe crawler wilt gebruiken en klik op **Volgende** om te beginnen met de configuratie van de nieuwe crawler.

Er wordt een wizard geopend voor het type crawler dat u wilt maken. Volg de instructies in de wizard om de crawler te maken. Als u meer informatie wilt over de opties die u voor het type crawler kunt opgeven, klikt u op de verschillende pagina's in de wizard op **Help**.

De nieuwe crawler wordt weergegeven op de pagina Crawler naast de andere crawlers die tot dezelfde collectie behoren. U kunt op elk gewenst moment opties selecteren om de crawlereigenschappen en crawlruimte te bewerken om wijzigingen in de crawler aan te brengen.

---

## Crawlereigenschappen bewerken

U kunt informatie over de crawler wijzigen en de manier aanpassen waarop gegevens worden gecrawld. U kunt bijvoorbeeld de manier wijzigen waarop de crawler systeemresources gebruikt.

### Voordat u begint


Om de crawlereigenschappen te bewerken, moet u de rol van beheerder van enterprise search hebben of beheerder zijn van de collectie die eigenaar is van de crawler.

### Over deze taak

Als u meer informatie wilt over de typen wijzigingen die u kunt aanbrengen, klikt u tijdens het bewerken van de crawlereigenschappen op **Help**. Welke eigenschappen kunnen worden bewerkt, is afhankelijk van het crawlertype.

### Procedure

Ga als volgt te werk om de eigenschappen voor een crawler te bewerken:

1. Open een collectie en selecteer de pagina Crawler. Zoek de crawler die u wilt bewerken en klik op  **Crawlereigenschappen**.
2. Wijzig de crawlereigenschappen en klik vervolgens op **OK**.
3. Om de wijzigingen te activeren, moet u de crawler stoppen en opnieuw starten. (Als u alleen de crawlerbeschrijving wijzigt, hoeft u de crawler niet opnieuw te starten.)

---

## Een crawlruimte bewerken

U kunt informatie over de gegevensbronnen wijzigen die met de crawler worden gecrawld. U kunt bijvoorbeeld gegevensbronnen toevoegen, gegevensbronnen verwijderen, het crawlrooster wijzigen en de regels voor het crawlen van documenten in een bepaalde gegevensbron wijzigen.

### Voordat u begint


Om een crawlruimte te bewerken, moet u de rol van beheerder van enterprise search hebben of beheerder zijn van de collectie die eigenaar is van de crawler.

### Over deze taak

Als u meer informatie wilt over de wijzigingen die u kunt aanbrengen voor het type crawler dat u beheert, klikt u tijdens het bewerken van de crawlruimte op **Help**.

### Procedure

Ga als volgt te werk om een crawlruimte te bewerken:

1. Open een collectie en selecteer de pagina Crawler. Zoek de crawler die u wilt bewerken en klik op  **Crawlruimte**.
2. Wijzig de crawlruimte door de opties te selecteren die u wilt wijzigen.  
Welke opties kunnen worden gekozen, is afhankelijk van het crawlertype. Voor bepaalde opties, zoals het toevoegen van gegevensbronnen aan de collectie, verschijnt een wizard voor het crawlertype om u te helpen bij het wijzigen van de crawlruimte.
3. Om de wijzigingen te activeren, moet u de crawler stoppen en opnieuw starten.

---

## Een crawler wissen

Als u een crawler wist, wordt alle informatie over de crawler van het enterprise search-systeem verwijderd. Informatie die eerder door de crawler is gecrawld, blijft echter in de hoofdindex staan totdat deze opnieuw wordt opgebouwd.

### Voordat u begint

Om een crawler te wissen, moet u de rol van beheerder van enterprise search hebben of beheerder zijn van de collectie die eigenaar is van de crawler.


### Over deze taak

Het wissen van een crawler kan veel tijd in beslag nemen. Nadat u hebt bevestigd dat u een crawler wilt wissen, wist het systeem alle crawlergegevens.

**Tip:** Omdat deze taak enige tijd in beslag neemt, ziet u mogelijk een bericht over een timeout van de aangevraagde bewerking hoewel het proces nog op de achtergrond wordt uitgevoerd. Om te bepalen of de taak is voltooid, klikt u op **Vernieuwen** in de beheerconsole (klik niet op **Vernieuwen** in de webbrowser). Het wissen is voltooid wanneer de naam van de crawler niet langer in de lijst van crawlers verschijnt.

### Procedure

Ga als volgt te werk om een crawler te wissen:

1. Open een collectie en selecteer de pagina Crawler.
2. Zoek de crawler die u wilt wissen en klik op  **Wissen**.

---

## Crawlerroosters

Crawlers die u maakt voor webbronnen worden doorlopend uitgevoerd. Als u een webcrawler hebt gestart, hoeft u deze meestal niet te stoppen, tenzij u wijzigingen hebt aangebracht in de configuratie van de crawler. Voor alle andere crawlertypen geeft u een crawlingrooster op wanneer u de crawler configureert.

Voor bepaalde typen gegevensbronnen wordt op basis van een enkel rooster bepaald wanneer de crawler de gegevensbronnen in de crawlruimte bezoekt. Voor andere typen gegevensbronnen kunt u verschillende roosters voor specifieke gegevensbronnen opgeven. U kunt bijvoorbeeld verschillende roosters opgeven voor het crawlen van elke Lotus Notes-database die door de crawler wordt gecrawld.

Tijdens het configureren van het rooster geeft u het crawltype op dat moet worden uitgevoerd. U kunt een volledige crawl van alle documenten in de crawlruimte plannen, een crawl plannen waarmee alle wijzigingen in de crawlruimte worden opgenomen (zoals nieuwe documenten, gewijzigde documenten en gewiste documenten), of een crawl plannen waarmee alleen nieuwe en gewijzigde documenten worden opgenomen. Een volledige crawl neemt de meeste tijd in beslag. Een crawl waarmee gewiste documenten worden verwijderd, neemt meer tijd in beslag dan een crawl waarmee gewiste documenten worden genegeerd.

Bij het bewerken van de crawlruimte van een crawler kunt u een tweede rooster voor het crawlen opgeven. U kunt bijvoorbeeld een rooster configureren waarmee alle documenten in de crawlruimte elke zaterdagavond worden gecrawld, en een tweede rooster configureren die regelmatig wordt uitgevoerd voor het crawlen van nieuwe en gewijzigde documenten.

Als u meerdere roosters voor crawlers maakt, kunt u nauwkeuriger bepalen wanneer de crawler de doelbronnen bezoekt. Als u bijvoorbeeld databases in verschillende tijdzones wilt crawlen, kunt u de crawler plannen op tijden waarop de gebruikers klaar zijn met werken.

---

## Content Edition-crawlers

Als u IBM WebSphere Information Integrator Content Edition-repository's in een enterprise search-collectie wilt opnemen, moet u een Content Edition-crawler configureren.

U kunt met de Content Edition-crawler de volgende repositorytypen crawlen:

- Documentum, FileNet Panagon Content Services
- FileNet P8 Content Manager
- Hummingbird Document Management (DM)
- Microsoft SharePoint
- OpenText Livelink
- Portal Document Manager (PDM)

Bij het configureren van de crawler moet u opgeven hoe de crawler alle repository's in de crawlruimte crawlt. Daarnaast moet u de itemklassen selecteren die in elke repository moeten worden gecrawld.



## Configuratie van de crawlerserver

Hoe u de omgeving voorbereidt op het crawlen van repository's is afhankelijk van het feit of u de directe werkstand of de serverwerkstand wilt gebruiken wanneer u verbinding maakt met de te crawlen gegevens. Als u de directe werkstand wilt gebruiken, moet u een connector configureren in WebSphere Information Integrator Content Edition. Als u de serverwerkstand wilt gebruiken, moet u een script uitvoeren op de crawlerserver. Dit script, dat bij WebSphere Information Integrator OmniFind Edition wordt geleverd, stelt de Content Edition-crawler in staat om te communiceren met WebSphere Information Integrator Content Edition-servers.

Als u de serverwerkstand gebruikt, moet u voordat u een Content Edition-crawler maakt, de taak uitvoeren die van toepassing is op uw omgeving:

- "De crawlerserver configureren op UNIX voor WebSphere II Content Edition" op pagina 43.
- "De crawlerserver configureren op Windows voor WebSphere II Content Edition" op pagina 44.

Voor uitgebreide instructies bij het configureren van het enterprise search-systeem voor het doorzoeken van WebSphere Information Integrator Content Edition-repository's raadpleegt u het IBM developerWorks-artikel *Search WebSphere Portal Document Manager using WebSphere Information Integrator OmniFind Edition* op <http://www-128.ibm.com/developerworks/db2/library/techarticle/dm-0606lee/>.

## Configuratieoverzicht

Als u een Content Edition-crawler wilt maken of wijzigen, meldt u zich aan bij de beheerconsole van enterprise search. U moet de rol van beheerder van enterprise search hebben of beheerder zijn van de collectie die eigenaar is van de crawler.

Bij het definiëren van de crawler moet u de volgende procedure uitvoeren. Hierbij wordt u geholpen door de wizard:

- Geef eigenschappen op die bepalen hoe de crawler werkt en op welke manier de systeemresources worden gebruikt. De crawlereigenschappen bepalen hoe de crawler alle repository's in de crawlruimte crawlt.
- Geef aan of de crawler gebruikmaakt van de directe werkstand of de serverwerkstand om toegang te krijgen tot de repository's. Voor de serverwerkstand moet u ook informatie opgeven aan de hand waarvan de crawler toegang kan krijgen tot de webtoepassingenserver.
- Selecteer de repository's die u wilt crawlen.
- Geef de gebruikers-ID's en wachtwoorden op waarmee de crawler toegang kan krijgen tot de geselecteerde repository's.
- Geef een rooster op voor het crawlen van de repository's.
- Selecteer de itemklassen die u in elke repository wilt crawlen.
- Geef opties op waarmee de eigenschappen van de itemklassen doorzoekbaar worden gemaakt. U kunt bijvoorbeeld bepaalde typen documenten uitsluiten van de crawlruimte of opgeven dat u een bepaalde versie van een repository wilt crawlen.
- Configureer beveiligingsopties op documentniveau. Als bij het maken van de collectie de beveiliging is ingeschakeld, kan de crawler beveiligingsgegevens koppelen aan documenten in de index. Met behulp van deze gegevens kunnen zoekprogramma's toegangsbesturing uitvoeren op basis van de toegangslijsten of beveiligingstokens.

Voor itemklassen van het type Documentum, FileNet Panagon Content Services, Hummingbird DM, Portal Document Manager en SharePoint kunt u ook een optie selecteren waarmee de legitimatiegegevens van gebruikers worden gecontroleerd wanneer een query wordt ingediend. Op deze manier worden de legitimatiegegevens vergeleken met de huidige toegangslijsten die door de oorspronkelijke gegevensbron worden beheerd en niet met de geïndexeerde beveiligingsgegevens. Een dergelijke controle is niet mogelijk voor de overige typen repository's.

Als u bij het definiëren van de crawler meer informatie wenst over de velden in de wizard en over de manier waarop u de informatie moet opgeven die de crawler nodig heeft om informatie te crawlen, klikt u in de wizard op **Help**.

## Directe werkstand-toegang tot WebSphere II Content Edition-repository's

U kunt de Content Edition-crawler zodanig configureren, dat u via de directe werkstand toegang tot de repository's kunt krijgen.

### Over deze taak

In de directe werkstand gebruikt de crawler een WebSphere Information Integrator Content Edition-connector die tijdens de installatie van WebSphere II OmniFind Edition op de crawlerserver is geïnstalleerd. De crawler gebruikt contentintegratie-API's om direct een verbinding te maken met de repository's die moeten worden gecrawld. Niet alle functies voor contentintegratieservers zijn beschikbaar wanneer de contentintegratieserver in de directe werkstand wordt uitgevoerd. Raadpleeg de documentatie bij WebSphere Information Integrator Content Edition voor informatie over het uitvoeren van de contentintegratieserver in de directe werkstand en over de manier waarop de functies verschillen van contentintegratieservers die in de serverwerkstand worden uitgevoerd.

In deze procedure vindt u de stappen die moeten worden uitgevoerd om de toegang via de directe werkstand in te stellen. Voor uitgebreide instructies raadpleegt u het IBM developerWorks-artikel *Search WebSphere Portal Document Manager using WebSphere Information Integrator OmniFind Edition* op <http://www-128.ibm.com/developerworks/db2/library/techarticle/dm-0606lee/>.

### Procedure

Ga als volgt te werk om het systeem zodanig te configureren, dat de crawler toegang tot de repository's kan krijgen via de directe werkstand:

1. Controleer of de omgevingsvariabelen VBR\_HOME en JAVA\_HOME in het bestand *iice\_install\_root/bin/config.sh* (onder UNIX) of in het bestand *iice\_install\_root\bin\config.bat* (onder Microsoft Windows) de juiste directory aangeven.
2. Als u de beheerconsole van WebSphere Information Integrator Content Edition wilt configureren om te worden uitgevoerd in de directe werkstand, voegt u de Java-systeemeigenschap **vbr.as.operationMode=direct** toe aan het bestand *iice\_install\_root/bin/Admin.bat* (onder UNIX) of het bestand *iice\_install\_root\bin\Admin.bat* (onder Windows).
3. Start de beheerconsole van WebSphere Information Integrator Content Edition in de directe werkstand en configureer de connector voor de WebSphere II OmniFind Edition-crawlerserver.
4. Selecteer de opties voor de directe werkstand wanneer u de beheerconsole van enterprise search gebruikt voor de configuratie van de Content Edition-crawler.

## Serverwerkstand-toegang tot WebSphere II Content Edition-repository's

U kunt de Content Edition-crawler zodanig configureren, dat u via de serverwerkstand toegang tot de repository's kunt krijgen.

In de serverwerkstand is de WebSphere Information Integrator Content Edition-connector die de crawler gebruikt om toegang tot de gegevens te krijgen, als een toepassingenserver in WebSphere Application Server geïnstalleerd en krijgt de crawler toegang tot de repository's via de server. Op deze manier kunt u optimaal profiteren van de J2EE-toepassingenserveromgeving.

Voordat u de crawler configureert voor toegang tot de WebSphere Information Integrator Content Edition-repository's via de serverwerkstand, moet u een script uitvoeren op de crawlerserver. Met dit script, dat wordt geleverd bij WebSphere Information Integrator OmniFind Edition, kan de Content Edition-crawler toegang krijgen tot de repository's op de server.

Voer de taak uit die van toepassing is op uw omgeving:

- "De crawlerserver configureren op UNIX voor WebSphere II Content Edition".
- "De crawlerserver configureren op Windows voor WebSphere II Content Edition" op pagina 44.

### De crawlerserver configureren op UNIX voor WebSphere II Content Edition

Als u WebSphere II OmniFind Edition installeert op een computer met IBM AIX, Linux of de Solaris-besturingsomgeving en u de Content Edition-crawler configureert voor gebruik van de serverwerkstand tijdens het benaderen van repository's, moet u een script uitvoeren om de crawlerserver te configureren. Met het script wordt de Content Edition-crawler ingeschakeld voor het benaderen van WebSphere Information Integrator Content Edition-repository's.

#### Over deze taak

De Content Edition-crawler gebruikt Java-bibliotheken van WebSphere Information Integrator Content Edition als Java-client. In de serverwerkstand zijn voor deze Java-bibliotheken EJB-gerelateerde Java-bibliotheken van WebSphere Application Server vereist. Om ervoor te zorgen dat de Content Edition-crawler kan samenwerken met de Java-bibliotheken, moet u een installatiescript uitvoeren waarmee WebSphere II OmniFind Edition op de crawlerserver wordt geïnstalleerd nadat u WebSphere Application Server hebt geïnstalleerd.

WebSphere Information Integrator Content Edition wordt gelijktijdig met WebSphere II OmniFind Edition op de crawlerserver geïnstalleerd. Als u de Content Edition-crawler in de serverwerkstand wilt gebruiken, moet u het bestand `vbr_access_services.jar` vanaf de WebSphere Information Integrator Content Edition-server kopiëren naar de crawlerserver.

#### Procedure

Ga als volgt te werk om de crawlerserver zodanig te configureren, dat hiermee WebSphere Information Integrator Content Edition-repository's kunnen worden gecrawld:

1. Als WebSphere II OmniFind Edition is geïnstalleerd in een configuratie met meerdere servers, moet u de WebSphere Application Server Java-bibliotheken installeren en een bind uitvoeren op de bibliotheken.
2. Voer op de crawlerserver het installatiescript uit voor de Content Edition-crawler:
  - a. Meld u aan als enterprise search-beheerder.
  - b. Start het volgende script (geïnstalleerd in de directory `$ES_INSTALL_ROOT/bin`) en volg de instructies:  
`escrvbr.sh`
3. Ga naar de crawlerserver, stop het enterprise search-systeem en start dit vervolgens opnieuw:

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

4. Kopieer het bestand `vbr_access_services.jar` vanaf de WebSphere Information Integrator Content Edition-server naar de crawlerserver.

**Bronlocatie voor kopiëren:**

Het bestand `vbr_access_services.jar` bevindt zich op de volgende standaardlocatie:

*was\_installatie\_root*/installedApps/*server\_naam*/*programma\_naam*

*was\_installatie\_root* is de installatiedirectory van WebSphere Application Server, *server\_naam* is de naam die u voor de server hebt opgegeven en *programma\_naam* is de naam die u hebt opgegeven voor het WebSphere Information Integrator Content Edition-programma in WebSphere Application Server.

**Doellocatie voor kopiëren:**

De doeldirectory op de crawlerserver is *iice\_installatie\_root*/lib, waarbij *iice\_installatie\_root* de installatiedirectory van WebSphere Information Integrator Content Edition is op de crawlerserver.

## De crawlerserver configureren op Windows voor WebSphere II Content Edition

Als u WebSphere II OmniFind Edition installeert op een Microsoft Windows-computer en u de Content Edition-crawler configureert voor de serverwerkstand tijdens het benaderen van repository's, moet u een script uitvoeren om de crawlerserver te configureren. Met het script wordt de Content Edition-crawler ingeschakeld voor het benaderen van WebSphere Information Integrator Content Edition-repository's.

### Over deze taak

De Content Edition-crawler gebruikt Java-bibliotheken van WebSphere Information Integrator Content Edition als Java-client. In de serverwerkstand zijn voor deze Java-bibliotheken EJB-gerelateerde Java-bibliotheken van WebSphere Application Server vereist. Om ervoor te zorgen dat de Content Edition-crawler kan samenwerken met de Java-bibliotheken, moet u een installatiescript uitvoeren waarmee WebSphere II OmniFind Edition op de crawlerserver wordt geïnstalleerd nadat u WebSphere Application Server hebt geïnstalleerd.

WebSphere Information Integrator Content Edition wordt gelijktijdig met WebSphere II OmniFind Edition op de crawlerserver geïnstalleerd. Als u de Content Edition-crawler in de serverwerkstand wilt gebruiken, moet u het bestand

vbr\_access\_services.jar vanaf de WebSphere Information Integrator Content Edition-server kopiëren naar de crawlerserver.

### Procedure

Ga als volgt te werk om de crawlerserver zodanig te configureren, dat hiermee WebSphere Information Integrator Content Edition-repository's kunnen worden gecrawld:

1. Als WebSphere II OmniFind Edition is geïnstalleerd in een configuratie met meerdere servers, moet u de WebSphere Application Server Java-bibliotheken installeren en een bind uitvoeren op de bibliotheken.
2. Voer op de crawlerserver het installatiescript uit voor de Content Edition-crawler:
  - a. Meld u aan met het enterprise search-beheerders-ID (dit gebruikers-ID is opgegeven tijdens de installatie van WebSphere II OmniFind Edition).
  - b. Start het volgende script (geïnstalleerd in de directory %ES\_INSTALL\_ROOT%\bin) en volg de instructies:  
escrvbr.vbs
3. Ga naar de crawlerserver, stop het enterprise search-systeem en start dit vervolgens opnieuw:

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

4. Kopieer het bestand vbr\_access\_services.jar vanaf de WebSphere Information Integrator Content Edition-server naar de crawlerserver.

#### Bronlocatie voor kopiëren:

Het bestand vbr\_access\_services.jar bevindt zich op de volgende standaardlocatie:

```
was_installatie_root\installedApps\server_naam\programma_naam
```

*was\_installatie\_root* is de installatiedirectory van WebSphere Application Server, *server\_naam* is de naam die u voor de server hebt opgegeven en *programma\_naam* is de naam die u hebt opgegeven voor het WebSphere Information Integrator Content Edition-programma in WebSphere Application Server.

#### Doellocatie voor kopiëren:

De doeldirectory op de crawlerserver is *iice\_installatie\_root*\lib, waarbij *iice\_installatie\_root* de installatiedirectory van WebSphere Information Integrator Content Edition is op de crawlerserver.

---

## DB2-crawlers

U gebruikt de DB2-crawler om IBM DB2 Universal Database-databases (DB2 UDB) in een collectie op te nemen.

Als u IBM WebSphere Information Integrator gebruikt om roepnaamtabellen voor de volgende typen databasesystemen te maken en gebruiken, kunt u de DB2-crawler gebruiken om de de tabellen te crawlen op basis van de roepnamen:

- CA-Datacom
- IBM DB2 UDB for z/OS
- DB2 UDB for iSeries
- IBM Informix
- IMS

- Oracle
- Microsoft SQL Server
- Software AG Adabas
- Sybase
- VSAM

Voor elke databaseserver die u wilt crawlen, moet u een aparte crawler configureren. Bij het configureren van de crawler moet u opgeven hoe de crawler alle databases op dezelfde server moet crawlen. Daarnaast moet u de tabellen selecteren die in elke database moeten worden gecrawld.

De tabellen die u voor de crawl selecteert, moeten databasetabellen, roepnaamtabellen of views zijn. De DB2-crawler ondersteunt geen samengevoegde tabellen.

## Configuratie van de crawlerserver

U kunt de databasetabellen pas crawlen als DB2 Administration Client op de crawlerserver is geïnstalleerd. Vervolgens moet u een script op de crawlerserver uitvoeren. Dit script, dat bij WebSphere Information Integrator OmniFind Edition wordt geleverd, zorgt voor de communicatie tussen de DB2-crawler en de databaseservers.

Voordat u met behulp van de beheerconsole van enterprise search een DB2-crawler configureert, voert u de taak uit die van toepassing is op uw omgeving:

- “De crawlerserver configureren op UNIX voor DB2-crawlers” op pagina 47.
- “De crawlerserver configureren op Windows voor DB2-crawlers” op pagina 48.

## Event Publishing

Als u gebruikmaakt van WebSphere Information Integrator Event Publisher Edition en u de te crawlen databases koppelt aan wachtrijtoewijzingen voor Event Publishing, kan de DB2-crawler de toewijzingen gebruiken om de crawlruimte bij te werken wanneer er wijzigingen in de databasetabellen worden aangebracht.

Een wachtrijtoewijzing van Event Publishing benoemt een WebSphere MQ-wachtrij die XML-berichten ontvangt wanneer wijzigingen in een databasetabel worden gepubliceerd. De crawler luistert naar de wachtrij voor informatie over tabelwijzigingen en verwerkt deze wijzigingen in de crawlruimte. (Wanneer een crawler een tabel voor het eerst crawlt, worden alle documenten gecrawld.)

In Event Publishing kunnen nieuwe en gewijzigde documenten sneller worden doorzocht dan de documenten die met crawler worden gecrawld op basis van het crawlerrooster.

Als bepaalde of alle tabellen zijn geconfigureerd voor het gebruik van Event Publishing, kunt u gegevens opgeven die de crawler tijdens het configureren toegang geven tot WebSphere MQ en de wachtrijtoewijzingen van Event Publishing.

Daarnaast moet u ervoor zorgen dat WebSphere MQ en WebSphere Information Integrator Event Publisher Edition zijn geconfigureerd op de server die moet worden gecrawld en dat de WebSphere MQ-clientmodule op de crawlerserver is geconfigureerd. Voer de volgende procedure uit als u Event Publishing wilt gebruiken in combinatie met een DB2-crawler:

- “WebSphere MQ configureren voor DB2-crawlers” op pagina 52.

- “WebSphere Information Integrator Event Publisher Edition configureren voor DB2-crawlers” op pagina 49.

## Configuratieoverzicht

Als u een DB2-crawler wilt maken of wijzigen, meldt u zich aan bij de beheerconsole van enterprise search. U moet de rol van beheerder van enterprise search hebben of beheerder zijn van de collectie die eigenaar is van de crawler.

Bij het definiëren van de crawler moet u de volgende procedure uitvoeren. Hierbij wordt u geholpen door de wizard:

- Geef eigenschappen op die bepalen hoe de crawler werkt en op welke manier de systeemresources worden gebruikt. De eigenschappen bepalen hoe de crawler alle databases op een bepaalde databaseserver crawlt.

- Geef informatie op over de databasetypen die u wilt crawlen.

Als u van plant bent de databases op afstand te crawlen die niet zijn opgenomen in de catalogus op de lokale databaseserver, moet u DB2 Administration Server op de server op afstand starten voordat u de DB2-crawler kunt gebruiken om deze databases te crawlen. Daarnaast moet u de hostnaam en de poort van de databaseserver op afstand opgeven wanneer u de crawler configureert.

- Selecteer de databases die u wilt crawlen.
- Geef de gebruikers-ID's en wachtwoorden op waarmee de crawler toegang kan krijgen tot de toegangsbesturing.
- Geef een rooster op voor het crawlen van de databases.
- Selecteer de tabellen die in elke database moeten worden gecrawld.

**Waarschuwing:** Omwille van de prestaties (en om te voorkomen dat er een time-out optreedt bij het configureren van de crawler) dient u alléén te kiezen voor het crawlen van alle tabellen als de database niet veel tabellen bevat of als de tabellen niet veel kolommen bevatten. Als u nu een bepaalde hoeveelheid te crawlen tabellen selecteert, kunt u de crawlruimte later bewerken en meer tabellen aan de collectie toevoegen.

- Selecteer de tabellen die moeten worden gecrawld wanneer wijzigingen in de tabellen naar een wachtrij voor Event Publishing worden gepubliceerd en geef informatie op waarmee de crawler toegang kan krijgen tot de wachtrij voor Event Publishing.
- Geef opties op waarmee de kolommen in bepaalde tabellen doorzoekbaar worden gemaakt. U kunt bijvoorbeeld instellen dat bepaalde kolommen in zoekopdrachten met parameters kunnen worden opgenomen en aangeven welke kolommen in de zoekresultaten kunnen worden afgebeeld.
- Configureer beveiligingsopties op documentniveau. Als bij het maken van de collectie de beveiliging is ingeschakeld, kan de crawler beveiligingsgegevens koppelen aan documenten in de index. Met behulp van die beveiligingsgegevens kunnen zoekprogramma's toegangsbesturing uitvoeren op basis van opgeslagen toegangslijsten of beveiligingstokens.

Als u bij het definiëren van de crawler meer informatie wenst over de velden in de wizard en over de manier waarop u de informatie moet opgeven die de crawler nodig heeft om informatie te crawlen, klikt u in de wizard op **Help**.

## De crawlerserver configureren op UNIX voor DB2-crawlers

Als u WebSphere II OmniFind Edition installeert op een computer waarop IBM AIX, Linux of de Solaris-besturingsomgeving wordt uitgevoerd, moet u een script uitvoeren om de crawlerserver te configureren. Met het script wordt de DB2-

crawler ingeschakeld voor communicatie met de databaseservers. Als u gebruikmaakt van Event Publishing, wordt met het script toegang van de crawler tot de WebSphere MQ-wachtrijmanagers en -wachtrijen ingesteld.

### Over deze taak

Om ervoor te zorgen dat de DB2-crawler de databasetabellen kan crawlen, moet u een installatiescript uitvoeren (escrdb2.sh) waarmee WebSphere II OmniFind Edition op de crawlerserver wordt geïnstalleerd.

Voordat u dit script uitvoert, moet u controleren of DB2 Administration Client op de crawlerserver is geïnstalleerd.

Als u werkt met Event Publishing, moet u de WebSphere MQ 5.3-modules voor Java Messaging op de crawlerserver installeren, zodat de DB2-crawler toegang kan krijgen tot de WebSphere MQ-wachtrijmanagers en -wachtrijen. Nadat u de WebSphere MQ-modules hebt geïnstalleerd, moet u het installatiescript escrdb2.sh uitvoeren.

### Procedure

Ga als volgt te werk om de crawlerserver te configureren zodat gegevens kunnen worden gecrawld door DB2-crawlers:

1. Optioneel: als u van plan bent Event Publishing te gebruiken, installeert u de WebSphere MQ 5.3-modules voor Java Messaging op de crawlerserver:
  - a. Meld u aan als de hoofdgebruiker en geef de volgende opdracht op:

```
export LD_ASSUME_KERNEL=2.4.19
```
  - b. Plaats de cd van WebSphere MQ.
  - c. Ga naar de directory waarin de MQ-modules voor Java Messaging zijn opgeslagen.
  - d. Geef de volgende opdracht op om de modules te installeren:

```
rpm -i MQSeriesJava-5.3.0-1.i386.rpm
```
2. Voer op de crawlerserver het installatiescript voor de DB2-crawler uit:
  - a. Meld u aan als de enterprise search-beheerder (dit gebruikers-ID is opgegeven tijdens de installatie van WebSphere II OmniFind Edition).
  - b. Start het volgende script (geïnstalleerd in de directory \$ES\_INSTALL\_ROOT/bin directory) en volg de instructies:

```
escrdb2.sh
```
3. Ga naar de crawlerserver, stop het enterprise search-systeem en start dit vervolgens opnieuw:

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

## De crawlerserver configureren op Windows voor DB2-crawlers

Als u WebSphere II OmniFind Edition installeert op een Microsoft Windows-computer, moet u een script uitvoeren om de crawlerserver te configureren. Met het script wordt de DB2-crawler ingeschakeld voor communicatie met de databaseservers. Als u gebruikmaakt van Event Publishing, wordt met het script toegang van de crawler tot de WebSphere MQ-wachtrijmanagers en -wachtrijen ingesteld.

### Over deze taak



Om ervoor te zorgen dat de DB2-crawler databasetabellen kan crawlen, moet u een installatiescript uitvoeren (escrdb2.vbs) waarmee WebSphere II OmniFind Edition op de crawlerserver wordt geïnstalleerd.

Voordat u dit script uitvoert, moet u controleren of DB2 Administration Client op de crawlerserver is geïnstalleerd.

Als u werkt met Event Publishing, moet u de WebSphere MQ 5.3-modules voor Java Messaging op de crawlerserver installeren, zodat de DB2-crawler toegang kan krijgen tot de WebSphere MQ-wachtrijmanagers en -wachtrijen. Nadat u de WebSphere MQ-modules hebt geïnstalleerd, moet u het installatiescript escrdb2.vbs uitvoeren.

### Procedure

Ga als volgt te werk om de crawlerserver te configureren zodat gegevens kunnen worden gecrawld door DB2-crawlers:

1. Optioneel: als u van plan bent Event Publishing te gebruiken, installeert u de WebSphere MQ 5.3-modules voor Java Messaging op de crawlerserver:
  - a. Plaats de cd van WebSphere MQ.
  - b. Start het installatieprogramma van WebSphere MQ.
  - c. Selecteer in het venster Choose Product Features als installatieoptie **Java Messaging**.
2. Voer op de crawlerserver het installatiescript voor de DB2-crawler uit:
  - a. Meld u aan met het enterprise search-beheerders-ID (dit gebruikers-ID is opgegeven tijdens de installatie van WebSphere II OmniFind Edition).
  - b. Start het volgende script (geïnstalleerd in de directory %ES\_INSTALL\_ROOT%\bin) en volg de instructies:  
escrdb2.vbs
3. Ga naar de crawlerserver, stop het enterprise search-systeem en start dit vervolgens opnieuw:

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

## WebSphere Information Integrator Event Publisher Edition configureren voor DB2-crawlers

Voordat u een DB2-crawler configureert voor het gebruik van Event Publishing, moet u zorgen dat IBM WebSphere Information Integrator Event Publisher Edition is geconfigureerd op de server die moet worden gecrawld.

### Over deze taak

Volg de onderstaande richtlijnen wanneer u WebSphere Information Integrator Event Publisher Edition configureert voor gebruik van de DB2-crawler:

- Zowel de gewijzigde als ongewijzigde kolommen in de brontabellen moeten worden geselecteerd voor publicatie.
- Gewiste rijen in de brontabellen moeten worden geselecteerd voor publicatie.
- Een wachtrij voor Event Publishing kan niet gemeenschappelijk worden gebruikt tussen meerdere databases.
- Een enkele database kan over meerdere wachtrijtoewijzingen en wachtrijen beschikken.

- In een tabel moet slechts één XML-publicatie aan een publicatiewachtrij-toewijzing zijn gekoppeld. (In een tabel mag niet meer dan één XML-publicatie aan één publicatiewachtrijtoewijzing zijn gekoppeld. In een tabel mag meer dan één XML-publicatie aanwezig zijn als elke XML-publicatie aan een andere publicatiewachtrijtoewijzing is gekoppeld.)

## Procedure

Voer de volgende stappen uit om een databaseserver te configureren zodat de DB2-crawler toegang kan krijgen tot de tabelupdates die naar een Event Publishing-wachtrij zijn gepubliceerd. (Raadpleeg de documentatie bij WebSphere Information Integrator Publisher Edition voor hulp bij deze stappen.)

1. Installeer WebSphere Information Integrator Event Publisher Edition op de databaseserver die u wilt crawlen.
2. Start het Replication Center Launchpad:
  - AIX, Linux of Solaris**  
db2rc
  - Windows-opdrachtaanwijzing**  
Klik op **Start** → **IBM DB2 Replication Center**.
3. Maak Q Capture-controletabellen:
  - a. Selecteer **Event Publishing** als view in het startvenster, selecteer **Create Q Capture Control Tables** en klik vervolgens op **Next**.
  - b. Selecteer in de lijst van beschikbare databaseservers in het veld **Q Capture server** de server die u als Q Capture-server wilt gebruiken en klik op **OK**.
  - c. Geef een gebruikers-ID en wachtwoord op met toegangsmachtiging tot de geselecteerde Q Capture-server. Wijzig het Q Capture-schema of accepteer de standaardschemanaam en klik op **Next**.
  - d. Geef de namen op van de wachtrijmanager en beheerwachtrij en start de wachtrij opnieuw die u hebt opgegeven tijdens het configureren van WebSphere MQ op deze databaseserver. Klik vervolgens op **Next**.
  - e. Klik op **Finish**. Klik nadat de pagina met berichten en SQL-scripts is weergegeven op **Close**.
  - f. Selecteer als verwerkingsoptie **Run now** en klik op **OK**. Als het bericht is weergegeven waarin wordt aangegeven dat de SQL-scripts zijn uitgevoerd, klikt u op **Close**.
4. Maak een XML-publicatie:
  - a. Selecteer in Replication Center Launchpad de optie **Event publishing** als view in het startvenster, selecteer **Create an XML Publication** en klik op **Next**.
  - b. Klik op de pagina Start op **Next**.
  - c. Bevestig op de pagina Server and Queue Map dat de Q Capture-server en het Q Capture-schema juist zijn, klik op de optie naast het veld **Publishing queue map** en klik op **New** om een publicatiewachtrijtoewijzing te maken.
  - d. Voer een naam in voor de wachtrijtoewijzing op de pagina General.
  - e. Geef op de pagina Properties de naam op van de verzendwachtrij (bijvoorbeeld de naam van de gegevenswachtrij die u hebt opgegeven tijdens het configureren van WebSphere MQ op deze server), selecteer **Row operation** of **Transaction** voor het type berichtinhoud, schakel de aankruisvakjes voor het verzenden van heartbeatberichten en het toevoegen van JMS-berichtheaders uit en klik vervolgens op **OK**.
  - f. Klik nadat de pagina met berichten en SQL-scripts is weergegeven op **Close**.

- g. Selecteer als verwerkingsoptie **Run now** en klik op **OK**. Als het bericht is weergegeven waarin wordt aangegeven dat de SQL-scripts zijn uitgevoerd, klikt u op **Close**.
  - h. Selecteer op de pagina Select Publishing Queue Map de wachtrijtoewijzing die u hebt gemaakt en klik op **OK**.
  - i. Bevestig op de pagina Server and Queue Map dat de naam van de wachtrijtoewijzing juist is en klik op **Next**.
  - j. Klik op de pagina Source Table op **Add**, klik op **Retrieve All** en selecteer de tabel die u wilt inschakelen voor Event Publishing. Klik op **OK** en vervolgens op **Next**.
  - k. Selecteer op de pagina Columns and Rows de kolommen die door de DB2-crawler moeten worden gecrawld (of selecteer alle kolommen) en selecteer de sorteerkolommen. Selecteer op de pagina waarop u de te crawlen rijen (of alle rijen) hebt geselecteerd, de optie voor het publiceren van brontabelverwijderingen. Als u deze opties hebt geconfigureerd, klikt u op **Next**.
  - l. Selecteer op de pagina Message Content de optie waarmee zowel gewijzigde als ongewijzigde kolommen in de kolomgegevens worden opgenomen en selecteer de optie voor alleen nieuwe gegevenswaarden. Zorg dat het aankruisvakje voor het automatisch starten van XML-publicaties is ingeschakeld en klik op **Next**.
  - m. Klik op de pagina Review and complete XML publications op **Next**.
  - n. Klik op de pagina Summary op **Finish**. Klik nadat de pagina met berichten en SQL-scripts is weergegeven op **Close**.
  - o. Selecteer als verwerkingsoptie **Run now** en klik op **OK**. Als het bericht is weergegeven waarin wordt aangegeven dat de SQL-scripts zijn uitgevoerd, klikt u op **Close**.
5. Start de Q Capture-server:
- a. Sluit het Replication Center Launchpad en start het Replication Center.
  - b. Klik in de objectstructuur op **Q Replication** → **Definitions** → **Q Capture Servers**.
  - c. Klik met de rechtermuisknop op het pictogram voor de Q Capture-server die u hebt geconfigureerd en selecteer **Enable Database for Q Replication**.
  - d. Klik nadat het waarschuwingsbericht is weergegeven op **OK**.
  - e. Nadat de pagina met DB2-berichten is weergegeven, klikt u op **Close**.
  - f. Klik in de objectstructuur met de rechtermuisknop op het pictogram voor de Q Capture-server en selecteer **Start Q Capture program**.
  - g. Selecteer voor de verwerkingsoptie de waarde **Run now**, geef de systeemnaam op, voer het gebruikers-ID en het wachtwoord voor de DB2-gebruiker in, geef het pad op voor de directory waarin de logbestanden zijn opgeslagen en voer de DB2-instancenaam in. Klik vervolgens op **OK**.
  - h. Als het bericht is weergegeven waarin wordt aangegeven dat de opdracht is verzonden, klikt u op **Close**.
  - i. Klik in de objectstructuur met de rechtermuisknop op het pictogram voor de Q Capture-server en selecteer **Check status**.
- De Q Capture-serverstatus wordt afgebeeld. Als er fouten optreden, verschijnt er een statusbericht waarin wordt aangegeven dat wordt aangenomen dat de server down is. Om de logbestanden te bekijken en de oorzaak van de fouten vast te stellen, geeft u de volgende opdracht op een opdrachtregel op:
- ```
asnqcap Capture_Server=capture server name LOGSTDOUT=y
```

## WebSphere MQ configureren voor DB2-crawlers

Voordat u een DB2-crawler configureert voor gebruik van Event Publishing, moet u ervoor zorgen dat IBM WebSphere MQ is geconfigureerd op de server waarnaar de crawler luistert.

### Voordat u begint

Zorg dat DB2 UDB, WebSphere Information Integrator Event Publisher Edition en WebSphere MQ op de doeldatabaseserver zijn geïnstalleerd.

### Beperkingen

Als de doeldatabaseserver is geïnstalleerd op een Linux-computer, moeten alle gebruikers van DB2 Universal Database (DB2 UDB), WebSphere MQ en WebSphere II OmniFind Edition de volgende omgevingsvariabele instellen:

```
export LD_ASSUME_KERNEL=2.4.19
```

Met deze omgevingsvariabele kunnen de LinuxThread-threadingimplementaties worden geëxporteerd naar elke shell waar de installatie is uitgevoerd, WebSphere MQ-besturingsopdrachten zijn opgegeven of WebSphere MQ-programma's worden uitgevoerd. Voor WebSphere MQ moet deze omgevingsvariabele worden geëxporteerd.

DB2-crawlers die gebruikmaken van Event Publishing maken verbinding met WebSphere MQ-wachtrijen via een clientverbinding. Om clientverbindingen mogelijk te maken, meldt u zich aan als WebSphere MQ-beheerder en voert u de volgende opdracht uit om het CCSID van de wachtrijmanager in te stellen op 819:

```
runmqsc wachtrij_manager_naam  
ALTER QMGR CCSID(819)  
END
```

### Over deze taak

De DB2-crawler ondersteunt de werkstand voor clientverbindingen met de WebSphere MQ-server. De crawler luistert naar XML-berichten die naar een Event Publishing-wachtrij zijn gepubliceerd. De crawler kan niet luisteren naar XML-berichten die via meerdere wachtrijen zijn overgebracht.

Nadat u WebSphere MQ hebt geconfigureerd, gebruikt de DB2-crawler de naam van de wachtrijmanager, de wachtrijnaam, de naam van de serverhost, het poortnummer van de server en de naam van het serverkanaal om de XML-berichten op te halen uit een publicatiewachtrij. De crawler analyseert de berichten en werkt de crawlruimte bij met gegevens over de bijgewerkte tabellen.

### Procedure

Voer de volgende stappen uit om een databaseserver te configureren zodat de DB2-crawler kan luisteren naar een Event Publishing-wachtrij. (Raadpleeg de documentatie bij WebSphere MQ voor hulp bij deze stappen.)

1. Meld u aan als WebSphere MQ-beheerder en geef de volgende opdrachten op om een wachtrijmanager en wachtrijen te maken.
  - a. Geef de volgende opdracht op een opdrachtregel op:  
crtmqm QM1

- b. Nadat het bericht Setup completed is weergegeven, geeft u de volgende opdracht op:  
strmqm QM1
  - c. Nadat het bericht 'QM1' started is weergegeven, geeft u de volgende opdracht op:  
runmqsc QM1
  - d. Nadat het bericht Starting MQSC for queue manager QM1 is weergegeven, geeft u de volgende opdracht op om een beheerwachtrij te maken:  
DEFINE QLOCAL('ASN.QM1.ADMINQ')
  - e. Nadat het bericht WebSphere MQ queue created is weergegeven, geeft u de volgende opdracht op om een wachtrij voor opnieuw starten te maken:  
DEFINE QLOCAL(' ASN.QM1.RESTARTQ')
  - f. Nadat het bericht WebSphere MQ queue created opnieuw is weergegeven, geeft u de volgende opdracht op om een gegevenswachtrij te maken:  
DEFINE QLOCAL(' ASN.QM1.DATAQ')
  - g. Nadat het bericht WebSphere MQ queue created opnieuw is weergegeven, geeft u de volgende opdracht op om af te sluiten:  
end
2. Geef de volgende opdracht op om de MQ-listener te starten op de databaseserver (de MQ-listener moet worden uitgevoerd wanneer u een DB2-crawler maakt die gebruikmaakt van Event Publishing). In dit voorbeeld is 1414 het poortnummer en SYSTEM.DEF.SVRCONN het standaardkanaal van de server:  
runmqtsr -m QM1 -t TCP -p 1414 &
3. Geef de volgende opdrachten op om een DB2 UDB-gebruikers toegang te verlenen tot de wachtrijmanager en de wachtrijen via de MQI (Message Queuing Interface) voor Event Publishing (in dit voorbeeld is het gebruikers-ID 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. Geef de volgende opdrachten op voor het gebruikers-ID dat wordt gebruikt om een DB2-crawler te maken en uit te voeren met Event Publishing. Met deze opdrachten verleent u het gebruikers-ID toegang tot de wachtrijmanager en de wachtrijen via de MQI (Message Queuing Interface) voor Event Publishing. In dit voorbeeld is het gebruikers-ID 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
```

## DB2-databases op een klassieke gegevensbronserver crawlen

De DB2-crawler kan een DB2-database op een klassieke gegevensbronserver crawlen via WebSphere Information Integrator Classic Federation.

### Over deze taak

Om een DB2-database op een klassieke gegevensbronserver te crawlen, moet u de database federaliseren met een DB2-database op de niet-klassieke gegevensbronserver. Hiervoor kunt u de ODBC-wrapper van WebSphere Information Integrator Classic Federation gebruiken.

## Procedure

Ga als volgt te werk om de database te federaliseren:

1. Installeer WebSphere Information Integrator Classic Federation op de klassieke gegevensbronsserver.
2. Installeer de WebSphere Information Integrator Classic Federation-clientmodule op de niet-klassieke gegevensbronsserver waarop de database is geïnstalleerd die u wilt federaliseren met de database op de klassieke gegevensbronsserver.
3. Configureer de ODBC-driver van WebSphere Information Integrator Classic Federation om een verbinding te maken met de klassieke gegevensbronsserver.
4. Maak een verbinding met de federatieve database en maak de ODBC-wrapper voor federatie met de database op de klassieke gegevensbronsserver.
5. Configureer de DB2-crawler zodat de federatieve database op de niet-klassieke gegevensbronsserver kan worden gecrawld. Op deze manier kan de database op de klassieke gegevensbronsserver worden gecrawld via WebSphere Information Integrator Classic Federation.

---

## DB2 Content Manager-crawlers

Als u itemtypen van IBM DB2 Content Manager wilt opnemen in een enterprise search-collectie, moet u een DB2 Content Manager-crawler configureren.

### Configuratie van de crawlerserver

Voordat u een DB2 Content Manager-server kunt crawlen, moet u een bepaald script uitvoeren op de crawlerserver. Dit script, dat bij WebSphere Information Integrator OmniFind Edition wordt geleverd, stelt de DB2 Content Manager-crawler in staat om te communiceren met DB2 Content Manager-servers.

Voordat u met behulp van de beheerconsole van enterprise search een DB2 Content Manager-crawler configureert, voert u de taak uit die van toepassing is op uw omgeving:

- “De crawlerserver configureren op UNIX for DB2 Content Manager-crawlers” op pagina 55.
- “De crawlerserver configureren op Windows for DB2 Content Manager-crawlers” op pagina 57.

### Configuratieoverzicht

U kunt met de DB2 Content Manager-crawler een willekeurig aantal DB2 Content Manager-servers crawlen. Bij het configureren van de crawler moet u opgeven hoe de crawler alle DB2 Content Manager-servers in de crawlruimte moet crawlen. Bovendien moet u opgeven welke specifieke itemtypen er op elke server moeten worden gecrawld.

Als u een DB2 Content Manager-crawler wilt maken of wijzigen, meldt u zich aan bij de beheerconsole van enterprise search. U moet de rol van beheerder van enterprise search hebben of beheerder zijn van de collectie die eigenaar is van de crawler.

Bij het definiëren van de crawler moet u de volgende procedure uitvoeren. Hierbij wordt u geholpen door de wizard:

- Geef eigenschappen op die bepalen hoe de crawler werkt en op welke manier de systeemresources worden gebruikt. De eigenschappen bepalen hoe de crawler alle itemtypen op alle DB2 Content Manager-servers in de crawlruimte crawlt.
- Selecteer de DB2 Content Manager-servers die u wilt crawlen.
- Geef de gebruikers-ID's en wachtwoorden op waarmee de crawler toegang kan krijgen tot de content op de DB2 Content Manager-servers.
- Geef een rooster op voor het crawlen van de servers.
- Geef op welke itemtypen er op elk van de DB2 Content Manager-servers moeten worden gecrawld.
- Geef opties op om de attributen in bepaalde itemtypen doorzoekbaar te maken. U kunt bijvoorbeeld bepaalde typen documenten uitsluiten van de crawlruimte en opgeven welke attributen kunnen worden afgebeeld in de zoekresultaten.
- Configureer beveiligingsopties op documentniveau. Als bij het maken van de collectie de beveiliging is ingeschakeld, kan de crawler beveiligingsgegevens koppelen aan documenten in de index. Met behulp van die beveiligingsgegevens kunnen zoekprogramma's toegangsbesturing uitvoeren op basis van opgeslagen toeganglijsten of beveiligingstokens.  
U kunt ook een optie selecteren waarmee de legitimatiegegevens van gebruikers worden gecontroleerd wanneer een query wordt ingediend. Op deze manier worden de legitimatiegegevens vergeleken met de huidige toeganglijsten die door de oorspronkelijke gegevensbron worden beheerd en niet met de geïndexeerde beveiligingsgegevens.

Als u bij het definiëren van de crawler meer informatie wenst over de velden in de wizard en over de manier waarop u de informatie moet opgeven die de crawler nodig heeft om informatie te crawlen, klikt u in de wizard op **Help**.

## De crawlerserver configureren op UNIX for DB2 Content Manager-crawlers

Als u WebSphere II OmniFind Edition installeert op een computer met IBM AIX, Linux of de Solaris-besturingsomgeving, moet u een script uitvoeren om de crawlerserver te configureren. Met dit script wordt de DB2 Content Manager-crawler ingeschakeld voor communicatie met IBM DB2 Content Manager-servers.

### Over deze taak

De DB2 Content Manager-crawler gebruikt de Java-connector voor DB2 Content Manager versie 8 om toegang te krijgen tot DB2 Content Manager-servers. U installeert deze connector door een van de volgende producten te installeren op de crawlerserver:

- IBM DB2 Information Integrator for Content, Versie 8.3 voor AIX, Solaris of Linux
- IBM DB2 Information Integrator for Content, Versie 8.2 voor AIX of Solaris
- IBM DB2 Content Manager Toolkit, Versie 8.2 voor Linux

Om ervoor te zorgen dat de DB2 Content Manager-crawler kan samenwerken met DB2 Content Manager moet u een installatiescript uitvoeren waarmee WebSphere II OmniFind Edition op de crawlerserver wordt geïnstalleerd na de installatie van de connector.

### Procedure

Ga als volgt te werk om de crawlerserver zodanig te configureren, dat hiermee DB2 Content Manager-servers kunnen worden gecrawld:

1. Installeer de Java-connector voor DB2 Content Manager versie 8 op de crawler-server:
  - a. Meld bij de crawlerserver aan als de hoofdgebruiker:
 

```
su - root
```
  - b. Voer het bestand db2profile uit. Bijvoorbeeld:
 

```
./home/db2inst/sqllib/db2profile
```
  - c. Exporteer de omgevingsvariabele JAVAHOME. Bijvoorbeeld:
 

```
export JAVAHOME=/usr/IBMJava2-141
```
  - d. Voeg de Java-directory toe aan de omgevingsvariabele PATH:
 

```
export PATH=$PATH:$JAVAHOME/bin
```
  - e. Plaats de installatie-cd van DB2 Information Integrator for Content en voer de installatiewizard uit.
  - f. Voer de volgende acties uit in het venster Component Selection. (Als u met DB2 Information Integrator for Content versie 8.3 werkt, kunt u dit venster oproepen met de optie voor de aangepaste installatie [Custom install].)
    - 1) Selecteer **Local connectors** in de lijst **Components** en selecteer vervolgens **Content Manager V8 connector** in de lijst **Subcomponents**.
    - 2) Selecteer **Connector toolkits and samples** in de lijst **Components** en selecteer **Content Manager V8 connector** in de lijst **Subcomponents**.
  - g. Geef een databasenaam, gebruikersnaam en wachtwoord op voor de DB2 Content Manager-bibliotheek en accepteer de standaardinstellingen voor de overige vensters.
2. Meld u bij de crawlerserver aan met een gebruikers-ID dat in de groep met DB2-beheerders is opgenomen.
3. Neem de DB2 Content Manager-bibliotheekserverdatabase op afstand op in de catalogus en controleer of de crawlerserver verbinding kan maken met de DB2 Content Manager-server:

```
db2 catalog tcpip node knooppunt remote hostnaam server poort
db2 catalog database database_naam as alias at node knooppunt
```

Hierbij geldt het volgende:

*knooppunt*

De korte hostnaam van de DB2 Content Manager-server (bijvoorbeeld *ibmes*).

*hostnaam*

De volledige hostnaam van de DB2 Content Manager-server (bijvoorbeeld *ibmes.ibm.com*).

*poort*

Het poortnummer van de DB2 Content Manager-server.

*database\_naam*

De naam van de DB2 Content Manager-database (bijvoorbeeld *ICMNLSDB*).

*alias*

De alias van de DB2 Content Manager-database (bijvoorbeeld *CMSVR*)

4. Optioneel: meld u aan als de hoofdgebruiker en test de databaseverbinding:

```
./Information_Integrator_for_Content_installatie_directory/bin/cmbenv81.sh
cd Information_Integrator_for_Content_installatie_directory/samples/java/icm
```



- ```
javac *.java
java SConnectDisconnectICM ICMdatabase_naam CMbeheerders_ID CMbeheerders_wachtwoord
```
5. Voer op de crawlerserver het installatiescript uit voor de DB2 Content Manager-crawler:
    - a. Ga naar de directory ES\_INSTALL\_ROOT/bin:
 

```
cd $ES_INSTALL_ROOT/bin
```
    - b. Start het volgende script en volg de instructies op het scherm:
 

```
escrcm.sh
```
  6. Ga naar de crawlerserver, stop het enterprise search-systeem en start dit vervolgens opnieuw:
 

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

## De crawlerserver configureren op Windows for DB2 Content Manager-crawlers

Als u WebSphere II OmniFind Edition installeert op een Microsoft Windows-computer, moet u een script uitvoeren om de crawlerserver te configureren. Met dit script wordt de DB2 Content Manager-crawler ingeschakeld voor communicatie met IBM DB2 Content Manager-servers.

### Over deze taak

De DB2 Content Manager-crawler gebruikt de Java-connector voor DB2 Content Manager versie 8 om toegang te krijgen tot DB2 Content Manager-servers. U installeert deze connector door IBM DB2 Information Integrator for Content Versie 8.2 of Versie 8.3 voor Windows te installeren op de crawlerserver. Om ervoor te zorgen dat de DB2 Content Manager-crawler kan samenwerken met DB2 Content Manager moet u een installatiescript uitvoeren waarmee WebSphere II OmniFind Edition op de crawlerserver wordt geïnstalleerd na de installatie van de connector.

### Procedure

Ga als volgt te werk om de crawlerserver zodanig te configureren, dat hiermee DB2 Content Manager-servers kunnen worden gecrawld:

1. Installeer de Java-connector voor DB2 Content Manager versie 8 op de crawlerserver:
  - a. Plaats de installatie-cd van DB2 Information Integrator for Content. Het installatieprogramma wordt automatisch gestart. De installatiewizard van DB2 Content Manager Enterprise Information Portal verschijnt.
  - b. Voer de volgende acties uit in het venster Component Selection. (Als u met DB2 Information Integrator for Content versie 8.3 werkt, kunt u dit venster oproepen met de optie voor de aangepaste installatie [Custom install].)
    - 1) Selecteer **Local connectors** in de lijst **Components** en selecteer vervolgens **Content Manager V8 connector** in de lijst **Subcomponents**.
    - 2) Selecteer **Connector toolkits and samples** in de lijst **Components** en selecteer **Content Manager V8 connector** in de lijst **Subcomponents**.
  - c. Geef een databasenaam, gebruikersnaam en wachtwoord op voor de DB2 Content Manager-bibliotheek en accepteer de standaardinstellingen voor de overige vensters.

2. Neem de DB2 Content Manager-bibliotheekserverdatabase op afstand op in de catalogus en controleer of de crawlerserver verbinding kan maken met de DB2 Content Manager-server. Geef de volgende opdrachten op bij de opdracht-aanwijzing op de crawlerserver:

```
db2 catalog tcpip node knooppunt remote hostnaam server poort  
db2 catalog database database_naam as alias at node knooppunt
```

Hierbij geldt het volgende:

*knooppunt*

De korte hostnaam van de DB2 Content Manager-server (bijvoorbeeld *ibmes*).

*hostnaam*

De volledige hostnaam van de DB2 Content Manager-server (bijvoorbeeld *ibmes.ibm.com*).

*poort*

Het poortnummer van de DB2 Content Manager-server.

*database\_naam*

De naam van de DB2 Content Manager-database (bijvoorbeeld *ICMNLSDDB*).

*alias*

De alias van de DB2 Content Manager-database (bijvoorbeeld *CMSVR*).

3. Optioneel: test de databaseverbinding door een opdrachtaanwijzing te openen en de volgende opdrachten op te geven:

```
cmbenv81.bat  
cd Information_Integrator_for_Content_installatie_directory\samples\java\icm  
javac *.java  
java SConnectDisconnectICM ICMdatabase_naam CMbeheerders_ID CMbeheerders_wachtwoord
```

4. Voer op de crawlerserver het installatiescript uit voor de DB2 Content Manager-crawler:

- a. Ga naar de directory *ES\_INSTALL\_ROOT*\bin:

```
cd %ES_INSTALL_ROOT%\bin
```

- b. Start het volgende script en volg de instructies op het scherm:

```
escrcm.vbs
```

5. Ga naar de crawlerserver, stop het enterprise search-systeem en start dit vervolgens opnieuw:

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

---

## Domino Document Manager-crawlers

Als u bibliotheken en kabinetten van Domino Document Manager wilt opnemen in een enterprise search-collectie, moet u een Domino Document Manager-crawler configureren.

### Configuratie van de crawlerserver

Als de Domino Document Manager-server die u wilt crawlen, gebruikmaakt van het NRPC-protocol (Notes Remote Procedure Call), moet u een installatiescript op de crawlerserver uitvoeren. Dit script, dat bij WebSphere Information Integrator

OmniFind Edition wordt geleverd, zorgt voor de communicatie tussen de Domino Document Manager-crawler en de servers die NRPC gebruiken.

Als de Domino Document Manager-server die u wilt crawlen, gebruikmaakt van het protocol DIIOP (Domino Internet Inter-ORB Protocol), hoeft u geen installatiescript op de crawlerserver uit te voeren. U moet de Domino Document Manager-server echter zodanig configureren, dat de Domino Document Manager-crawler toegang tot de server kan krijgen.

**Belangrijk:** Als de Domino Document Manager-server gebruik maakt van DIIOP en u de crawler configureert voor het gebruik van HTTPS of DIIOP over SSL zodat de transmissie tussen de crawler en de server versleuteld wordt, moet u het bestand TrustedCerts.class (bijvoorbeeld c:\certs of /data/certs) kopiëren van de Domino Document Manager-server naar de crawlerserver. In een configuratie met twee of vier servers moet u het bestand TrustedCerts.class tevens kopiëren naar de servers waarop de zoekcomponent is geïnstalleerd. Zorg ervoor dat het bestand zich op de crawlerserver en de zoekservers op dezelfde locatie bevindt. U geeft het pad voor het bestand TrustedCerts.class op wanneer u de crawler configureert.

Als WebSphere II OmniFind Edition is geïnstalleerd op een IBM AIX-systeem, moet u ervoor zorgen dat de I/O Completion Port-module is geïnstalleerd en beschikbaar is op de crawlerserver.

Voordat u met behulp van de beheerconsole van enterprise search een Domino Document Manager-crawler configureert, voert u de taken uit die van toepassing zijn op uw omgeving:

- “De crawlerserver configureren op UNIX voor het crawlen van Lotus Domino-bronnen” op pagina 68.
- “De crawlerserver configureren op Windows voor het crawlen van Lotus Domino-bronnen” op pagina 71.
- “Servers configureren die het DIIOP-protocol gebruiken” op pagina 72.
- “De I/O Completion Port op AIX configureren voor het crawlen van Lotus Domino-bronnen” op pagina 73.

## Beveiliging op documentniveau

Als collectiebeveiliging is ingeschakeld en de server die u wilt crawlen, gebruikmaakt van het NRPC-protocol, moet u een Lotus Domino Trusted Server op de crawlerserver configureren. Trusted Server wordt gebruikt om toegangsbesturing op documentniveau af te dwingen. Voordat u de collectie doorzoekbaar maakt voor gebruikers, moet u de volgende taken uitvoeren:

- Lotus Domino Trusted Servers configureren voor het controleren van legitimatiegegevens van gebruikers.
- Algemene beveiliging inschakelen in WebSphere Application Server en het zoekprogramma configureren voor het gebruik van beveiliging. Met deze stap zorgt u ervoor dat de legitimatiegegevens van gebruikers worden gecontroleerd wanneer gebruikers proberen het zoekprogramma te gebruiken. Deze legitimatiegegevens kunnen vervolgens op de zoekservers worden gebruikt om de machtigingen van gebruikers voor Lotus Domino-documenten te controleren.

## Configuratieoverzicht

U kunt met de Domino Document Manager-crawler een willekeurig aantal Domino Document Manager-bibliotheken crawlen. Als u wilt crawlen vanaf één Domino

Document Manager-server, selecteert u de bibliotheken tijdens het maken van de crawler. Tijdens het bewerken van de crawlruimte kunt u de documenten die u in dezelfde crawlruimte wilt opnemen, vanaf een andere Domino Document Manager-server toevoegen. Bij het maken of bewerken van de crawler kunt u aangeven of u alle kabinetten in de bibliotheken die u voor crawlen hebt geselecteerd, wilt crawlen of dat u specifieke kabinetten wilt crawlen.

Als u een Domino Document Manager-crawler wilt maken of wijzigen, meldt u zich aan bij de beheerconsole van enterprise search. U moet de rol van beheerder van enterprise search hebben of beheerder zijn van de collectie die eigenaar is van de crawler.

Bij het definiëren van de crawler moet u de volgende procedure uitvoeren. Hierbij wordt u geholpen door de wizard:

- Geef eigenschappen op die bepalen hoe de crawler werkt en op welke manier de systeemresources worden gebruikt. De crawlereigenschappen bepalen hoe de crawler alle documenten in de crawlruimte crawlt.
- Geef de Domino Document Manager-server en het communicatieprotocol op.
- Als de server is geconfigureerd voor gebruik van het DIIOP-protocol, kunt u opgeven hoe de crawler verbinding met Domino-objecten moet maken. U kunt bijvoorbeeld opties opgeven voor het gebruik van HTTPS of SSL (Secure Sockets Layer) om de communicatie te coderen.
- Selecteer de bibliotheken die u wilt crawlen.
- Geef een rooster op voor het crawlen van de bibliotheken.
- Selecteer de documenten die u wilt crawlen. De crawler kan alle kabinetten in een bibliotheek crawlen, of alleen de documenten die zich in de geselecteerde kabinetten bevinden.
- Geef opties op om de velden in de verschillende bibliotheken en kabinetten doorzoekbaar te maken. U kunt bijvoorbeeld bepaalde velden uitsluiten van de crawlruimte en opties opgeven voor het doorzoeken van bijlagen.
- Configureer beveiligingsopties op documentniveau. Als bij het maken van de collectie de beveiliging is ingeschakeld, kan de crawler beveiligingsgegevens koppelen aan documenten in de index. Met behulp van die beveiligingsgegevens kunnen zoekprogramma's toegangsbesturing uitvoeren op basis van opgeslagen toegangslijsten of beveiligingstokens.

U kunt ook een optie selecteren waarmee de legitimatiegegevens van gebruikers worden gecontroleerd wanneer een query wordt ingediend. Op deze manier worden de legitimatiegegevens vergeleken met de huidige toegangslijsten die door de oorspronkelijke gegevensbron worden beheerd en niet met de geïndexeerde beveiligingsgegevens.

Als u bij het definiëren van de crawler meer informatie wenst over de velden in de wizard en over de manier waarop u de informatie moet opgeven die de crawler nodig heeft om informatie te crawlen, klikt u in de wizard op **Help**.

#### **Verwante onderwerpen**

“Tips voor het crawlen van Lotus Domino-databases” op pagina 67

Lees de richtlijnen voor het crawlen van Lotus Domino-databases door voordat u een Notes-crawler configureert.

“Beveiliging op documentniveau afdwingen voor Lotus Domino-documenten” op pagina 259

Als de te crawlen Domino-server gebruikmaakt van het NRPC-protocol (Notes Remote Procedure Call), moet u de crawlerserver zodanig configureren, dat toegangsbesturing op documentniveau kan worden afgedwongen.

### Verwante taken

“De crawlerserver configureren op UNIX voor het crawlen van Lotus Domino-bronnen” op pagina 68

Als u WebSphere II OmniFind Edition installeert op een computer waarop IBM AIX, Linux of de Solaris-besturingsomgeving wordt uitgevoerd en u servers wilt crawlen die gebruikmaken van het NRPC-protocol (Notes Remote Procedure Call), moet u een script uitvoeren om de crawlerserver te configureren. Met het script wordt communicatie tussen Notes-, QuickPlace- en Domino Document Manager-crawlers en de databaseservers ingesteld.

“De crawlerserver configureren op Windows voor het crawlen van Lotus Domino-bronnen” op pagina 71

Als u WebSphere II OmniFind Edition installeert op een Microsoft Windows-computer en u servers wilt crawlen die gebruikmaken van het NRPC-protocol (Notes Remote Procedure Call), moet u een script uitvoeren om de crawlerserver te configureren. Met het script wordt communicatie tussen Notes-, QuickPlace- en Domino Document Manager-crawlers en de databaseservers ingesteld.

“Servers configureren die het DIIOP-protocol gebruiken” op pagina 72

Om servers te crawlen die gebruikmaken van het DIIOP-protocol (Domino Internet Inter-ORB Protocol), moet u de server zodanig configureren dat de Notes-, QuickPlace- en Domino Document Manager-crawlers het protocol kunnen gebruiken.

“De I/O Completion Port op AIX configureren voor het crawlen van Lotus Domino-bronnen” op pagina 73

U kunt de Notes-, QuickPlace- of Domino Document Manager-crawlers pas op een IBM AIX-systeem gebruiken als de IOCP-module (I/O Completion Port) is geïnstalleerd en is geconfigureerd voor gebruik door de crawler.

“Lotus Domino Trusted Servers configureren voor het controleren van legitimatiegegevens van gebruikers” op pagina 260

Om beveiliging af te dwingen voor documenten die zijn gecrawld met een Notes-crawler die gebruikmaakt van het NRPC-protocol (Notes Remote Procedure Call), moeten de te crawlen Domino-servers worden geconfigureerd als Lotus Domino Trusted Servers.

---

## Exchange Server-crawlers

Als u openbare mappen van Microsoft Exchange Server in een enterprise search-collectie wilt opnemen, moet u een Exchange Server-crawler configureren.

U kunt met de Exchange Server-crawler een willekeurig aantal mappen en sub-mappen crawlen op Exchange Server Public Folder Server. Tijdens het maken van een crawler selecteert u de content die u wilt crawlen op een Public Folder Server. Op een later tijdstip kunt u de crawlruimte bewerken en content vanaf een andere Public Folder Server toevoegen.

Om een Exchange Server-crawler te maken of wijzigen, meldt u zich aan bij de beheerconsole van enterprise search. U moet de rol van beheerder van enterprise search hebben of beheerder zijn van de collectie die eigenaar is van de crawler.

Bij het definiëren van de crawler moet u de volgende procedure uitvoeren. Hierbij wordt u geholpen door de wizard:

- Geef eigenschappen op die bepalen hoe de crawler werkt en op welke manier de systeemresources worden gebruikt. De crawlereigenschappen bepalen hoe de crawler alle submappen op alle servers in de crawlruimte crawlt.

- Geef informatie op over de Exchange Server Public Folder Server die u wilt crawlen.  
U moet een gebruikers-ID en een wachtwoord opgeven zodat de crawler toegang heeft tot de content op de server. Als de server gebruikmaakt van het SSL-protocol (Secure Sockets Layer), kunt u opties opgeven waarmee de crawler toegang heeft tot het sleutelruimtebestand op de crawlerserver.
- Geef een rooster op voor het crawlen van de Public Folder Server.
- Selecteer de mappen en submappen die u wilt crawlen.
- Geef opties op waarmee de documenten in de submappen doorzoekbaar worden gemaakt. Zo kunt u bijvoorbeeld bepaalde typen documenten uitsluiten uit de crawlruimte.
- Configureer beveiligingsopties op documentniveau. Als bij het maken van de collectie de beveiliging is ingeschakeld, kan de crawler beveiligingsgegevens koppelen aan documenten in de index. Met behulp van die beveiligingsgegevens kunnen zoekprogramma's toegangsbesturing uitvoeren op basis van opgeslagen toegangslijsten of beveiligingstokens.

Als u bij het definiëren van de crawler meer informatie wenst over de velden in de wizard en over de manier waarop u de informatie moet opgeven die de crawler nodig heeft om informatie te crawlen, klikt u in de wizard op **Help**.

#### Verwante taken

“Toegang beheren tot beveiligde Exchange Server-documenten” op pagina 258  
Als u een Exchange Server-crawler wilt gebruiken om documenten te crawlen die door een firewall worden beveiligd, moet u controleren of de crawlerserver toegang kan krijgen tot de server met de openbare map van Microsoft Exchange Server.

---

## JDBC-databasecrawlers

U gebruikt de JDBC database-crawler om databases die kunnen worden benaderd met het JDBC-protocol (Java Database Connectivity) op te nemen in een enterprise search-collectie.

### Ondersteunde databasesystemen en -drivers

Voor elk type databasesysteem dat u wilt crawlen, moet u een aparte crawler configureren. Als u een crawler maakt, geeft u opties op voor het crawlen van één database. Later kunt u databases van hetzelfde type toevoegen aan de crawlruimte.

Om het JDBC-protocol te gebruiken voor het crawlen van tabellen in een database, moet de juiste JDBC-driver op de crawlerserver staan. De JDBC database-crawler ondersteunt de volgende databasesystemen en JDBC-drivers van het type 4:

| Database-systeem          | Naam JDBC type 4-driver                                                                                                    | Klassenpaden standaard JDBC-driver                                                                                                              |
|---------------------------|----------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------|
| IBM DB2 UDB 8.2           | com.ibm.db2.jcc.DB2Driver                                                                                                  | <i>db2_install_root</i> /java/db2jcc.jar<br><i>db2_install_root</i> /java/db2jcc_license_cu.jar                                                 |
| Oracle 9i en 10g          | Oracle.jdbc.driver.OracleDriver                                                                                            | <i>oracle_home</i> /jdbc/lib/ojdbc14.jar                                                                                                        |
| Microsoft SQL Server 2000 | com.microsoft.jdbc.sqlserver.SQLServerDriver                                                                               | <i>mssql_jdbc_home</i> /lib/mssqlserver.jar<br><i>mssql_jdbc_home</i> /lib/msbase.jar<br><i>mssql_jdbc_home</i> /lib/msutil.jar                 |
| Microsoft SQL Server 2005 | com.microsoft.sqlserver.jdbc.SQLServerDriver (de JDBC-driver voor SQL Server 2005 wordt niet ondersteund op AIX-systemen.) | <i>install_dir</i> /sqljdbc_1.0/loc/sqljdbc.jar<br>waarbij <i>loc</i> uw locale aangeeft, zoals <i>install_dir</i> /sqljdbc_1.0/enu/sqljdbc.jar |

## JDBC database-crawlers versus DB2-crawlers

Als u momenteel met de DB2-crawler werkt, wilt u deze mogelijk blijven gebruiken. U kunt de gegevens die voor een DB2-crawler zijn opgeslagen, niet migreren naar een JDBC database-crawler.

Gebruik in de volgende situaties de DB2-crawler in plaats van de JDBC database-crawler:

- U wilt DB2 UDB-databases crawlen met een JDBC type 2-driver.
- U wilt Oracle- en SQL Server-databases crawlen die federatief worden gebruikt met een DB2 UDB-database. Met de DB2-crawler hebt u met een roepnaam toegang tot al deze databasetypen.
- U wilt databases van het type DB2 UDB for z/OS, DB2 UDB for iSeries, Informix, Sybase, VSAM, IMS, CA-Datacom of Software AG Adabas crawlen. U moet deze typen databases federatief gebruiken met een DB2 UDB-database en benaderen met de DB2-crawler via een roepnaam.
- U wilt Event Publishing gebruiken om de enterprise search-index bij te werken wanneer de updates voor een database worden gepubliceerd.

## Configuratieoverzicht

Als u een JDBC database-crawler wilt maken of wijzigen, meldt u zich aan bij de beheerconsole van enterprise search. U moet de rol van beheerder van enterprise search hebben of beheerder zijn van de collectie die eigenaar is van de crawler.

Bij het definiëren van de crawler moet u de volgende procedure uitvoeren. Hierbij wordt u geholpen door de wizard:

- Geef eigenschappen op die bepalen hoe de crawler werkt en op welke manier de systeemresources worden gebruikt. De crawlereigenschappen bepalen hoe de crawler alle databases in de crawlruimte crawlt.
- geef het type database op dat u wilt crawlen.
- Selecteer de database die u wilt crawlen en geef indien nodig een gebruikers-ID en wachtwoord op waarmee de crawler toegang tot de database kan krijgen.
- Geef een rooster op voor het crawlen van de database.
- Selecteer de tabellen die u wilt crawlen.

- Waarschuwing:** Omwille van de prestaties (en om te voorkomen dat er een time-out optreedt bij het configureren van de crawler) dient u alléén te kiezen voor het crawlen van alle tabellen als de database niet veel tabellen bevat of als de tabellen niet veel kolommen bevatten. Als u nu een bepaalde hoeveelheid te crawlen tabellen selecteert, kunt u de crawlruimte later bewerken en meer tabellen aan de collectie toevoegen.
- Geef opties op waarmee de kolommen in bepaalde tabellen doorzoekbaar worden gemaakt. U kunt bijvoorbeeld instellen dat bepaalde kolommen in zoekopdrachten met parameters kunnen worden opgenomen en aangeven welke kolommen in de zoekresultaten kunnen worden afgebeeld.
  - Configureer beveiligingsopties op documentniveau. Als bij het maken van de collectie de beveiliging is ingeschakeld, kan de crawler beveiligingsgegevens koppelen aan documenten in de index. Met behulp van die beveiligingsgegevens kunnen zoekprogramma's toegangsbesturing uitvoeren op basis van opgeslagen toeganglijsten of beveiligingstokens.

Als u bij het definiëren van de crawler meer informatie wenst over de velden in de wizard en over de manier waarop u de informatie moet opgeven die de crawler nodig heeft om informatie te crawlen, klikt u in de wizard op **Help**.

---

## NNTP-crawlers

Als u artikelen van NNTP-nieuwsgroepen in een enterprise search-collectie wilt opnemen, moet u een NNTP-crawler configureren.

U kunt met de NNTP-crawler een willekeurig aantal NNTP-servers crawlen. Tijdens het configureren van de crawler selecteert u de nieuwsgroepen op de NNTP-server die u wilt crawlen. Tijdens het bewerken van de crawlruimte kunt u andere NNTP-servers toevoegen die door de crawler moeten worden gecrawld.

Wanneer u hebt aangegeven welke nieuwsgroepen moeten worden gecrawld, kunt u de groepen selecteren die u wilt opnemen in of uitsluiten van de crawlruimte. Met dit ontwerp kunt u op een eenvoudige manier toestaan dat de crawler het merendeel van de nieuwsgroepen op een server kan crawlen en voorkomen dat de crawler de nieuwsgroepen crawl't die gebruikers niet mogen doorzoeken.

Zo kunt u regels opgeven waarmee alle nieuwsgroepen op een bepaalde NNTP-server moeten worden opgenomen en vervolgens opgeven dat nieuwsgroepen op die server moeten worden uitgesloten als in de namen van de nieuwsgroepen de tekst privé voorkomt.

Om een NNTP-crawler te maken of wijzigen, meldt u zich aan bij de beheerconsole van enterprise search. U moet de rol van beheerder van enterprise search hebben of beheerder zijn van de collectie die eigenaar is van de crawler.

Bij het definiëren van de crawler moet u de volgende procedure uitvoeren. Hierbij wordt u geholpen door de wizard:

- Geef eigenschappen op die bepalen hoe de crawler werkt en op welke manier de systeemresources worden gebruikt. De crawlereigenschappen bepalen hoe de crawler alle nieuwsgroepen in de crawlruimte crawl't.
- Geef aan welke NNTP-server moet worden gecrawld. Als de server is beveiligd met een wachtwoord, moet u het gebruikers-ID en het wachtwoord opgeven dat de crawler moet gebruiken om toegang te krijgen tot de nieuwsgroepen op de server.
- Geef een rooster op voor het crawlen van de server.
- Geef patronen op voor het opnemen van nieuwsgroepen en het uitsluiten van bepaalde nieuwsgroepen van de crawlruimte.
- Geef aan of de crawler automatisch de taal en codetabel moet vaststellen van de artikelen die moeten worden gecrawld of dat de crawler een specifieke taal en codetabel moet gebruiken.
- Configureer beveiligingsopties op documentniveau. Als bij het maken van de collectie de beveiliging is ingeschakeld, kan de crawler beveiligingsgegevens koppelen aan documenten in de index. Met behulp van die beveiligingsgegevens kunnen zoekprogramma's toegangsbesturing uitvoeren op basis van opgeslagen toegangslijsten of beveiligingstokens.

Als u bij het definiëren van de crawler meer informatie wenst over de velden in de wizard en over de manier waarop u de informatie moet opgeven die de crawler nodig heeft om informatie te crawlen, klikt u in de wizard op **Help**.



---

## Notes-crawlers

Als u IBM Lotus Notes-databases wilt opnemen in een enterprise search-collectie, moet u een Notes-crawler configureren.

### Configuratie van de crawlerserver

Als de Lotus Notes-server die u wilt crawlen, gebruikmaakt van het NRPC-protocol (Notes Remote Procedure Call), moet u een installatiescript op de crawlerserver uitvoeren. Dit script, dat bij WebSphere Information Integrator OmniFind Edition wordt geleverd, zorgt voor de communicatie tussen de Notes-crawler en de servers die NRPC gebruiken.

Als de Lotus Notes-server die u wilt crawlen, gebruikmaakt van het DIIOP-protocol (Domino Internet Inter-ORB Protocol), hoeft u geen installatiescript op de crawlerserver uit te voeren. U moet de Lotus Notes-server echter zodanig configureren, dat de Notes-crawler toegang tot de server kan krijgen.

**Belangrijk:** Als de Lotus Notes-server gebruik maakt van DIIOP en u de crawler configureert voor het gebruik van HTTPS of DIIOP over SSL zodat de transmissie tussen de crawler en de server versleuteld wordt, moet u het bestand `TrustedCerts.class` (bijvoorbeeld `c:\certs` of `/data/certs`) kopiëren van de Lotus Notes Manager-server naar de crawlerserver. In een configuratie met twee of vier servers moet u het bestand `TrustedCerts.class` tevens kopiëren naar de servers waarop de zoekcomponent is geïnstalleerd. Zorg ervoor dat het bestand zich op de crawlerserver en de zoekservers op dezelfde locatie bevindt. U geeft het pad voor het bestand `TrustedCerts.class` op wanneer u de crawler configureert.

Als WebSphere II OmniFind Edition is geïnstalleerd op een IBM AIX-systeem, moet u ervoor zorgen dat de I/O Completion Port-module is geïnstalleerd en beschikbaar is op de crawlerserver.

Voordat u met behulp van de beheerconsole van enterprise search een Notes-crawler configureert, voert u de taken uit die van toepassing zijn op uw omgeving:

- “De crawlerserver configureren op UNIX voor het crawlen van Lotus Domino-bronnen” op pagina 68.
- “De crawlerserver configureren op Windows voor het crawlen van Lotus Domino-bronnen” op pagina 71.
- “Servers configureren die het DIIOP-protocol gebruiken” op pagina 72.
- “De I/O Completion Port op AIX configureren voor het crawlen van Lotus Domino-bronnen” op pagina 73.

### Beveiliging op documentniveau

Als collectiebeveiliging is ingeschakeld en de server die u wilt crawlen, gebruikmaakt van het NRPC-protocol, moet u een Lotus Domino Trusted Server op de crawlerserver configureren. Trusted Server wordt gebruikt om toegangsbesturing op documentniveau af te dwingen. Voordat u de collectie doorzoekbaar maakt voor gebruikers, moet u de volgende taken uitvoeren:

- Lotus Domino Trusted Servers configureren voor het controleren van legitimatiegegevens van gebruikers.
- Algemene beveiliging inschakelen in WebSphere Application Server en het zoekprogramma configureren voor het gebruik van beveiliging. Met deze stap zorgt u ervoor dat de legitimatiegegevens van gebruikers worden gecontroleerd wan-

neer gebruikers proberen het zoekprogramma te gebruiken. Deze legitimatiegegevens kunnen vervolgens op de zoekservers worden gebruikt om de machtigingen van gebruikers voor Lotus Domino-documenten te controleren.

## Configuratieoverzicht

U kunt met de Notes-crawler een willekeurig aantal standaard-Lotus Notes-databases (.nsf-bestanden) crawlen. Als u wilt crawlen vanaf één Lotus Notes-server, selecteert u de databases of directory's tijdens het maken van de crawler. Tijdens het bewerken van de crawlruimte kunt u de documenten die u in dezelfde crawlruimte wilt opnemen, vanaf een andere Lotus Notes-server toevoegen. Bij het maken of bewerken van de crawler kunt u aangeven of u alle databases of directory's op de server wilt crawlen of dat u specifieke databases, views en mappen wilt crawlen.

Als u een Notes-crawler wilt maken of wijzigen, meldt u zich aan bij de beheerconsole van enterprise search. U moet de rol van beheerder van enterprise search hebben of beheerder zijn van de collectie die eigenaar is van de crawler.

Bij het definiëren van de crawler moet u de volgende procedure uitvoeren. Hierbij wordt u geholpen door de wizard:

- Geef eigenschappen op die bepalen hoe de crawler werkt en op welke manier de systeemresources worden gebruikt. De crawlereigenschappen bepalen hoe de crawler alle documenten in de crawlruimte crawlt.
- Geef de hostnaam, de poort en het communicatieprotocol van de Lotus Notes-server op.
- Als de server is geconfigureerd voor gebruik van het DIIOP-protocol, kunt u opgeven hoe de crawler verbinding met Domino-objecten moet maken. U kunt bijvoorbeeld opties opgeven voor het gebruik van HTTPS of SSL (Secure Sockets Layer) om de communicatie te coderen.
- Selecteer de databases of directory's die u wilt crawlen.
- Geef een rooster op voor het crawlen van de databases of directory's.
- Selecteer de documenten die u wilt crawlen. U kunt alle documenten in een directory, alle documenten in een database of de documenten in geselecteerde views en mappen van een database crawlen.
- Geef opties op om de velden in de verschillende databases, views en mappen doorzoekbaar te maken. U kunt bijvoorbeeld bepaalde velden uitsluiten van de crawlruimte en opties opgeven voor het doorzoeken van bijlagen.
- Configureer beveiligingsopties op documentniveau. Als bij het maken van de collectie de beveiliging is ingeschakeld, kan de crawler beveiligingsgegevens koppelen aan documenten in de index. Met behulp van die beveiligingsgegevens kunnen zoekprogramma's toegangsbesturing uitvoeren op basis van opgeslagen toeganglijsten of beveiligingstokens.

U kunt ook een optie selecteren waarmee de legitimatiegegevens van gebruikers worden gecontroleerd wanneer een query wordt ingediend. Op deze manier worden de legitimatiegegevens vergeleken met de huidige toeganglijsten die door de oorspronkelijke gegevensbron worden beheerd en niet met de geïndexeerde beveiligingsgegevens.

Als u bij het definiëren van de crawler meer informatie wenst over de velden in de wizard en over de manier waarop u de informatie moet opgeven die de crawler nodig heeft om informatie te crawlen, klikt u in de wizard op **Help**.

### Verwante onderwerpen

“Tips voor het crawlen van Lotus Domino-databases”

Lees de richtlijnen voor het crawlen van Lotus Domino-databases door voordat u een Notes-crawler configureert.

“Beveiliging op documentniveau afdwingen voor Lotus Domino-documenten” op pagina 259

Als de te crawlen Domino-server gebruikmaakt van het NRPC-protocol (Notes Remote Procedure Call), moet u de crawlerserver zodanig configureren, dat toegangsbesturing op documentniveau kan worden afgedwongen.

#### **Verwante taken**

“De crawlerserver configureren op UNIX voor het crawlen van Lotus Domino-bronnen” op pagina 68

Als u WebSphere II OmniFind Edition installeert op een computer waarop IBM AIX, Linux of de Solaris-besturingsomgeving wordt uitgevoerd en u servers wilt crawlen die gebruikmaken van het NRPC-protocol (Notes Remote Procedure Call), moet u een script uitvoeren om de crawlerserver te configureren. Met het script wordt communicatie tussen Notes-, QuickPlace- en Domino Document Manager-crawlers en de databaseservers ingesteld.

“De crawlerserver configureren op Windows voor het crawlen van Lotus Domino-bronnen” op pagina 71

Als u WebSphere II OmniFind Edition installeert op een Microsoft Windows-computer en u servers wilt crawlen die gebruikmaken van het NRPC-protocol (Notes Remote Procedure Call), moet u een script uitvoeren om de crawlerserver te configureren. Met het script wordt communicatie tussen Notes-, QuickPlace- en Domino Document Manager-crawlers en de databaseservers ingesteld.

“Servers configureren die het DIIOP-protocol gebruiken” op pagina 72

Om servers te crawlen die gebruikmaken van het DIIOP-protocol (Domino Internet Inter-ORB Protocol), moet u de server zodanig configureren dat de Notes-, QuickPlace- en Domino Document Manager-crawlers het protocol kunnen gebruiken.

“De I/O Completion Port op AIX configureren voor het crawlen van Lotus Domino-bronnen” op pagina 73

U kunt de Notes-, QuickPlace- of Domino Document Manager-crawlers pas op een IBM AIX-systeem gebruiken als de IOCP-module (I/O Completion Port) is geïnstalleerd en is geconfigureerd voor gebruik door de crawler.

“Lotus Domino Trusted Servers configureren voor het controleren van legitimatiegegevens van gebruikers” op pagina 260

Om beveiliging af te dwingen voor documenten die zijn gecrawld met een Notes-crawler die gebruikmaakt van het NRPC-protocol (Notes Remote Procedure Call), moeten de te crawlen Domino-servers worden geconfigureerd als Lotus Domino Trusted Servers.

## **Tips voor het crawlen van Lotus Domino-databases**

Lees de richtlijnen voor het crawlen van Lotus Domino-databases door voordat u een Notes-crawler configureert.

- Notes-databases die zijn gebaseerd op standaardjablonen (zoals een discussie-database), zijn het meest geschikt om te crawlen.
- De Notes-crawler is van toepassing op de volgende veldtoewijzingsregels:
  - De namen van de belangrijkste velden uit de Domino-standaardjablonen worden eerst geregistreerd.
  - De waarden uit de Notes-velden die in de tabel met toewijzingsregels zijn opgegeven, worden als documentsamenvattingen in de zoekresultaten gebruikt.

- De waarden uit de Notes-velden die niet zijn opgegeven in de tabel met toewijzingsregels, worden niet gebruikt in de documentsamenvattingen.
- De waarden uit de Notes-velden die zijn toegewezen aan het titelveld, worden als documenttitel in de zoekresultaten gebruikt.
- De velden in de volgende tabel worden standaard toegewezen aan de zoekveldnamen:

Tabel 2. Standaardveldtoewijzingsregels

| Notes-databaseveldnaam | Zoekveldnaam |
|------------------------|--------------|
| Title                  | Title        |
| EventTitle             | Title        |
| Subject                | Title        |
| Body                   | Body         |
| Mission                | Body         |
| From                   | Maker        |
| Author                 | Maker        |
| Keywords               | Categories   |
| Categories             | Categories   |
| TeamRoomName           | Organization |
| TeamName               | Organization |
| Department             | Organization |

- De Notes-crawler kan alle veldtypen crawlen, met uitzondering van velden die zijn berekend voor weergave.
- Statische tekst en afbeeldingen die op een Notes-formulier zijn geplaatst, worden niet gecrawld.
- Als u alle velden wilt crawlen en de veldgegevens wilt maximaliseren, selecteert u tijdens het configureren van de crawler het aankruisvakje **Alle** onder **Crawlen** (u kunt het veld **Alle velden crawlen, behalve** gebruiken om het aantal velden te beperken dat kan worden gecrawld).

Als u wilt voorkomen dat overbodige velden worden gecrawld, heft u de selectie van het aankruisvakje **Crawlen** op voor alle velden, met uitzondering van de velden die aan zoekvelden zijn toegewezen.

#### Verwante onderwerpen

“Domino Document Manager-crawlers” op pagina 58

Als u bibliotheken en kabinetten van Domino Document Manager wilt opnemen in een enterprise search-collectie, moet u een Domino Document Manager-crawler configureren.

“Notes-crawlers” op pagina 65

Als u IBM Lotus Notes-databases wilt opnemen in een enterprise search-collectie, moet u een Notes-crawler configureren.

“QuickPlace-crawlers” op pagina 75

Als u Lotus QuickPlace-ruimten en -rooms in een enterprise search-collectie wilt opnemen, moet u een QuickPlace-crawler configureren.

## De crawlerserver configureren op UNIX voor het crawlen van Lotus Domino-bronnen

Als u WebSphere II OmniFind Edition installeert op een computer waarop IBM AIX, Linux of de Solaris-besturingsomgeving wordt uitgevoerd en u servers wilt

crawlers die gebruikmaken van het NRPC-protocol (Notes Remote Procedure Call), moet u een script uitvoeren om de crawlerserver te configureren. Met het script wordt communicatie tussen Notes-, QuickPlace- en Domino Document Manager-crawlers en de databaseservers ingesteld.

### **Beperkingen**

Domino Server kan niet tegelijkertijd op dezelfde computer worden uitgevoerd met een Notes-, QuickPlace- of Domino Document Manager-crawler die is geconfigureerd voor het gebruik van het NRPC-protocol. Als u probeert een van deze crawlers te starten terwijl Domino Server wordt uitgevoerd, treedt er een fout op en wordt de crawler gestopt.

### **Over deze taak**

De crawlers die het NRPC-protocol gebruiken, gebruiken Domino-bibliotheken als client. U kunt deze bibliotheken installeren door Lotus Domino Server op de crawlerserver te installeren. Om ervoor te zorgen dat de crawlers kunnen samenwerken met de Domino-bibliotheken moet u een installatiescript uitvoeren waarmee WebSphere II OmniFind Edition op de crawlerserver wordt geïnstalleerd nadat u de Domino-bibliotheken hebt geïnstalleerd.

### **Procedure**

Ga als volgt te werk om de crawlerserver zodanig te configureren, dat hiermee Lotus Notes-, Lotus QuickPlace- en Domino Document Manager-servers kunnen worden gecrawld:

1. Maak de gebruiker notes en de groep notes op de crawlerserver:
  - a. Meld u aan als de hoofdgebruiker:  
su - root
  - b. Voeg een gebruiker toe:  
useradd notes
  - c. Voeg een wachtwoord toe voor deze gebruiker:  
passwd notes  
U wordt gevraagd het wachtwoord te wijzigen.
2. Installeer Lotus Domino Server op de crawlerserver:
  - a. Plaats de cd van Domino Server en installeer het product. (Als u geen cd hebt, kunt u de image downloaden.)
  - b. Ga naar de map waarin het besturingssysteem is geïnstalleerd.  
  
AIX: cd /mnt/cdrom/aix  
Linux: cd /mnt/cdrom/linux  
Solaris: cd /mnt/cdrom/solaris
  - c. Start het installatieprogramma:  
./install
  - d. Beantwoord de prompts en accepteer de standaardwaarden of geef andere installatie-instellingen op (zoals de paden voor de installatie- en gegevensdirectory's).  
Raadpleeg de documentatie bij Domino als u hulp nodig hebt bij het installeren van Domino Server.
3. Voer op de crawlerserver het installatiescript in dat door WebSphere II OmniFind Edition wordt geleverd:

- a. Meld u aan als de enterprise search-beheerder (dit gebruikers-ID is opgegeven tijdens de installatie van WebSphere II OmniFind Edition).
- b. Start het volgende script, dat in de directory \$ES\_INSTALL\_ROOT/bin is geïnstalleerd:  
escrnote.sh
- c. Beantwoord de prompts:
  - Beantwoord de volgende prompt met Y als Domino Server in de standaarddirectory is geïnstalleerd en beantwoord de prompt met N als dit niet het geval is:  
Het directorypad voor Lotus Notes /opt/lotus/notes/latest/linux is gevonden. Is dit het juiste pad voor Lotus Notes?

Het standaardpad voor AIX is /opt/lotus/notes/latest/ibmpow.  
 Het standaardpad voor Linux is /opt/lotus/notes/latest/linux.  
 Het standaardpad voor Solaris is /opt/lotus/notes/latest/sunspa.

- Als Domino Server niet in de standaarddirectory op de crawlersserver is geïnstalleerd, geeft u als antwoord op de volgende prompt op waar Domino is geïnstalleerd:  
Geef het pad op voor de Lotus Notes-directory

Op een Linux-computer kunt u bijvoorbeeld /opt/lotus/notes/latest/linux opgeven.

- Beantwoord de volgende prompt met Y als de Domino Server-gegevensdirectory in de standaarddirectory is geïnstalleerd en beantwoord de prompt met N als dit niet het geval is:  
Het gegevensdirectorypad voor Lotus Notes/local/notesdata is gevonden. Is dit het juiste pad van de Lotus Notes-gegevensdirectory?

Het standaardpad is /local/notesdata.

- Als de Domino Server-gegevensdirectory niet is geïnstalleerd op de standaardlocatie op de crawlersserver, geeft u het Domino-gegevenspad op als antwoord op de volgende prompt:  
Geef het pad op voor de Lotus Notes-gegevensdirectory.

4. Ga naar de crawlersserver, stop het enterprise search-systeem en start dit vervolgens opnieuw:

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

#### **Verwante onderwerpen**

“Domino Document Manager-crawlers” op pagina 58

Als u bibliotheken en kabinetten van Domino Document Manager wilt opnemen in een enterprise search-collectie, moet u een Domino Document Manager-crawler configureren.

“Notes-crawlers” op pagina 65

Als u IBM Lotus Notes-databases wilt opnemen in een enterprise search-collectie, moet u een Notes-crawler configureren.

“QuickPlace-crawlers” op pagina 75

Als u Lotus QuickPlace-ruimten en -rooms in een enterprise search-collectie wilt opnemen, moet u een QuickPlace-crawler configureren.

## De crawlerserver configureren op Windows voor het crawlen van Lotus Domino-bronnen

Als u WebSphere II OmniFind Edition installeert op een Microsoft Windows-computer en u servers wilt crawlen die gebruikmaken van het NRPC-protocol (Notes Remote Procedure Call), moet u een script uitvoeren om de crawlerserver te configureren. Met het script wordt communicatie tussen Notes-, QuickPlace- en Domino Document Manager-crawlers en de databaseservers ingesteld.

### Beperkingen

Lotus Domino Server en de Lotus Notes-client kunnen niet tegelijkertijd op dezelfde computer worden uitgevoerd met een Notes-, QuickPlace- of Domino Document Manager-crawler die is geconfigureerd voor gebruik van het NRPC-protocol. Als u probeert een van deze crawlers te starten terwijl Domino Server wordt uitgevoerd, treedt er een fout op en wordt de crawler gestopt.

### Over deze taak

De crawlers die het NRPC-protocol gebruiken, gebruiken Domino-client-bibliotheken. U kunt deze bibliotheken installeren door Lotus Domino Server op de crawlerserver te installeren. Om ervoor te zorgen dat de crawlers kunnen samenwerken met de Domino-bibliotheken moet u een installatiescript uitvoeren waarmee WebSphere II OmniFind Edition op de crawlerserver wordt geïnstalleerd nadat u de Domino-bibliotheken hebt geïnstalleerd.

### Procedure

Ga als volgt te werk om de crawlerserver zodanig te configureren, dat hiermee Lotus Notes-, Lotus QuickPlace- en Domino Document Manager-servers kunnen worden gecrawld:

1. Meld u bij de crawlerserver aan met een gebruikers-ID dat in de groep met beheerders is opgenomen. Zorg dat het gebruikers-ID machtiging heeft om Lotus Notes te installeren.
2. Installeer Lotus Notes:
  - a. Plaats de cd van Domino Server. (Als u geen cd hebt, kunt u de image downloaden.)
  - b. Start het installatieprogramma: setup.exe
  - c. Beantwoord de prompts en accepteer de standaardwaarden of geef andere installatie-instellingen op (zoals de paden voor de installatie- en gegevensdirectory's).  
Raadpleeg de documentatie bij Lotus Domino als u hulp nodig hebt.
3. Voer op de crawlerserver het installatiescript uit dat door WebSphere II OmniFind Edition wordt geleverd:
  - a. Meld u aan met het enterprise search-beheerders-ID (dit gebruikers-ID is opgegeven tijdens de installatie van WebSphere II OmniFind Edition).
  - b. Start het volgende script, dat in de directory %ES\_INSTALL\_ROOT%\bin is geïnstalleerd:  
escrnote.vbs
  - c. Beantwoord de prompts:
    - Beantwoord de volgende prompt met Y als Lotus Notes in de standaard-directory is geïnstalleerd en beantwoord de prompt met N als dit niet het geval is:

Het directorypad voor Lotus Notes c:\lotus\notes is gevonden.  
Is dit het juiste pad voor Lotus Notes?

Het standaardinstallatiepad op een Windows-computer is c:\lotus\notes of c:\lotus\domino.

- Als Lotus Notes niet in de standaarddirectory op de crawlerserver is geïnstalleerd, geeft u als antwoord op de volgende prompt op waar Lotus Notes is geïnstalleerd:

Geef het pad op voor de Lotus Notes-directory

- Beantwoord de volgende prompt met Y als de Lotus Notes-gegevensdirectory in de standaarddirectory is geïnstalleerd en beantwoord de prompt met N als dit niet het geval is:

Het directorypad voor Lotus Notes c:\lotus\notes\data is gevonden.  
Is dit het juiste pad van de Lotus Notes-gegevensdirectory?

Het standaardpad op een Windows-computer is c:\lotus\notes\data of c:\lotus\domino\data.

- Als de Lotus Notes-gegevensdirectory niet is geïnstalleerd op de standaardlocatie op de crawlerserver, geef u het gegevensdirectorypad op als antwoord op de volgende prompt:

Geef het pad op voor de Lotus Notes-gegevensdirectory.

4. Ga naar de crawlerserver, stop het enterprise search-systeem en start dit vervolgens opnieuw:

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

#### **Verwante onderwerpen**

“Domino Document Manager-crawlers” op pagina 58

Als u bibliotheken en kabinetten van Domino Document Manager wilt opnemen in een enterprise search-collectie, moet u een Domino Document Manager-crawler configureren.

“Notes-crawlers” op pagina 65

Als u IBM Lotus Notes-databases wilt opnemen in een enterprise search-collectie, moet u een Notes-crawler configureren.

“QuickPlace-crawlers” op pagina 75

Als u Lotus QuickPlace-ruimten en -rooms in een enterprise search-collectie wilt opnemen, moet u een QuickPlace-crawler configureren.

## **Servers configureren die het DIIOOP-protocol gebruiken**

Om servers te crawlen die gebruikmaken van het DIIOOP-protocol (Domino Internet Inter-ORB Protocol), moet u de server zodanig configureren dat de Notes-, QuickPlace- en Domino Document Manager-crawlers het protocol kunnen gebruiken.

### **Voordat u begint**

Op de server die u wilt crawlen, moeten de DIIOOP- en HTTP-taken worden uitgevoerd.

### **Procedure**

Ga als volgt te werk om servers te configureren die gebruikmaken van het DIIOOP-protocol:



1. Configureer het serverdocument:
  - a. Open het document server op de Lotus Notes-, Lotus QuickPlace- of Domino Document Manager-server die u wilt crawlen. Dit document is opgeslagen in de Domino-directory.
  - b. Vouw op de pagina Configuration de sectie **server** uit.
  - c. Geef in het gebied **Programmability Restrictions** op de pagina Security de beveiligingsbeperkingen voor uw omgeving in de volgende velden op:
    - **Run restricted Lotus Script/Java agents**
    - **Run restricted Java/Javascript/COM**
    - **Run unrestricted Java/Javascript/COM**

U kunt bijvoorbeeld een sterretje (\*) opgeven om onbeperkte toegang toe te staan voor Lotus Script/Java-agents, en de gebruikersnamen opgeven die in de Domino-directory voor de Java/Javascript/COM-beperkingen zijn geregistreerd.

**Belangrijk:** De crawler die u configureert om deze server te crawlen met het DIIOP-protocol, moet in staat zijn de gebruikersnamen te gebruiken die u in deze velden opgeeft.

- d. Open de pagina Internet Protocol, open vervolgens de pagina HTTP en stel de optie **Allow HTTP clients to browse database** in op **Yes**.
2. Configureer het gebruikersdocument:
  - a. Open het document user op de Lotus Notes-, Lotus QuickPlace- of Domino Document Manager-server die u wilt crawlen. Dit document is opgeslagen in de Domino-directory.
  - b. Geef in het veld **Internet password** op de pagina Basics een wachtwoord op.  
Wanneer u de beheerconsole van enterprise search gebruikt om opties te configureren voor het crawlen van deze server, voert u dit gebruikers-ID en wachtwoord in op de pagina waarop u de te crawlen server aangeeft. De crawler gebruikt deze legitimatiegegevens om toegang tot de server te krijgen.
3. Start de DIIOP-taak opnieuw op de server.

#### **Verwante onderwerpen**

“Domino Document Manager-crawlers” op pagina 58

Als u bibliotheken en kabinetten van Domino Document Manager wilt opnemen in een enterprise search-collectie, moet u een Domino Document Manager-crawler configureren.

“Notes-crawlers” op pagina 65

Als u IBM Lotus Notes-databases wilt opnemen in een enterprise search-collectie, moet u een Notes-crawler configureren.

“QuickPlace-crawlers” op pagina 75

Als u Lotus QuickPlace-ruimten en -rooms in een enterprise search-collectie wilt opnemen, moet u een QuickPlace-crawler configureren.

## **De I/O Completion Port op AIX configureren voor het crawlen van Lotus Domino-bronnen**

U kunt de Notes-, QuickPlace- of Domino Document Manager-crawlers pas op een IBM AIX-systeem gebruiken als de IOCP-module (I/O Completion Port) is geïnstalleerd en is geconfigureerd voor gebruik door de crawler.

**Over deze taak**

Zonder de IOCP-module kan het ontdekkingsproces niet worden uitgevoerd tijdens het maken van een crawler. Het volgende foutbericht wordt afgebeeld:

```
FFQM0105E Fout ontvangen van de server -  
Bericht: FFQG0024E Er is een onverwachte uitzondering afgevangen: ontdekken
```

Het volgende bericht, dat de ENOEXEC-fout bevat, wordt naar het bestand \$ES\_NODE\_ROOT/logs/system\_jjjjmmdd.log geschreven. (Bepaalde berichttekst is over meerdere regels verdeeld zodat de tekst beter leesbaar is.)

```
5/20/05 18:08:52.423 JST [Error] [ES_ERR_EXCEPTION_DEFAULT_MESSAGE] [] [discovery]  
iies10.yamato.ibm.com:0:2108088751:control:ComponentDiscoveryW.java:  
com.ibm.es.control.discovery.server.ComponentDiscoveryW.discover:86  
FFQ00277E Uitzondering opgetreden met het detail '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'  
and a stack trace of '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$NativeLibrary.load(Native Method)  
at java.lang.ClassLoader.loadLibrary0(ClassLoader.java:2120)  
at java.lang.ClassLoader.loadLibrary(ClassLoader.java:1998)  
at java.lang.Runtime.loadLibrary0(Runtime.java:824)  
at java.lang.System.loadLibrary(System.java:908)  
at lotus.domino.NotesThread.load(NotesThread.java:306)  
at lotus.domino.NotesThread.checkLoaded(NotesThread.java:327)  
at lotus.domino.NotesThread.sinitThread(NotesThread.java:181)  
at com.ibm.es.crawler.discovery.notes.NotesLibrary$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)
```

## Procedure

Ga als volgt te werk om de IOCP-module te installeren en te controleren of deze op de juiste wijze op de crawlerserver is geïnstalleerd:

U moet de volgende taken uitvoeren

1. Installeer de IOCP-module (bos.iocp.rte) vanaf de cd voor AIX op de crawlerserver.

Nadat u de IOCP-module hebt geïnstalleerd en voordat u een Notes-, QuickPlace- of Domino Document Manager-crawler maakt, moet u een softwarefix op de module toepassen. Raadpleeg de informatie bij de volgende link voor instructies:

<http://www.ibm.com/support/docview.wss?uid=swg21086556>

2. Geef de volgende opdracht op om te controleren of de IOCP-module op de crawlerserver is geïnstalleerd:

```
$ ls1pp -l bos.iocp.rte
```

De uitvoer van de opdracht ls1pp moet er ongeveer als volgt uitzien:

| Fileset                 | Level    | State     | Description              |
|-------------------------|----------|-----------|--------------------------|
| -----                   |          |           |                          |
| Path: /usr/lib/objrepos |          |           |                          |
| bos.iocp.rte            | 5.2.0.10 | COMMITTED | I/O Completion Ports API |

```
Path: /etc/objrepos
bos.iocp.rte 5.2.0.10 COMMITTED I/O Completion Ports API
```

3. Geef de volgende opdracht op om te controleren of de status van de IOCP-poort **Available** is:

```
$ lsdev -Cc iocp
```

De uitvoer van de opdracht lsdev moet er ongeveer als volgt uitzien:

```
iocp0 Available I/O Completion Ports
```

4. Als de IOCP-poortstatus **Defined** is, wijzigt u de status in **Available**:
  - a. Meld u bij de crawlerserver aan als hoofdgebruiker en geef de volgende opdracht op:

```
# smit iocp
```
  - b. Selecteer **Change / Show Characteristics of I/O Completion Ports** en wijzig **STATE to be configured at system restart** van **Defined** in **Available**.
  - c. Start de crawlerserver opnieuw op.
  - d. Geef de opdracht lsdev opnieuw op en bevestig dat de status van de IOCP-poort is gewijzigd in **Available**.

#### Verwante onderwerpen

“Domino Document Manager-crawlers” op pagina 58

Als u bibliotheken en kabinetten van Domino Document Manager wilt opnemen in een enterprise search-collectie, moet u een Domino Document Manager-crawler configureren.

“Notes-crawlers” op pagina 65

Als u IBM Lotus Notes-databases wilt opnemen in een enterprise search-collectie, moet u een Notes-crawler configureren.

“QuickPlace-crawlers”

Als u Lotus QuickPlace-ruimten en -rooms in een enterprise search-collectie wilt opnemen, moet u een QuickPlace-crawler configureren.

---

## QuickPlace-crawlers

Als u Lotus QuickPlace-ruimten en -rooms in een enterprise search-collectie wilt opnemen, moet u een QuickPlace-crawler configureren.

### Configuratie van de crawlerserver

Als de QuickPlace-server die u wilt crawlen, gebruikmaakt van het NRPC-protocol (Notes Remote Procedure Call), moet u een installatiescript op de crawlerserver uitvoeren. Dit script, dat bij WebSphere Information Integrator OmniFind Edition wordt geleverd, zorgt voor de communicatie tussen de QuickPlace-crawler en de servers die NRPC gebruiken.

Als de QuickPlace-server die u wilt crawlen, gebruikmaakt van het DIIOP-protocol (Domino Internet Inter-ORB Protocol), hoeft u geen installatiescript op de crawlerserver uit te voeren. U moet de QuickPlace-server echter zodanig configureren, dat de QuickPlace-crawler toegang tot de server kan krijgen.

Als de QuickPlace-server die u wilt crawlen, gebruikmaakt van een LDAP-server (Lightweight Directory Access Protocol), moet de QuickPlace-server worden geconfigureerd voor gebruik met het DIIOP-protocol (de QuickPlace-crawler kan het NRPC-protocol niet gebruiken voor het crawlen van LDAP-gegevens). Als u de LDAP-server als tweede Domino-server wilt gebruiken, moet u ook een Directory Assistance-database en de QuickPlace-server configureren.

**Belangrijk:** Als de QuickPlace-server gebruik maakt van DIIOP en u de crawler configureert voor het gebruik van HTTPS of DIIOP over SSL zodat de transmissie tussen de crawler en de server versleuteld wordt, moet u het bestand TrustedCerts.class (bijvoorbeeld c:\certs of /data/certs) kopiëren van de QuickPlace Notes Manager-server naar de crawlerserver. In een configuratie met twee of vier servers moet u het bestand TrustedCerts.class tevens kopiëren naar de servers waarop de zoekcomponent is geïnstalleerd. Zorg ervoor dat het bestand zich op de crawlerserver en de zoekservers op dezelfde locatie bevindt. U geeft het pad voor het bestand TrustedCerts.class op wanneer u de crawler configureert.

Als WebSphere II OmniFind Edition is geïnstalleerd op een IBM AIX-systeem, moet u ervoor zorgen dat de I/O Completion Port-module is geïnstalleerd en beschikbaar is op de crawlerserver.

Voordat u met behulp van de beheerconsole van enterprise search een QuickPlace-crawler configureert, voert u de taken uit die van toepassing zijn op uw omgeving:

- “De crawlerserver configureren op UNIX voor het crawlen van Lotus Domino-bronnen” op pagina 68.
- “De crawlerserver configureren op Windows voor het crawlen van Lotus Domino-bronnen” op pagina 71.
- “Servers configureren die het DIIOP-protocol gebruiken” op pagina 72.
- “De QuickPlace-server configureren voor gebruik van de beveiligingsoptie Lokale gebruiker” op pagina 261.
- “Directory Assistance op een QuickPlace-server configureren” op pagina 262.
- “De I/O Completion Port op AIX configureren voor het crawlen van Lotus Domino-bronnen” op pagina 73.

## Beveiliging op documentniveau

Als collectiebeveiliging is ingeschakeld en de server die u wilt crawlen, gebruikmaakt van het NRPC-protocol, moet u een Lotus Domino Trusted Server op de crawlerserver configureren. Trusted Server wordt gebruikt om toegangsbesturing op documentniveau af te dwingen. Voordat u de collectie doorzoekbaar maakt voor gebruikers, moet u de volgende taken uitvoeren:

- Lotus Domino Trusted Servers configureren voor het controleren van legitimatiegegevens van gebruikers.
- Algemene beveiliging inschakelen in WebSphere Application Server en het zoekprogramma configureren voor het gebruik van beveiliging. Met deze stap zorgt u ervoor dat de legitimatiegegevens van gebruikers worden gecontroleerd wanneer gebruikers proberen het zoekprogramma te gebruiken. Deze legitimatiegegevens kunnen vervolgens op de zoekservers worden gebruikt om de machtigingen van gebruikers voor Lotus Domino-documenten te controleren.

## Bijlagen crawlen

In Lotus QuickPlace kunt u Microsoft Office-documenten importeren en publiceren. (De opties die u selecteert bij het importeren zijn Imported Page, Microsoft Word Page, Microsoft Excel Page, Microsoft PowerPoint Page en Multiple Imported Pages). De QuickPlace-crawler kan deze typen geïmporteerde documenten crawlen, echter uitsluitend onder de volgende voorwaarden:

- De server gebruikt het DIIOP-protocol.
- U hebt het crawlen van bijlagen ingeschakeld tijdens het configureren van crawl-opties voor de crawler.

- U hebt de crawler geconfigureerd voor het crawlen van het veld "\$FILE" of voor alle velden.

## Configuratieoverzicht

U kunt met de QuickPlace-crawler een willekeurig aantal QuickPlace-ruimten crawlen. Als u wilt crawlen vanaf één QuickPlace-server, selecteert u de ruimten tijdens het maken van de crawler. Tijdens het bewerken van de crawlruimte kunt u de documenten die u in dezelfde crawlruimte wilt opnemen, vanaf een andere QuickPlace-server toevoegen. Bij het maken of bewerken van de crawler kunt u aangeven of u alle rooms in de ruimten die u voor crawlen hebt geselecteerd, wilt crawlen of dat u specifieke rooms wilt crawlen.

**Beperking:** Wanneer u een gebruikers-ID opgeeft voor de crawler, moet u zorgen dat het ID voldoende machtigingen heeft zodat toegang tot alle QuickPlace-ruimten in het Domino-domein kan worden verkregen. Hiertoe configureert u de QuickPlace-server en wijst u een gebruikers-ID toe aan de gereserveerde groep QuickPlaceAdministratorsSUGroup. De groepsnaam bevat geen spaties en is hoofdlettergevoelig.

Als u een QuickPlace-crawler wilt maken of wijzigen, meldt u zich aan bij de beheerconsole van enterprise search. U moet de rol van beheerder van enterprise search hebben of beheerder zijn van de collectie die eigenaar is van de crawler.

Bij het definiëren van de crawler moet u de volgende procedure uitvoeren. Hierbij wordt u geholpen door de wizard:

- Geef eigenschappen op die bepalen hoe de crawler werkt en op welke manier de systeemresources worden gebruikt. De crawlereigenschappen bepalen hoe de crawler alle documenten in de crawlruimte crawlt.
- Geef de QuickPlace-server en het communicatieprotocol op.
- Als de server is geconfigureerd voor gebruik van het DIIOP-protocol, kunt u opgeven hoe de crawler verbinding met Domino-objecten moet maken. U kunt bijvoorbeeld opties opgeven voor het gebruik van HTTPS of SSL (Secure Sockets Layer) om de communicatie te coderen.
- Geef informatie op over de gebruikersdirectory die aan de server is gekoppeld (deze informatie is nodig in de crawler zodat de toeganginstellingen kunnen worden gehandhaafd wanneer gebruikers de collectie doorzoeken).
- Selecteer de ruimten die u wilt crawlen.
- Geef een rooster op voor het crawlen van de ruimten.
- Selecteer de documenten die u wilt crawlen. De crawler kan alle rooms in een ruimte crawlen, of alleen de documenten die zich in de geselecteerde rooms bevinden.
- Geef opties op om de velden in de verschillende ruimten en rooms doorzoekbaar te maken. U kunt bijvoorbeeld bepaalde velden uitsluiten van de crawlruimte en opties opgeven voor het doorzoeken van bijlagen.
- Configureer beveiligingsopties op documentniveau. Als bij het maken van de collectie de beveiliging is ingeschakeld, kan de crawler beveiligingsgegevens koppelen aan documenten in de index. Met behulp van die beveiligingsgegevens kunnen zoekprogramma's toegangsbesturing uitvoeren op basis van opgeslagen toegangslijsten of beveiligingstokens.

U kunt ook een optie selecteren waarmee de legitimatiegegevens van gebruikers worden gecontroleerd wanneer een query wordt ingediend. Op deze manier

worden de legitimatiegegevens vergeleken met de huidige toegangslijsten die door de oorspronkelijke gegevensbron worden beheerd en niet met de geïndexeerde beveiligingsgegevens.

Als u bij het definiëren van de crawler meer informatie wenst over de velden in de wizard en over de manier waarop u de informatie moet opgeven die de crawler nodig heeft om informatie te crawlen, klikt u in de wizard op **Help**.

#### **Verwante onderwerpen**

“Tips voor het crawlen van Lotus Domino-databases” op pagina 67

Lees de richtlijnen voor het crawlen van Lotus Domino-databases door voordat u een Notes-crawler configureert.

“Beveiliging op documentniveau afdwingen voor Lotus Domino-documenten” op pagina 259

Als de te crawlen Domino-server gebruikmaakt van het NRPC-protocol (Notes Remote Procedure Call), moet u de crawlerserver zodanig configureren, dat toegangsbesturing op documentniveau kan worden afgedwongen.

#### **Verwante taken**

“De crawlerserver configureren op UNIX voor het crawlen van Lotus Domino-bronnen” op pagina 68

Als u WebSphere II OmniFind Edition installeert op een computer waarop IBM AIX, Linux of de Solaris-besturingsomgeving wordt uitgevoerd en u servers wilt crawlen die gebruikmaken van het NRPC-protocol (Notes Remote Procedure Call), moet u een script uitvoeren om de crawlerserver te configureren. Met het script wordt communicatie tussen Notes-, QuickPlace- en Domino Document Manager-crawlers en de databaseservers ingesteld.

“De crawlerserver configureren op Windows voor het crawlen van Lotus Domino-bronnen” op pagina 71

Als u WebSphere II OmniFind Edition installeert op een Microsoft Windows-computer en u servers wilt crawlen die gebruikmaken van het NRPC-protocol (Notes Remote Procedure Call), moet u een script uitvoeren om de crawlerserver te configureren. Met het script wordt communicatie tussen Notes-, QuickPlace- en Domino Document Manager-crawlers en de databaseservers ingesteld.

“Servers configureren die het DIIOP-protocol gebruiken” op pagina 72

Om servers te crawlen die gebruikmaken van het DIIOP-protocol (Domino Internet Inter-ORB Protocol), moet u de server zodanig configureren dat de Notes-, QuickPlace- en Domino Document Manager-crawlers het protocol kunnen gebruiken.

“De I/O Completion Port op AIX configureren voor het crawlen van Lotus Domino-bronnen” op pagina 73

U kunt de Notes-, QuickPlace- of Domino Document Manager-crawlers pas op een IBM AIX-systeem gebruiken als de IOCP-module (I/O Completion Port) is geïnstalleerd en is geconfigureerd voor gebruik door de crawler.

“Lotus Domino Trusted Servers configureren voor het controleren van legitimatiegegevens van gebruikers” op pagina 260

Om beveiliging af te dwingen voor documenten die zijn gecrawld met een Notes-crawler die gebruikmaakt van het NRPC-protocol (Notes Remote Procedure Call), moeten de te crawlen Domino-servers worden geconfigureerd als Lotus Domino Trusted Servers.

“De QuickPlace-server configureren voor gebruik van de beveiligingsoptie Lokale gebruiker” op pagina 261

Als u van plan bent een QuickPlace-crawler te configureren voor gebruik van

de optie Lokale gebruiker om beveiliging te implementeren, moet u de Domino-directory op de Lotus QuickPlace-server configureren voordat u de crawler maakt.

“Directory Assistance op een QuickPlace-server configureren” op pagina 262  
Als u van plan bent een QuickPlace-crawler te configureren voor gebruik van een LDAP-directory om beveiliging te implementeren, moet u een Directory Assistance-database op de Lotus QuickPlace-server maken voordat u de crawler configureert.

---

## UNIX-bestandssysteemcrawlers

Als u documenten die in AIX-, Linux- of Solaris-bestandssystemen zijn opgeslagen in een enterprise search-collectie wilt opnemen, moet u een UNIX-bestandssysteem-crawler configureren.

U kunt met de UNIX-bestandssysteem-crawler een willekeurig aantal bestandssystemen crawlen. Tijdens het configureren van de crawler selecteert u de lokale en niet-lokale directory's en subdirectory's die u wilt crawlen.

Als u de crawlerserver op een Windows-computer hebt geïnstalleerd, kunt u deze server niet gebruiken om AIX-, Linux- of Solaris-bestandssysteembronnen te crawlen (de UNIX-bestandssysteem-crawler verschijnt dan niet in de lijst van beschikbare crawlertypen).

De UNIX-bestandssysteem-crawler crawlt documenten volgende leesmachtigingen die voor de enterprise search-beheerder zijn opgegeven.

Als u een UNIX-bestandssysteem-crawler wilt maken of wijzigen, meldt u zich aan bij de beheerconsole van enterprise search. U moet de rol van beheerder van enterprise search hebben of beheerder zijn van de collectie die eigenaar is van de crawler.

Bij het definiëren van de crawler moet u de volgende procedure uitvoeren. Hierbij wordt u geholpen door de wizard:

- Geef eigenschappen op die bepalen hoe de crawler werkt en op welke manier de systeemresources worden gebruikt. De crawlereigenschappen bepalen hoe de crawler alle subdirectory's in de crawlruimte crawlt.
- Geef een rooster op voor het crawlen van de bestandssystemen.
- Selecteer de subdirectory's en de niveaus van de subdirectory's die u door de crawler wilt laten crawlen.
- Geef opties op waarmee de documenten in de subdirectory's doorzoekbaar worden gemaakt. Zo kunt u bijvoorbeeld bepaalde typen documenten uitsluiten uit de crawlruimte.
- Configureer beveiligingsopties op documentniveau. Als bij het maken van de collectie de beveiliging is ingeschakeld, kan de crawler beveiligingsgegevens koppelen aan documenten in de index. Met behulp van die beveiligingsgegevens kunnen zoekprogramma's toegangsbesturing uitvoeren op basis van opgeslagen toegangslijsten of beveiligingstokens.

Als u bij het definiëren van de crawler meer informatie wenst over de velden in de wizard en over de manier waarop u de informatie moet opgeven die de crawler nodig heeft om informatie te crawlen, klikt u in de wizard op **Help**.

---

## Web-crawlers

Als u pagina's van websites in een enterprise search-collectie wilt opnemen, moet u een webcrawler configureren.

Met de webcrawler kunt u een willekeurig aantal HTTP-servers (Hypertext Transfer Protocol) en beveiligde HTTPS-servers (Secure HTTP) crawlen. De crawler bezoekt een website en leest de gegevens op de site. De crawler volgt vervolgens de links in documenten om aanvullende documenten te crawlen. De webcrawler kan links crawlen en extraheren van afzonderlijke pagina's of van *framesets* (pagina's die zijn gemaakt met HTML-frames).

De gecrawlde gegevens kunnen zijn opgeslagen in een van de verschillende algemene indelingen en kunnen afkomstig zijn van diverse bronnen op intranet of internet. Voorbeelden van algemene indelingen zijn HTML, PDF, Microsoft Word, Lotus WordPro, Extensible Markup Language (XML), enzovoort.

Als u een webcrawler wilt maken of wijzigen, meldt u zich aan bij de beheerconsole van enterprise search. U moet de rol van beheerder van enterprise search hebben of beheerder zijn van de collectie die eigenaar is van de crawler.

Bij het definiëren van de crawler moet u de volgende procedure uitvoeren. Hierbij wordt u geholpen door de wizard:

- Geef eigenschappen op die bepalen hoe de crawler werkt en op welke manier de systeemresources worden gebruikt. De crawlereigenschappen bepalen hoe de crawler alle webpagina's in de crawlruimte crawlt.
- Geef regels op om het bezoek aan bepaalde websites toe te staan en te verbieden. Bij het opgeven van crawlregels kunt u de regels testen en controleren of de crawler de sites kan bezoeken die u in de crawlruimte wilt opnemen.
- Geef opties op om bepaalde bestandstypen op te nemen en bestanden met bepaalde extensies uit te sluiten.
- Geef aan op welke manier soft error-pagina's in de crawler moeten worden verwerkt.
- Configureer beveiligingsopties op documentniveau. Als bij het maken van de collectie de beveiliging is ingeschakeld, kan de crawler beveiligingsgegevens koppelen aan documenten in de index. Met behulp van die beveiligingsgegevens kunnen zoekprogramma's toegangsbesturing uitvoeren op basis van opgeslagen toegangslijsten of beveiligingstokens.
- Geef opties op voor het crawlen van websites die met een wachtwoord zijn beveiligd (op de webservers die moeten worden gecrawld, moet voor de wachtwoordprompt HTTP Basic Authentication of moeten HTML-formulieren worden gebruikt).
- Geef opties op voor het crawlen van websites die worden bestuurd door een proxyserver.

Als u bij het definiëren van de crawler meer informatie wenst over de velden in de wizard en over de manier waarop u de informatie moet opgeven die de crawler nodig heeft om informatie te crawlen, klikt u in de wizard op **Help**.

## Configuratie van de gebruikersagent

Als u een website wilt crawlen waarop het Robots Exclusion Protocol wordt gebruikt, zorgt u dat de naam van de gebruikersagent die u voor de webcrawler hebt geconfigureerd voor toegang tot de website, is toegestaan in het bestand robots.txt op de website.



Wanneer u het enterprise search-systeem start, laadt de webcrawler de naam van de gebruikersagent die u voor de crawler hebt gedefinieerd. Voordat de crawler een pagina downloadt van een website die nog niet eerder is bezocht (of die al enige tijd niet is bezocht), probeert de crawler eerst het bestand robots.txt te downloaden. Dit bestand bevindt zich in de hoofddirectory van de website.

Als het bestand robots.txt niet bestaat, gelden voor de website geen beperkingen voor het crawlen. Als het bestand wél bestaat, wordt in het bestand aangegeven welke gedeelten van de site (directory's) niet beschikbaar zijn voor crawlers. In het bestand robots.txt vindt u de machtigingen voor crawlers die worden aangegeven op basis van de naam van de gebruikersagent.

Het Robots Exclusion Protocol is niet verplicht. De webcrawler van enterprise search voldoet in de volgende gevallen aan het protocol:

- Als het bestand robots.txt een vermelding bevat voor de naam van de gebruikersagent die voor de webcrawler is geconfigureerd, voldoet de webcrawler aan de beperkingen voor die gebruikersagent.
- Als de naam van de gebruikersagent niet voorkomt in het bestand robots.txt, de laatste vermelding User-agent: \* is (wat staat voor een willekeurige gebruikersagent) en de beperking Disallow: / is (wat betekent dat crawlen niet is toegestaan, beginnend bij de root van de website), kan de webcrawler niet crawlen op de site.
- Als de naam van de gebruikersagent niet voorkomt in het bestand robots.txt is, de laatste vermelding User-agent: \* is en de beperking Allow: / is, kan de webcrawler de site crawlen.

Websitebeheerders nemen vaak een definitieve vermelding op waarmee de toegang wordt geblokkeerd voor alle crawlers waaraan niet expliciet toegang is verleend. Als u een nieuwe webcrawler configureert en weet dat bepaalde websites die u wilt crawlen, gebruikmaken van het Robots Exclusion Protocol, vraagt u of de websitebeheerder een vermelding voor uw crawler opneemt in de robots.txt-bestanden.

Zorg dat u dezelfde naam van de gebruikersagent in de eigenschappen van de webcrawler opgeeft en in alle robots.txt-bestanden die deel uitmaken van de gewenste websites.

Als geen van de websites die u wilt crawlen, gebruikmaakt van het Robots Exclusion Protocol, is de waarde die u voor de eigenschap van de gebruikersagent opgeeft, niet van belang. Bepaalde toepassingenservers, JSP's en servlets passen hun antwoord echter aan de naam van de gebruikersagent aan. Zo bestaan er verschillende antwoorden voor het verwerken van browserincompatibiliteit. De naam van de gebruikersagent die u voor de webcrawler opgeeft, kan van belang zijn in een dergelijke situatie, ongeacht het Robots Exclusion Protocol. Als u deze typen sites wilt crawlen, neemt u contact op met de websitebeheerders om te controleren of de webcrawler is toegestaan op de site.

## **Gebruik van het Robots Exclusion Protocol in de webcrawler**

De webcrawler probeert het Robots Exclusion Protocol na te leven en geen websites te crawlen als in de regels van het bestand robots.txt op de server is aangegeven dat crawlen niet is toegestaan.

Het downloaden is probleemloos uitgevoerd als de crawler het bestand robots.txt vanaf een webserver kan ophalen of kan bevestigen dat het bestand robots.txt niet

bestaat. Het downloaden is niet uitgevoerd als de crawler de regels niet kan ophalen of niet kan bevestigen dat het bestand robots.txt bestaat.

Een probleemloze download houdt echter niet in dat de crawler is gemachtigd om te crawlen omdat het crawlen kan worden verboden op basis van de regels in het bestand robots.txt. Als een download mislukt, kan het crawlen tijdelijk worden verboden omdat de crawler niet kan vaststellen welke regels actief zijn.

In het volgende overzicht vindt u de stappen die de crawler uitvoert tijdens het downloaden van het bestand robots.txt:

1. Als de crawler een nieuwe site ontdekt, wordt geprobeerd het IP-adres van de server te verkrijgen. Als deze poging mislukt, is crawlen niet mogelijk.
2. Wanneer ten minste één IP-adres beschikbaar is, probeert de crawler het bestand robots.txt te downloaden met behulp van HTTP (of HTTPS) GET.
3. Als er een timeout optreedt voor de socketverbinding, deze wordt verbroken of wanneer er een andere low level-fout is opgetreden (zoals een SSL-certificaat-probleem), logt de crawler het probleem en wordt de poging herhaald voor elk bekend IP-adres voor de doelserver.
4. Als de crawler alle adressen heeft geprobeerd en er geen verbinding is gemaakt, wacht de crawler twee seconden en worden alle adressen nogmaals geprobeerd.
5. Als er een verbinding is gemaakt en de HTTP-headers zijn uitgewisseld, wordt de retourstatus onderzocht. Als de statuscode 500 of hoger is, wordt dit door de crawler geïnterpreteerd als een slechte verbinding en probeert de crawler een ander IP-adres. Voor de overige statuswaarden probeert de crawler geen alternatieve IP-adressen en voert de crawler de stap uit die vereist is voor de bijbehorende statuscode.

Als de crawler een HTTP-statuscode lager dan 500 ontvangt of als de crawler alle IP-adressen tweemaal heeft geprobeerd, gaat de crawler als volgt te werk:

1. Als er geen HTTP-status lager dan 500 is ontvangen, wordt de site voorlopig gediskwalificeerd.
2. Als de HTTP-status 400, 404 of 410 is ontvangen, kan de site zonder regels worden gecrawld.
3. Als een HTTP-status tussen 200 en 299 is ontvangen, is de stap die moet worden genomen, afhankelijk van de volgende voorwaarden:
  - Als de content is afgekapt, wordt de site voorlopig gediskwalificeerd.
  - Als de content zonder fouten is geanalyseerd, kan de site worden gecrawld met behulp van de aangetroffen regels.
  - Als de content met fouten is geanalyseerd, kan de site worden gecrawld zonder regels.
4. Als er een andere HTTP-status is geretourneerd, is de site voorlopig gediskwalificeerd.

Als de crawler probeert het bestand robots.txt voor een site te downloaden, wordt een permanente tijdsaanduiding bijgewerkt voor die site, ook wel de robotdatum genoemd. Als een site is gediskwalificeerd omdat de informatie in het bestand robots.txt niet beschikbaar is, wordt de telling van het aantal fouten voor de robot verhoogd.

Als het herhalingsinterval is bereikt, probeert de crawler opnieuw de informatie in het bestand robots.txt voor de mislukte site op te halen. Als het aantal achtereenvolgende mislukte pogingen het maximum aantal toegestane pogingen heeft

bereikt, stopt de crawler met het ophalen van het bestand robots.txt voor de site en wordt de site gediskwalificeerd voor crawlen.

Als een site is gekwalificeerd voor crawlen (als de controle van het bestand robots.txt probleemloos is uitgevoerd), wordt de telling van het aantal mislukte pogingen opnieuw op nul ingesteld. De crawler gebruikt de resultaten van de download totdat het interval voor het controleren van de regels is verstreken. Op dat moment moet de site opnieuw worden gekwalificeerd.

**Tip:**

- Als een server content retourneert maar syntaxisfouten bevat, als de server een ander robotprotocol gebruikt dan versie 1994 of als de content andere informatie bevat dat robotregels (zoals een soft error-pagina), gaat de crawler te werk alsof er geen bestand met regels aanwezig is en wordt de site gecrawld. Meestal is deze actie juist, omdat collectiebeheerders de sitecontent of het standaardservergedrag niet besturen. Als een webserverbeheerder niet wil dat een site wordt gecrawld en geen bestand met regels installeert, kan de collectiebeheerder de site vanuit de webcrawler blokkeren door het domein, het IP-adres of de HTTP-prefix van de site op te geven in de crawlerregels.
- Als een server statuscode 302 of een andere omleidingscode retourneert, interpreteert de crawler de code alsof voor de site een robots.txt-bestand aanwezig is dat moet worden gebruikt, maar het bestand zich niet op de juiste locatie bevindt (de siteroot). De webserverbeheerder moet het bestand naar de juiste locatie verplaatsen zodat de webcrawler de regels in het bestand kan volgen.
- Als er problemen met certificaten zijn opgetreden (als het certificaat bijvoorbeeld verouderd is, de certificaatgever niet betrouwbaar is of het certificaat zelfondertekend is terwijl de crawler niet is geconfigureerd voor het accepteren van zelfondertekende certificaten), kan de crawler geen verbinding met de site maken en wordt de site gediskwalificeerd. Deze problemen zorgen er waarschijnlijk toch al voor dat andere pagina's van de site niet kunnen worden gecrawld. Om de site in te schakelen om te worden gecrawld, moet de collectiebeheerder zelfondertekende certificaten inschakelen, de machtiging van de site toevoegen aan het betrouwbare sleutelruimtebestand of de webserverbeheerder vragen om een up-to-date certificaat.
- De webcrawler is mogelijk geconfigureerd voor het gebruik van HTTP Basic Authentication (inclusief HTTP Basic Proxy Authentication). Indien op de juiste wijze geconfigureerd, is er ook verificatie vereist voor downloads van het bestand robots.txt. Met de statuscode 403, 407 of andere met verificatie samenhangende reacties, worden problemen met de machtiging aangegeven, waarna de crawler de site diskwalificeert. (Alleen HTTP Basic Authentication wordt ondersteund.)
- Als het bestand robots.txt voor een site de maximale lengte voor een robotpagina overschrijdt, kan de collectiebeheerder de geconfigureerde maximumlengte vergroten (de standaardwaarde van één miljoen bytes zou voldoende moeten zijn).

Als hulp bij het oplossen van problemen kunt u een siterapport aanvragen wanneer u de webcrawler bewaakt. Selecteer opties om de content van het bestand robots.txt te bekijken (om te controleren of de toegang van de webcrawler tot de site is verboden op basis van de regels), om de datum en tijd te bekijken waarop de crawler voor het laatst heeft geprobeerd het bestand robots.txt te downloaden (de crawler probeert het bestand pas weer te downloaden als het herhalingsinterval is verstreken) en te bekijken hoeveel opeenvolgende pogingen de crawler heeft gedaan om het bestand robots.txt (zonder resultaat) te downloaden. Klik tijdens het bewaken van de webcrawler op **Help** voor meer informatie over deze siterapportopties en over de manier waarop u de resultaten moet interpreteren.

Informatie over het Robots Exclusion Protocol vindt u op <http://www.robotstxt.org/wc/exclusion.html>.

## Ondersteuning voor JavaScript

De webcrawler voor enterprise search kan bepaalde links (URL's) vinden die zijn opgenomen in de gedeelten met JavaScript van webdocumenten.

De webcrawler kan zowel relatieve als absolute links vinden. Als een HTML-document een BASE-element bevat, gebruikt de crawler dat element om relatieve links om te zetten. Anders gebruikt de crawler de eigen URL van het document.

Ondersteuning voor JavaScript is beperkt tot de extractie van links. De crawler analyseert JavaScript niet, bouwt geen DOM (Document Object Model), interpreteert geen JavaScript-instructies en voert deze ook niet uit. De crawler zoekt in de documentcontent (inclusief, maar niet beperkt tot de JavaScript-gedeelten) naar reeksen die lijken op URL's in JavaScript-instructies. Dit houdt twee zaken in:

- Bepaalde URL's worden gevonden die door de striktere HTML-parser worden genegeerd. De crawler wijst alle items af die syntactisch gezien geen geldige URL's zijn. Het kan echter zijn dat bepaalde geldige URL's die door het scanproces worden geretourneerd, minder interessant zijn voor zoekacties.
- De documentcontent die door JavaScript wordt gegenereerd, bijvoorbeeld wanneer een gebruiker een pagina bekijkt met een browser en de browser JavaScript uitvoert, is niet zichtbaar voor de webcrawler en wordt dus ook niet in de index opgenomen.

## Regels voor het beperken van de webcrawlruijnte

Om ervoor te zorgen dat gebruikers alleen toegang hebben tot de websites die u instelt, geeft u regels op om beperkingen op te leggen voor de webcrawler.

Tijdens het crawlen van een webpagina ontdekt de webcrawler links naar andere pagina's en worden deze links in een wachtrij gezet zodat deze later kunnen worden gecrawld. Het crawlen en ontdekken kan net zolang worden herhaald als door de tijd- en geheugenresources is toegestaan. Bij het configureren van een webcrawler geeft u op waar de crawler moet beginnen met crawlen. Op basis van deze eerste URL's (die ook wel *start-URL's* worden genoemd) heeft de webcrawler via directe of indirecte links toegang tot alle documenten op het web.

Om de crawlruijnte te beperken, kunt u de webcrawler zodanig configureren, dat bepaalde URL's grondig kunnen worden gecrawld en de links worden genegeerd die buiten het interessegebied liggen. Omdat de crawler standaard alle ontdekte URL's accepteert, moet u regels opgeven waarmee de URL's worden aangegeven die u in de collectie wilt opnemen en de rest van de pagina's wordt verwijderd.

U kunt op verschillende manieren aangeven welke URL's door de webcrawler moeten worden gecrawld. U kunt het volgende opgeven:

- Een lijst van start-URL's waar de crawler moet beginnen met crawlen.
- Drie typen crawlregels: domein, IP-adres (Internet Protocol) en URL-prefix
- Een lijst van MIME-typen voor de documenten die u wilt opnemen
- Een lijst van bestandsextensies voor de documenten die u wilt uitsluiten
- Het maximum aantal directory's in een URL-pad

Crawlregels hebben de volgende indeling:  
actie type doel

actie is verbieden of toestaan; type is domein, IP-adres of URL-prefix (HTTP of HTTPS); en doel is afhankelijk van de waarde van type. Om doelen aan te geven die voldoen aan een patroon, kunt u een sterretje (\*) gebruiken (hiervoor gelden beperkingen).

## Domeinregels

Het doel van een domeinregel is een DNS-domeinnaam. U kunt bijvoorbeeld opgeven dat het volledige domein `www.ibm.com` kan worden gecrawld:

```
allow domain www.ibm.com
```

U kunt een sterretje als jokerteken opgeven, zodat de regel van toepassing is op elke hostnaam die eindigt op de rest van het patroon. U kunt bijvoorbeeld opgeven dat de domeinen die met de server beginnen en eindigen op `ibm.com`, niet moeten worden gecrawld:

```
forbid domain server*.ibm.com
```

De hostnamen zijn hoofdlettergevoelig, ongeacht of u een expliciete domeinnaam of een domeinnaampatroon opgeeft. Bijvoorbeeld: `*.user.ibm.com` voldoet aan `joe.user.ibm.com` en `mary.smith.user.ibm.com`, maar niet aan `joe.user.IBM.com`.

Een domeinregel waarmee geen poortnummer wordt aangegeven, is van toepassing op alle poorten in dat domein. In het volgende voorbeeld zijn alle poorten in het domein `sales` toegestaan:

```
allow domain sales.ibm.com
```

Als met een domeinregel een poortnummer wordt aangegeven, is de regel alleen van toepassing op die poort. In het volgende voorbeeld is alleen poort 443 in het domein `sales` toegestaan:

```
allow domain sales.ibm.com:443
```

## Prefixregels

Met een prefixregel bepaalt u op welke manier URL's worden gecrawld die met een bepaalde tekenreeks beginnen. Het doel is één URL, die doorgaans een of meer sterretjes bevat om een patroon aan te geven. Vaak wordt een sterretje gebruikt als het laatste teken in de prefixtekenreeks.

Een prefixregel maakt het mogelijk om een gehele website of een gedeelte van een website te crawlen. U kunt een directorypad of -patroon opgeven en vervolgens alles vanaf dat punt in de directorystructuur toestaan of verbieden. In het volgende voorbeeld werken de regels samen om toe te staan dat de crawler alles in de directory `public` van `sales.ibm.com` kan crawlen, en om toegang van de crawler tot de andere pagina's op de site te verbieden:

```
allow prefix http://sales.ibm.com/public/*  
forbid prefix http://sales.ibm.com/*
```

Bij het opgeven van prefixregels kunt u meerdere sterretjes opgeven en kunt u deze op een willekeurige plek in de prefixtekenreeks invoeren, met uitzondering van de laatste positie. Met de volgende regel wordt bijvoorbeeld verboden dat de crawler de documenten op het hoogste niveau van de directory van de site `sales.ibm.com` kan crawlen als de directorynaam eindigt op `fs`. (Bijvoorbeeld wanneer er bestandssysteeminstallaties zijn die geen nuttige informatie voor de zoekindex bevatten.)

```
forbid http://sales.ibm.com/*fs/*
```

## Adresregels

Met een adresregel kunt u het crawlen van volledige hosts of netwerken beheren door een IP-adres (alleen IPv4) en netwerkmasker als het doel op te geven. Bijvoorbeeld:

```
allow address 9.0.0.0 255.0.0.0
```

Met het netwerkmasker kunt u patroonmatching opgeven. Een adresregel is alleen van toepassing op een kandidaat-IP-adres als het IP-adres in de regel en het kandidaat-IP-adres identiek zijn, behalve waar sprake is van maskering met nullen in het netmasker. Met de adresregel wordt een patroon gedefinieerd en met het netwerkmasker worden de significante bits in het adrespatroon gedefinieerd. Een nul in het netwerkmasker fungeert als jokerteken en geeft aan dat elke waarde die in diezelfde bitpositie in het adres wordt opgegeven, overeenkomt.

In het vorige voorbeeld is de regel voor toestaan van toepassing op elk IP-adres met de waarde 9 voor het eerste octet en met een willekeurige waarde voor de laatste drie octets.

De regel `forbid address 0.0.0.0 0.0.0.0` is uitermate geschikt om als laatste adres in de lijst van adressen op te nemen. Alle IP-adressen voldoen aan deze regel omdat het netwerkmasker alle bits insignificant maakt (de regel verbiedt alle adressen die niet zijn toegestaan door een eerdere regel in de lijst van regels).

**Beperkingen voor proxyservers:** Als u van plan bent websites te crawlen die worden bestuurd door een proxyserver, moet u geen regels voor IP-adressen opgeven. Een proxyserver wordt meestal gebruikt wanneer een gebruikersagent (browser of crawler) geen directe toegang heeft tot de netwerken waarop de webserver zijn geïnstalleerd. Een HTTP-proxyserver kan bijvoorbeeld HTTP-aanvragen van een crawler aan een webserver doorgeven en de antwoorden weer doorgeven aan de crawler.

Wanneer een webcrawler gebruikmaakt van een proxyserver, is het IP-adres van de proxyserver het enige IP-adres dat voor de crawler bestaat voor andere hosts. Als regels voor IP-adressen worden gebruikt om de crawler te beperken tot een subnet van IP-adressen, worden hierdoor vrijwel alle URL's geclassificeerd met de statuscode 760 (wat aangeeft dat ze verboden zijn door de webruimte).

## Volgorde van crawlregels

De crawler past de crawlregels op verschillende momenten tijdens het ontdekken en crawlen van URL's toe. De volgorde van de regels is van belang, maar dan alleen binnen de regels van elk type. Het maakt verschil of een adresregel vóór of na een andere adresregel wordt toegepast, maar het maakt geen verschil of een adresregel vóór of na een prefixregel wordt toegepast omdat de crawler de regels niet op hetzelfde moment toepast.

Binnen de set regels voor een type test de crawler het domein, het adres of de URL van een kandidaat ten opzichte van elke regel, waarbij wordt begonnen met de eerste opgegeven regel totdat er een regel wordt gevonden die van toepassing is. De actie die is opgegeven voor de eerste regel die van toepassing is, wordt gebruikt.

De afhankelijkheid van de volgorde heeft voor de meeste crawlregels een typische structuur als gevolg:

- De set domeinregels begint doorgaans met regels voor verbieden, waarmee afzonderlijke domeinen uit de crawlruimte worden verwijderd. De collectiebeheerder kan bijvoorbeeld vaststellen dat bepaalde domeinen geen nuttige informatie bevatten.
- De lijst van regels voor verbieden wordt doorgaas gevolgd door een reeks regels voor toestaan (met jokertekens), waarmee de crawler elk domein kan bezoeken dat eindigt op een van de domeinnamen op hoog niveau waarmee het intranet van een bedrijf wordt aangegeven, zoals \*.ibm.com en \*.lotus.com.

Eindig de set van domeinregels met de volgende standaardregel, waarmee de domeinen worden verwijderd die op basis van een eerdere regel niet zijn toegestaan:

```
forbid domain *
```

Deze laatste regel is van essentieel belang omdat hiermee wordt voorkomen dat de het volledige internet in de crawlruimte wordt opgenomen.

- De set van adresregels begint doorgaans met een klein aantal regels voor toestaan waarmee de crawler de netwerken op hoog niveau (klasse-A, klasse-B of klasse-C) kan crawlen die een bedrijfsintranet omvatten.

De adresregels eindigen meestal met de volgende regel, waarmee wordt voorkomen dat de crawler websites crawlt die zich buiten het bedrijfsnetwerk bevinden.

```
forbid 0.0.0.0 0.0.0.0
```

- De set van prefixregels is doorgaans het grootst omdat deze gedetailleerde specificaties van de toegestane en verboden gebieden bevat die zijn opgegeven als structuren en substructuren. Aangeraden wordt om eerst de strenger gelokaliseerde gebieden toe te staan of te verbieden en vervolgens de tegenovergestelde regel in een algemener patroon op te geven om de andere gebieden toe te staan of te verbieden.

Het prefixgedeelte eindigt niet altijd met een bepaalde regel. Met de voorgestelde laatste domein- en adresregels kunt u ervoor zorgen dat de crawler niet buiten het bedrijfsnetwerk crawlt en dat dit efficiënter gebeurt dan met het testen van URL-prefixen.

De crawler kan prefixregels efficiënter toepassen als u de regels op basis van actie groepeerd (verbieden of toestaan). In plaats van korte reeksen met regels voor toestaan en verbieden op te geven die elkaar afwisselen, geeft u een lange reeks met regels op die een bepaalde actie voorschrijven en geeft u vervolgens een lange reeks met regels op die de andere actie voorschrijven. U kunt regels voor toestaan en verbieden samenvoegen om zo de gewenste crawlruimte te krijgen. Als u de regels voor toestaan en verbieden groepeerd, kunnen de prestaties van de crawler worden verbeterd.

## Bestandsextensies, MIME-typen en maximale crawldiepte

Met deze opties beschikt u over extra manieren om de content voor de crawlruimte op te geven. U kunt bepaalde documenttypen uitsluiten op basis van de bestandsextensies van documenten en u kunt bepaalde documenttypen opnemen op basis van het MIME-type van het document. Als u opgeeft welke MIME-typen door de crawler moeten worden gecrawld, moet u er rekening mee houden dat het MIME-type in webdocumenten vaak onjuist is ingesteld.

De maximale crawldiepte is het aantal schuine strepen in een URL vanaf de siteroot. Met deze optie kunt u voorkomen dat de crawler terechtkomt in recursieve bestandssysteemstructuren met een oneindige diepte. De crawldiepte komt

niet overeen met de niveaus waardoor de crawler bladert tijdens het volgen van links tussen documenten.

## Start-URL's

Start-URL's zijn de URL's waar de crawler begint met crawlen en deze URL's worden in de crawl ingevoegd wanneer de crawler wordt gestart. Als de start-URL's al zijn ontdekt, worden ze niet eerder gecrawld of opnieuw gecrawld dan de andere websites die op basis van de crawlregels zijn toegestaan.

Een start-URL is van belang de eerste keer dat een webcrawler wordt gestart en de crawlruimte leeg is. Daarnaast is een start-URL belangrijk wanneer u een URL die niet eerder is ontdekt, aan de lijst van start-URL's in een crawlruimte toevoegt.

Start-URL's moeten volledige URL's zijn en niet alleen de domeinnamen. U moet het protocol opgeven en als de poort niet 80 is, moet u ook het poortnummer opgeven.

De volgende URL's zijn geldige start-URL's:

```
http://w3.ibm.com/  
http://sales.ibm.com:9080/
```

De volgende URL is geen geldige start-URL:

```
www.ibm.com
```

U moet de start-URL's in de crawlruimte opnemen. De crawler kan niet beginnen met het crawlen van de opgegeven start-URL als de crawlregels verbieden dat de URL kan worden gecrawld.

## URL-verbindingen met de webcrawler testen

Als u in de webcrawler de URL's hebt opgegeven die moeten worden gecrawld, kunt u de configuratie van de crawlregels testen.

Bij het opgeven van de te crawlen domeinen, HTTP-prefixen of IP-adressen kunt u op **Testen** klikken of u kunt de pagina URL's testen selecteren om te testen of de crawler verbinding kan maken met de start-URL's of met de URL's die u opgeeft.

De testresultaten laten zien of de crawler toegang heeft tot URL's met de naam van de gebruikersagent die is opgegeven in de crawlereigenschappen. Daarnaast kan worden aangegeven dat een bepaalde URL niet kan worden gecrawld omdat er een uitsluitingsregel van kracht is. (Het zou bijvoorbeeld kunnen zijn dat een document niet wordt gecrawld omdat het een extensie heeft die is uitgesloten van de crawlruimte.)

Nadat een site minimaal één keer gecrawld is, kunt u de test gebruiken om aanvullende informatie te verkrijgen. Het testrapport kan bijvoorbeeld de meest recente HTTP-statuscode bevatten (welke aangeeft of de crawl van een URL al dan niet gelukt is), kan aangeven wanneer de URL voor het laatst is gecrawld en wanneer de volgende crawl volgens planning plaatsvindt, en kan aangeven of de gebruikersagent gebruik maakt van het actuele bestand robots.txt van de server.

## Intervalinstellingen voor opnieuw crawlen in de webcrawler

Als u de manier wilt beïnvloeden waarop de webcrawler URL's opnieuw bezoekt, geeft u bepaalde opties op in de eigenschappen van de webcrawler.



De meeste andere crawlertypen in een enterprise search-systeem worden uitgevoerd op basis van roosters die door de beheerder worden opgegeven. Een webcrawler daarentegen wordt continu uitgevoerd nadat deze is gestart. Als u wilt instellen hoe vaak de crawler opnieuw URL's bezoekt die al eerder zijn gecrawld, geeft u minimum- en maximumintervallen op voor opnieuw crawlen.

Als u de beheerconsole van enterprise search gebruikt om een webcrawler te maken of de eigenschappen van de webcrawler te bewerken, kunt u een optie selecteren waarmee u geavanceerde eigenschappen kunt configureren. Op de pagina Geavanceerde eigenschappen van webcrawler geeft u de minimum- en maximumintervallen voor opnieuw crawlen op. De webcrawler gebruikt de waarden die u opgeeft om een interval te berekenen voor het opnieuw crawlen van gegevens.

De eerste keer dat een pagina wordt gecrawld, gebruikt de crawler voor het instellen van een datum voor opnieuw crawlen de datum en tijd waarop de pagina is gecrawld en het gemiddelde van de opgegeven minimum- en maximumintervallen. De pagina zal niet opnieuw worden gecrawld vóór die datum. Wanneer de pagina na die datum opnieuw wordt gecrawld, is afhankelijk van de belasting van de crawler en de verhouding tussen oude en nieuwe URL's in de crawlruimte.

Telkens wanneer de pagina opnieuw wordt gecrawld, controleert de crawler of de content is gewijzigd. Als de content is gewijzigd, zal het volgende interval voor opnieuw crawlen korter zijn dan het vorige interval, maar niet korter dan het opgegeven minimuminterval voor opnieuw crawlen. Als de content niet is gewijzigd, zal het volgende interval voor opnieuw crawlen langer zijn dan het vorige interval, maar niet langer dan het opgegeven maximuminterval voor opnieuw crawlen.

## Opties voor het bezoeken van URL's met de webcrawler

U kunt instellen dat de webcrawler bepaalde URL's zo snel mogelijk bezoekt.

Als u de crawlruimte moet vernieuwen met gegevens van bepaalde websites, kunt u de crawler bewaken, de optie **(Opnieuw) te bezoeken URL's** selecteren en vervolgens de URL's of URL-patronen opgeven van de pagina's die (opnieuw) moeten worden gecrawld.

Als de afdeling Communicatie van uw bedrijf bijvoorbeeld een webpagina aan het intranet toevoegt of wijzigingen in een pagina aanbrengt om een belangrijke beleidsregel te publiceren, kunt u de URL van de nieuwe of gewijzigde pagina opgeven. Als de crawler actief is, wordt de opgegeven URL in de wachtrij gezet zodat de URL kan worden gecrawld de eerstvolgende keer dat de crawler controleert of er pagina's in de wachtrij aanwezig zijn die kunnen worden bezocht (standaard elke tien minuten). Als de crawler niet actief is, wordt de opgegeven URL in de wachtrij gezet zodat de URL kan worden gecrawld zodra de crawler wordt gestart.

Zorg dat in de crawlregels een regel is opgenomen voor het bezoeken van de URL's die u opgeeft. De crawler kan de URL's die u opgeeft, eerder bezoeken dan normaal het geval zou zijn. Ook in dit geval geldt dat een URL pas kan worden gecrawld als er een crawlregel bestaat waarmee wordt aangegeven dat de URL kan worden gecrawld.

De nieuw gecrawelde gegevens zijn pas beschikbaar om te worden doorzocht nadat er de volgende keer een hoofdindex is opgebouwd.

## De manier waarop de webcrawler omgaat met soft error-pagina's

U kunt de webcrawler configureren voor het afhandelen van aangepaste pagina's die door websitebeheerders worden gemaakt als zij geen standaardfoutcodes willen gebruiken in het antwoord op aanvragen voor bepaalde pagina's.

Als een HTTP-server de pagina die door een client wordt aangevraagd niet kan verzenden, stuurt de server normaal gesproken een antwoord dat uit een header en een statuscode bestaat. De statuscode geeft aan wat het probleem is (fout 404 geeft bijvoorbeeld aan dat het bestand niet is gevonden). Sommige websitebeheerders maken speciale pagina's met een gedetailleerde uitleg van het probleem en configureren de HTTP-server zodanig dat deze pagina's in plaats van de standaardfoutcodes worden verzonden. Deze aangepaste pagina's worden *soft error-pagina's* genoemd.

Soft error-pagina's kunnen de resultaten van de webcrawler vertekenen. Bijvoorbeeld: in plaats van een header die een probleem aangeeft, kan het zijn dat de crawler een soft error-pagina ontvangt plus de statuscode 200, die aangeeft dat een geldige HTML-pagina is gedownload. Deze gedownloade soft error-pagina heeft echter niets te maken met de aangevraagde URL en de content ervan is, elke keer dat deze pagina wordt teruggestuurd in plaats van de aangevraagde pagina, nagenoeg identiek. Deze irrelevante en vrijwel identieke pagina's vervuilen de index en de zoekresultaten.

Om deze situatie te omzeilen, kunt u tijdens het configureren van de webcrawler opties opgeven voor het afhandelen van soft error-pagina's. De webcrawler heeft de volgende informatie nodig over alle websites die soft error-pagina's verzenden:

- Een URL-patroon voor de site waarvoor de soft error-pagina's worden gebruikt. Dit URL-patroon bestaat uit het protocol (HTTP of HTTPS), de hostnaam, het poortnummer (als het standaardpoortnummer niet wordt gebruikt) en de padnaam. U kunt een sterretje (\*) als jokerteken gebruiken, waarmee een of meer tekens in het patroon overeenkomen, tot het eerstvolgende niet-jokerteken. Het patroon dat u opgeeft, is hoofdlettergevoelig.
- Een titelpatroon voor tekst die overeenkomt met de <TITLE>-tag van een HTML-document. U kunt een sterretje (\*) gebruiken als jokerteken om dit patroon op te geven. Het patroon dat u opgeeft, is hoofdlettergevoelig.
- Een contentpatroon voor tekst die overeenkomt met de content van een HTML-document. Als er een <BODY>-tag aanwezig is, is de content is niet alleen de content van de <BODY>-tag. De content is alles dat volgt op de HTTP-header van het bestand. U kunt een sterretje (\*) gebruiken als jokerteken om dit patroon op te geven. Het patroon dat u opgeeft, is hoofdlettergevoelig.
- Een geheel getal waarmee de statuscode wordt aangegeven die wordt gebruikt voor documenten waarvan de opgegeven URL-, titel- en contentpatronen overeenkomen.

### Voorbeeld

Met de volgende configuratie moet de webcrawler alle geldige HTML-pagina's (statuscode 200) vergelijken die vanaf de website [http://www.mijnsite.com/hr/\\*](http://www.mijnsite.com/hr/*) zijn teruggestuurd naar de opgegeven titel- en contentpatronen. Als de <TITLE>-tag van een pagina begint met "Sorry, de pagina" en de content van het document

alles bevat (\*), wordt de pagina in de crawler op dezelfde manier afgehandeld als met statuscode 404 (de pagina is niet gevonden).

Tabel 3. Voorbeeld van soft error-pagina

| URL-patroon                  | Titelpatroon      | Contentpatroon | HTTP-statuscode |
|------------------------------|-------------------|----------------|-----------------|
| http://www.mijnsite.com/hr/* | Sorry, de pagina* | *              | 404             |

Voor het afhandelen van verschillende statuscodes kunt u meerdere vermeldingen voor dezelfde website maken. Voor elke statuscode van dezelfde website is een eigen vermelding in de configuratie van de webcrawler nodig.

## Jokertekens gebruiken

De URL-, titel- en contentpatronen zijn geen expressies. Het jokerteken komt overeen met alle tekens in het patroon tot het eerstvolgende niet-jokerteken. Bijvoorbeeld:

\*404 komt overeen met *alle tekens*404

404: \* komt overeen met 404: *alle tekens*

http://\*.mijnsite.com/\* komt overeen met *http://alle hosts.mijnsite.com/alle bestanden*

\* komt overeen met *alle tekens*

## Invloed op prestaties

Wanneer u opties voor het afhandelen van soft error-pagina's configureert, heeft de crawler meer tijd nodig voor de verwerking omdat alle gecrawelde pagina's moeten worden gecontroleerd. Er is meer verwerkingstijd nodig om de patroonmatches te controleren en te bepalen of een pagina of een vervangende statuscode moet worden verzonden.

## Ondersteuning voor het crawlen van beveiligde websites

Als u legitimatiegegevens opgeeft in de beheerconsole van enterprise search, kunt u instellen dat de webcrawler toegang heeft tot beveiligde content, zoals documenten waarvoor een toegangswachtwoord is vereist.

Als een webserver voor het beperken van de toegang tot websites gebruikmaakt van HTTP Basic Authentication of HTML Form-Based Authentication, kunt u in de configuratie van de webcrawler legitimatiegegevens opgeven aan de hand waarvan de crawler toegang kan krijgen tot de pagina's van met een wachtwoord beveiligde websites. Daarnaast kunt u opties opgeven voor het handmatig configureren van cookiebestanden.

### Door HTTP Basic Authentication beveiligde websites

Als een webserver voor het beperken van de toegang tot websites gebruikmaakt van HTTP Basic Authentication, kunt u legitimatiegegevens opgeven aan de hand waarvan de webcrawler toegang kan krijgen tot pagina's die met een wachtwoord zijn beveiligd.

Om vast te stellen of een gebruiker (of clienttoepassing) toegang heeft tot de pagina's van een website, gebruiken veel webserver de clientverificatieprocedure HTTP Basic Authentication om de identiteit van de gebruiker te bepalen. Deze interactie is meestal interactief:

- Wanneer een HTTP-gebruikersagent (zoals een webbrowser) een pagina aanvraagt die is beveiligd met behulp van HTTP Basic Authentication, retourneert

de webserver statuscode 401, waarmee wordt aangegeven dat de aanvrager niet is gemachtigd om de aangevraagde pagina weer te geven.

- De webserver probeert via de aanvrager de legitimatiegegevens te verkrijgen die kunnen worden gebruikt om te controleren of de gebruiker toegang heeft.
- De webbrowser geeft een dialoogvenster weer waarin de gebruiker wordt gevraagd een gebruikersnaam, wachtwoord en andere informatie in te voeren die nodig is voor de legitimatiegegevens van de gebruiker.
- De webbrowser codeert de legitimatiegegevens en gebruikt deze gegevens wanneer de aanvraag voor de beveiligde pagina opnieuw wordt ingediend.
- Als de legitimatiegegevens geldig zijn, antwoordt de webserver met retourcode 200 en wordt de content van de aangevraagde pagina weergegeven.
- Voor de volgende aanvragen voor pagina's van dezelfde webserver worden meestal dezelfde legitimatiegegevens gebruikt, zodat de bevoegde gebruiker meer beveiligde pagina's kan bekijken zonder hiervoor steeds de legitimatiegegevens op te geven.

Als de identiteit van een gebruiker is vastgesteld, wisselen de webserver en de HTTP-gebruikersagent meestal tokens uit, die ook wel *cookies* worden genoemd. Met cookies zorgt u ervoor dat de aanmeldstatus van de gebruiker bekend is tussen de verschillende HTTP-aanvragen.

Omdat de webcrawler niet interactief wordt uitgevoerd, moet u voordat het crawlproces start, de legitimatiegegevens opgeven aan de hand waarvan de met een wachtwoord beveiligde pagina's kunnen worden gecrawld. Tijdens het maken van een webcrawler of het bewerken van de crawlruimte geeft u informatie op over alle beveiligde websites die u wilt crawlen.

Om deze informatie te kunnen opgeven, moet u nauw samenwerken met de beheerders van de websites of webserver die worden beveiligd met behulp van HTTP Basic Authentication. De beheerders moeten u de beveiligingsvereisten leveren voor de websites die u wilt crawlen, inclusief de informatie die moet worden gebruikt om de identiteit van de webcrawler te verifiëren en te kunnen vaststellen of de crawler bevoegd is om de beveiligde pagina's te crawlen.

Als tijdens het maken van de collectie beveiliging is ingeschakeld voor de collectie, kunt u beveiligingstokens opgeven (zoals gebruikers-ID's, groeps-ID's of gebruikersrollen), om de toegang tot documenten te beheren wanneer u de crawler configureert. De webcrawler koppelt deze beveiligingstokens aan elk document dat in het bestandssysteem voor de opgegeven hoofd-URL wordt gecrawld. De tokens worden gebruikt naast de beveiligingstokens op documentniveau die u voor de volledige webcrawlruimte hebt geconfigureerd.

De volgorde van de URL's is van belang. Nadat u informatie over een met een wachtwoord beveiligde website hebt opgegeven, moet u de site op de gewenste positie in de crawlvolgorde plaatsen. Zet de meer specifieke URL's boven aan de lijst, en de meer algemene URL's onderaan. Wanneer de webcrawler een kandidaat-URL evalueert, worden de verificatiegegevens gebruikt die zijn opgegeven voor de eerste URL in de lijst die overeenkomt met de kandidaat-URL.

### **Door Form-Based Authentication beveiligde websites**

Als een webserver voor het beperken van de toegang tot websites gebruikmaakt van Forms-Based Authentication, kunt u legitimatiegegevens opgeven aan de hand waarvan de webcrawler toegang kan krijgen tot pagina's die met een wachtwoord zijn beveiligd.

Om vast te stellen of een gebruiker (of clienttoepassing) toegang heeft tot de pagina's van een website, gebruiken veel webserver HTML-formulieren om de identiteit van de gebruiker te bepalen. Deze interactie is meestal interactief:

- Wanneer een HTTP-gebruikersagent (zoals een webbrowser) een pagina aanvraagt die is beveiligd met behulp van Form-Based Authentication, controleert de webserver of de aanvraag een cookie bevat waarmee de identiteit van de gebruiker kan worden vastgesteld.
- Als er geen cookie aanwezig is, wordt de gebruiker gevraagd de beveiligingsgegevens op een formulier in te voeren. Wanneer de gebruiker het formulier indient, retourneert de webserver de vereiste cookies en kan de aanvraag voor de met een wachtwoord beveiligde pagina worden voortgezet.
- Latere aanvragen met de vereiste cookie kunnen ook worden voortgezet. De bevoegde gebruiker kan aanvullende beveiligde content bekijken zonder dat voor elke aanvraag opnieuw een formulier hoeft te worden ingevoerd en legitimatiegegevens hoeven te worden opgegeven.

Omdat de webcrawler niet interactief wordt uitgevoerd, moet u voordat het crawlproces start, de legitimatiegegevens opgeven aan de hand waarvan de met een wachtwoord beveiligde pagina's kunnen worden gecrawld. Tijdens het maken van een webcrawler of het bewerken van de crawlruimte geeft u informatie op over alle beveiligde websites die u wilt crawlen.

De velden die u opgeeft, komen overeen met de velden die een interactieve gebruiker invult wanneer de webbrowser hierom vraagt, of met verborgen of statische velden die vereist zijn voor de aanmelding.

Om deze informatie te kunnen opgeven, moet u nauw samenwerken met de beheerders van de websites of webserver die worden beveiligd met behulp van Form-Based Authentication. De beheerders moeten u de beveiligingsvereisten leveren voor de websites die u wilt crawlen, inclusief de informatie die moet worden gebruikt om de identiteit van de webcrawler te verifiëren en te kunnen vaststellen of de crawler bevoegd is om de beveiligde pagina's te crawlen.

De volgorde van de URL-patronen is van belang. Nadat u informatie over een met een wachtwoord beveiligde website hebt opgegeven, moet u de site op de gewenste positie in de crawlvolgorde plaatsen. Zet de meer specifieke URL-patronen bovenaan in de lijst en de meer algemene URL-patronen onderaan. Wanneer de webcrawler een kandidaat-URL evalueert, worden de formuliergegevens gebruikt die zijn opgegeven voor het eerste URL-patroon in de lijst dat overeenkomt met de kandidaat-URL.

## Websites die worden bediend door proxyservers

Als de webcrawler geen directe toegang tot een netwerk heeft, kunt u de crawler zodanig configureren dat deze een HTTP-proxyserver gebruikt om toegang te krijgen tot de content die moet worden gecrawld.

Als toegang tot een TCP/IP-netwerk niet beschikbaar is op de computer waarop de crawler moet worden uitgevoerd of als toegang alleen beschikbaar is voor processen die hiervoor zijn gemachtigd, kunt u de webcrawler configureren voor gebruik van een HTTP-proxyserver. Een HTTP-proxy is een proces dat op een opgegeven poort luistert voor HTTP-opdrachten. De proxyserver geeft de aanvragen door aan de webserver, waarna de antwoorden van de webserver worden doorgestuurd naar de aanvragende client (de webcrawler). Een proxyserver kan op dezelfde computer als de webcrawler of op een andere computer worden uitgevoerd.

Als u crawlt zonder gebruik te maken van een proxyserver, worden URL-aanvragen direct naar de host verzonden. Als u crawlt met een proxyserver, worden aanvragen naar de proxyserver verzonden.

Tijdens het maken van een webcrawler of het bewerken van de crawlruimte geeft u informatie op over de proxyserver die de webcrawler moet gebruiken bij het crawlen van pagina's in het domein van de proxyserver. Voordat u een proxyserver aan de crawlruimte toevoegt, moet u de namen opgeven van de domeinen die door de proxyserver worden bediend, de hostnaam of het IP-adres van de proxyserver en het poortnummer van de proxyserver.

Als voor de proxyserver verificatie is vereist, moet u de gebruikersnaam en het wachtwoord opgeven dat de webcrawler kan gebruiken om toegang te krijgen tot de pagina's die door de proxyserver worden bediend. De webcrawler ondersteunt alleen HTTP Basic Proxy Authentication, zoals beschreven in RFC2616 (<http://rfc.net/rfc2616.html>). Overige verificatietypen, waaronder NTLM (Windows NT LAN Manager), worden niet ondersteund.

Nadat u een proxyserver hebt toegevoegd, moet u deze selecteren en op de gewenste positie in de crawlvolgorde plaatsen. Zet de meer specifieke domeinnamen bovenaan in de lijst, en de meer algemene onderaan. Wanneer de webcrawler een kandidaat-URL evalueert, worden de proxyservergegevens gebruikt die zijn opgegeven voor het eerste domein in de lijst die overeenkomt met de kandidaat-URL. (URL's die niet overeenkomen met een proxyregel worden gezien als URL's die direct toegankelijk zijn voor de crawler.)

## Beheer van cookies

In de meeste gevallen vindt het beheer van cookies automatisch plaats en is er geen interventie van de enterprise search-beheerder vereist. Indien noodzakelijk kunt u handmatig cookies opgeven voor een webcrawlingsessie.

Cookies zijn verborgen tokens die door webserver aan gebruikersagents worden geretourneerd als onderdeel van HTTP-responsheaders. Cookies zijn alleen van belang voor de webserver die ze verstrekt en worden gebruikt om de status tussen HTTP-aanvragen te beheren. Tijdens de clientverificatie kan de webserver bijvoorbeeld een cookie retourneren waarmee de server kan bepalen of een geverifieerde gebruiker al is aangemeld. Via de cookie kan de gebruiker extra pagina's op die webserver aanvragen, zonder dat de gebruiker zich hiervoor opnieuw hoeft aan te melden.

De webcrawler behoudt de cookies die van webserver zijn ontvangen en gebruikt deze gedurende de crawlerinstance. De cookies worden opgeslagen in het bestand `cookies.ini`, dat aan het einde van elke crawlersessie opnieuw wordt geschreven door de crawler. Wanneer de webcrawler stopt, worden alle niet-verlopen cookies opgeslagen, waarna ze opnieuw worden geladen wanneer de volgende sessie wordt gestart.

Als u handmatig cookies opgeeft, slaat u deze indien nodig in een apart bestand op en voegt u ze samen met de cookies in het bestand `cookies.ini`. Niet-verlopen cookies worden niet gewist door de crawler. Als er echter een probleem optreedt als gevolg waarvan de hele cookieverzameling niet kan worden geschreven, wilt u waarschijnlijk voorkomen dat de handmatig opgegeven cookies verloren gaan. Voordat u een crawlersessie start, moet u in dit geval de handmatig opgegeven cookies samenvoegen met de cookies die automatisch in de crawler worden behouden.

## Cookie-indeling

De cookies die u wilt samenvoegen met het enterprise search-bestand cookies.ini, moeten in een bepaalde indeling zijn opgeslagen.

- Elke cookie moet op een aparte regel zijn geplaatst. Lege regels en commentaar zijn toegestaan, maar worden niet bewaard in het bestand cookies.ini.
- Elke cookie moet in de volgende indeling zijn opgeslagen:

*CookieN(cookie\_lengte,URL\_lengte)cookie\_tekst,validatie\_URL*

### Cookie

Een vereist trefwoord waarmee het begin van een cookievermelding wordt aangegeven.

Het trefwoord Cookie kan geen spaties bevatten en aan het woord moet één cijfer worden toegevoegd, 0, 1 of 2. Met het cijfer wordt het cookietype aangegeven: versie-0 (Netscape), versie-1 (RFC2109) of versie-2 (RFC2965).

Poortlijsten worden niet ondersteund in RFC2965-cookies.

### *cookie\_lengte*

De lengte (in tekens) van de bijbehorende cookietekst.

### *URL\_lengte*

De lengte (in tekens) van de bijbehorende validatie-URL.

### *cookie\_tekst*

De content van de cookie die naar de oorspronkelijke webserver moet worden verzonden. Met deze tekenreeks (die het rechterdeel van de instructie Set-Cookie in een HTTP-responsheader voorstelt) worden de naam, het waardepaar en andere content van de cookie aangegeven (zoals het pad, de beveiligingsinstellingen, enzovoort) die met de cookie moeten worden verzonden. Als scheidingsteken van deze tekenreeks wordt de komma gebruikt (,).

### *validatie\_URL*

De URL waarop deze cookie is ontdekt. Deze URL wordt gebruikt om te bepalen naar welke locatie de cookie moet worden verzonden (door de aanwezigheid van bijvoorbeeld een domeinnaam en padnaam). De validatie-URL moet voldoen aan de beveiligings- en privacybeperkingen voor cookies die op de oorspronkelijke webserver zijn ingesteld.

In het volgende voorbeeld zijn de cookies op twee regels opgegeven, zodat u ze beter kunt lezen. U moet cookies op één regel opgeven:

```
Cookie0(53,40)ASPSESSIONIDQSQTACSD=SLNSIDFNLSIDNFLSINFLSNL;path=/  
https://www.ibm.com:443/help/solutions/
```

## Cookies configureren voor de webcrawler

U kunt handmatig cookies opgeven voor een webcrawlingsessie en deze samenvoegen met de cookies die door de webcrawler worden onderhouden.

### Voordat u begint

Om handmatig cookies voor de webcrawler te configureren, moet u een enterprise search-beheerder zijn.

### Procedure

Ga als volgt te werk om handmatig cookies voor een webcrawler te configureren:

1. Open de beheerconsole van enterprise search, bewaak de collectie waarvoor u cookies wilt opgeven en stop de webcrawler.

2. Meld u bij de crawlerserver aan als enterprise search-beheerder. (Dit gebruikers-ID is opgegeven tijdens de installatie van WebSphere II OmniFind Edition.)
3. Ga naar de gegevensdirectory voor de crawler die u wilt configureren, waarbij *crawler\_sessie\_ID* het ID is dat door het enterprise search-systeem aan de crawlersessie is toegewezen. Bijvoorbeeld:  
`ES_NODE_ROOT/data/col_56092.WEB_88534`
4. Open het bestand cookies.ini, voeg de gewenste cookie-items toe aan de items die al worden weergegeven, sla het bestand op en sluit het bestand vervolgens af. Zorg dat de cookies reeds aanwezige cookies niet worden overschreven.
5. Ga naar de beheerconsole van enterprise search en start de webcrawler opnieuw die u hebt gestopt.

## Configuratie van algemene webcrawlruijnte

U kunt een algemene crawlruijnte configureren voor webcrawlers, zodat u het verwijderen van URL's uit de index beter kunt beheeren.

Elke webcrawler wordt geconfigureerd met een crawlruijnte waarmee de URL's worden gedefinieerd die al dan niet moeten worden gecrawld. De ontdekte URL's die in de crawlruijnte zijn opgenomen, worden (in een database) bewaard voor latere crawling. De URL's die niet in de crawlruijnte zijn opgenomen, worden gewist. Als de crawler met een lege database start, blijven de crawlruijntedefinitie en de database consistent terwijl de crawler actief is.

Soms wordt een crawler gestopt en wordt de crawlruijnte verkleind (bijvoorbeeld als gevolg van nieuwe regels waarmee bepaalde pagina's niet mogen worden gecrawld). Wanneer de crawler opnieuw wordt gestart, zijn de crawlruijntedefinitie en de database niet meer consistent. De database bevat URL's (waarvan een aantal wel en een aantal niet wordt gecrawld) die niet aanwezig zijn in de nieuwe, kleinere crawlruijnte.

Als een collectie slechts één webcrawler bevat, kan de webcrawler de consistentie herstellen door de HTTP-statuscodes voor deze URL's te wijzigen in 760 (wat aangeeft dat de URL's moeten worden uitgesloten) en een aanvraag in te dienen voor het verwijderen van de pagina's die nu van de index zijn uitgesloten.

Als u de crawlruijnte verdeelt tussen twee of meer webcrawlers (om er bijvoorbeeld voor te zorgen dat een aantal pagina's vaker wordt gecrawld dan de rest), behoudt elke webcrawler onafhankelijke databasetabellen (die in eerste instantie leeg zijn) en crawlt elke crawler een ander deel van de webcrawlruijnte. De crawlruijnte van de oorspronkelijke crawler wordt vervolgens verkleind tot de ruimte die over is nadat de delen die door de andere crawlers worden gecrawld, zijn verwijderd. Als de oorspronkelijke crawler de consistentie probeert te herstellen door de verplaatste pagina's uit de index te verwijderen, ontstaat er een probleem. Omdat de verplaatste pagina's nu door andere crawlers worden gecrawld, moeten de pagina's in de index aanwezig blijven.

Door een algemene crawlruijnte op een hoger niveau te configureren, kunt u aangegeven welke URL's niet door de oorspronkelijke crawler moeten worden gecrawld, maar ook niet uit de index moeten worden verwijderd. De URL's die niet meer in de crawlruijnte van een crawler aanwezig zijn, worden via de ontdekkingsprocessen gemarkeerd voor uitsluiting en worden tijdens het volgende crawlproces uit de index verwijderd.



De algemene crawlruimte wordt gedefinieerd via het configuratiebestand `global.rules`, dat moet zijn opgenomen in de `crawlerconfiguratedirectory` (als het bestand `global.rules` aanwezig is, is de functie voor de algemene crawlruimte ingeschakeld). Als het bestand bestaat, wordt dit tijdens de initialisatie van de crawler gelezen. Als het bestand niet bestaat, functioneert de crawler met een crawlruimte die uit één niveau bestaat en worden de documenten indien nodig uit de index verwijderd om consistentie tussen de crawlruimtedefinitie en de database te waarborgen.

Als er een algemene crawlruimte aanwezig is, worden de URL's opgenomen of uitgesloten zoals eerder aangegeven, maar wordt de verwijdering van een URL uit de index alleen aangevraagd als de URL niet is opgenomen in een webcrawlruimte.

Het bestand `global.rules` heeft dezelfde syntaxis als het lokale bestand `crawl.rules`, met als uitzondering dat het alleen domeinnaamregels kan bevatten. Deze beperking zorgt ervoor dat een crawlruimte alleen tussen crawlers kan worden gepartitioneerd op basis van DNS-hostnamen en niet op basis van IP-adressen of HTTP-prefixpatronen. URL's die zijn uitgesloten op basis van URL-prefix- of IP-adresregels in de lokale crawlruimte (zoals gedefinieerd in het bestand `crawl.rules`) worden niet beïnvloed door de algemene crawlruimte; dergelijke URL's worden nog steeds uitgesloten.

De algemene crawlruimte wordt alleen gebruikt om te voorkomen dat URL's op basis van een lokale domeinregel voor de crawlruimte van een crawler van de index worden uitgesloten. De volgende regels zijn van toepassing (in de weergegeven volgorde):

1. Als een URL van de crawlerdatabase wordt uitgesloten op basis van een lokale prefixregel of adresregel, wordt statuscode 760 aan de URL toegewezen en wordt de URL uit de index verwijderd. De URL wordt dan niet meer gecrawld.
2. Als een URL van de crawlerdatabase wordt uitgesloten op basis van een lokale domeinregel en er geen algemene crawlruimte aanwezig is, wordt statuscode 760 aan de URL toegewezen en wordt de URL uit de index verwijderd. De URL wordt dan niet meer gecrawld.
3. Als een URL van de crawlerdatabase wordt uitgesloten op basis van een lokale domeinregel, maar expliciet wordt toegestaan door een regel in de algemene crawlruimte, wordt statuscode 761 aan de URL toegewezen. De crawler zal de URL niet opnieuw crawlen, maar de URL wordt niet uit de index verwijderd (aangenomen wordt dat de URL is de lokale crawlruimte van een andere crawler is opgenomen).
4. Als een URL van de crawlerdatabase wordt uitgesloten op basis van een lokale domeinregel en de URL niet expliciet wordt toegestaan door een regel in de algemene crawlruimte, wordt statuscode 760 aan de URL toegewezen en wordt de URL uit de index verwijderd.

Omdat de algemene crawlruimte alleen wordt geraadpleegd om te voorkomen dat de URL's worden verwijderd die al zijn uitgesloten door de lokale crawlruimte, wordt de URL niet meer gecrawld als er geen regel van toepassing is op een kandidaat-URL.

Het bestand `global.rules` moet aanwezig zijn in de `directory master_config` van elke crawler die de algemene crawlruimte gemeenschappelijk gebruikt. U moet alle exemplaren van het bestand `global.rules` en het afzonderlijke, lokale bestand `crawl.rules` aanpassen zodat deze consistent zijn.

## De instructies no-follow en no-index

U kunt de kwaliteit van zoekopdrachten verbeteren door instructies voor de webcrawler op te geven waarmee wordt bepaald of de links op pagina's moeten worden gevolgd en of pagina's moeten worden geïndexeerd.

Bepaalde webpagina's bevatten instructies van het type no-follow of no-index, waarmee robots (zoals de webcrawler) worden geïnstrueerd de gevonden links op deze pagina's niet te volgen, de content van deze pagina's niet in de index op te nemen of om beide acties niet uit te voeren.

Als u deze instructies op de juiste wijze beheert, kunt u de kwaliteit van de crawl-opdrachten verbeteren. Het kan bijvoorbeeld voorkomen dat bepaalde directorypagina's duizenden links bevatten, terwijl er geen andere nuttige content aanwezig is. Deze pagina's moeten worden gecrawld en de links moeten worden gevolgd, terwijl het indexeren van de directorypagina's zelf geen nut heeft.

Ook kan het voorkomen dat u niet wilt dat de crawler zich dieper in de hiërarchie verplaatst, terwijl de gewenste pagina's op de eindpunten links bevatten en geen instructies van het type no-follow. Omdat enkele van deze pagina's automatisch worden gegenereerd, hebben zij geen eigenaars die de vereiste instructies zouden kunnen invoegen.

Als u regels wilt opgeven voor het crawlen van dergelijke pagina's, maakt of bewerkt u het configuratiebestand `followindex.rules`. Houd bij het opgeven van regels in dit bestand rekening met de volgende richtlijnen:

- De regels die u configureert, moeten URL-prefixen aangeven (u kunt geen websites opgeven met behulp van IP-adressen of DNS-hostnamen).
- De URL-prefixen kunnen sterretjes (\*) als jokertekens bevatten om meerdere sites met dezelfde URL's toe te staan of te verbieden.
- De volgorde is van essentieel belang (de crawler past de eerste regel toe die overeenkomt met een kandidaat-URL).
- De regels, waarmee het volgen of indexeren expliciet wordt toegestaan of verboden, overschrijven de andere instellingen, waaronder die in het doeldocument.

## Instructies van het type no-follow en no-index in webpagina's

U kunt regels in een configuratiebestand opgeven om te bepalen of de webcrawler de links naar pagina's moet volgen of pagina's moet indexeren die instructies van het type no-follow of no-index bevatten.

### Voordat u begint

Om instructies van het type no-follow en no-index voor de webcrawler op te geven, moet u een enterprise search-beheerder zijn. De instructies die u opgeeft, overschrijven de instructies die aanwezig zijn in de te crawlen pagina's.

### Procedure

Ga als volgt te werk om instructies van het type no-follow en no-index te overschrijven:

1. Open de beheerconsole van enterprise search, bewaak de collectie waarvoor u regels wilt configureren en stop de webcrawler.
2. Meld u bij de crawlerserver aan als enterprise search-beheerder. Dit gebruikers-ID is opgegeven bij de installatie van WebSphere II OmniFind Edition.

3. Ga naar de configuratiedirectory voor de crawler die u wilt configureren, waarbij *crawler\_sessie\_ID* het ID is dat door het enterprise search-systeem aan de crawlersessie is toegewezen. Bijvoorbeeld:  
`ES_NODE_ROOT/master_config/col_56092.WEB_88534`
4. Maak of bewerk het bestand `followindex.rules`.
5. Voer regels voor de crawler in de volgende indeling in, waarbij *URLprefix* de eerste tekens zijn van de websites waarvoor u het volgen of indexeren wilt toestaan of verbieden:  

```
forbid follow URLprefix
allow follow URLprefix
forbid index URLprefix
allow index URLprefix
```
6. Sla het bestand op en sluit dit af.
7. Ga naar de beheerconsole van enterprise search en start de webcrawler opnieuw die u hebt gestopt.

## Configureren welke datum de webcrawler gebruikt voor gecrawelde documenten

U kunt in een configuratiebestand een optie opgeeft waarmee u bepaalt welke datum de webcrawler gebruikt als de datum van een gecrawld document.

### Voordat u begint

Standaard stelt de webcrawler de waarde voor het veld `Date` in gecrawelde documenten in op de datum waarop dat document is gecrawld. Desgewenst kunt u de webcrawler echter ook zodanig configureren dat deze datum wordt ingesteld op de datum en tijd waarop het document voor het laatst is gewijzigd, zoals gemeld door de webserver.

De datum `Laatst Gewijzigd` is wellicht niet voor alle documenten beschikbaar. Als u de crawler instelt op het gebruik van deze waarde en de webserver voor een bepaald document geen gegevens voor `Laatst Gewijzigd` retourneert, stelt de crawler de waarde van het veld `Date` voor dat document in op de datum en tijd waarop het document is gecrawld.

Tabel 4. Hoe de webserver het veld `Date` voor gecrawelde documenten configureert

| Crawler-configuratie                                               | Door webserver geretourneerde gegevens                                                        | Metagegevensveld <code>Date</code> voor gecrawelde documenten           |
|--------------------------------------------------------------------|-----------------------------------------------------------------------------------------------|-------------------------------------------------------------------------|
| Standaard-configuratie                                             | De webserver retourneert een waarde <code>Laatst Gewijzigd</code> voor een gecrawld document  | Datum en tijd gecrawld                                                  |
|                                                                    | De webserver retourneert geen waarde <code>Laatst Gewijzigd</code> voor een gecrawld document | Datum en tijd gecrawld                                                  |
| Geconfigureerd om datum <code>Laatst Gewijzigd</code> te gebruiken | De webserver retourneert een waarde <code>Laatst Gewijzigd</code> voor een gecrawld document  | Waarde van <code>Laatst Gewijzigd</code> zoals gemeld door de webserver |
|                                                                    | De webserver retourneert geen waarde <code>Laatst Gewijzigd</code> voor een gecrawld document | Datum en tijd gecrawld                                                  |

Om op te geven welke datum de webcrawler moet gebruiken voor gecrawelde documenten, moet u enterprise search-beheerder zijn.

## Procedure

U kunt een webcrawler als volgt zodanig configureren dat de datum en tijd voor Laatste Gewijzigd, zoals gemeld door de webserver, wordt gebruikt:

1. Meld u als enterprise search-beheerder aan bij de crawlerserver. Dit gebruikers-ID is opgegeven bij de installatie van WebSphere II OmniFind Edition.
2. Open het bestand `crawl.properties` voor de webcrawler die u wilt configureren in een editor. Dit bestand bevindt zich in de volgende directory, waarbij *crawlersessie-ID* een ID is dat door het enterprise search-systeem is toegewezen aan de crawlersessie.

```
ES_NODE_ROOT/master_config/crawlersessie-ID/crawl.properties
```

Bijvoorbeeld:

```
/home/esadmin/master_config/col_00112.WEB_23344/crawl.properties  
C:\Program Files\IBM\esadmin\master_config\col_55667.WEB_78899\crawl.properties
```

3. Voeg de volgende regel toe en sla het bestand op:  
`which_date=LastModified`
4. Ga naar de beheerconsole van enterprise search en start de webcrawler opnieuw.

Als een webserver gegevens van het type Laatste Gewijzigd terugstuurt voor documenten die zijn gecrawld met deze webcrawler, gebruikt de crawler de geretourneerde datum en tijd voor het veld Date van de gecrawelde documenten.

---

## Web Content Management-crawlers

Als u IBM Workplace Web Content Management-documenten in een enterprise search-collectie wilt opnemen, moet u een Web Content Management-crawler configureren.

### Configuratie van de WebSphere Portal-server

Als u Workplace Web Content Management op een WebSphere Portal-server versie 6 installeert, kunt u met de Web Content Management-crawler Web Content Management-sites crawlen. U kunt opties voor het crawlen van deze sites configureren die afwijken van de opties die u opgeeft voor de portalsites die door een WebSphere Portal-crawler worden gecrawld.

Voordat u een Web Content Management-crawler maakt, moet u een script uitvoeren om de enterprise search-omgeving in WebSphere Portal in te stellen. Dit script (`wp6_install.sh` onder AIX, Linux of Solaris, of `wp6_install.bat` onder Windows) wordt tijdens de installatie van WebSphere Information Integrator OmniFind Edition op de zoekservers geïnstalleerd.

### URL-indeling

Bij het opgeven van de URL's die moeten worden gecrawld, moet u de volgende indeling gebruiken:

```
http_protocol://portal_hostnaam:poort_nummer/portal_prefix  
/basis_servlet_pad_WCM_zoekopdracht/searchseed?site=WCM_site_naam&lib=WCM_bibliotheek_naam
```

In het volgende voorbeeld ziet u een URL voor de site in het standaardinstallatiepad van Workplace Web Content Management in WebSphere Portal:

<http://portal.server.ibm.com:80/wps/wcmsearchseed/searchseed?site=SiteTest01&lib=Web+Content>

Als de naam van de site of bibliotheek spaties bevat, moet u de spaties vervangen door het plusteken (+). Web Content moet bijvoorbeeld worden vervangen door Web+Content.

## Configuratieoverzicht

U kunt met de Web Content Management-crawler een willekeurig aantal Web Content Management-sites crawlen. Bij het configureren van de crawler moet u de URL's opgeven van de sites die u wilt crawlen. De crawler downloadt vervolgens de pagina's die als links zijn opgenomen op de opgegeven sites.

De te crawlen sites moeten toegankelijk zijn met behulp van hetzelfde ID en hetzelfde wachtwoord van de WebSphere Portal-beheerder. Als u sites wilt crawlen die verschillende legitimatiegegevens gebruiken, moet u een aparte Web Content Management-crawler configureren.

Als u een Web Content Management-crawler wilt maken of wijzigen, meldt u zich aan bij de beheerconsole van enterprise search. U moet de rol van beheerder van enterprise search hebben of beheerder zijn van de collectie die eigenaar is van de crawler.

Bij het definiëren van de crawler moet u de volgende procedure uitvoeren. Hierbij wordt u geholpen door de wizard:

- Geef eigenschappen op die bepalen hoe de crawler werkt en op welke manier de systeemresources worden gebruikt. De crawlereigenschappen bepalen hoe de crawler alle documenten in de crawlruimte crawlt.
- Geef de URL's op voor de sites die moeten worden gecrawld en geef de informatie op die de crawler nodig heeft om verbinding met de sites te maken. Tijdens het maken of bewerken van de crawler kunt u testen of de crawler verbinding kan maken met de URL's die voor deze crawler worden geconfigureerd. Voordat u de crawler start, wordt u via een bericht geïnformeerd of de crawler toegang kan krijgen tot de documenten die moeten worden gecrawld.
- Geef opties voor beveiliging op documentniveau op. Als bij het maken van de collectie de beveiliging is ingeschakeld, kan de crawler beveiligingsgegevens koppelen aan documenten in de index. Met behulp van deze gegevens kan de toegangsbesturing worden afgedwongen op basis van opgeslagen toegangslijsten of beveiligingstokens.

U kunt ook een optie selecteren waarmee de legitimatiegegevens van gebruikers worden gecontroleerd wanneer een query wordt ingediend. Op deze manier worden de legitimatiegegevens vergeleken met de huidige toegangslijsten die door de oorspronkelijke gegevensbron worden beheerd en niet met de geïndexeerde beveiligingsgegevens.

**Belangrijk:** Als u beveiligde Web Content Management-pagina's wilt doorzoeken, moet u in WebSphere Portal zoekopdrachten opgeven met behulp van de portlet Zoeken voor enterprise search. Zoekopdrachten die via het voorbeeldzoekprogramma ESSearchApplication worden opgegeven, beschikken niet over de juiste legitimatiegegevens en kunnen de gebruikersmachtigingen voor het openen van documenten niet controleren.

- Geef informatie op waarmee de crawler kan communiceren met een proxy-server (als de Web Content Management-sites gebruikmaken van proxy servers).

- Geef verificatiegegevens op waarmee de crawler toegang kan krijgen tot documenten met SSO-beveiliging (Single Sign-On).
- Geef informatie op over het sleutelruimtebestand, zodat de crawler het SSL-protocol (Secure Sockets Layer) kan gebruiken voor het maken van een verbinding met Web Content Management-sites.
- Geef de taal en de codetabel op van de documenten die moeten worden gecrawld.
- Geef opties op voor het crawlen en doorzoeken van metagegevens in Web Content Management-documenten.
- Geef roosters op voor het crawlen van de Web Content Management-sites.

Als u bij het definiëren van de crawler meer informatie wenst over de velden in de wizard en over de manier waarop u de informatie moet opgeven die de crawler nodig heeft om informatie te crawlen, klikt u in de wizard op **Help**.

#### **Verwante onderwerpen**

“Integratie van enterprise search met WebSphere Portal” op pagina 271  
 U kunt de zoekmogelijkheden van IBM WebSphere Portal uitbreiden door de zoekportlet voor enterprise search-portlets in WebSphere Portal te implementeren en WebSphere Portal zodanig te configureren, dat de zoekportlet als standaardzoekmachine wordt gebruikt.

#### **Verwante taken**

“Enterprise search installeren in WebSphere Portal versie 6” op pagina 279  
 Om een enterprise search-systeem te integreren met WebSphere Portal versie 6 of later, gebruikt u het script wp6\_install.

---

## **WebSphere Portal-crawlers**

Als u pagina's van een IBM WebSphere Portal-site wilt opnemen in een enterprise search-collectie, moet u een WebSphere Portal-crawler configureren.

### **Configuratie van de WebSphere Portal-server**

Voordat u een WebSphere Portal-crawler maakt, moet u een script uitvoeren om de enterprise search-omgeving in WebSphere Portal in te stellen. Voor de verschillende versies van WebSphere Portal zijn verschillende scripts nodig. De scripts worden op de zoekservers geïnstalleerd tijdens de installatie van WebSphere Information Integrator OmniFind Edition.

- Voor WebSphere Portal versie 5.1.0 of later voert u het script wp5\_install.sh (op AIX-, Linux- en Solaris-systemen) of het script wp5\_install.bat uit (op Windows-systemen).
- Voor WebSphere Portal versie 6 voert u het script wp6\_install.sh (op AIX-, Linux- en Solaris-systemen) of het script wp6\_install.bat uit (op Windows-systemen).

### **Configuratieoverzicht**

U kunt de WebSphere Portal-crawler gebruiken om een enkele WebSphere Portal-site te crawlen. Bij het configureren van de crawler moet u de URL opgeven van de portalsite die u wilt crawlen. De crawler downloadt vervolgens de portlets en pagina's die als links zijn opgenomen op de opgegeven portal-URL. Als u een andere portalsite wilt crawlen, maakt u een andere crawler.

Als u een WebSphere Portal-crawler wilt maken of wijzigen, meldt u zich aan bij de beheerconsole van enterprise search. U moet de rol van beheerder van enterprise search hebben of beheerder zijn van de collectie die eigenaar is van de crawler.

Bij het definiëren van de crawler moet u de volgende procedure uitvoeren. Hierbij wordt u geholpen door de wizard:

- Geef eigenschappen op die bepalen hoe de crawler werkt en op welke manier de systeemresources worden gebruikt. De crawlereigenschappen bepalen hoe de crawler alle pagina's op de site crawlt.
- Geef de URL op van de portalsite die moet worden gecrawld en geef de informatie die de crawler nodig heeft om verbinding met de site te maken. Omdat deze typen URL's lang kunnen zijn en gecodeerde niet-ASCII-tekens kunnen bevatten, kunt u de URL vanaf de WebSphere Portal-server kopiëren en in de beheerconsole van enterprise search plakken.

Tijdens het maken of bewerken van de crawler kunt u testen of de crawler verbinding kan maken met de URL's die voor deze crawler worden geconfigureerd. Voordat u de crawler start, wordt u via een bericht geïnformeerd of de crawler toegang kan krijgen tot de documenten die moeten worden gecrawld.

- Geef opties voor beveiliging op documentniveau op. Als bij het maken van de collectie de beveiliging is ingeschakeld, kan de crawler beveiligingsgegevens koppelen aan documenten in de index. Met behulp van deze gegevens kan de toegangsbesturing worden afgedwongen op basis van opgeslagen toegangslijsten of beveiligingstokens.

U kunt ook een optie selecteren waarmee de legitimatiegegevens van gebruikers worden gecontroleerd wanneer een query wordt ingediend. Op deze manier worden de legitimatiegegevens vergeleken met de huidige toegangslijsten die door de oorspronkelijke gegevensbron worden beheerd en niet met de geïndexeerde beveiligingsgegevens.

**Belangrijk:** Als u beveiligde WebSphere Portal-pagina's wilt doorzoeken, moet u in WebSphere Portal zoekopdrachten opgeven met behulp van de portlet Zoeken voor enterprise search. Zoekopdrachten die via het voorbeeldzoekprogramma ESSearchApplication worden opgegeven, beschikken niet over de juiste legitimatiegegevens en kunnen de gebruikersmachtigingen voor het openen van documenten niet controleren.

- Geef informatie op waarmee de crawler kan communiceren met een proxyserver (als de WebSphere Portal-site voor het verzorgen van de pagina's gebruikmaakt van een proxyserver).
- Geef verificatiegegevens op waarmee de crawler toegang kan krijgen tot documenten met SSO-beveiliging (Single Sign-On).
- Geef informatie op over het sleutelruimtebestand, zodat de crawler het SSL-protocol (Secure Sockets Layer) kan gebruiken voor het maken van een verbinding met WebSphere Portal-site.
- Geef de taal en de codetabel op van de documenten die moeten worden gecrawld.
- Geef opties op voor het crawlen en doorzoeken van metagegevens in WebSphere Portal-documenten.
- Geef roosters op voor het crawlen van de WebSphere Portal-documenten.

Als u bij het definiëren van de crawler meer informatie wenst over de velden in de wizard en over de manier waarop u de informatie moet opgeven die de crawler nodig heeft om informatie te crawlen, klikt u in de wizard op **Help**.

### Verwante onderwerpen

“Integratie van enterprise search met WebSphere Portal” op pagina 271  
U kunt de zoekmogelijkheden van IBM WebSphere Portal uitbreiden door de zoekportlet voor enterprise search-portlets in WebSphere Portal te implementeren en WebSphere Portal zodanig te configureren, dat de zoekportlet als standaardzoekmachine wordt gebruikt.

### Verwante taken

“De URL kopiëren om te crawlen vanaf WebSphere Portal”

Om te voorkomen dat u spelfouten maakt bij het invoeren van URL's kunt u de URL van de site die u wilt crawlen, kopiëren en in het juiste veld plakken tijdens het configureren van een WebSphere Portal- of Web Content Management-crawler.

“Enterprise search installeren in WebSphere Portal versie 6” op pagina 279

Om een enterprise search-systeem te integreren met WebSphere Portal versie 6 of later, gebruikt u het script wp6\_install.

“Enterprise search installeren in WebSphere Portal versie 5.1” op pagina 273

Om een enterprise search-systeem te integreren met WebSphere Portal versie 5.1.0 of later, gebruikt u het script wp5\_install.

## De URL kopiëren om te crawlen vanaf WebSphere Portal

Om te voorkomen dat u spelfouten maakt bij het invoeren van URL's kunt u de URL van de site die u wilt crawlen, kopiëren en in het juiste veld plakken tijdens het configureren van een WebSphere Portal- of Web Content Management-crawler.

### Over deze taak

Tijdens het maken van een WebSphere Portal- of Web Content Management-crawler geeft u de URL op van de site op de WebSphere Portal-server die u wilt crawlen. Omdat de URL's lang zijn en vaak gecodeerde niet-ASCII-tekenen bevatten, kunt u de volgende procedure gebruiken om de URL vanaf de WebSphere Portal-server te kopiëren en in de beheerconsole van enterprise search te plakken.

### Procedure

Ga als volgt te werk om de URL op te geven die met de crawler moet worden gecrawld:

1. Als u in de beheerconsole van enterprise search de URL's wilt opgeven die moeten worden gecrawld, moet u ervoor zorgen dat de WebSphere Portal-server is gestart. Vervolgens moet u zich bij WebSphere Portal aanmelden als beheerder.
2. Als u werkt met WebSphere Portal versie 5.1, voert u de volgende stappen uit op de WebSphere Portal-server:
  - a. Klik op **Administration** in de rechterbovenhoek.
  - b. Klik op **Portal Settings** in het navigatiegebied links en klik vervolgens op **Search Administration**.
  - c. Klik op **PortalCollection** in het gebied Search Collections op de pagina Manage Search Collections. Als er meer collecties beschikbaar zijn, kunt u een andere collectie selecteren.
  - d. Klik in het gebied Content Sources in the Collection op **Add Content Source**.
  - e. Klik bij **Crawl source type** op **Portal site**. De URL van de site verschijnt in het veld **Collect documents linked from this URL**.



- f. Kopieer de URL naar het klembord. U kunt de URL markeren en de Ctrl-toets ingedrukt houden terwijl u op Insert drukt.
3. Als u werkt met WebSphere Portal versie 6, voert u de volgende stappen uit op de WebSphere Portal-server:
  - a. Klik op **Administration** in de linkerbenedenhoek.
  - b. Klik op **Manage Search** in het navigatiegebied links.
  - c. Klik op de pagina Manage Search op de link **Search Collections**.
  - d. Klik op de pagina Manage Search op **Default Portal Search Service** bij de opties voor Search service.
  - e. Klik in de tabel Search Collections op de collectie **Portal Content**.
  - f. Klik in de tabel Content Sources op het pictogram Edit uiterst rechts (het pictogram met het potlood), naast **Portal Content Source**.
  - g. Klik voor **Content source type** op **Portal Site**. De URL van de site verschijnt in het veld **Collect documents linked from this URL**.
  - h. Kopieer de URL naar het klembord. U kunt de URL markeren en de Ctrl-toets ingedrukt houden terwijl u op Insert drukt.
4. Ga terug naar de beheerconsole van enterprise search en plak de gekopieerde URL in het veld voor de site-URL.

#### Verwante onderwerpen

“WebSphere Portal-crawlers” op pagina 102

Als u pagina’s van een IBM WebSphere Portal-site wilt opnemen in een enterprise search-collectie, moet u een WebSphere Portal-crawler configureren.

---

## Windows-bestandssysteemcrawler

Als u documenten die zijn opgeslagen in een Microsoft Windows-bestandssysteem wilt opnemen in een enterprise search-collectie, moet u een Windows-bestandssysteem-crawler configureren.

U kunt met de Windows-bestandssysteem-crawler een willekeurig aantal Windows-bestandssystemen crawlen. Tijdens het configureren van de crawler selecteert u de lokale en niet-lokale directory’s en subdirectory’s die u wilt crawlen.

Als u de crawlerserver op een UNIX-, Linux- of Solaris-computer installeert, kunt u deze server niet gebruiken om Windows-bestandssysteembronnen te crawlen (de Windows-bestandssysteem-crawler verschijnt dan niet in de lijst van beschikbare crawlertypen).

### Crawlen in gemeenschappelijke netwerkdirectory’s

De Windows-bestandssysteem-crawler crawlt documenten op basis van de leesmachtigingen die voor de enterprise search-beheerder zijn opgegeven. De beheerder is de IBM WebSphere Information Integrator OmniFind Edition-services-account.

U kunt een gebruikers-ID en wachtwoord opgeven voor de directory’s die moeten worden gecrawld. Het gebruikers-ID en wachtwoord worden echter alleen gebruikt om verbinding te maken met de gemeenschappelijke netwerkdirectory’s. De crawler crawlt de bestanden volgens de leesmachtigingen die voor de desbetreffende gebruiker zijn ingesteld voor gemeenschappelijke netwerkdirectory’s, niet voor lokale stations. De crawler kan het gebruikers-ID gebruiken om verbinding te maken met één server of met één netwerkdirectory tegelijk.

Verbindingen met netwerkdirectory's worden pas verbroken als u de IBM WebSphere Information Integrator OmniFind Edition-service opnieuw start. Als een verbinding tot stand is gebracht, kunt u toegang tot de directory krijgen met een onjuist ID en wachtwoord. Een dergelijke verbinding is echter alleen toegestaan voor de ontdek- en crawlersessies van Windows-bestandssysteem die worden beheerd door het enterprise search-systeem. Om eventuele beveiligingsrisico's te voorkomen, moet u ervoor zorgen dat de machtigingen voor de enterprise search-beheerdersaccount (waarmee de IBM WebSphere Information Integrator OmniFind Edition-service wordt uitgevoerd) op de juiste manier zijn ingesteld.

U kunt problemen met het maken van verbinding met een netwerkdirectory in de toekomst voorkomen door hetzelfde gebruikers-ID en wachtwoord voor dezelfde netwerkdirectory op te geven. Als u een verkeerd gebruikers-ID en wachtwoord opgeeft en de IBM WebSphere Information Integrator OmniFind Edition-service opnieuw start, kan de Windows-bestandssysteem-crawler mogelijk niet crawlen omdat deze verbinding met de directory probeert te maken met behulp van onjuiste legitimatiegegevens. Er kan worden gecrawld als de netwerkverbinding tot stand wordt gebracht door een andere Windows-bestandssysteem-crawler die het juiste gebruikers-ID en wachtwoord gebruikt.

## Configuratieoverzicht

Als u een Windows-bestandssysteem-crawler wilt maken of wijzigen, meldt u zich aan bij de beheerconsole van enterprise search. U moet de rol van beheerder van enterprise search hebben of beheerder zijn van de collectie die eigenaar is van de crawler.

Bij het definiëren van de crawler moet u de volgende procedure uitvoeren. Hierbij wordt u geholpen door de wizard:

- Geef eigenschappen op die bepalen hoe de crawler werkt en op welke manier de systeemresources worden gebruikt. De crawlereigenschappen bepalen hoe de crawler alle subdirectory's in de crawlruimte crawlt.
- Geef een rooster op voor het crawlen van de bestandssystemen.
- Selecteer de subdirectory's die u wilt crawlen.

U kunt opgeven hoeveel niveaus van subdirectory's in de crawler moeten worden gecrawld. Om bestandssystemen op afstand te crawlen, moet u een gebruikers-ID en een wachtwoord opgeven waarmee de crawler toegang krijgt tot de gegevens.

- Geef opties op waarmee de documenten in de subdirectory's doorzoekbaar worden gemaakt. U kunt bijvoorbeeld bepaalde typen documenten uitsluiten van de crawlruimte of een gebruikers-ID en wachtwoord opgeven waarmee de crawler toegang krijgt tot de bestanden in een bepaalde subdirectory.
- Configureer beveiligingsopties op documentniveau. Als bij het maken van de collectie de beveiliging is ingeschakeld, kan de crawler beveiligingsgegevens koppelen aan documenten in de index. Met behulp van die beveiligingsgegevens kunnen zoekprogramma's toegangsbesturing uitvoeren op basis van opgeslagen toeganglijsten of beveiligingstokens.

U kunt ook een optie selecteren waarmee de legitimatiegegevens van gebruikers worden gecontroleerd wanneer een query wordt ingediend. Op deze manier worden de legitimatiegegevens vergeleken met de huidige toeganglijsten die door de oorspronkelijke gegevensbron worden beheerd en niet met de geïndexeerde beveiligingsgegevens.

Om beveiliging op documentniveau af te dwingen, moet u ervoor zorgen dat de accountgegevens van de gebruiker en het domein op de juiste wijze zijn geconfigureerd op de crawlerserver.

Als u bij het definiëren van de crawler meer informatie wenst over de velden in de wizard en over de manier waarop u de informatie moet opgeven die de crawler nodig heeft om informatie te crawlen, klikt u in de wizard op **Help**.

#### **Verwante onderwerpen**

“Beveiliging op documentniveau afdwingen voor documenten van het Windows-bestandssysteem” op pagina 263

Om in te stellen dat de actuele legitimatiegegevens moeten worden gevalideerd wanneer een gebruiker documenten doorzoekt die zijn gecrawld door een Windows-bestandssysteem-crawler, moet u de domeinaccountgegevens op zowel de crawlerserver als de Microsoft Windows-server configureren.

---

## **Ondersteuning voor Data Listener-programma's configureren**

U kunt enterprise search uitbreiden door een externe crawler te maken met behulp van de Data Listener-API. Met de aangepaste Data Listener-programma's kunnen gegevens aan collecties worden toegevoegd, gegevens uit collecties worden verwijderd of webcrawlers opdracht worden gegeven URL's (opnieuw) te bezoeken.

#### **Voordat u begint**

Om Data Listener-programma's te configureren, moet u de rol van beheerder van enterprise search hebben.

#### **Over deze taak**


Een Data Listener-programma maakt het mogelijk om typen gegevensbronnen te crawlen die niet kunnen worden gecrawld met de standaardcrawlers van enterprise search. Voordat u een Data Listener-programma kunt gebruiken, moet u legitimatiegegevens configureren waarmee het programma toegang kan krijgen tot de collecties en deze kan bijwerken.

Als het Data Listener-clientprogramma verbinding maakt met de Data Listener, moeten het clientprogramma-ID en wachtwoord, en het ID van de bij te werken collectie worden doorgegeven. Deze informatie moet overeenkomen met de informatie die u voor het programma in de beheerconsole configureert.



De Data Listener wordt automatisch gestart bij het starten van het enterprise search-systeem. Als u het poortnummer wijzigt nadat u het programma in de beheerconsole hebt geconfigureerd, moet u de Data Listener opnieuw starten.

#### **Procedure**

Ga als volgt te werk om Data Listener-programma's te configureren:

1. Klik op **Systeem** om de view Systeem te openen.
2. Klik op  **Wijzigen** om naar de view Systeem wijzigen te gaan.
3. Klik op de pagina Data Listener op **Data Listener-programma's configureren**.
4. Geef op de pagina Data Listener-programma's het aantal threads op dat de Data Listener kan maken voor verwerkingsopdrachten van clientprogramma's en het poortnummer waar de Data Listener naar opdrachten luistert. Geef ook

het maximum aantal documenten (per collectie) op dat in een tijdelijke opslagplek kan worden bewaard voordat de documenten door de parser worden geanalyseerd.

5. Klik op **Data Listener-programma toevoegen** om informatie over een client-programma toe te voegen.
6. Geef op de pagina Data Listener-programma toevoegen verificatiegegevens op waarmee de Data Listener-clientprogramma's toegang kunnen krijgen tot de collecties in enterprise search. De ID-s van de Data Listener-clientprogramma's moeten uniek zijn binnen het enterprise search-systeem.
7. Selecteer de collecties die door het Data Listener-programma kunnen worden bijgewerkt:
  - Klik op **Alle collecties** als u wilt instellen dat het programma alle collecties kan bijwerken.
  - Klik op **Bepaalde collecties** als u wilt instellen dat het programma alleen de collecties kan bijwerken die u opgeeft.  
Als u deze optie selecteert, wordt er een lijst met collecties afgebeeld. Schakel het aankruisvakje **Selecteren** in voor elke collectie die met het programma kan worden bijgewerkt.
8. Klik op **OK**.
9. Als u het poortnummer voor de Data Listener wijzigt of het aantal documenten dat op een tijdelijke opslagplek kan worden bewaard, moet u de Data Listener opnieuw starten:
  - a. Klik op  **Monitor** om de bewakingsview van het systeem te wijzigen.
  - b. Klik op de pagina Data Listener op  **Opnieuw starten**.

#### Verwante onderwerpen

Data listener

#### Verwante taken

"De Data Listener bewaken" op pagina 317

U kunt de Data Listener bewaken om de status ervan te bekijken en om gegevens over de activiteiten van de Data Listener-clientprogramma's weer te geven.

---

## Aangepaste crawler plug-ins

Bij het configureren van de eigenschappen voor crawlers, kunt u de Java-klasse opgeven die moet worden gebruikt om toegangsbesturing op documentniveau af te dwingen. Daarnaast kunt u de Java-klasse gebruiken om de index bij te werken door metagegevens en documentcontent toe te voegen, te wijzigen of te verwijderen. Door een plugin te schrijven kunt u tevens de mogelijkheden van de crawler voor het crawlen van archiefbestanden uitbreiden.

Een plug-in bevat een Java-klasse die wordt aangeroepen voor elk document dat de crawler crawlt. De document-ID (URI) wordt via de enterprise search-index, beveiligingstokens, metagegevens en documentcontent aan de Java-klasse doorgegeven. De klasse kan nieuwe of gewijzigde beveiligingstokens, metagegevens en content terugsturen, of de klasse kan beveiligingstokens, metagegevens en content verwijderen.

Als alle documenten in de crawlruimte één keer zijn gecrawld, wordt de plug-in alleen aangeroepen voor nieuwe of gewijzigde documenten. Als u de beveiligingstokens, metagegevens of documentcontent wilt wijzigen voor documenten die in de enterprise search-index zijn opgenomen, maar die niet zijn bijge-

werkt in de oorspronkelijke gegevensbron, start u een volledige crawl van alle documenten in de crawlruimte en reorganiseert u vervolgens de hoofdindex.

## **Plug-ins gebruiken om beveiliging af te dwingen**

U kunt beveiliging op documentniveau afdwingen door een of meer beveiligingstokens (een door komma's gescheiden tekenreeks) te koppelen aan alle documenten die door een crawler worden gecrawld. Groeps-ID's worden vaak als beveiligingstokens gebruikt.

Standaard wordt aan elk document een openbaar token toegewezen, zodat het document voor iedereen beschikbaar is. Het openbare token kan worden vervangen door een waarde die door de beheerder wordt geleverd of door een waarde die uit een veld in het gecrawelde document wordt gehaald.

Met behulp van de plug-in kunt u uw eigen bedrijfsregels toepassen om de waarde van de beveiligingstokens voor gecrawelde documenten te bepalen. De beveiligingstokens die aan elk document zijn gekoppeld, zijn opgeslagen in de index. De tokens worden gebruikt om documenten te filteren die overeenkomen met de beveiligingstokens en ervoor te zorgen dat alleen de documenten waarvoor een gebruiker weergavemachtigingen heeft, in de zoekresultaten worden geretourneerd.

## **Plug-ins gebruiken om metagegevens toe te voegen, te wijzigen of te verwijderen**

Voor alle gecrawelde documenten worden documentmetagegevens gemaakt (zoals de datum waarop een document voor het laatst is gewijzigd). Met behulp van de plug-in kunt u uw eigen bedrijfsregels toepassen om de waarde van de metagegevens te bepalen die voor elk document wordt geïndexeerd.

De metagegevens worden gemaakt als naamwaardeparen. Gebruikers kunnen de metagegevens zoeken met een vrije zoekopdracht of met een zoekopdracht waarbij de veldnaam van de metagegevens wordt opgegeven.

## **Plug-ins gebruiken om documentcontent toe te voegen, te wijzigen of te verwijderen**

Documentcontent bestaat uit de delen van een document met doorzoekbare content en content die kan worden opgenomen in de dynamische document-samenvattingen in de zoekresultaten. Met behulp van de crawlerplug-in kunt u uw eigen bedrijfsregels toepassen om de waarde van de content te bepalen die voor elk document wordt geïndexeerd.

## **Webcrawlerplug-ins**

Met de API's voor de webcrawler kunt u bepalen hoe documenten worden gecrawld en hoe documenten worden voorbereid op analyses. U kunt bijvoorbeeld velden toevoegen aan de HTTP-opdrachtheader die wordt gebruikt wanneer de crawler een document aanvraagt. Als een document is gecrawld (en de analyse en tokenisatie dus zijn uitgevoerd), kunt u de content, beveiligingstokens en metagegevens wijzigen. Daarnaast kunt u voorkomen dat het document naar de parser wordt verzonden.

## Plugins voor archiefbestanden

Door een plugin te schrijven kunt u de mogelijkheden van de crawler voor het crawlen van archiefbestanden uitbreiden, en het mogelijk maken dat er ook andere archiefbestanden dan ZIP en TAR worden gecrawld. Zo kunt u bijvoorbeeld een plugin schrijven voor het crawlen van documenten in LZH-indeling.

### Verwante onderwerpen

Crawler-plugins

### Verwante verwijzing

“Ondersteuning voor het crawlen van gearchiveerde bestanden”

De crawlers in enterprise search kunnen de bestanden uit een archiefbestand (zoals een ZIP- of TAR-bestand) extraheren, zodat de individuele bestanden in het archief kunnen worden geïndexeerd en doorzocht.

---

## Ondersteuning voor het crawlen van gearchiveerde bestanden

De crawlers in enterprise search kunnen de bestanden uit een archiefbestand (zoals een ZIP- of TAR-bestand) extraheren, zodat de individuele bestanden in het archief kunnen worden geïndexeerd en doorzocht.

### Ondersteunde archiefbestandsindelingen

De volgende archiefbestandsindelingen worden ondersteund:

Tabel 5. Ondersteunde archiefbestandsindelingen in enterprise search-crawlers

| Bestandsex-<br>tensie | MIME-type       | Gegevens- | Opmerkingen                                                                                                                                                                                                                                                                                                      |
|-----------------------|-----------------|-----------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| .zip, .ZIP            | application/zip | zip       | <ul style="list-style-type: none"><li>Afhankelijk van de mogelijkheden van het pakket java.utl.zip</li><li>Ondersteuning voor verkleinde compressie (methode 8):<ul style="list-style-type: none"><li>– Geen ondersteuning voor gecodeerde bestanden</li><li>– Geen ondersteuning voor zip64</li></ul></li></ul> |
| .tar                  | application/tar | tar       | Ondersteunde tar-indelingen: <ul style="list-style-type: none"><li>• GNU tar 1.13</li><li>• POSIX 1003.1-1998 (ustar)</li><li>• POSIX 1003.1-2001 (pax)</li></ul>                                                                                                                                                |
| .tar, .gz, .tgz       | n.v.t.          | tgz       | Afhankelijk van de mogelijkheden van het pakket java.utl.zip                                                                                                                                                                                                                                                     |

### Beperkingen en richtlijnen

Er is geen ondersteuning voor recursieve bestandsextractie. Als een archiefbestand een ander archiefbestand bevat, wordt het archiefbestand dat in het oorspronkelijke archiefbestand is opgenomen, niet gecrawld. In het volgende voorbeeld worden Folder2/Archive2.zip en de content van Folder2/Archive2.zip (de documenten in Folder3) niet gecrawld:

```
Archive1.zip
  Folder1/PowerPoint.ppt
  Folder2/Text.txt
  Folder2/Archive2.zip
    Folder3/Excel.xls
    Folder3/MSWord.doc
```

Automatische detectie van codetabellen is niet beschikbaar voor bestanden die uit een archiefbestand zijn geëxtraheerd. Bij het extraheren van de bestanden gebruikt de crawler de codetabelinstelling die is opgegeven voor het gebruik van platte tekst en onbekende documenttypen. Wanneer u de beheerconsole van enterprise search gebruikt om de taal- en codetabelinstellingen voor een crawler te configureren, moet u de codetabel opgeven die de crawler moet gebruiken voor documenten met platte tekst en voor documenten waarvan de codetabellen niet automatisch kunnen worden gedetecteerd.

Om te bepalen wanneer de bestanden in een archiefbestand opnieuw moeten worden gecrawld, gebruikt de crawler de wijzigingsdatum in de koptekstgegevens van het archiefitem voor elk bestand. Tijdens het bewaken van een crawler worden bij de cijfers voor de gecrawelde documenten (waaronder het aantal ingevoegde, bijgewerkte en verwijderde documenten) gegevens afgebeeld over de bestanden die uit archiefbestanden zijn geëxtraheerd.

Om crawlers in staat te stellen archiefbestanden in andere indelingen voor archiefbestanden te doorzoeken (bijvoorbeeld LZH-bestanden), moet u een crawlerplugin schrijven en de crawler zodanig configureren dat deze plugin wordt gebruikt.

## Migratie

Om de crawlers die aanwezig waren voordat WebSphere Information Integrator OmniFind Edition versie 8.4 werd geïnstalleerd, in te schakelen voor het crawlen van archiefbestanden, moet u de crawlruimte van de crawler bewerken. Zorg dat de instellingen voor het uitsluiten van MIME-typen en bestandsextensies geen verwijzingen bevatten naar .zip-, .tar-, .tgz- of .gz-bestanden.

**Belangrijk:** Als u deze instellingen wijzigt voor een Content Edition- of DB2 Content Manager-crawler, moet u alle documenten opnieuw crawlen, zodat de wijzigingen kunnen worden toegepast.

### Verwante onderwerpen

“Aangepaste crawler plug-ins” op pagina 108

Bij het configureren van de eigenschappen voor crawlers, kunt u de Java-klasse opgeven die moet worden gebruikt om toegangsbesturing op documentniveau af te dwingen. Daarnaast kunt u de Java-klasse gebruiken om de index bij te werken door metagegevens en documentcontent toe te voegen, te wijzigen of te verwijderen. Door een plugin te schrijven kunt u tevens de mogelijkheden van de crawler voor het crawlen van archiefbestanden uitbreiden.

---

## URI-indelingen in een enterprise search-index

Met de URI (Uniform Resource Identifier) van documenten in een enterprise search-index wordt het type crawler aangegeven waarmee het document aan de collectie is toegevoegd.

U kunt URI's of URI-patronen opgeven tijdens het configureren van categorieën, scopes en snellinks voor een collectie. Daarnaast kunt u de URI opgeven wanneer u documenten uit de index wilt verwijderen of wanneer u gedetailleerde statusinformatie over een URI wilt bekijken.

Om de URI's of URI-patronen voor een document vast te stellen, doorzoekt u de collectie. U kunt op de URI's in de zoekresultaten klikken om de documenten op te halen waarin u geïnteresseerd bent. U kunt de URI vanuit de zoekresultaten kopiëren voor gebruik in de beheerconsole van enterprise search. U kunt bijvoor-

beeld een URI-patroon opgeven om automatisch de documenten die aan het URI-patroon voldoen, aan een categorie in enterprise search te koppelen.

## Archiefbestanden

Voor documenten die uit een archiefbestand (zoals een .zip- of .tar-bestand) zijn geëxtraheerd en vervolgens zijn gecrawld, is de URI-indeling als volgt:

*Oorspronkelijke\_URI*(?|&)ArchiveEntry=*Item\_Naam*(&ArchiveEntry=*Item\_Naam*)

### Parameters

*Oorspronkelijke\_URI*

De locatie van het archiefbestand in de gegevensbron.

*Item\_Naam*

De naam van het archiefitem met URL-codering in het archiefbestand.

### Voorbeelden

file:///d:/Archive1.zip

file:///d:/Archive1.zip?ArchiveEntry=Folder1/PowerPoint.ppt

file:///d:/Archive1.zip?ArchiveEntry=Folder2/Text.txt

## Content Edition-crawlers

De URI-indeling voor documenten die zijn gecrawld door een Content Edition-crawler in de serverwerkstand is:

vbr://*Server\_Naam*/*Systeem\_ID\_Repository*/*Persistent\_ID\_Repository*  
/*Item\_ID*/*Versie\_ID*  
/*Item\_Type*/[?Page=*Pagina\_Nummer*&] *JNDI\_properties*

De URI-indeling voor documenten die zijn gecrawld door een Content Edition-crawler in de directe werkstand is:

vbr:///Systeem\_ID\_Repository/Persistent\_ID\_Repository  
/Item\_ID/Versie\_ID  
/Item\_Type/[?Page=Pagina\_Nummer]

### Parameters

URL-codering wordt op alle velden toegepast.

*Server\_Naam*

De naam van de WebSphere Information Integrator Content Edition-server.

*Systeem\_ID\_Repository*

Het systeem-ID voor de repository.

*Persistent\_ID\_Repository*

Het persistent ID voor de repository.

*Item\_ID*

Het ID voor het item.

*Versie\_ID*

Het ID voor de versie. Als het versie-ID leeg is, wordt met deze waarde de laatste versie van het document aangegeven.

*Item\_Type*

Het type van het item (CONTENT of FOLDER).

*Pagina\_Nummer*

Het paginanummer.



### *JNDI\_properties*

De JNDI-eigenschappen voor de J2EE-programmaclient. Er zijn twee typen eigenschappen:

#### **java.naming.factory.initial**

De naam van de klasse voor de programmaserver die wordt gebruikt om de EJB-handle te maken.

#### **java.naming.provider.url**

De URL van de naamgevingservice voor de programma-server die wordt gebruikt om de EJB-handle aan te vragen.

### **Voorbeelden**

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
```

### **DB2-crawlers**

De URI-indeling voor documenten die zijn gecrawld door een DB2-crawler is:

```
db2://Database_Naam/Tabel_Naam  
/Kolom_Naam1_Uniek_ID/Waarde1_Uniek_ID  
[/Kolom_Naam2_Uniek_ID/Waarde2_Uniek_ID/...  
/Kolom_NaamN_Uniek_ID/WaardeN_Uniek_ID]
```

#### **Parameters:**

URL-codering wordt op alle velden toegepast.

##### *Database\_Naam*

De interne naam van de database of het alias voor de database.

##### *Tabel\_Naam*

De naam van de doeltabel, inclusief de naam van het schema.

##### *Kolom\_1\_Uniek\_ID*

De naam van de eerste kolom voor het unieke ID in de tabel.

##### *Waarde1\_Uniek\_ID*

De waarde van de eerste kolom voor het unieke ID.

##### *Kolom\_NaamN\_Uniek\_ID*

De naam van de *ne* kolom voor het unieke ID in de tabel.

##### *WaardeN\_Uniek\_ID*

De waarde van de *ne* kolom voor het unieke ID.

### **Voorbeelden**

Lokale, gecatalogiseerde database:

```
db2://LOCALDB/SCHEMA1.TABLE1/MODEL/ThinkPadA20
```

Niet-lokale, niet-gecatalogiseerde database:

```
db2://myserver.mycompany.com:50001/REMOTEDB/SCHEMA2.TABLE2/NAME/DAVID
```

## DB2 Content Manager-crawlers

De URI-indeling voor documenten die zijn gecrawld door een DB2 Content Manager-crawler is:

*cm://Server\_Naam/Item\_Type\_Naam/PID*

### Parameters

URL-codering wordt toegepast op de parameter *PID*.

*Server\_Naam*

De naam van de IBM DB2 Content Manager-bibliotheekserver.

*Item\_Type\_Naam*

De naam van het doelitemtype.

*PID*

Het persistent ID van DB2 Content Manager.

### Voorbeeld

*cm://cmsrvctg/ITEMTYPE1/92+3+ICM8+icm1sdb12+ITEMTYPE159+26+A1001001A03F27B94411D1831718+A03F27B+94411D183171+14+1018*

## Domino Document Manager-crawlers

De URI-indeling voor documenten die zijn gecrawld door een Domino Document Manager-crawler is:

*dominodoc://Server\_Naam:Poort\_Nummer/Database\_Replica\_ID/Database\_Pad\_en\_Naam/Universeel\_ID\_View/Universeel\_ID\_Document/?AttNo=Bijlage\_Nummer&AttName=Bestands\_Naam\_Bijlage*

### Parameters

URL-codering wordt op alle velden toegepast.

*Server\_Naam*

De naam van de Domino Document Manager-server.

*Poort\_Nummer*

Optioneel: het poortnummer voor de Domino Document Manager-server.

*Database\_Replica\_ID*

Het ID voor de databasereplica.

*Database\_Pad\_en\_Naam*

Het pad en de bestandsnaam voor de document-NSF-database op de Domino Document Manager-doelserver.

*Universeel\_ID\_View*

Het ID dat wordt gebruikt om de Domino Document Manager-documenten te crawlen.

*Universeel\_ID\_Document*

Het ID dat in het gecrawde document is gedefinieerd.

*Bijlage\_Nummer*

Optioneel: een opeenvolgend nummer voor elke bijlage (beginnend met nul).

*Bestands\_Naam\_Bijlage*

Optioneel: de oorspronkelijke naam van het bijlagebestand.

### Voorbeelden

Een Domino Document Manager-document:

*dominodoc://dominodocsvr.ibm.com/49256D3A000A20DE/domdoc%2FADMN-6FAJXL.nsf/8178B1C14B1E9B6B8525624F0062FE9F/0205F44FA3F45A9049256DB20042D226*

Een documentbijlage:

```
dominodoc://dominodocsvr.ibm.com/49256D3A000A20DE/domdoc%2FADMN-6FAJXL.nsf  
/8178B1C14B1E9B6B8525624F0062FE9F/0205F44FA3F45A9049256DB20042D226  
?AttNo=0&AttName=AttachedFile.doc
```

## Exchange Server-crawlers

De URI-indeling voor documenten die zijn gecrawld door een Exchange Server-crawler is:

```
exchange://OWA_pad[?useSSL=true]
```

### Parameters

*OWA\_Pad*

Het OWA-pad (Outlook Web Access), zonder het protocol.

**useSSL=true**

Wordt toegevoegd als het protocol van het oorspronkelijke OWA-pad HTTPS is.

### Voorbeelden

Lopende tekst in document:

```
exchange://exchangesvr.ibm.com/public/RootFolder1/Folder1/Document.EML
```

Documentbijlage:

```
exchange://exchangesvr.ibm.com/public/RootFolder1/Folder1/Document.EML/  
AttachedFile.doc
```

Ingeschakeld voor SSL:

```
exchange://exchangesvr.ibm.com/public/TeamRoom/Folder1/Document.EML  
?useSSL=true
```

## JDBC database-crawlers

De URI-indeling voor documenten die zijn gecrawld door een JDBC database-crawler is:

```
jdbc://DB_URL/Tabel_Naam  
/Kolom_Naam1_Uniek_ID/Waarde1_Uniek_ID  
/[Kolom_Naam2_Uniek-ID/Waarde2_Uniek_ID  
/.../Kolom_NaamN_Uniek_ID/WaardeN_Uniek_ID]
```

### Parameters

URL-codering wordt op alle velden toegepast.

*DB\_URL* De URL voor de database.

*Tabel\_Naam*

De naam van de doeltabel, inclusief de naam van het schema.

*Kolom\_1\_Uniek\_ID*

De naam van de eerste kolom voor het unieke ID in de tabel.

*Waarde1\_Uniek\_ID*

De waarde van de eerste kolom voor het unieke ID.

*Kolom\_NaamN\_Uniek\_ID*

De naam van de *ne* kolom voor het unieke ID in de tabel.

*WaardeN\_Uniek\_ID*

De waarde van de *ne* kolom voor het unieke ID.

### Voorbeelden:

DB2 UDB-database:

```
jdbc:db2://host01.svl.ibm.com:50000/SAMPLE/DB2INST1.ORG/DEPTNUMB/51
```

Oracle-database:

```
jdbc:oracle:thin:@/host01.svl.ibm.com:1521:ora/SCOTT.EMP/EMPNO/7934
```

MS SQL Server 2000-database:

```
jdbc:microsoft:sqlserver://host01.svl.ibm.com:1433;  
DatabaseName=Northwind/dbo.Region/RegionID/100
```

MS SQL Server 2005-database:

```
jdbc:sqlserver://host01.svl.ibm.com:1433;  
DatabaseName=Northwind/dbo.Region/RegionID/100
```

### Notes-crawlers

De URI-indeling voor documenten die zijn gecrawld door een Notes-crawler is:

```
domino://Server_Naam[:Poort_Nummer]/Database_Replica_ID/Database_Pad_en_Naam  
/[Universeel_ID_View]/Universeel_ID_Document  
[?AttNo=Bijlage_Nummer&AttName=Bestands_Naam_Bijlage]
```

#### Parameters

URL-codering wordt op alle velden toegepast.

*Server\_Naam*

De naam van de Lotus Notes-server.

*Poort\_Nummer*

Het poortnummer voor de Lotus Notes-server. Het poortnummer is optioneel.

*Database\_Replica\_ID*

Het ID voor de databasereplica.

*Database\_Pad\_en\_Naam*

Het pad en de bestandsnaam voor de NSF-database op de Lotus Notes-doelserver.

*Universeel\_ID\_View*

Het universele ID voor de view dat voor de doeldatabase is gedefinieerd. Dit ID is alleen opgegeven als het document is geselecteerd vanuit een view of map. Als u geen view of map hebt aangewezen voor het crawlen (als u bijvoorbeeld hebt aangegeven dat alle documenten in een database moeten worden gecrawld), is het universele ID voor de view niet opgegeven.

*Universeel\_ID\_Document*

Het universele document-ID dat is gedefinieerd in het document dat door de crawler is gecrawld.

*Bijlage\_Nummer*

Een opeenvolgend nummer voor elke bijlage (beginnend met nul). Het bijlagenummer is optioneel.

*Bestands\_Naam\_Bijlage*

De oorspronkelijke naam van het bijlagebestand. De bijlagebestandsnaam is optioneel.

### Voorbeelden

Een document is geselecteerd voor crawlen per view of map:

domino://dominosvr.ibm.com/49256D3A000A20DE/Database.nsf/  
8178B1C14B1E9B6B8525624F0062FE9F/0205F44FA3F45A9049256DB20042D226

Een document is niet geselecteerd voor crawlen per view of map:

domino://dominosvr.ibm.com/49256D3A000A20DE/Database.nsf//  
0205F44FA3F45A9049256DB20042D226

Een documentbijlage:

domino://dominosvr.ibm.com/49256D3A000A20DE/Database.nsf//  
0205F44FA3F45A9049256DB20042D226?AttNo=0&AttName=AttachedFile.doc

## QuickPlace-crawlers

De URI-indeling voor documenten die zijn gecrawld door een QuickPlace-crawler is:

quickplace://Server\_Naam:Poort\_Nummer/Database\_Replica\_ID/Database\_Pad\_en\_Naam  
/Universeel\_ID\_View/Universeel\_ID\_Document  
/?AttNo=Bijlage\_Nummer&AttName=Bestands\_Naam\_Bijlage

### Parameters

URL-codering wordt op alle velden toegepast.

*Server\_Naam*

De naam van de Lotus QuickPlace-server.

*Poort\_Nummer*

Optioneel: het poortnummer voor de QuickPlace-server.

*Database\_Replica\_ID*

Het ID voor de databasereplica.

*Database\_Pad\_en\_Naam*

Het pad en de bestandsnaam voor de document-NSF-database op de QuickPlace-doelserver.

*Universeel\_ID\_View*

Het ID dat is gebruiker om QuickPlace-documenten te crawlen.

*Universeel\_ID\_Document*

Het ID dat in het gecrawlde document is gedefinieerd.

*Bijlage\_Nummer*

Optioneel: een opeenvolgend nummer voor elke bijlage (beginnend met nul).

*Bestands\_Naam\_Bijlage*

Optioneel: de oorspronkelijke naam van het bijlagebestand.

### Voorbeelden

Een document:

quickplace://ltsvr.ibm.com/49257043000214B3/QuickPlace%5Csamplplace  
%5CPageLibrary4925704300021490.nsf  
/A7986FD2A9CD47090525670800167225  
/2B02B1DE3A82B2CE49257043001C2498

Een paginabijlage:

quickplace://ltsvr.ibm.com/49257043000214B3/QuickPlace%5Csamplplace  
%5CPageLibrary4925704300021490.nsf  
/A7986FD2A9CD47090525670800167225  
/2B02B1DE3A82B2CE49257043001C2498  
?AttNo=0&AttName==QPCons3.ppt

## UNIX-bestandssysteem-crawlers

De URI-indeling voor documenten die zijn gecrawld door een UNIX-bestands-systeem-crawler is:

`file:///Directory_Naam/Bestands_Naam`

### Parameters

URL-codering wordt op alle velden toegepast.

*Directory\_Naam*

De naam van het absolute pad voor de directory.

*Bestands\_Naam*

De naam van het bestand.

### Voorbeeld

`file:///home/user/test.doc`

## Web Content Management-crawlers

De URI-indeling voor WebSphere Content Management-documenten die zijn gecrawld door een Web Content Management-crawler is:

`wcm://Pagina_URL?pageID=Pagina_ID[&useSSL=true]`

### Parameters

URL-codering wordt op alle velden toegepast.

*Pagina\_URL*

De URL voor het document (uniek voor elk document).

*Pagina\_ID*

Het pagina-ID.

**useSSL** Als het protocol HTTPS is, wordt `useSSL=true` aan de URI toegevoegd. Anders wordt `useSSL` weggelaten.

### Voorbeelden

HTTP-protocol:

`wcm://wp6server.ibm.com:9081/wps/wcm/myconnect/Web+Content/Site01/SiteArea01/ContentTest01?pageID=6QReDeJ9DI3R0663E03Q06L1E2MR47MHOC3Q862RD6J0863B0GJS86J9E0`

HTTPS-protocol:

`wcm://wp6server.ibm.com:9444/wps/wcm/myconnect/Web+Content/Site01/SiteArea01/ContentTest01?pageID=6QReDeJ9DI3R0663E03Q06L1E2MR47MHOC3Q862RD6J0863B0GJS86J9E0&useSSL=true`

## WebSphere Portal-crawlers: WebSphere Portal versie 5

De URI-indeling voor WebSphere Portal versie 5-documenten die zijn gecrawld door een WebSphere Portal-crawler is:

`wps://Pagina_URL?portletDefID=Portlet_Def_ID&portletID=Portlet_ID&pageID=Pagina_ID[&useSSL=true]`

### Parameters

URL-codering wordt op alle velden toegepast.

*Pagina\_URL*

De URL voor het document (uniek voor elk document).

*Portlet\_Def\_ID*

Het portletdefinitie-ID.

*Portlet\_ID*  
Het portlet-ID.

*Pagina\_ID*  
Het pagina-ID.

**useSSL** Als het protocol HTTPS is, wordt useSSL=true aan de URI toegevoegd. Anders wordt useSSL weggelaten.

#### Voorbeelden

Lopende tekst in document:

```
wps://wpsserver.ibm.com:9081/wps/myportal!/ut/p/kcxml/04_Sj9SPykssy0x+LKnPy1vM0Y_QjzKCN4g3cQbJgQio-pFQAW99X4_83FT9AP2C5IhyR0dFRQD8qHRj/delta/base64xml/L01DU1kvd0NrQUpORUEvNFBVR0VoQSEvN18wXzZPLzZfMF80RA!!?portletDefID=3_0_3S&pageID=6_0_6J
```

#### Voorbeelden

Ingeschakeld voor SSL:

```
wps://wpsserver.ibm.com:9081/wps/myportal!/ut/p/kcxml/04_Sj9SPykssy0x+LKnPy1vM0Y_QjzKCN4g3cQbJgQio-pFQAW99X4_83FT9AP2C5IhyR0dFRQD8qHRj/delta/base64xml/L01DU1kvd0NrQUpORUEvNFBVR0VoQSEvN18wXzZPLzZfMF80RA!!?portletDefID=7_0_A4&pageID=6_0_6J&useSSL=true
```

## WebSphere Portal-crawlers: WebSphere Portal versie 6

De URI-indeling voor WebSphere Portal versie 6-documenten die zijn gecrawld door een WebSphere Portal-crawler is:

```
wp6://Pagina_URL?portletURL=Portlet_URL?portletDefID=Portlet_Def_ID&pageID=Pagina_ID[&useSSL=true]
```

#### Parameters

URL-codering wordt op alle velden toegepast.

*Pagina\_URL*  
De URL voor het document (uniek voor elk document).

*Portlet\_URL*  
De unieke URL voor het document.

*Portlet\_Def\_ID*  
Het portletdefinitie-ID.

*Pagina\_ID*  
Het pagina-ID.

**useSSL** Als het protocol HTTPS is, wordt useSSL=true aan de URI toegevoegd. Anders wordt useSSL weggelaten.

#### Voorbeelden

HTTP-protocol:

```
wp6://wp6server.ibm.com:9081/wps/myportal!/ut/p/c1/04_SB8K8xLLM9MSSzPy8xBz9CP0os3gjE59gQwMLQ0P_IDMnAyNHA3f3UEsTD1NjA6B8pFm8AQ7gaEBAadzjIPrz6_Tzyc1P1C3IjDHQdFRUBTu-saA!!/d12/d0/Y2BkbGBgY1rDwMDEJ1XAwMggYxZvZOITbGhgYWjobuhmaGDka0Bu5uHqFRpkaAAAEisaBQ!!?portletURL=/wps/myportal!/ut/p/c1/04_SB8K8xLLM9MSSzPy8xBz9CP0os3gjE59gQwMLQ0P_IDMnAyNHA3f3UEsTD1NjA6B8pFm8AQ7gaEBAadzjIPrz6_Tzyc1P1C3IjDHQdFRUBTu-saA!!/d12/d0/Y2BiUZnBwMqsyabykYGBmS2tcDoDE40MebyRiU-woYGFOaG7oZuhgZGjgbuZh6uXY4ihAYOMGQ7Z0CDBAwDcXPkM&portletDefID=6_24LS10811G1F102A0G6HEJUR10&pageID=3_24LS108110R6B02A0GGU94LN00
```

HTTPS-protocol:

```
wp6://wp6server.ibm.com:9444/wps/myportal!/ut/p/c1/04_SB8K8xLLM9MSSzPy
8xBz9CP0os3gjE59gQwMLQ0P_IDMnAyNHA3f3UEsTD1NjA6B8pFm8AQ7gaEBAzjIPrz6_
Tzyc1P1C3IjDHQdFRUBTu-saA!!/d12/d0/Y2BkbGBgY1rDwMDEJ1XAwMggYxZvZOITbGh
gYWjobuhmaGDka0Bu5uHq5RtqaAAA50L41Q!!
?portletUrl=/wps/myportal!/ut/p/c1/04_SB8K8xLLM9MSSzPy8xBz9CP0os3gjE59
gQwMLQ0P_IDMnAyNHA3f3UEsTD1NjA6B8pFm8AQ7gaEBAzjIPrz6_Tzyc1P1C3IjDHQdF
RUBTu-saA!!/d12/d0/Y2BiUZnBwMqsyBykYGBmS2tcDoDE40MebyRiU-woYGFoaG7oZu
hgZGjgbuZh6uXWZihAYOMGQ5Z31BDAAk73P2
&portletDefID=6_24LS10811G1F102A0G6HEJMU10
&pageID=3_24LS108110R6B02A0GGU94T410&useSSL=true
```

## Windows-bestandssysteem-crawlers

De URI-indeling voor documenten die zijn gecrawld door een Windows-bestands-systeem-crawler is:

```
file:///Directory_Naam/Bestands_Naam
file:///Netwerk_Map_Naam/Directory_Naam/Bestands_Naam
```

### Parameters

URL-codering wordt op alle velden toegepast.

#### *Directory\_Naam*

De naam van het absolute pad voor de directory.

#### *Bestands\_Naam*

De naam van het bestand.

#### *Netwerk\_Map\_Naam*

Alleen voor documenten op niet-lokale servers is dit de naam van de gemeenschappelijke map in een Windows-netwerk.

### Voorbeelden

Lokaal bestandssysteem:

```
file:///d:/directory/test.doc
```

Netwerkbestandssysteem:

```
file:///filesvr.ibm.com/directory/file.doc
```

### Verwante taken

“Categorieën configureren” op pagina 128

U kunt een onbeperkt aantal categorieën voor een collectie maken en elke categorie kan een onbeperkt aantal regels bevatten. Op basis van de regels wordt bepaald welke documenten automatisch aan de categorie moet worden gekoppeld.

“Scopes configureren” op pagina 173

Bij het configureren van een scope voor een enterprise search-collectie geeft u de URI's of URI-patronen op voor een reeks documenten in de index die door de gebruikers kunnen worden doorzocht.

“URI's uit de index verwijderen” op pagina 175

Om te voorkomen dat gebruikers bepaalde documenten in een collectie doorzoeken, kunt u de URI's van die documenten uit de index verwijderen.

“Snellinks configureren” op pagina 186

Als u een snellink wilt maken voor een enterprise search-collectie, moet u de URI van een document koppelen aan de trefwoorden op basis waarvan het document in de zoekresultaten wordt opgenomen.

“Gegevens over een URI bekijken” op pagina 301

U kunt gedetailleerde gegevens over een URI bekijken. U ziet actuele en historische informatie over de manier waarop het document dat door deze URI wordt vertegenwoordigd, is gecrawld, geïndexeerd en doorzocht.



“Rapporten over gewiste documenten bekijken” op pagina 318  
U kunt gedetailleerde informatie bekijken over documenten die uit een enterprise search-systeem zijn gewist. Deze informatie is alleen beschikbaar als documenttracering voor de collectie is ingeschakeld.



---

## Beheer van enterprise search-parser

Om de vindbaarheid van documenten te verbeteren, kunt u opties opgeven voor de manier waarop documenten en metagegevens worden geanalyseerd en gecategoriseerd voordat ze aan de enterprise search-index worden toegevoegd.

U kunt de volgende opties opgeven voor het analyseren van de documentcontent en het optimaliseren van de vindbaarheid van informatie:

### **Opties configureren voor het analyseren van Chinese, Japanse en Koreaanse documenten**

U kunt opties opgeven voor het gebruik van n-gramsegmentering waarmee u documenten kunt analyseren die zijn geschreven in het Chinees, Japans en Koreaans. Daarnaast kunt u nieuwe-regeltekens verwijderen uit de witruimten in Chinese en Japanse documenten.

### **Native XML-zoekopdrachten inschakelen**

Als uw collectie XML-documenten bevat, kunt u instellen dat de documenten kunnen worden gezocht met behulp van de native XML-querysyntaxis, zoals XPath- en XML-fragmenten. Met een native XML-zoekopdracht kunnen gebruikers query's opgeven op basis van de relatie tussen verschillende XML-elementen.

### **Categorieën configureren**

U kunt documenten groeperen die hetzelfde URI-patroon gebruiken of die bepaalde woorden in categorieën bevatten. Wanneer gebruikers de collectie doorzoeken, kunnen ze ervoor kiezen in de zoekresultaten alleen de documenten weer te geven die deel uitmaken van bepaalde categorieën.

### **Zoekvelden configureren**

U kunt elementen in XML-documenten toewijzen aan zoekvelden in de index. Daarnaast kunt u metagegevenselementen in HTML-documenten toewijzen aan zoekvelden. Als u zoekvelden in de zoekindex maakt, kunnen gebruikers query's uitvoeren op bepaalde delen van XML- en HTML-documenten zodat nauwkeurigere zoekresultaten worden verkregen.

### **Tekstverwerkingsopties configureren**

Als er aangepaste tekstanalyseprogramma's aan het enterprise search-systeem zijn toegevoegd, kunt u er een selecteren om bij een collectie te gebruiken. Als u een analyseprogramma aan een collectie hebt gekoppeld, kunt u opties opgeven voor het toewijzen van content zodat deze taalkundig kan worden geanalyseerd en geannoteerd. Daarnaast kunt u opgeven hoe de resultaten van de analyse moeten worden toegewezen aan de enterprise search-index of aan de JDBC-databasetabellen.

### **Velden toewijzen aan wegingsklassen**

U kunt opgeven dat documenten met velden die voldoen aan de zoektermen, een hogere ranking in de zoekresultaten krijgen dan andere documenten die aan de zoektermen voldoen. Als u velden toewijst aan wegingsklassen, kunt u aangeven welke content- en metagegevensvelden moeten worden gewogen. Daarnaast kunt u de scores configureren die voor elke wegingsklasse worden gebruikt voor de ranking van documenten.

### **Verwante onderwerpen**

Taalkundige ondersteuning voor semantische zoekopdrachten

Tekstanalyse in enterprise search

Basisconcepten gebruikt bij verwerking van tekstanalyse

Programma's voor semantische zoekopdrachten

Zoektermen in semantische zoekopdrachten

"Taal- en codetabelondersteuning voor enterprise search" op pagina 153

De taalkundige verwerking voor enterprise search wordt door de parser en de zoekservers verschillend afgehandeld.

"Detectie van documentindelingen voor enterprise search" op pagina 145

Voor enterprise search wordt een standaardtoewijzing van de URL-extensie en het MIME-type gebruikt om het documenttype vast te stellen en het parsertype dat voor elk document moet worden gebruikt.

---

## Werken met categorieën

Categorieën stellen u in staat om documenten die dezelfde eigenschappen hebben, te groeperen en om alleen documenten te doorzoeken en op te halen die voldoen aan de criteria op lid te zijn van die groep.

Als u documenten aan categorieën koppelt en uw zoekprogramma's deze functie ondersteunen, kunnen gebruikers een subset van de collectie doorzoeken door de naam van de categorie op te geven. Als ze de hele collectie doorzoeken, kunnen de gebruikers de zoekresultaten verfijnen en alleen door die documenten bladeren die tot dezelfde categorie behoren als een van de resultaatdocumenten.

Bij het configureren van een categorie geeft u regels op waarmee u de crawler opdracht geeft documenten aan de categorie te koppelen. U kunt documenten groeperen die hetzelfde URI-patroon hebben of documenten groeperen op basis van hun content (bijvoorbeeld documenten waarin bepaalde woorden wel, of juist niet, voorkomen).

Voor het maken en beheren van categorieën maakt u gebruik van de beheerconsole van enterprise search:

- U selecteert het type categorisering op het moment dat u de collectie maakt. U kunt kiezen of u geen categorieën of regelgebaseerde categorieën gebruikt.
- Wanneer u de parsingregels voor de collectie opgeeft, kunt u het type categorisering indien nodig wijzigen. Als u het type categorisering wijzigt nadat er al documenten zijn gecrawld en geïndexeerd, is de kwaliteit van de zoekresultaten lager totdat u alle documenten opnieuw hebt gecrawld en de hoofdindex opnieuw hebt opgebouwd.
- Als u ervoor kiest om regelgebaseerde categorieën te gebruiken, kunt u via de beheerconsole de categoriestructuur, de categorieën en de categorieregels beheren. Als u categorieën of categorieregels wijzigt nadat er al documenten zijn gecrawld en geïndexeerd, is de kwaliteit van de zoekresultaten lager totdat u alle documenten opnieuw hebt gecrawld en de hoofdindex opnieuw hebt opgebouwd.

## Regelgebaseerde categorieën

U kunt regels opstellen die bepalen welke documenten er bij een categorie in een enterprise search-collectie horen.

U kunt categorieregels maken voor collecties die u in enterprise search maakt en voor regelgebaseerde categorieën die u vanuit IBM WebSphere Portal-collecties

migreert. Om regels voor het categoriseren van documenten te configureren, moet u bij het maken van de collectie of bij het opstellen van de parsingregels voor de collectie opgeven dat u gebruik wilt maken van regelgebaseerde categorieën.

De parser gebruikt de regels die u opstelt voor het koppelen van documenten aan een of meer categorieën:

- Als een document aan minimaal één regel in een categorie voldoet, koppelt de parser dat document aan die categorie.
- Als een document aan minimaal één regel in verschillende categorieën voldoet, koppelt de parser dat document aan al die categorieën.
- Als een document aan geen enkele regel voor een categorie voldoet, koppelt de parser dat document niet aan de categorie. De gebruikers kunnen wel naar een dergelijk document zoeken en ze kunnen het ook ophalen als ze de collectie doorzoeken, maar als ze een categorie doorzoeken, wordt het document niet gevonden en niet opgehaald.

Bij het werken met de categoriestructuur (of taxonomie) van een collectie kunt u zelf bepalen op welke plaats in de hiërarchie van categorieën u een nieuwe categorie wilt toevoegen. U kunt de categoriestructuur ook gebruiken om categorieën te selecteren die u wilt wijzigen en om regels voor het categoriseren van documenten toe te voegen, te wissen of te wijzigen.

Bij het configureren van een regel voor het categoriseren van documenten moet u kiezen of enterprise search het document bij een bepaalde categorie moet indelen op basis van de URI van het document of de content ervan.

## Regels voor URI-patronen

Een URI-regel geldt voor de URI van het document. U geeft een gedeeltelijke URI op (een patroon) en alle documenten die het opgegeven patroon in hun URI hebben, voldoen aan de regel:

Als u bijvoorbeeld opgeeft dat de tekst van de regel `/hr/` is, dan voldoet de eerste van de onderstaande URI's wel aan de regel, maar de tweede niet:

```
file:///corporate/hr/doktersverklaring.doc  
http://bedrijf.com/human resources/doktersverklaring.htm
```

Omdat alle URI's worden opgevat als patronen, worden eventueel aan het begin of eind van het patroon geplaatste jokertekens (sterretjes) door het systeem genegeerd. Bijvoorbeeld: `*/hr/*` en `/hr/` leveren dezelfde set URI's op.

Regels voor URI-patronen zijn niet hoofdlettergevoelig. Als een URI spaties bevat, moet het URI-patroon voldoen aan de regels van enterprise search voor het coderen van URI's. Het volgende voorbeeld laat juiste en onjuiste manieren zien van het opgeven van de URI voor het pad van een Windows-bestandssysteem:

```
Onjuiste URI: file:///c:/program files/  
Juiste URI:  
file:///c:/program+files/
```

## regels voor documentcontent

Regels voor documentcontent worden op dezelfde manier uitgedrukt als query's. Als het document aan de query voldoet, voldoet het ook aan de regel. Bij het opge-

ven van een regel, geeft u op welke woorden of woordcombinaties het document moet bevatten (of juist niet mag bevatten) en kiest u een taal voor de regels voor het vinden van de stam.

Bijvoorbeeld: de volgende regel bepaalt dat een document, als het ofwel het woord hr ofwel de woordcombinatie human resources bevat, aan de regel voldoet:

```
hr "human resources"
```

Een ander voorbeeld: de volgende regel bepaalt dat een document, als de titel het woord "health" maar niet de woordcombinatie "employee benefits" bevat, aan de regel voldoet:

```
+title:health -title:"employee benefits"
```

Contentregels ondergaan dezelfde taalkundige normalisaties als query's met de SI-API (Search and Index API). De syntaxis voor contentregels omvat echter slechts een subset van de bewerkingen die beschikbaar zijn in de syntaxis voor SI-API-query's. Alleen de volgende query-operators zijn toegestaan in contentregels:

- + Een plusteken vóór een term om aan te geven dat die term aanwezig moet zijn in het document.
  - Een minteken vóór een term om aan te geven dat die term niet aanwezig mag zijn in het document.
  - " "
- Haakjes rond twee of meer woorden om aan te geven dat de exacte woordcombinatie in het document aanwezig moet zijn.

*field\_name:*

Als u vóór een woord of woordcombinatie een veldnaam plaatst, geeft u daarmee aan dat dat woord of die woordcombinatie moet voorkomen (of juist niet) in het genoemde veld.

Alle content- en metagegevensvelden in de collectie die zijn geconfigureerd als doorzoekbare velden, worden ondersteund. De volgende SI-API-trefwoorden en -typen voor velden worden echter niet ondersteund:

- site:
- url:
- link:
- docID:
- samegroupas:
- parametrische velden
- beveiligingstokens
- attributen (zoals \$source, \$language, \$doctype, etc.)

## Categoriestructuren

In een categoriestructuur ziet u alle regelgebaseerde categorieën in een collectie. Via de categoriestructuur kunt u categorieën maken en wissen, en wijzigingen aanbrengen in de regels waarmee documenten worden gekoppeld aan categorieën.

De categoriestructuur (ook wel de taxonomie genoemd) is hiërarchisch geordend. De boomstructuur begint met de hoofd- of hoofdcategorie; alle andere categorieën stammen van die hoofdcategorie af. U kunt een willekeurig aantal categorieën en subcategorieën nesten om de gebruikers zo veel mogelijk keus te geven voor het doorbladeren en ophalen van documenten.

Als een document bijvoorbeeld voldoet aan de regels van verschillende categorieën, wordt het aan al die categorieën gekoppeld. Als de gebruikers een categorie doorzoeken, of bij het doorkijken van de zoekresultaten door documenten bladeren die tot dezelfde categorie behoren, verhoogt het feit dat het document tot verschillende categorieën behoort de kans dat het door de gebruikers wordt gevonden.

Via het beheer van de categoriestructuur kunt u bepalen welke documenten tot een of meer categorieën behoren. Dit doet u door nieuwe categorieën in te voegen, of te "nesten", onder bestaande categorieën. Als u een categorie maakt, geeft u op of die categorie moet worden gemaakt op hoofdniveau of als subcategorie van een bestaande categorie. U kunt de categoriestructuur ook gebruiken om categorieën van de collecties te wissen en om de regels voor het koppelen van documenten aan categorieën te wijzigen. Het wijzigen van een categorie behelst het hernoemen van de categorie, het toevoegen of wissen van categoriseringsregels en het wijzigen van de content van afzonderlijke regels.

Bij het beheer van de categoriestructuur kunt u de volgende beschrijvingen van het zoek- en bladergedrag als richtlijn gebruiken:

- Als een gebruiker een categorie van een hoog niveau doorzoekt, wordt die categorie samen met alle subcategorieën doorzocht voor documenten die voldoen aan de zoekcriteria. Als een gebruiker een categorie doorzoekt die geen subcategorieën heeft, wordt alleen die categorie doorzocht.
- Als een gebruiker door de zoekresultaten bladert en een optie kiest om alleen in documenten van een bepaalde categorie te bladeren, worden alleen de documenten in die categorie afgebeeld. De namen van alle subcategorieën worden in de zoekresultaten eveneens afgebeeld, zodat de gebruiker van de ene categorie naar de andere kan gaan en de subsets van documenten een voor een kan bekijken.

## Een type categorisering selecteren

Bij het selecteren van een type categorisering geeft u de aanpak op die u wilt volgen bij het koppelen van documenten aan categorieën in de collectie.

### Voordat u begint

Om het type categorisering te wijzigen, moet u de rol van beheerder van enterprise search hebben of beheerder zijn van de collectie die u wilt wijzigen.

### Over deze taak

Het type categorisering is opgegeven tijdens het maken van de collectie. Indien nodig kunt u de manier wijzigen waarop documenten voor een collectie worden gecategoriseerd. U kunt regelgebaseerde categorieën gebruiken die u speciaal voor een collectie configureert of u kunt geen categorieën gebruiken.

**Belangrijk:** Als u het type categorisering wijzigt nadat u gegevens hebt gecrawld en een index voor een collectie hebt opgebouwd, wordt de index instabiel. Om te garanderen dat de zoekresultaten correct zijn, moet u alle documenten in de collectie opnieuw crawlen en vervolgens de hoofdindex opnieuw opbouwen.

### Procedure

Ga als volgt te werk om het type categorisering te selecteren:

1. Open een collectie, selecteer de pagina Parser en klik vervolgens op **Type categorisering selecteren**.

2. Selecteer een van de volgende opties op de pagina Type categorisering selecteren:

**Geen** Selecteer deze optie als u voor deze collectie geen gebruik wilt maken van categorieën.

**Regelgebaseerd**

Selecteer deze optie als u documenten wilt indelen op basis van regels die u specifiek voor deze collectie opgeeft.

## Categorieën configureren

U kunt een onbeperkt aantal categorieën voor een collectie maken en elke categorie kan een onbeperkt aantal regels bevatten. Op basis van de regels wordt bepaald welke documenten automatisch aan de categorie moet worden gekoppeld.

### Voordat u begint

Om categorieën te configureren, moet u de rol van beheerder van enterprise search hebben of beheerder zijn van de collectie waarvan de categorieën deel uitmaken.

Voor het categorietype moet de optie voor het gebruik van regelgebaseerde categorieën zijn geselecteerd.

Als u voorbeelden wilt van de manier waarop u regels kunt opgeven voor het koppelen van documenten aan categorieën klikt u tijdens het maken of bewerken van een categorie op **Help**.

### Over deze taak

Als uw zoekprogramma's ondersteuning bieden voor categorieën, kunnen gebruikers een subset van de collectie doorzoeken door de naam van de categorie op te geven. De gebruikers kunnen tevens een categorie in de zoekresultaten selecteren en het bladeren vervolgens beperken tot documenten die tot de geselecteerde categorie behoren.

**Belangrijk:** Als u de categorieën of categorieregels wijzigt nadat u gegevens hebt gecrawld en een index voor een collectie hebt opgebouwd, wordt de index instabiel. Om te garanderen dat de zoekresultaten correct zijn, moet u alle documenten in de collectie opnieuw crawlen en vervolgens de hoofdindex opnieuw opbouwen.

### Procedure

Ga als volgt te werk om een categorie te configureren:

1. Open een collectie, selecteer de pagina Parser en klik op **Categoriestructuur configureren**.
2. Selecteer op de pagina Categoriestructuur de locatie in de structuur waaraan u een categorie wilt toevoegen en klik op **Een categorie maken**.

Als u de root hebt geselecteerd, wordt de nieuwe categorie op hoofdniveau gemaakt. Als u een categorienaam selecteert, wordt de nieuwe categorie onder de geselecteerde categorie in de categoriestructuur genest.

Er verschijnt een wizard die u helpt bij het opgeven van de regels voor het koppelen van documenten aan de nieuwe categorie:

- a. Voer op de pagina Een categorie maken een beschrijvende naam in voor de categorie en klik vervolgens op **Volgende**.
- b. Klik op de pagina Categorieregels opstellen op **Regel toevoegen**.



- c. Voer in het veld **Regelnaam** op de pagina Categorieregels opstellen een unieke naam in voor de regel. Namen van categorieregels moeten uniek zijn binnen het totaal van alle categorieën in een collectie.
- d. Geef de regel op die u wilt gebruiken voor het koppelen van documenten aan deze categorie en klik vervolgens op **OK**.
- Als u de URI van een document wilt gebruiken om te bepalen of het document deel uitmaakt van de categorie, klikt u op **URI-patroon** en geeft u het URI-patroon op.  
Als de tekst die u opgeeft, aanwezig is in de URI, wordt het document aan deze categorie gekoppeld.  
Bijvoorbeeld: `file:///c:/program+files/finance`
  - Als u een query wilt uitvoeren op de doorzoekbare content om te bepalen of een document deel uitmaakt van de categorie, klikt u op **Document-inhoud**. Selecteer de taal van de documenten en geef vervolgens de woorden en woordcombinaties op die in de documentinhoud moeten worden weergegeven of verborgen.  
U drukt de regel in dezelfde indeling op als de query, met als uitzondering dat alleen de query-operatoren voor opnemen (+), uitsluiten (-), woordcombinatie (" ") en veldnaam (*veld\_naam:*) zijn toegestaan. N-gramsegmentering wordt niet ondersteund bij contentregels.  
Als de woorden die u opgeeft, in een document voorkomen (of juist niet), wordt dat document aan deze categorie gekoppeld.  
Bijvoorbeeld: `+financiën -boekhouding +titel:"fiscaal jaar"`
- e. Klik op **Voltooien**.  
De nieuwe categorie wordt op de pagina Categoriestructuur weergegeven met de overige categorieën die deel uitmaken van de collectie.

#### Verwante verwijzing

"URI-indelingen in een enterprise search-index" op pagina 111

Met de URI (Uniform Resource Identifier) van documenten in een enterprise search-index wordt het type crawler aangegeven waarmee het document aan de collectie is toegevoegd.

---

## Werken met XML-zoekvelden

U kunt XML-elementen aan zoekvelden toewijzen als u wilt instellen dat gebruikers bepaalde gedeelten van XML-documenten kunnen doorzoeken.

Gebruik de beheerconsole van enterprise search om XML-elementen aan zoekvelden toe te wijzen.

Meestal wordt alle tekst in een XML-document geïndexeerd. Als u XML-elementen toewijst aan zoekvelden, kunt u de structuurgegevens van XML-documenten gebruiken ter ondersteuning van specifiekere query's. U kunt de gegevens in een XML-element bijvoorbeeld doorzoekbaar maken op veldnaam en zorgen dat de gegevens in de zoekresultaten worden weergegeven door de XML-elementnaam aan een veldnaam toe te wijzen. U kunt de toewijzing nauwkeuriger maken door de kenmerken van het XML-element (en de waarden van deze kenmerken) op te geven als criteria voor het element op basis waarvan dit wordt aangemerkt als zoekveld.

## XML-zoekvelden

Met behulp van HTML-zoekvelden kunnen gebruikers query's uitvoeren op gedeelten van HTML-documenten.

XML-documenten worden steeds vaker gebruikt omdat ze zowel half-gestructureerde als niet-gestructureerde tekst kunnen bevatten. De structuur van XML is ingesloten en gebruikt een context die expliciet is gedefinieerd op basis van de XML-elementen die rond de tekst worden weergegeven. De naam van de auteur kan bijvoorbeeld als volgt worden afgebeeld:

```
<author>John Smith</author>
```

In deze context wordt met de tekst John Smith de auteur van een XML-document aangegeven.

In enterprise search kunt u XML-elementen koppelen of toewijzen aan veldnamen. Bij het configureren van analyseopties voor een collectie geeft u op welke XML-elementen aan welke zoekveldnamen moeten worden toegewezen. Als u XML-elementen aan zoekvelden toewijst, kunnen gebruikers waarden van deze elementen zoeken door de veldnamen in query's op te geven. Query's waarbij specifieke velden worden doorzocht, kunnen nauwkeurigere zoekresultaten opleveren dan query's in de vorm van vrije tekst, waarbij de volledige content van documenten wordt doorzocht.

Als uw collectie bijvoorbeeld XML-documenten bevat en u opgeeft dat de elementen <title> en <author> moeten worden gemarkeerd als zoekvelden in de index, kunnen gebruikers rechtstreeks query's uitvoeren op deze elementen. Bij een zoekopdracht naar author:Smith worden XML-documenten gevonden waarin Smith voorkomt in elementen die zijn toegewezen aan het veld author.

Stel dat het element <summary> informatie kan bevatten die nuttig is om in de zoekresultaten weer te geven. Als u het element <summary> toewijst aan een zoekveld en opgeeft dat de waarde van dit element in de zoekresultaten moet worden weergegeven, maakt de content van het element deel uit van het resultaatdocument.

Als u een XML-element toewijst aan een veldnaam, kan de tekst van het element worden doorzocht op basis van de veldnaam die u opgeeft. Als het XML-element kenmerken bevat, worden de kenmerkwaarden echter niet geïndexeerd en kunnen deze niet rechtstreeks worden doorzocht. Om een query op die kenmerkwaarden uit te voeren, moet u de analyseopties configureren en native XML-zoekopdrachten inschakelen. Als u native XML-zoekopdrachten hebt ingeschakeld, is de structuur van het XML-document beschikbaar in de index en kunt u een query uitvoeren op het document door XPath-querybeperkingen op te geven. Bijvoorbeeld:

```
@xpath::'//elementName[@attributeName="attributeValue"]
```

## XML-elementen aan zoekvelden toewijzen

Wanneer u een XML-element aan een zoekveld toewijst, geeft u aan welke XML-elementen gebruikers kunnen doorzoeken door een veldnaam in een query op te geven.

### Voordat u begint

Om XML-elementen aan zoekvelden toe te wijzen, moet u de rol van beheerder van enterprise search hebben of beheerder zijn van de collectie waarvan de XML-documenten deel uitmaken.

## Beperkingen

Per collectie kunnen verschillende XML-veldtoewijzingen bestaan, maar slechts één XML-hoofdelementtoewijzing. Met het hoofdelement van een XML-document worden de toewijzingen dienovereenkomstig toegepast.

## Over deze taak

Als u een XML-veldtoewijzing maakt of velden in een bestaande XML-veldtoewijzing toevoegt, wijzigt of wist, worden de wijzigingen van kracht op het moment dat u de parser opnieuw start. De nieuwe en gewijzigde toewijzingen zijn alleen van invloed op nieuwe gegevens die worden geanalyseerd nadat de parser opnieuw is gestart en zijn niet van invloed op de gegevens die al in de index zijn opgenomen. Om de documenten bij te werken die al in de index zijn opgenomen, moet u de documenten opnieuw crawlen en indexeren.

Voor deze taak wordt het volgende voorbeeld-XML-document gebruikt om aan te geven hoe u personeelsgegevens kunt toewijzen en kunt instellen dat gebruikers direct query's kunnen uitvoeren op bepaalde elementen.

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<personnel>
  <personnelrecord>
    <phone>5555</phone>
    <email>joe@us.ibm.com</email>
    <jobroles>Manager, architect
      <jobrole>Managing Search Development Group</jobrole>
      <jobrole>Architecting Search Technology</jobrole>
    </jobroles>
    <location>New York</location>
    <section id="expertise">
      <text>Linguistics</text>
    </section>
  </personnelrecord>
</personnel>
```

## Procedure

Ga als volgt te werk om de XML-elementen in dit voorbeeld aan zoekvelden toe te wijzen:

1. Open een collectie, selecteer de pagina Parser en klik op **XML-elementen toe-wijzen aan velden**.
2. Klik op de pagina XML-veldtoewijzingen op **XML-toewijzing maken**. De pagina XML-veldtoewijzing maken verschijnt.
3. Voer in het veld **Naam XML-hoofdelement** de naam in van het hoofdelement: personnel.

Zorg dat de naam die u hier opgeeft, exact overeenkomt met het hoofdelement in de XML-documenten die u wilt doorzoeken. Tijdens het analyseren en indexeren van XML-documenten selecteert enterprise search op basis van de naam van het hoofdelement de toewijzing die moet worden gebruikt.

4. Voer in het veld **Naam XML-toewijzing** een naam in voor deze set XML-veldtoewijzingsregels.

Nadat u een set XML-toewijzingsregels hebt gemaakt, wordt deze naam weer-gegeven op de pagina XML-veldtoewijzingen. Selecteer deze naam als u toewijzingsregels wilt toevoegen, wissen of wijzigen.

5. Wijs het XML-element jobrole toe aan het zoekveld jobrole:
  - a. Voer in het veld **Veldnaam** de waarde jobrole in.

**Tip:** Als u informatie wilt over de ASCII- en metatekens die niet worden ondersteund in veldnamen, klikt u op **Help**.

- b. Voer in het veld **Naam XML-element** de waarde `jobrole` in.
- c. Als u wilt instellen dat gebruikers query's kunnen uitvoeren op het veld `jobrole`, schakelt u het aankruisvakje **Veldzoekopdracht** in.
- d. Om ervoor te zorgen dat er alleen een match is als de zoektermen exact voldoen aan de volledige waarde van het veld `jobrole` (als er dus geen andere woorden aanwezig zijn in het veld), schakelt u het aankruisvakje **Volledige overeenkomst** in.
- e. Schakel het aankruisvakje **Sorteerbaar** in om in te stellen dat gebruikers de zoekresultaten kunnen sorteren op basis van de waarden in het veld `jobrole`.

Het veld wordt alleen als sorteerbaar veld afgebeeld als het zoekprogramma deze mogelijkheid ondersteunt. Het voorbeeldzoekprogramma voor enterprise search kent geen velden die worden toegewezen vanuit XML-elementen in de lijst van velden die u kunt selecteren voor het sorteren van zoekresultaten.

- f. Als de waarden van het veld `jobrole` in de zoekresultaten moeten worden weergegeven, schakelt u het aankruisvakje **Zoekresultaten** in.
6. Wijs het XML-element `jobroles` aan hetzelfde zoekveld toe:
- a. Klik op **Veld toevoegen** om een lege regel toe te voegen aan de lijst van veldtoewijzingsregels.
  - b. Voer in het veld **Veldnaam** de waarde `jobrole` in.
  - c. Voer in het veld **Naam XML-element** de waarde `jobroles` in.

**Tip:** De namen van XML-elementen hoeven niet overeen te komen met de namen van de zoekvelden. U kunt verschillende XML-elementen aan hetzelfde zoekveld toewijzen.

- d. Als u wilt instellen dat gebruikers query's kunnen uitvoeren op het veld `jobrole`, dat de zoektermen volledig moeten overeenkomen met de veldwaarde, dat u het veld wilt gebruiken om de zoekresultaten te sorteren en dat het veld in de zoekresultaten moet worden weergegeven, schakelt u de desbetreffende aankruisvakjes in.
7. Wijs het XML-element `section`, waaraan het kenmerk `ID` met de waarde `expertise` is gekoppeld, toe aan het zoekveld `expertise`:
- a. Klik op **Veld toevoegen** om een lege regel toe te voegen aan de lijst van veldtoewijzingsregels.
  - b. Voer in het veld **Veldnaam** de waarde `expertise` in.
  - c. Voer in het veld **Naam XML-element** de waarde `section` in.
  - d. Voer in het veld **Namen van XML-attributen** de waarde `id` in.
  - e. Voer in het veld **Waarde XML-kenmerk** de waarde `expertise` in.
  - f. Als u wilt instellen dat gebruikers query's kunnen uitvoeren op het veld `expertise`, schakelt u het aankruisvakje **Veldzoekopdracht** in.
  - g. Om ervoor te zorgen dat er alleen een match is als de zoektermen exact voldoen aan de volledige waarde van het veld `expertise` (als er dus geen andere woorden aanwezig zijn in het veld), schakelt u het aankruisvakje **Volledige overeenkomst** in.
  - h. Schakel het aankruisvakje **Sorteerbaar** in om in te stellen dat gebruikers de zoekresultaten kunnen sorteren op basis van de waarden in het veld `expertise` (indien het zoekprogramma deze mogelijkheid ondersteunt).

- i. Als de waarden van het veld `expertise` in de zoekresultaten moeten worden weergegeven, schakelt u het aankruisvakje **Zoekresultaten** in.

### Voorbeelden:

Als u alle personen in een organisatie wilt vinden die met zoekproducten werken, geeft u de volgende query op:

```
jobrole:search
```

Als u alle personen in een organisatie wilt vinden die ervaring hebben op taalkundig gebied, geeft u de volgende query op:

```
expertise:linguistics
```

---

## Werken met HTML-zoekvelden

U kunt HTML-metagegevens-elementen aan zoekvelden toewijzen als u wilt instellen dat gebruikers bepaalde secties met metagegevens in HTML-documenten kunnen doorzoeken.

Gebruik de beheerconsole om HTML-metagegevens-elementen aan zoekvelden toe te wijzen.

Als u HTML-metagegevens-elementen toewijst aan zoekvelden, kunnen gebruikers de HTML-documenten met nauwkeurigere query's doorzoeken.

### HTML-zoekvelden

Met behulp van HTML-zoekvelden kunnen gebruikers query's uitvoeren op de attributen van HTML-documenten.

De metagegevens-elementen in HTML-documenten zijn vergelijkbaar met documentattributen omdat ze informatie over het document bevatten, over de manier waarop het document is ingedeeld en de manier waarop het document toegankelijk is op het web. Bijvoorbeeld:

```
<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="openbaar" />
<meta name="abstract" content="In dit onderwerp wordt een IBM-product beschreven." />
<meta name="format" content="XHTML" />
```

Enterprise search kan de namen of HTML-metagegevens-elementen koppelen (of toewijzen) aan zoekveldnamen. Bij het configureren van de analyseopties voor een collectie geeft u op welke HTML-metagegevens-elementen aan welke zoekveldnamen moeten worden toegewezen. Als u HTML-metagegevens-elementen aan zoekvelden toewijst, kunnen gebruikers documenten met deze elementen zoeken door de namen van zoekvelden in query's op te geven. Query's waarbij specifieke velden worden doorzocht, kunnen nauwkeurigere zoekresultaten opleveren dan query's in de vorm van vrije tekst, waarbij de volledige content van documenten wordt doorzocht.

Als uw collectie bijvoorbeeld HTML-documenten bevat en u opgeeft dat de metagegevens-elementen `copyright` en `abstract` moeten worden geïndexeerd als zoekvelden, kunnen gebruikers query's uitvoeren op deze specifieke elementen.

Met de zoekopdracht copyright:IBM worden bijvoorbeeld HTML-documenten gevonden waarvoor de tekst IBM voorkomt in de copyright-metagegevens.

Tijdens het toewijzen van HTML-metagegevens-elementen aan zoekvelden geeft u op of u alle HTML-metagegevens-elementen wilt toewijzen, alleen de elementen in de Dublin Core-elementenset of alleen de HTML-metagegevens-elementen die u opgeeft. Voor een beschrijving van de elementen in de Dublin Core-metagegevens-elementenset raadpleegt u de website Dublin Core Metadata Initiative:

<http://dublincore.org/documents/dcmi-terms/#H2>

Als u toewijzingen maakt voor alle HTML-metagegevens-elementen of voor alle elementen in de Dublin Core-elementenset, komen de standaardzoekveldnamen overeen met de naam van het metagegevens-element. U kunt de standaardzoekveldnaam wijzigen en andere zoekopties opgeven voor bepaalde elementen door de elementen toe te voegen aan de lijst van elementen die u individueel wilt configureren.

## HTML-metagegevens-elementen toewijzen aan zoekvelden

Wanneer u een HTML-metagegevens-element aan een zoekveld toewijst, geeft u aan welke HTML-metagegevens-elementen gebruikers kunnen doorzoeken door een veldnaam in een query op te geven.

### Voordat u begint

Om HTML-metagegevens-elementen aan zoekvelden toe te wijzen, moet u de rol van beheerder van enterprise search hebben of beheerder zijn van de collectie waarvan de HTML-documenten deel uitmaken.

### Over deze taak

Als u een HTML-veldtoewijzing maakt of velden in een bestaande HTML-veldtoewijzing toevoegt, wijzigt of wist, worden de wijzigingen van kracht op het moment dat u de parser opnieuw start. De nieuwe en gewijzigde toewijzingen zijn alleen van invloed op nieuwe gegevens die worden geanalyseerd nadat de parser opnieuw is gestart en zijn niet van invloed op de gegevens die al in de index zijn opgenomen. Om de documenten bij te werken die al in de index zijn opgenomen, moet u de documenten opnieuw crawlen en indexeren.

### Procedure

Ga als volgt te werk om HTML-metagegevens-elementen aan zoekvelden toe te wijzen:

1. Open een collectie, selecteer de pagina Parser en klik op **XML-metagegevens toewijzen aan velden**.
2. Geef op de pagina HTML-veldtoewijzingen op welke HTML-metagegevens-elementen u aan de zoekvelden wilt toewijzen:
  - Als u alle HTML-elementen of alle elementen uit de Dublin Core-elementenset wilt toewijzen, geeft u op of alle gebruikers de velden moeten kunnen doorzoeken op basis van de veldnaam en geeft u op of de velden in de zoekresultaten moeten worden weergegeven. Daarnaast geeft u aan of de zoektermen van de gebruiker exact moeten voldoen aan de volledige waarde van het metagegevens-element (de woorden in het veld zijn de woorden die vol-

doen aan de zoektermen van de gebruiker) en of gebruikers de zoekresultaten kunnen sorteren op basis van dit veld.

In het systeem worden de namen van de metagegevens-elementen automatisch als de veldnamen gebruikt. Als u de standaardveldnamen wilt gebruiken voor specifieke elementen of als u andere zoekopties voor bepaalde elementen wilt gebruiken, kunt u een elementnaam toewijzen aan een unieke veldnaam en vervolgens de zoekopties opgeven die u voor dat element wilt gebruiken.

- Ga als volgt te werk om bepaalde metagegevens-elementen aan zoekvelden toe te wijzen:
  - a. Klik op **Veld toevoegen** om een lege regel toe te voegen aan de lijst van veldtoewijzingsregels.
  - b. Voer de naam in die u aan het HTML-metagegevens-element dat u toewijst, wilt koppelen. De gebruikers kunnen deze veldnaam opgeven wanneer ze een zoekopdracht opgeven voor de HTML-documenten in deze collectie.

**Tip:** Als u informatie wilt over de ASCII- en metatekens die niet worden ondersteund in veldnamen, klikt u op **Help**.

- c. Voer de naam in van het metagegevens-element dat u wilt toewijzen.
- d. Als u wilt instellen dat gebruikers query's kunnen uitvoeren op dit veld, schakelt u het aankruisvakje **Veldzoekopdracht** in.

**Tip:** Als dit aankruisvakje is uitgeschakeld, kan het veld niet worden gezocht via een veldzoekopdracht of via een vrije zoekopdracht. Deze actie kan nuttig zijn als u bijvoorbeeld de optie voor het opnemen van alle HTML-metagegevens-elementen of de Dublin Core-metagegevens-elementen hebt geselecteerd, maar wilt voorkomen dat bepaalde velden worden gezocht.

- e. Om in te stellen dat gebruikers query's op dit veld kunnen uitvoeren en om op te geven dat een document alleen overeenkomt als de zoektermen exact voldoen aan de volledige veldwaarde (als het veld geen andere woorden bevat), schakelt u het aankruisvakje **Volledige overeenkomst** in.
- f. Schakel het aankruisvakje **Sorteerbaar** in om in te stellen dat gebruikers de resultaten kunnen sorteren op basis van dit veld.

Het veld wordt alleen als sorteerbaar veld afgebeeld als het zoekprogramma deze mogelijkheid ondersteunt. Het voorbeeldzoekprogramma voor enterprise search kent geen velden die worden toegewezen vanuit HTML-metagegevens-elementen in de lijst van velden die u kunt selecteren voor het sorteren van zoekresultaten.

- g. Als het gegevenstype van dit veld DECIMAL, DOUBLE, INTEGER, SHORT, TIME of TIMESTAMP is en u wilt instellen dat gebruikers parametrische zoekopdrachten kunnen opgeven bij het zoeken naar dit veld, schakelt u het aankruisvakje **Parametrische zoekopdracht** in.
- h. Als dit veld in de zoekresultaten moet worden weergegeven, schakelt u het aankruisvakje **Zoekresultaten** in.

### Voorbeeld:

Gebruikers kunnen een query uitvoeren op de toegewezen veldnamen om HTML-documenten met bepaalde metagegevens te zoeken. Als u bijvoorbeeld het

metagegevens-element beschrijving hebt toegewezen aan het zoekveld abstract, kunnen gebruikers de volgende HTML-documenten met een beschrijving van Thinkpad-computers ophalen:

```
abstract:thinkpad
```

---

## Aangepaste tekstverwerking

U kunt de kwaliteit en nauwkeurigheid van zoekresultaten verbeteren door aangepaste tekstverwerkingsalgoritmen in enterprise search-collecties te integreren.

WebSphere Information Integrator OmniFind Edition biedt ondersteuning voor UIMA (IBM Unstructured Information Management Architecture), een framework voor het maken, vinden, samenstellen en instellen van functies voor tekstanalyses. Programmaontwikkelaars maken en testen analysealgoritmen voor de content die moet worden doorzocht en maken vervolgens een verwerkingsprogramma-archief (.pear-bestand) dat alle resources bevat die nodig zijn om het archief te kunnen gebruiken voor enterprise search. Als u collecties wilt doorzoeken met behulp van de aangepaste analysealgoritmen, moet u het archief (dat het tekstanalyseprogramma bevat) toevoegen aan het enterprise search-systeem.

De component voor analyselogica in een tekstanalyseprogramma wordt een *annotator* genoemd. Elke annotator voert bepaalde taalkundige analysetaken uit. Een tekstverwerkingsprogramma kan een willekeurig aantal annotators bevatten of kan zijn samengesteld uit verschillende tekstanalyseprogramma's die allemaal over eigen aangepaste annotators beschikken.

De gegevens die door de annotators worden geleverd, worden de *analyseresultaten* genoemd. Analyseresultaten, die overeenkomen met de gegevens waarnaar u wilt zoeken, worden geschreven naar een gegevensstructuur die ook wel de *Common Analysis Structure* wordt genoemd.

Als u de tekstverwerkingsopties voor een collectie wilt configureren, moet u de volgende taken uitvoeren:

- Selecteer het tekstanalyseprogramma dat u wilt gebruiken om voor het annoteren van de documenten in de collectie.
- Als uw collectie XML-documenten met betekenisvolle markup bevat en u deze markup wilt gebruiken in de aangepaste tekstanalyse, kunt u toewijzingsbestanden aan de collectie toewijzen en de uitvoer van de XML-toewijzing aan de Common Analysis Structure koppelen.

U kunt bijvoorbeeld de content van de elementen <Geadresseerde> en <Klant> toewijzen aan aantekeningen van het type Persoon in de Common Analysis Structure. Deze aantekeningen kunnen vervolgens worden gebruikt door de aangepaste annotators, die extra informatie kunnen vinden (zoals het geslacht van de persoon). U kunt ook aantekeningen van het type Persoon toewijzen aan de enterprise search-index, zodat gebruikers kunnen zoeken naar personen zonder dat ze de oorspronkelijke XML-elementen hoeven te kennen.

Als u wilt instellen dat gebruikers de oorspronkelijke XML-elementen in query's kunnen opgeven, hoeft u geen XML-toewijzingen te configureren. In plaats daarvan kunt u analyseopties configureren en native XML-zoekopdrachten inschakelen voor de collectie.

- Wijs de Common Analysis Structure toe aan de enterprise search-index, zodat de geannoteerde documenten kunnen worden doorzocht met behulp van semantische zoekopdrachten.



Afhankelijk van de entiteiten en relaties die door de annotators worden gevonden, kunnen gebruikers bijvoorbeeld zoeken naar begrippen die in dezelfde zin staan (zoals de naam van een specifieke persoon en die van een concurrent), of naar een trefwoord en een begrip (zoals de naam Alex en een telefoonnummer).

- Wijs de Common Analysis Structure toe aan een relationele database. U kunt gegevens toewijzen aan IBM DB2 Universal Database (DB2 UDB) of Oracle-tabellen. Met dit type toewijzing kunnen de resultaten van analyses in database-programma's worden gebruikt, zoals datamining programma's. Daarnaast kunt u met dit type toewijzing SQL-query's gebruiken om gegevens buiten enterprise search te doorzoeken.

#### **Verwante onderwerpen**

Integratie van aangepaste tekstanalyse

Basisconcepten gebruikt bij verwerking van tekstanalyse

Werkstroom voor integratie van aangepaste analyses

Algoritmen voor tekstanalyses

Programma's voor semantische zoekopdrachten

Zoektermen in semantische zoekopdrachten

## **Tekstanalyseprogramma's aan het systeem toevoegen**

Als u een aangepast tekstanalyseprogramma maakt, moet u dit aan het systeem toevoegen voordat u het programma kunt gebruiken in enterprise search. Collecties kunnen een tekstanalyseprogramma gebruiken voor het analyseren en annoteren van documenten en daarmee het de nauwkeurigheid van de zoekresultaten verbeteren.

### **Voordat u begint**

Om tekstanalyseprogramma's aan het systeem toe te voegen, moet u de rol van beheerder van enterprise search hebben.

### **Over deze taak**


Programmaontwikkelaars kunnen een verwerkingsprogramma-archief maken (.pear) dat voldoet aan het UIMA-framework voor tekstanalyses. Het archief bevat alle resources die nodig zijn voor het doorzoeken van enterprise search-collecties. Als u collecties wilt doorzoeken met behulp van de aangepaste analysealgoritmen, moet u het archief (dat het tekstanalyseprogramma bevat) toevoegen aan het enterprise search-systeem.

Nadat u een tekstanalyseprogramma aan het systeem hebt toegevoegd, kunt u de weergavenaam wijzigen en een optie selecteren om de XML-bron te bekijken. De XML-bron laat zien welke informatie door dit programma is geproduceerd.

Als aan de collectie al een tekstanalyseprogramma is gekoppeld, kunt u het tekstanalyseprogramma niet van het systeem verwijderen.

### **Procedure**

Ga als volgt te werk om een aangepast tekstanalyseprogramma aan het enterprise search-systeem toe te voegen:

1. Klik op **Systeem** om de view Systeem te openen.
2. Klik op  **Wijzigen** om naar de view Systeem wijzigen te gaan.
3. Klik op de pagina Parser op **Tekstanalyseprogramma's configureren**.

4. Klik op de pagina Tekstanalyseprogramma's op **Tekstanalyseprogramma toevoegen**.
5. Voer op de pagina Tekstanalyseprogramma toevoegen een beschrijvende naam in voor het nieuwe programma. Deze weergavenaam wordt gebruikt om het tekstanalyse aan te duiden in de beheerconsole.
6. Geef de locatie van het .pear-bestand op. Als het bestand kleiner is dan 8 MB, kan het op uw lokale computer staan en kunt u op Bladeren zoeken om het bestand te zoeken. Als het bestand groter is dan 8 MB, moet het bestand op de indexserver staan en moet u het volledige pad naar het bestand invoeren.
7. Klik op **OK**. Het tekstanalyseprogramma wordt weergegeven op de pagina Tekstanalyseprogramma's.

#### **Verwante onderwerpen**

Werkstroom voor integratie van aangepaste analyses

Integratie van aangepaste tekstanalyse

Basisconcepten gebruikt bij verwerking van tekstanalyse

XML-markup in analyses en zoekopdrachten

#### **Verwante taken**

Een toewijzingsbestand voor toewijzing van XML-elementen aan de Common Analysis Structure maken

## **Een tekstanalyse aan een collectie koppelen**

Als er aangepaste tekstanalyseprogramma's aan het enterprise search-systeem zijn gekoppeld, kunt u er een selecteren om bij een collectie te gebruiken. Gebruikers kunnen dan bij het doorzoeken van collecties semantische query's opgeven en de kwaliteit en nauwkeurigheid van de zoekresultaten verbeteren.

### **Voordat u begint**

Om een tekstanalyseprogramma aan een collectie te koppelen, moet u de rol van beheerder van enterprise search hebben of beheerder van de collectie zijn.

### **Over deze taak**

Als aan de collectie al een tekstanalyseprogramma is gekoppeld, worden de volgende acties uitgevoerd wanneer u een ander programma koppelt:

- Als u **Geen aangepaste analyse** selecteert, worden alle toewijzingen voor tekstanalyse die u eerder hebt gedefinieerd voor de collectie, gereset. Vanaf dat moment gebruikt de collectie de standaardwaarden van het systeem.
- Als u de naam van een ander aangepast analyseprogramma selecteert, blijven alle toewijzingen voor tekstanalyse die u eerder hebt gedefinieerd voor de collectie, behouden. Als u bijvoorbeeld programma\_1 verandert in programma\_2, dan neemt programma\_2 de voor programma\_1 gedefinieerde XML-toewijzingsbestanden over.

### **Procedure**

Ga als volgt te werk om een tekstanalyseprogramma aan een collectie te koppelen:

1. Open een collectie, selecteer de pagina Parser en klik op **Tekstverwerkingsopties configureren**.

2. Klik op **Tekstanalyseprogramma selecteren**. Als er geen aangepaste tekstanalyseprogramma's aan het enterprise search-systeem zijn gekoppeld of als de collectie gebruikmaakt van de standaardanalysealgoritmen, is de programmaam **Standaard**.
3. Selecteer op de pagina Tekstanalyseprogramma voor deze collectie selecteren de naam van het programma dat voor de collectie moet worden gebruikt. Als er geen tekstanalyseprogramma's beschikbaar zijn of als u **Geen aangepaste analyse** hebt geselecteerd, past de parser de standaardregels voor tekstanalyse toe tijdens het annoteren van documenten en het voorbereiden van documenten voor de index.

#### **Verwante onderwerpen**

Werkstroom voor integratie van aangepaste analyses

Integratie van aangepaste tekstanalyse

Basisconcepten gebruikt bij verwerking van tekstanalyse

## **XML-elementen toewijzen aan de Common Analysis Structure**

Als uw collectie XML-documenten met een betekenisvolle markup bevat en u deze markup wilt gebruiken om ervoor te zorgen dat gebruikers de enterprise search-index of relationele databasetabellen kunnen doorzoeken met semantische zoekopdrachten, kunt u de XML-elementen aan de Common Analysis Structure toewijzen.

#### **Voordat u begint**

Om XML-elementen toe te wijzen aan de Common Analysis Structure, moet u de rol van beheerder van enterprise search hebben of beheerder van de collectie zijn.

#### **Beperkingen**

De maximumgrootte van een toewijzingsbestand is 8 MB.

#### **Over deze taak**

U kunt aangepaste toewijzingsbestanden maken om aangepaste tekstanalyseprocessen in te schakelen voor toegang tot bepaalde elementen in XML-documenten, of om verschillende XML-elementen toe te wijzen aan een aangepast type voor gebruik in semantische zoekopdrachten. De toewijzingsbestanden moeten voldoen aan het UIMA-framework voor tekstanalyse.

Als u toewijzingsbestanden toevoegt aan een collectie waarvoor een aangepast tekstanalyseprogramma wordt gebruikt, zorgt u dat de XML-elementen in bron-documenten worden toegewezen aan aantekeningen in de Common Analysis Structure. Deze aantekeningen kunnen vervolgens worden gebruikt door het aangepaste tekstanalyseprogramma. U kunt de Common Analysis Structure toewijzen aan de index en instellen dat gebruikers query's kunnen uitvoeren op de aantekeningen terwijl ze de collectie doorzoeken met semantische zoekopdrachten.

U kunt bijvoorbeeld de content van de elementen Geadresseerde en Klant toewijzen aan aantekeningen van het type Persoon in de Common Analysis Structure. Deze aantekeningen kunnen vervolgens worden gebruikt door de aangepaste annotators, die extra informatie kunnen vinden (zoals het geslacht van de persoon). U kunt ook aantekeningen van het type Persoon toewijzen aan de enterprise search-index, zodat gebruikers kunnen zoeken naar personen zonder dat ze de oorspronkelijke XML-elementen hoeven te kennen.

Als u wilt instellen dat gebruikers de oorspronkelijke XML-elementen in query's kunnen opgeven, hoeft u geen toewijzingsbestanden te configureren. In plaats daarvan kunt u analyseopties configureren en native XML-zoekopdrachten inschakelen voor de collectie.

### Procedure

Ga als volgt te werk om XML-elementen aan de Common Analysis Structure toe te wijzen:

1. Open een collectie, selecteer de pagina Parser en klik op **Tekstverwerkings-opties configureren**.
2. Klik in het gebied **XML-elementen toewijzen aan de Common Analysis Structure** op **Toewijzing toevoegen**.
3. Voer op de pagina XML-elementen toevoegen aan de Common Analysis Structure een beschrijvende weergavenaam in voor het toewijzingsbestand.
4. Geef de locatie van het bestand op. Als het bestand is opgeslagen op het lokale systeem, kunt u op Bladeren klikken en het bestand opzoeken. Als het toewijzingsbestand is opgeslagen op de indexserver, moet u het volledige pad voor het bestand typen.
5. Klik op **OK**. Het nieuwe toewijzingsbestand wordt toegevoegd aan de pagina Opties voor tekstverwerking.

#### Verwante onderwerpen

Werkstroom voor integratie van aangepaste analyses

Integratie van aangepaste tekstanalyse

Basisconcepten gebruikt bij verwerking van tekstanalyse

XML-markup in analyses en zoekopdrachten

#### Verwante taken

Een toewijzingsbestand voor toewijzing van XML-elementen aan de Common Analysis Structure maken

## De Common Analysis Structure toewijzen aan de index

U kunt aangeven welke tekstanalyseresultaten uit een Common Analysis Structure aan de index moeten worden toegewezen en beschikbaar moeten worden gesteld aan gebruikers die met behulp van semantische zoekopdrachten query's uitvoeren op een collectie.

### Voordat u begint

Om de Common Analysis Structure toe te wijzen aan de index, moet u de rol van beheerder van enterprise search hebben of beheerder van de collectie zijn.

### Beperkingen

De maximumgrootte van een toewijzingsbestand is 8 MB.

### Over deze taak

Als u de Common Analysis Structure aan de enterprise search-index toewijst, kunnen gebruikers nauwkeurige semantische zoekopdrachten opgeven en wordt de kwaliteit van de zoekopdrachten verbeterd.

Afhankelijk van de entiteiten en relaties die door de annotators worden gevonden, kunnen gebruikers dan bijvoorbeeld zoeken naar begrippen die in dezelfde zin staan (zoals de naam van een specifieke persoon en die van een concurrent), of naar een trefwoord en een begrip (zoals de naam Alex en een telefoonnummer).

### Procedure

Ga als volgt te werk om de Common Analysis Structure aan de index toe te wijzen:

1. Open een collectie, selecteer de pagina Parser en klik op **Tekstverwerkings-opties configureren**.
2. Klik in het gebied **De Common Analysis Structure toewijzen aan de index** op **Toewijzingsbestand selecteren**.
3. Selecteer op de pagina Toewijzingsbestand voor deze collectie selecteren het toewijzingsbestand dat u wilt gebruiken voor de enterprise search-index:
  - Als u de standaardtoewijzingsregels voor de enterprise search-index wilt gebruiken, selecteert u **Standaard**.
  - Als u een aangepaste Common Analysis Structure aan de index wilt toewijzen, geeft u de locatie van het toewijzingsbestand op. Als het bestand is opgeslagen op het lokale systeem, kunt u op Bladeren klikken en het bestand opzoeken. Als het bestand is opgeslagen op de indexserver, moet u het volledige pad voor het bestand typen.
4. Klik op **OK**. Het toewijzingsbestand dat u hebt opgegeven, wordt op de pagina Opties voor tekstverwerking weergegeven.

#### Verwante onderwerpen

Werkstroom voor integratie van aangepaste analyses

Integratie van aangepaste tekstanalyse

Basisconcepten gebruikt bij verwerking van tekstanalyse

Indextoewijzing voor aangepaste-analyseresultaten

#### Verwante taken

Een toewijzingsbestand voor de toewijzing van de Common Analysis Structure aan een index maken

## De Common Analysis Structure toewijzen aan een relationele database

U kunt aangeven welke tekstanalyseresultaten uit een Common Analysis Structure moeten worden toegewezen aan een relationele database voor gebruik in database-programma's.

### Voordat u begint

Om de Common Analysis Structure toe te wijzen aan een relationele database, moet u de rol van beheerder van enterprise search hebben of beheerder van de collectie zijn.

### Beperkingen

De maximumgrootte van een toewijzingsbestand is 8 MB.

### Over deze taak

Door de Common Analysis Structure aan relationele databasetabellen toe te wijzen, kunnen de gegevens in databaseprogramma's worden gebruikt. Gebruikers kunnen bijvoorbeeld SQL-query's buiten enterprise search opgeven om de annotaties te zoeken die door de tekstanalyse zijn toegevoegd. Daarnaast kunt u de informatie gebruiken voor verdere tekstverwerking, bijvoorbeeld voor gebruik in dataminingprogramma's.

U kunt één toewijzingsbestand tegelijk gebruiken om een Common Analysis Structure aan een relationele database toe te wijzen. Als u een nieuw toewijzingsbestand uploadt, wordt het huidige toewijzingsbestand overschreven.

### Procedure

Ga als volgt te werk om de Common Analysis Structure aan een relationele database toe te wijzen:

1. Open een collectie, selecteer de pagina Parser en klik op **Tekstverwerkings-opties configureren**.
2. Klik in het gebied **De Common Analysis Structure toewijzen aan een relationele database** op **Toewijzing toevoegen**.
3. Voer op de pagina De Common Analysis Structure toewijzen aan een relationele database een beschrijvende weergavenaam in voor het toewijzingsbestand dat u wilt gebruiken om de informatie toe te wijzen aan een relationele database.
4. Geef de locatie van het toewijzingsbestand op. Als het bestand is opgeslagen op het lokale systeem, kunt u op Bladeren klikken en het bestand opzoeken. Als het bestand is opgeslagen op de indexserver, moet u het volledige pad voor het bestand typen.
5. Klik op **OK**. De weergavenaam van het toewijzingsbestand wordt weergegeven op de pagina Opties voor tekstverwerking.

#### Verwante onderwerpen

Werkstroom voor integratie van aangepaste analyses

Integratie van aangepaste tekstanalyse

Basisconcepten gebruikt bij verwerking van tekstanalyse

Databasetoewijzingen voor geselecteerde analyseresultaten

#### Verwante taken

Een toewijzingsbestand maken voor het toewijzen van de Common Analysis Structure aan een database

---

## Threads voor de parserservice configureren

Als er voldoende geheugenresources in uw systeem aanwezig zijn, kunt u het aantal threads vergroten dat in de parser beschikbaar is voor het analyseren van documenten.

### Voordat u begint

Als u een groot aantal collecties hebt, kan het verstandig zijn het aantal parserthreads te verhogen. Zorg dat uw systeem voldoende geheugen bevat voor extra threads. Een parser met één thread heeft 200 MB geheugen nodig. Voor elke extra thread wordt nog eens 50 MB verlangd.

Om het aantal threads te configureren dat voor de parser wordt gestart, moet u de rol van beheerder van enterprise search hebben of beheerder van een collectie zijn.

## Procedure

Ga als volgt te werk om het aantal parserthreads te configureren:

1. Open een collectie, selecteer de pagina Parser en klik op **Parseropties configureren**.
2. Geef op hoeveel parserthreads er maximaal kunnen worden gestart voor de parser en klik op **OK**.
3. Start de parser opnieuw.

---

## Geavanceerde analyse voor samengestelde termen inschakelen

U kunt de kwaliteit van zoekopdrachten verbeteren door in de parser het gebruik van geavanceerde analyse van samengestelde termen in te schakelen. Dit houdt in dat de samengestelde termen worden ontleed zodat elk deel als afzonderlijke term kan worden behandeld.

### Voordat u begint

Om opties op te geven voor het analyseren van samengestelde termen, moet u de rol van beheerder hebben of beheerder van de collectie zijn.

### Over deze taak

In sommige talen worden termen die uit meerdere woorden bestaan, samengevoegd tot één woord zonder spaties (*samengestelde* termen). Het geavanceerd analyseren en het ontleden van samengestelde termen is met name nuttig voor het doorzoeken van documenten in talen zoals het Duits, en is van essentieel belang voor het doorzoeken van documenten in talen zoals het Koreaans.

Als u geavanceerde analyse inschakelt voor samengestelde termen, kunnen gebruikers zoeken naar termen zonder dat ze jokertekens hoeven te gebruiken om samengestelde vormen van de zoektermen te vinden. Als u bijvoorbeeld de term *Organ* (*organ*) zoekt, kunnen documenten als resultaat worden gegeven waarin de term *Organ spender* (*organ donor*) voorkomt, terwijl documenten met de term *Organisation* (*organization*) niet worden geretourneerd. In tegenstelling tot de zoekopdracht met jokertekens *Organ\**, waarmee alles wat op *Organ* volgt als resultaat wordt gegeven, geldt de zoekopdracht hier alleen voor volledige taalkundige subwoorden binnen de grotere samengestelde termen.

De door de gebruiker gedefinieerde termen, zoals synoniemen en gewogen woorden, zijn ook van toepassing op samengestelde delen die als afzonderlijke woorden in de zoekopdracht worden gebruikt.

## Procedure

Ga als volgt te werk om geavanceerde analyse van samengestelde termen in te schakelen:

1. Open een collectie, selecteer Parser en klik op **Parseropties configureren**.
2. Schakel het aankruisvakje **Geavanceerde analyse voor samengestelde termen inschakelen** in en klik op **OK**.

### Verwante onderwerpen

Taalkundige ondersteuning voor semantische zoekopdrachten

Tekstanalyse in enterprise search

---

## Ondersteuning voor native XML-zoekopdrachten inschakelen

Als een collectie XML-documenten bevat, kunt u instellen dat gebruikers de XML-markup kunnen gebruiken tijdens het doorzoeken van documenten. Dit doet u door native XML-zoekopdrachten voor de collectie in te schakelen.

### Voordat u begint

Om ondersteuning in te schakelen voor het doorzoeken van XML-documenten met native XML-zoekopdrachten, moet u de rol van beheerder van enterprise search hebben of beheerder van de collectie zijn.

### Over deze taak

Een native XML-zoekopdracht, zoals XPath of XML-fragmenten, kan nauwkeurigere zoekresultaten opleveren omdat gebruik wordt gemaakt van de XML-markup van de documenten in de query. Gebruikers kunnen opgeven dat een zoekterm moet voorkomen in een bepaald XML-element of -kenmerk.

Facturen van een computerdealer die in XML-indeling zijn opgeslagen, kunnen bijvoorbeeld vermeldingen van het type <order> bevatten die de elementen <bedrijf> en <computertype> bevatten. Om de facturen op te halen met orders voor IBM-notebooks, geeft u een zoekopdracht met de trefwoorden IBM en notebook op om documenten weer te geven die Dell-notebookcomputers en IBM-desktopmodellen bevatten. Als u gebruikmaakt van een XML-zoekopdracht, kunt u opgeven dat de term IBM moet voorkomen in het element <bedrijf>, dat de term notebook moet voorkomen in het element <computertype> en dat beide elementen onder hetzelfde element <order> moeten voorkomen. Op deze manier worden facturen opgehaald die specifiek van toepassing zijn op IBM-notebooks.

### Procedure

Ga als volgt te werk om in te stellen dat gebruikers collecties kunnen doorzoeken met native XML-zoekopdrachten:

1. Open een collectie, selecteer de pagina Parser en klik op **Parseropties configureren**.
2. Selecteer **Gebruikers in staat stellen XML-documenten te doorzoeken met native XML-zoekopdrachten**.
3. Klik op **OK**.

#### Verwante onderwerpen

Taalkundige ondersteuning voor semantische zoekopdrachten

Tekstanalyse in enterprise search

Programma's voor semantische zoekopdrachten

Zoektermen in semantische zoekopdrachten



---

## Detectie van documentindelingen voor enterprise search

Voor enterprise search wordt een standaardtoewijzing van de URL-extensie en het MIME-type gebruikt om het documenttype vast te stellen en het parsertype dat voor elk document moet worden gebruikt.

In het configuratiebestand `parserTypes.cfg` kunt u de standaardtoewijzing van URL-extensies en MIME-typen aan parsertypen wijzigen en uitbreiden. Het bestand `parserTypes.cfg` bevat de regels voor de toewijzing van bestandsextensies of MIME-typen aan parsertypen. U kunt bijvoorbeeld de bestandsextensie `.content` toewijzen en opgeven dat documenten met dat type moeten worden geanalyseerd met behulp van de HTML-parser.

Documentindelingen beschikken over verschillende interne voorstellingen. In een enterprise search-systeem worden interne filters en filters van derden gebruikt voor het analyseren van documenten, en veel documenten worden geanalyseerd met parserservices die speciaal zijn ontworpen voor een bepaalde indeling.

Het detecteren van de documentindeling en de toewijzing van de parser gebeurt op de volgende manier:

1. Met het algoritme voor de detectie van de documentindeling wordt de extensie van de URL van het verwerkte document gecontroleerd.
2. Het systeem controleert het MIME-type van het document, dat deel uitmaakt van de metagegevens van de crawler.
3. Vervolgens wordt geprobeerd het juiste parsertype aan elk document toe te wijzen. Voor HTML-, tekst- (TXT) en XML-documenten wijst het systeem een parsertype toe dat specifiek is voor elk van deze documentindelingen. Voor alle overige documentindelingen wordt de Stellant-parser gebruikt. De Stellant-parser ondersteunt honderden documentindelingen, maar slechts een subset van de documentfilters is ingeschakeld voor enterprise search. U kunt echter de configuratiebestanden bewerken om in te stellen dat de overige documenttypen moeten worden geanalyseerd door de Stellant-parser.
4. Als het bestand `parserTypes.cfg` niet beschikbaar is, wordt de standaardtoewijzing gebruikt om het documenttype en de parser vast te stellen. Om het documenttype vast te stellen, worden de volgende stappen uitgevoerd:
  - a. De URL-extensie wordt vergeleken met de door de klant gedefinieerde extensieregels in het bestand `parserTypes.cfg`.
  - b. Het MIME-type wordt vergeleken met de door de klant gedefinieerde MIME-typeregels in het bestand `parserTypes.cfg`.
  - c. De URL-extensie wordt vergeleken met de standaardregels voor enterprise search.
  - d. Het MIME-type wordt vergeleken met de standaard-MIME-typeregels voor enterprise search.
5. Als de documentindeling van een document niet kan worden vastgesteld, wordt het document afgewezen. Mogelijk verschijnt dan een foutbericht waarin wordt aangegeven dat het documenttype niet wordt ondersteund.

Als Stellant is toegewezen als het parsertype, kan er een foutbericht verschijnen als de documentindeling niet kan worden herkend door Stellant. De fout kan optreden als:

- Het document beschadigd is.

- Het document is opgeslagen in een indeling die niet door Stellent wordt ondersteund. Om dit probleem op te lossen, moet u de afgewezen documentindelingen toevoegen aan het bestand `stellentTypes.cfg`. Daarnaast moet u het bestand `parserTypes.cfg` bijwerken en aangeven dat het MIME-type of de extensie van de afgewezen documentindelingen moeten worden gekoppeld aan de Stellent-parser.

#### Verwante onderwerpen

“Beheer van enterprise search-parser” op pagina 123

Om de vindbaarheid van documenten te verbeteren, kunt u opties opgeven voor de manier waarop documenten en metagegevens worden geanalyseerd en gecategoriseerd voordat ze aan de enterprise search-index worden toegevoegd.

---

## Documenttypen die standaard worden ondersteund

Bij het detecteren van de documentindeling worden slechts enkele documenttypen geëvalueerd.

De volgende documentindelingen zijn native typen die automatisch worden gedetecteerd en geanalyseerd door de ingebouwde collectieparserservices:

HTMLPlain text  
XML

Standaard worden de volgende documentindelingen geanalyseerd door de Stellent-parser:

123  
Excel  
Freelance  
Ichitaro  
Lotus WordPro  
MS Visio  
MS Word  
PDFPowerPoint  
Rich Text RTF

Als u de ondersteuning voor het analyseren van andere documenttypen wilt uitbreiden, moet u de configuratiebestanden bijwerken (`parserTypes.cfg` en `stellentTypes.cfg`) en regels opgeven voor de toewijzing van specifieke documenttypen aan `parserservices` of de Stellent-filter.

---

## Documenttypen die zijn gekoppeld aan collectie- en Stellent-parsers

Om ervoor te zorgen dat de documenten in een crawlruimte nauwkeurig en efficiënt worden geanalyseerd, kunt u configuratiebestanden maken om op te geven welke documenttypen door de collectieparser moeten worden geanalyseerd en welke moeten worden geanalyseerd door een Stellent-documentfilters.

In een enterprise search-collectie worden de meeste documentindelingen verwerkt door geïntegreerde HTML- of XML-parsers. Bepaalde typen documenten worden meestal niet geanalyseerd (zoals Postscript-documenten) en andere typen documenten worden verwerkt op basis van Stellent-analysefuncties (zoals de documenttypen Microsoft Word, Microsoft Excel, Microsoft PowerPoint, Lotus Freelance, Lotus 123, PDF, RT en Ichitaro).

Omdat de metagegevens misleidend kunnen zijn, kunnen documenten in platte tekst en HTML-documenten met fouten naar de Stellent-parser worden verzonden, en vervolgens worden teruggestuurd naar een van de geïntegreerde parsers. Een

dergelijke situatie kan van invloed zijn op de systeemprestaties. Voor andere documenten is het wellicht niet mogelijk het documenttype te detecteren, zodat de documenten worden overgeslagen. Om deze situatie te voorkomen, kunt u configuratiebestanden maken om te bepalen waar en hoe de verschillende documenttypen moeten worden geanalyseerd.

Het koppelen van documenttypen aan de collectie- en Stellent-parser bestaat uit de volgende taken:

1. Het configureren van documenttypen voor de collectieparser. Deze stap bestaat uit het maken van een configuratiebestand waarmee u documenttypen koppelt aan de parser die voor een collectie wordt gebruikt. U kunt een van deze configuratiebestanden per collectie maken.
2. Het configureren van documenttypen voor de Stellent-parser. Deze stap bestaat uit het maken van een configuratiebestand waarmee u documenttypen toewijst aan de Stellent-documentfilters die voor een collectie worden gebruikt. U kunt een van deze configuratiebestanden per collectie maken.
3. Het stoppen en opnieuw starten van de parser. Als u de wijzigingen wilt activeren, moet u de beheerconsole van enterprise search gebruiken om de collectie te bewaken waarvoor u documenttypen hebt geconfigureerd. Vervolgens moet u de parser stoppen en opnieuw starten.

## Documenttypen koppelen aan een collectieparser

Als u bepaalde typen documenten aan een collectieparser wilt koppelen, moet u het configuratiebestand `parserTypes.cfg` maken. Deze taak wordt niet ondersteund op de beheerconsole van enterprise search.

### Voordat u begint

Om deze taak uit te voeren, moet u zijn aangemeld als beheerder van enterprise search.

### Over deze taak

Als het configuratiebestand niet bestaat, worden de standaardparserserviceregels in de collectieparser gebruikt. Als het configuratiebestand wél bestaat, geven de regels in het bestand het volgende aan:

- Welke URL-extensie en welk MIME-type aan welk parsertype worden toegewezen.
- Hoe documenten worden geanalyseerd waarvan het type onbekend is als gevolg van onvolledige metagegevens.

Het bestand `parserTypes.cfg` bevat een reeks documentregels, waarbij elke documentregel een van de volgende regels bevat:

#### **EXTENSION** *extensie parser*

Alle documenten waarvan de URL eindigt met de opgegeven extensie worden door de opgegeven parser verwerkt. Neem geen punt op in de extensie. De vergelijking is niet hoofdlettergevoelig.

#### **CONTENTTYPE** *type/subtype parser*

Alle documenten waarvan het contenttype overeenkomt met het opgegeven type/subtype worden door de opgegeven parser verwerkt. Als het contenttype van het document `t/s` is, is er sprake van een overeenkomst als `t` overeenkomt met het type, als `s` overeenkomt met het subtype of het subtype een jokerteken is (het sterretje, `*`).

### UNKNOWN *parser*

Alle documenten waarvan de extensie en het contenttype onbekend zijn (niet beschikbaar zijn gemaakt door de crawler), worden door de opgegeven parser verwerkt.

### DEFAULT *parser*

Alle documenten die niet worden gedekt door de regels, worden door de opgegeven parser verwerkt.

In alle gevallen moet voor *parser* de waarde `html`, `xml`, `stellent` of `geen` zijn opgegeven, waarbij `geen` betekent dat documenten van dat type niet moeten worden geanalyseerd.

Als meerdere regels overeenkomen voor het document, heeft de meest specifieke regel voorrang en wordt niet gelet op de volgorde waarin de regels voorkomen:

- Een EXTENSION-regel is specifiekere dan een CONTENTTYPE-regel.
- Een CONTENTTYPE-regel met een subtype is specifiekere dan een CONTENTTYPE-regel met een jokerteken. Een regel voor contenttype `application/postscript` heeft voorrang op een regel voor `application/*`.
- Er kunnen niet twee regels voor dezelfde extensie of hetzelfde contenttype aanwezig zijn. Als dit wel het geval is, wordt op basis van de implementatie bepaald welke regel voorrang heeft.

### Procedure

Ga als volgt te werk om documenttypen aan de collectieparser te koppelen:

1. Meld u bij de indexserver aan als enterprise search-beheerder. Dit gebruikers-ID is opgegeven bij de installatie van WebSphere II OmniFind Edition.
2. Maak het configuratiebestand als volgt, waarbij *collectie\_ID* de collectie aangeeft die u wilt configureren:  
`ES_NODE_ROOT/master_config/collectie_ID.parserdriver/parserTypes.cfg`
3. Gebruik een teksteditor om het bestand te bewerken en geef de parserservice-regels op. Sla het bestand op en sluit dit vervolgens af.
4. Om de wijzigingen te activeren, moet u de beheerconsole van enterprise search gebruiken om de parser voor de collectie te bewaken, waarna u de parser stopt en opnieuw start.

### Voorbeeld

In dit voorbeeld worden alle documenten met de extensie `txt`, `htm` of `html`, met een contenttype dat begint met `text/` of met een onbekende extensie of met een onbekend contenttype door de geïntegreerde HTML-parser verwerkt. Met de geïntegreerde XML-parser worden alle documenten met de extensie `xml` of met het contenttype `text/xml` verwerkt. Alle overige documenten, waaronder de documenten met een contenttype dat begint met `application/`, worden naar de Stellent-parser verzonden.

```
EXTENSION doc stellent
EXTENSION txt html
EXTENSION htm html
EXTENSION html html
EXTENSION xml xml
EXTENSION ps none
CONTENTTYPE text/xml xml
```

```
CONTENTTYPE text/* html
CONTENTTYPE application/* stellent
UNKNOWN html
DEFAULT stellent
```

## Standaardserviceregels voor collectieparsers

Als u geen configuratiebestand maakt voor het toewijzen van bestandstypen en contenttypen aan de parser voor een collectie, worden voor het analyseren van documenten standaardregels gebruikt.

De volgende standaardregels worden door de collectieparser gebruikt:

```
EXTENSION pdf stellent
EXTENSION ppt stellent
EXTENSION prz stellent
EXTENSION lwp stellent
EXTENSION doc stellent
EXTENSION rtf stellent
EXTENSION xls stellent
EXTENSION 123 stellent
EXTENSION vsd stellent
EXTENSION vdx stellent
EXTENSION jxw stellent
EXTENSION jsw stellent
EXTENSION jtw stellent
EXTENSION jaw stellent
EXTENSION juw stellent
EXTENSION jbw stellent
EXTENSION jvw stellent
EXTENSION jfw stellent
EXTENSION jtt stellent
EXTENSION jtd stellent
EXTENSION jttd stellent
EXTENSION jtdd stellent
EXTENSION jtddx stellent
EXTENSION ps none
EXTENSION xml xml
EXTENSION txt text
EXTENSION htm html
EXTENSION html html
EXTENSION shtml html
EXTENSION xhtml html
EXTENSION asp html

CONTENTTYPE application/postscript none
CONTENTTYPE application/* stellent
CONTENTTYPE text/rtf stellent
CONTENTTYPE text/richtext stellent
CONTENTTYPE text/xml xml
CONTENTTYPE text/html html
CONTENTTYPE text/plain text

UNKNOWN none
DEFAULT none
```

## Documenttypen koppelen aan een Stellent-parser

Om op te geven welke typen documenten door Stellent-documentfilters moeten worden geanalyseerd, maakt u het configuratiebestand `stellenttypes.cfg`. Deze taak wordt niet ondersteund op de beheerconsole van enterprise search.

### Voordat u begint

Om deze taak uit te voeren, moet u zijn aangemeld als beheerder van enterprise search.

### Over deze taak

In het configuratiebestand `stellentypes.cfg` wordt het volgende aangegeven:

- Welke bestandstypen door de Stellent-parser moeten worden geanalyseerd. Een bestandstype komt overeen met een van de bestandstypen die door de Stellent-bibliotheek wordt herkend.
- Welke bestandstypen moeten worden teruggestuurd naar de collectieparser voor verwerking met een van de geïntegreerde parsers. (Deze actie is nodig omdat het kan voorkomen dat de collectieparser een document ongewenst naar de Stellent-parser terugstuurt als gevolg van misleidende metagegevens.)
- Welke bestandstypen moeten worden afgewezen omdat deze niet worden ondersteund in enterprise search.

Als u een configuratiebestand hebt opgegeven en dit niet bestaat, wordt de parser niet gestart. Als er geen configuratiebestand is opgegeven voor de eigenschap `OutsideInSupportedTypes` in het bestand `stellent.properties`, worden de standaard-analyseregels voor Stellent-parsers gebruikt.

In het configuratiebestand worden de documenttypen weergegeven en wordt aangegeven op elke manier deze worden afgehandeld. Het bestand bevat een reeks documentregels, waarbij elke documentregel een van de volgende regels bevat:

```
accept DEFAULT
accept ALL doctype
accept type doctype
native DEFAULT
native type doctype
reject type
```

Hierbij geldt het volgende:

#### *doctype*

De waarde die moet worden gebruikt voor het `doctype`-querytoken. Documenten kunnen op basis van het documenttype worden doorzocht. Een gebruiker kan bijvoorbeeld `$doctype::pdf` opgeven om PDF-documenten te doorzoeken.

*type* Een van de filtertypewaarden in de Stellent-bibliotheek. *doctype* is de waarde die voor het `doctype`-token moet worden gebruikt als een regel wordt toegepast.

#### **DEFAULT**

In de lijst van geaccepteerde of native typen (afhankelijk van het type van de regel), is de standaardlijst opgenomen. Met deze optie kunt u de standaardconfiguratie uitbreiden in plaats van deze te vervangen.

**Alle** Alle typen die niet expliciet zijn opgenomen, worden geaccepteerd met het opgegeven `doctype`-token.

De regels in het configuratiebestand worden als volgt verwerkt:

1. Als er een `reject`-regel is voor *type*, wordt het documenttype niet geaccepteerd.
2. Als er een `native` regel is voor *type* (inclusief de standaardanalyseregels als `native DEFAULT` is opgegeven), wordt het document teruggestuurd naar de geïntegreerde parser, samen met de waarde voor het token *doctype* dat door

deze regel is opgegeven. De waarde van *doctype* moet txt, htm of xml zijn (respectievelijk platte tekst, HTML of XML).

3. Als er een accept-regel is voor *type* (inclusief de standaardlijst als accept DEFAULT is opgegeven), wordt het documenttype geaccepteerd.
4. Als accept ALL is opgegeven, wordt het documenttype geaccepteerd.
5. Als dit niet het geval is, wordt het document afgewezen en niet geanalyseerd.

Als het documenttype wordt geaccepteerd, wordt de waarde voor *doctype* gebruikt die in de toegepaste regel is opgegeven. Deze waarde wordt samen met de geanalyseerde content naar de collectieparser teruggestuurd.

## Procedure

Ga als volgt te werk om documenttypen aan de Stellent-parser te koppelen:

1. Meld u bij de indexserver aan als enterprise search-beheerder. Dit gebruikers-ID is opgegeven bij de installatie van WebSphere II OmniFind Edition.
2. Bewerk het bestand `ES_NODE_ROOT/master_config/collectie_ID.stellent/stellent.properties`, waarbij *collectie\_ID* de collectie aangeeft die u wilt configureren.
3. Geef voor de eigenschap `OutsideInSupportedTypes` het absolute pad op naar het configuratiebestand dat u aan het maken bent.

U kunt bijvoorbeeld het volgende configuratiebestand voor één collectie maken en het bestand opslaan bij de andere collectiespecifieke bestanden:

```
ES_NODE_ROOT/master_config/collectie_ID.stellent/stellenttypes.cfg
```

U kunt ook het volgende configuratiebestand maken, zodat u dezelfde instellingen voor alle collecties kunt gebruiken en het bestand opslaat bij de andere bestanden op systeemniveau. (Als u deze benadering gebruikt, moet u zorgen dat u dit pad in het bestand `stellent.properties` opgeeft voor alle collecties, zie stap 2.)

```
ES_INSTALL_ROOT/default_config/stellent/stellenttypes.cfg
```

4. Gebruik een teksteditor om het configuratiebestand te maken en geef de Stellent-analyseregels op. Sla het bestand op en sluit dit vervolgens af.
5. Om de wijzigingen te activeren, moet u de beheerconsole van enterprise search gebruiken om de parser voor de collectie te bewaken, waarna u de parser stopt en opnieuw start.





---

## Taal- en codetabelondersteuning voor enterprise search

De taalkundige verwerking voor enterprise search wordt door de parser en de zoekservers verschillend afgehandeld.

Vanwege de taalkundige verwerking maakt de parser geen onderscheid tussen talen en locales. Als een gebruiker een collectie doorzoekt die documenten in meerdere talen bevat, kunnen de zoekresultaten op de zoekservers echter worden beperkt tot een bepaalde taal of locale.

Als de metagegevens in een Engelstalig document bijvoorbeeld en\_US aangeven voor de documentlocale, wordt het document geïndexeerd als Engelstalig document (en) en als een document dat de locale voor de Verenigde Staten gebruikt voor Engels (en\_US). Met dit type indexering kunnen localespecifieke gegevens zoals getallen, datums en tijden op de juiste wijze worden weergegeven. Bij het zoeken naar de collectie kan het document worden gevonden ongeacht het feit of de gebruiker zoekt naar documenten met de locale en of en\_US.

Als een document alleen is geïndexeerd aan de hand de taalcode, zoals en, wordt het document alleen op basis van de taalcode en niet op basis van de locale geïndexeerd. Als gebruikers in de collectie bijvoorbeeld documenten met en\_US zoeken, wordt het document niet gevonden.

Een enterprise search-systeem biedt taalkundige ondersteuning voor de volgende talen en taalcodes met twee codes (zoals is vastgelegd in de ISO 639-standaard):

### Talen met eenvoudige tekst:

- en=Engels
- sq=Albanees az=Azerbeidzjaans (Latijns)
- bg=Bulgaars
- be=Wit-Russisch
- ca=Catalaans
- hr=Kroatisch
- cs=Tsjechisch
- da=Deens
- nl=Nederlands
- et=Estlands
- fi=Fins
- fr=Frans
- de=Duits
- el=Grieks
- hu=Hongaars
- is=IJslands
- id=Indonesisch
- in=Indonesisch
- it=Italiaans
- kk=Kazakstaans
- lv=Lets
- lt=Litouws
- lo=Laotiaans
- mk=Macedonisch
- ms=Maleis
- mt=Maltees
- no=Noors
- nb=Noors (Bokmal)
- pl=Pools
- pt=Portugees
- ro=Roemeens
- ru=Russisch

sr=Servisch (Cyrillisch)  
sh=Servisch (Latijns)  
sk=Slowaaks  
sl=Sloveens  
es=Spaans  
sv=Zweeds  
tr=Turks  
uk=Oekraïens  
cy=Welsh

#### **Ideografische talen:**

Voor vereenvoudigd en traditioneel Chinees worden in plaats van de codes met twee tekens uitgebreide taalcodes gebruikt.

zh-CN=Chinees (Vereenvoudigd)  
zh-TW=Chinees (Traditioneel)  
ja=Japans  
ko=Koreaans

#### **Talen met ingewikkelde tekst:**

ar=Arabisch  
as=Assamees  
bn=Bengali  
gu=Gujarati  
iw=Hebreeuws  
he=Hebreeuws  
hi=Hindi  
kn=Kannada  
ml=Malayalam  
mr=Marathi  
or=Oriya  
pa=Punjabi  
ta=Tamil  
te=Telugu  
th=Thais  
ur=Oerdoe  
vi=Vietnamees

In het enterprise search-systeem kan het merendeel van deze talen automatisch worden gedetecteerd. Daarnaast kan ook de codetabel die in documenten met platte tekst wordt gebruikt, automatisch worden gedetecteerd. Als u een expliciete taal of codetabel wilt gebruiken, kunt u tijdens het configureren van een crawler automatische taal- en codetabeldetectie uitschakelen.

#### **Verwante onderwerpen**

“Beheer van enterprise search-parser” op pagina 123

Om de vindbaarheid van documenten te verbeteren, kunt u opties opgeven voor de manier waarop documenten en metagegevens worden geanalyseerd en gecategoriseerd voordat ze aan de enterprise search-index worden toegevoegd.

---

## **Automatische taaldetectie**

In een enterprise search-systeem kunnen documenten in praktisch elke taal worden verwerkt.

Als een document in een van de volgende talen is geschreven, kan de taal automatisch worden gedetecteerd. Als u weet welke taal voor uw documenten wordt gebruikt, kunt u de gewenste taal opgeven tijdens het configureren van een crawler zodat de taal niet automatisch wordt gedetecteerd.

Arabisch  
Bulgaars  
Tsjechisch  
Chinees (Vereenvoudigd)

Chinees (Traditioneel)  
Deens  
Nederlands  
Engels  
Fins  
Frans (Canada)  
Frans (Frankrijk)  
Duits (Duitsland)  
Duits (Zwitserland) Grieks  
Hebreeuws  
Hongaars  
IJslands  
Italiaans  
Japans  
Koreaans  
Noors (Bokmal)  
Pools  
Portugees (Brazilië)  
Portugees (Portugal)  
Roemeens  
Russisch  
Spaans  
Zweeds  
Thais  
Turks

---

## Automatische codetabeldetectie

Een enterprise search-systeem ondersteunt documenten met verschillende code-tabellen.

Voor tekstbestanden kunnen de volgende codetabellen automatisch worden gedetecteerd. Voor de overige documentindelingen worden de metagegevens in het document gebruikt voor de detectie van de codetabel (bijvoorbeeld de HTML-metagegevens-elementen). Als u weet welke codetabel voor uw documenten wordt gebruikt, kunt u de gewenste codetabel opgeven tijdens het configureren van een crawler zodat de codetabel niet automatisch wordt gedetecteerd.

### Unicode-coderingen:

UTF-8  
UTF-16BE  
UTF-16LE

### Multiple-bytecoderingen:

Shift-JIS  
ISO-2022-CN  
ISO-2022-JP  
ISO-2022-KR  
GB18030  
EUC-JP  
EUC-KR

### Single-bytecoderingen:

ISO-8859-1: Deens, Nederlands, Duits, Engels, Frans, Italiaans,  
Noors, Portugees, Spaans, Zweeds  
ISO-8859-2: Tsjechisch, Hongaars, Pools, Roemeens  
ISO-8859-5: Russisch  
ISO-8859-6: Arabisch  
ISO-8859-7: Grieks  
ISO-8859-8: Hebreeuws, (Hebreeuws visuele ordening)  
ISO-8859-9: Turks  
Windows-1250: Tsjechisch, Hongaars, Pools, Roemeens  
Windows-1251: Russisch  
Windows-1252: Deens, Nederlands, Duits, Engels, Frans, Italiaans,  
Noors, Portugees, Spaans, Zweeds

Windows-1253: Grieks  
Windows-1254: Turks  
Windows-1255: Hebreeuws  
Windows-1256: Arabisch  
KOI8-R: Russisch

Tekensetdetectie is een onnauwkeurig proces. Tijdens de codetabeldetectie wordt geprobeerd de tekenset (charset) te identificeren die de meeste overeenkomsten vertoont met de kenmerken van de bytegegevens. Omdat dit proces gedeeltelijk statistisch van aard is, kan de juistheid van de resultaten niet worden gegarandeerd.

De beste resultaten worden verkregen als de invoergegevens in hoofdzakelijk één taal zijn opgegeven. Daarnaast is een minimum van een paar honderd bytes platte tekst in de taal nodig.

Als de gedetecteerde codering en de ondersteunde codering niet overeenkomen, wordt de standaardcodetabel voor de collectie gebruikt.

---

## Taalkundige analyse van Chinese, Japanse en Koreaanse documenten

Om de vindbaarheid van Chinese, Japanse en Koreaanse documenten te verbeteren, kunt u taalkundige analyseopties opgeven.

Voor Chinese, Japanse en Koreaanse documenten kunt u opgeven dat de parser de n-gramsegmenteringsmethode voor lexicale analyses moet gebruiken. Voor Chinese en Japanse documenten kunt u ook instellen dat de parser nieuwe-regeltekens uit spaties moet verwijderen.

### N-gram-segmentering

Bij het maken van een collectie selecteert u het type lexicale analyse dat u wilt gebruiken voor het analyseren van documenten die zijn geschreven in talen waarin geen spaties worden gebruikt als scheidingstekens tussen woorden.

Witruimtesegmentering van Unicode gebruikt spaties om woorden van elkaar te scheiden. N-gram-segmentering beschouwt overlappende reeksen van een bepaald aantal tekens als een enkel woord. Voor talen als Chinees, Japans en Koreaans, die geen spaties gebruiken om woorden van elkaar te scheiden, levert de methode n-gram-segmentering betere zoekresultaten op dan de witruimtesegmentering van Unicode.

Kies tijdens het maken van een collectie de segmenteringsmethode die u wilt gebruiken voor het analyseren van documenten. Als de collectie is gemaakt, kunt u de instelling bekijken in de analyseopties, maar niet wijzigen.

### Witruimten uit tekst verwijderen

U kunt de parser zodanig configureren, dat witruimten uit de tekst worden verwijderd.

#### Voordat u begint

Om deze taak uit te voeren, moet u zijn aangemeld als beheerder van enterprise search.

#### Over deze taak

Als u deze optie inschakelt voor een collectie, verwijdert de parser de witruimte-tekens waarmee de tekens worden gescheiden die uit twee letters bestaan. U kunt de witruimtetekens bijvoorbeeld verwijderen als uw documenten zijn geschreven in een taal waarin geen witruimten worden gebruikt om de grenzen van woorden te scheiden, zoals het Chinees of Japans.

Als u de parser configureert om witruimtetekens te verwijderen, kunt u aangeven of u alleen de witruimten tussen DBCS-tekens (Double-Byte Character Set) wilt verwijderen of alle witruimten wilt verwijderen, ongeacht de tekencontext. U kunt deze laatste optie bijvoorbeeld kiezen als u Engelse tekst in een Japans document opneemt, en u ook de witruimten uit de Engelse tekst wilt verwijderen.

De parser verwijdert de volgende tekens:

- Tab (0x09)
- LF of Line Feed (0x0A)
- CR of Carriage Return (0x0D)

Om de wijzigingen te activeren, moet u de parser stoppen en opnieuw starten. Om de wijzigingen toe te passen op de documenten in de index, moet u de documenten opnieuw crawlen en de hoofdindex opnieuw opbouwen.

### Procedure

Ga als volgt te werk om witruimten uit de tekst te verwijderen:

1. Meld u bij de indexserver aan als enterprise search-beheerder. Dit gebruikers-ID is opgegeven bij de installatie van WebSphere II OmniFind Edition.
2. Gebruik een teksteditor om het volgende bestand te bewerken, waarbij *collectie\_ID* het ID is dat voor de collectie is opgegeven tijdens het maken van de collectie (of dat door het systeem is toegewezen):

```
ES_NODE_ROOT/master_config/collectie_ID.parserdriver/collection.properties
```

3. Geef aan op welke manier witruimten moeten worden verwijderd:
  - Als u de witruimte tussen DBCS-tekens wilt verwijderen, stelt u de waarde van de eigenschap `removeCjNewlineChars` in op `true`:

```
removeCjNewlineChars=true
```

- Als u de witruimten overall in de documenten wilt verwijderen, stelt u de waarde van de eigenschap `removeCjNewlineChars` in op `true` en stelt u de waarde van de eigenschap `removeCjNewlineCharsMode` in op `all`:

```
removeCjNewlineChars=true  
removeCjNewlineCharsMode=all
```



---

## Beheer van enterprise search-index

Om ervoor te zorgen dat gebruikers altijd toegang hebben tot de laatste informatie, maakt enterprise voor elke collectie een index. Deze index wordt beheerd door op regelmatige basis de content bij te werken.

Om de gegevens doorzoekbaar te maken die door crawlers worden verzameld, moet u indexen maken. Wanneer u een collectie voor het eerst maakt, maakt enterprise search een index voor alle gegevens die zijn gecrawld. Als de crawlers nieuwe en gewijzigde gegevensbronnen crawlen, wordt voor de nieuwe content een update gemaakt. Uiteindelijk moet de update worden samengevoegd in de basisindex. Dit samenvoegingsproces wordt ook wel het opbouwen van de hoofdindex genoemd. Bij het bijwerken van de index of het opbouwen van de hoofdindex wordt de nieuwe content naar de zoekservers gekopieerd en wordt de content beschikbaar gemaakt om te worden doorzocht.

Gegevens worden door de crawlers continu of op gezette tijden verzameld. Als u de index regelmatig bijwerkt, kunnen gebruikers altijd de meest recente gegevens doorzoeken. Een index die continu wordt bijgewerkt, moet uiteindelijk opnieuw worden opgebouwd. Naarmate een index groter wordt, worden meer systeemresources in beslag genomen. Voor de beste prestaties moet u de hoofdindex regelmatig opbouwen.

Hoe vaak u een index opbouwt, is afhankelijk van:

- De systeemresources (bestandssysteemruimte, processorsnelheid en geheugen)
- Het aantal documenten dat (opnieuw) moet worden gecrawld
- Het type gegevens dat moet worden gecrawld
- Hoe vaak de categorieregels worden gewijzigd (de wijzigingen worden pas van kracht als de index is opgebouwd)
- Hoe vaak het starten van de crawler wordt afgedwongen (in plaats van op de geplande tijd te worden uitgevoerd)
- Hoe vaak externe crawlers URI's verwijderen of toevoegen (deze typen crawlers werken interactief samen met enterprise search via de Data Listener-API)

Voor collecties met miljoenen documenten die voornamelijk uit webdocumenten bestaan, moet u de hoofdindex ongeveer eenmaal per dag opbouwen en de index elk uur of elke twee uur bijwerken.

Om een actuele, doorzoekbare index te beheren, voert u de volgende taken uit:

- Roosters opgeven voor het opbouwen van de index
- Het indexrooster wijzigen
- Het indexrooster in- en uitschakelen
- Gelijktijdige bouwopdrachten voor de index configureren

Om opties op te geven die van invloed zijn op de gebruikersweergave van de index, kunt u een van de volgende taken uitvoeren:

- Ondersteuning voor jokertekens in query's configureren
- Scopes configureren om het aantal documenten te beperken dat gebruikers kunnen doorzoeken
- Documenten van dezelfde bron samenvoegen in de zoekresultaten

- URI's uit de index verwijderen

#### Verwante taken

“De indexactiviteit voor een collectie bewaken” op pagina 313

U kunt de index voor een collectie bewaken als u de voortgang van het opbouwen van een index wilt bekijken, het indexrooster wilt in- of uitschakelen of indexactiviteiten wilt starten en stoppen.

“De indexeerwachtrij van enterprise search bewaken” op pagina 314

U kunt de status van het opbouwen van een index in de indexeerwachtrij bekijken, het opbouwen van een index stoppen of een index uit de wachtrij wissen.

---

## Rooster opstellen voor het opbouwen van de index

U kunt roosters opstellen voor het opbouwen van de hoofdindex en het bijwerken van de index met nieuwe content.

### Voordat u begint

Om het rooster voor het opbouwen van de index te plannen, moet u de rol van beheerder van enterprise search zijn of beheerder van de collectie zijn.

### Over deze taak

Om te garanderen dat gebruikers altijd toegang hebben tot de meest recente informatie in de bronnen die ze doorzoeken, kunt u het beste een rooster opgeven waarmee de index regelmatig wordt opgebouwd. Bij het opbouwen van de hoofdindex wordt de volledige index opnieuw opgebouwd. De indexeringsprocessen lezen alle gegevens die door de crawlers zijn verzameld en door de parser zijn geanalyseerd. Bij het opbouwen van een delta-index wordt de informatie doorzoekbaar gemaakt die is gecrawld sinds de hoofdindex voor het laatst is opgebouwd.

Standaard is de optie voor het plannen van opbouwopdrachten van indexen geselecteerd. Met deze optie worden taken gepland voor het opbouwen van de hoofd- en delta-indexen wanneer het enterprise search-systeem wordt gestart. U kunt het aankruisvakje **Inschakelen bij starten van systeem** op elk gewenst moment uitschakelen om te voorkomen dat een geplande opbouwopdracht voor de index wordt uitgevoerd. Zo kunt u het rooster uitschakelen om problemen op te lossen.

### Procedure

Ga als volgt te werk om opbouwopdrachten voor een index te plannen:

1. Open een collectie in een editor, selecteer de pagina Index en klik op **Rooster voor bouwen van index opstellen**.
2. Om op te geven hoe vaak de index moet worden bijgewerkt met nieuwe inhoud, geeft u in het gebied **Geef een rooster op voor het bouwen van een delta-index** op de pagina Rooster voor bouwen van index opstellen de volgende opties op:
  - a. Geef in de velden **Jaar, Maand, Dag, Uur** en **Minuut** in het gebied **Starten op** aan wanneer de de delta-index voor het eerst moet worden gebouwd.
  - b. Geef in de velden **dagen, uren** en **minuten** in het gebied **Interval voor bijwerken** aan hoe vaak de delta-indexen moeten worden opgebouwd.

U wordt aangeraden delta-indexen regelmatig op te bouwen, bijvoorbeeld elke uur of om het uur. Afhankelijk van hoe vaak er wijzigingen in de



inhoud optreden, kunt u kiezen voor een hogere of een lagere frequentie. U kunt bijvoorbeeld opgeven: elk uur (0 dagen en 1 uur) of elke 12 uur (0 dagen en 12 uur).

3. Om op te geven hoe vaak de index volledig opnieuw moet worden opgebouwd, geeft u de volgende opties op in het gebied **Geef een rooster op voor het bouwen van een hoofdindex**:
  - a. Geef in de velden **Jaar, Maand, Dag, Uur** en **Minuut** in het gebied **Starten op** aan wanneer de de hoofdindex voor het eerst moet worden opgebouwd.
  - b. Geef in de velden **dagen, uren** en **minuten** in het gebied **Interval voor bijwerken** aan hoe vaak de hoofdindex moeten worden opgebouwd.

U wordt aangeraden de index regelmatig volledig op te bouwen, bijvoorbeeld elke 24 uur. Afhankelijk van hoe vaak er wijzigingen in de inhoud optreden, kunt u kiezen voor een hogere of een lagere frequentie. Zo kunt u bijvoorbeeld opgeven: elke 12 uur (0 dagen en 12 uur) of elke tweeënhalve dag (2 dagen en 12 uur).
4. Klik op **OK**.

## Het rooster voor indexering wijzigen

U kunt het rooster voor het opbouwen van de index wijzigen.

### Voordat u begint

Als u een indexrooster wilt wijzigen, moet u de rol van beheerder van enterprise search hebben of beheerder van de collectie zijn.

### Procedure

Ga als volgt te werk om het indexrooster te wijzigen:

1. Open een collectie in een editor, selecteer de pagina **Index** en wijzig de gewenste waarden in de velden **Maand, Dag, Jaar** en **Uur**. Geef aan hoe vaak updates in de index moeten worden aangebracht en hoe vaak de index moet worden opgebouwd.
2. Klik op **Aanbrengen**.

## Indexroosters in- en uitschakelen

U kunt de roosters voor het opbouwen van de index in- en uitschakelen.

### Voordat u begint




Als u een indexrooster wilt in- of uitschakelen, moet u de rol van beheerder van enterprise search hebben of beheerder van de collectie zijn.

### Over deze taak

U kunt een rooster voor een index uitschakelen om te voorkomen dat het geplande opbouwproces voor een index wordt uitgevoerd. U kunt voorkomen dat een index op de geplande datum en tijd wordt gebouwd als u bijvoorbeeld eerst bepaalde problemen wilt oplossen.

U kunt een rooster in- of uitschakelen tijdens het bewerken van een collectie en tijdens het bewaken van een collectie.

### Procedure

1. Ga als volgt te werk om het rooster voor een index in of uit te schakelen tijdens het bewerken van een collectie:
  - a. Breng de gewenste wijzigingen aan in de collectie.
  - b. Schakel op de pagina Index het aankruisvakje **Inschakelen bij starten van systeem** in of uit als u het opbouwen van de index wilt inschakelen of uitschakelen.
  - c. Schakel het aankruisvakje **Inschakelen bij starten van systeem** in of uit om het rooster voor het opbouwen van de hoofdindex in of uit te schakelen.
  - d. Klik op **Aanbrengen**.
2. Ga als volgt te werk om het rooster voor een index in of uit te schakelen tijdens het bewaken van een collectie:
  - a. Bewaak de collectie die u wilt wijzigen.
  - b. Als een index is gepland en u wilt voorkomen dat de index op de geplande datum en tijd wordt opgebouwd, klikt u op de pagina Index op  **Rooster uitschakelen**. De index wordt pas opgebouwd als u het rooster inschakelt of op  **Start** klikt om het proces voor het opbouwen van de index te starten.
  - c. Als een index is gepland maar het rooster voor het opbouwen is uitgeschakeld, klikt u op  **Rooster inschakelen**.  
De index wordt in de wachtrij gezet om opnieuw te worden opgebouwd op de datum en tijd die u in het indexrooster hebt opgegeven.

## Indexen gelijktijdig opbouwen

U kunt het gebruik van de indexeringsresources bepalen door op te geven voor hoeveel collecties de bouwopdrachten voor de index gelijktijdig kan worden verwerkt. Als er voldoende systeemresources aanwezig zijn, kunt u de kwaliteit van zoekopdrachten verbeteren door in te stellen dat de index op hetzelfde moment moet worden bijgewerkt als het moment waarop de hoofdindex wordt opgebouwd.

### Voordat u begint

Om opties voor het opbouwen van de index voor het systeem op te geven, moet u de rol van beheerder van enterprise search hebben.

### Over deze taak

In enterprise search kunt u meerdere indexen tegelijkertijd opbouwen door resources gemeenschappelijk te gebruiken tussen collecties. Op deze manier kunt u opdrachten voor het opbouwen van indexen voor verschillende collecties parallel verwerken. Door de processen gemeenschappelijk te gebruiken, kunt u voorkomen dat de hele wachtrij geblokkeerd raakt omdat er een zeer grote index wordt opgebouwd.

Als het opbouwen van een index wordt aangevraagd of is gepland, wordt die opdracht toegevoegd aan de indexeerwachtrij en wordt gewacht totdat de opdracht kan worden verwerkt. Omdat elke collectie een eigen index heeft, kan het gebeuren dat er meerdere bouwopdrachten voor indexen tegelijk in de wachtrij staan. Bij het configureren van de indexopties voor het systeem geeft u op hoeveel collecties resources gemeenschappelijk kunnen gebruiken en hoeveel opdrachten parallel kunnen worden verwerkt.


Daarnaast kunt u opgeven dat de opdrachten voor het bijwerken van een index op hetzelfde moment moeten worden verwerkt als het moment waarop de hoofdindex voor de collectie wordt opgebouwd. Als u deze optie selecteert, worden de zoekservers vernieuwd met de meest recente documenten (van de delta-index) terwijl de tragere bouw van de hoofdindex nog aan de gang is. Het opbouwen van de index neemt echter veel resources in beslag. Tijdens het opbouwen van een index wordt aanzienlijk veel systeemgeheugen en schijfruimte in beslag genomen. Als u deze optie inschakelt en er onvoldoende ruimte of geheugen aanwezig is, kan dit nadelige gevolgen hebben voor de algehele systeemprestaties.

Als u het aantal gelijktijdige opbouwopdrachten voor indexen verhoogt, worden de opdrachten die al in de wachtrij zijn gezet, niet automatisch gestart. De wijziging is van invloed op nieuwe opbouwopdrachten van indexen die in de wachtrij worden gezet nadat u de waarde hebt gewijzigd.

Als u het aantal gelijktijdige opbouwopdrachten voor indexen verlaagt, worden de opdrachten die al in de wachtrij zijn gezet, niet automatisch gestopt. De wijzigingen worden actief wanneer de huidige opbouwopdracht voor de index stopt, waarna de opdrachten in de wachtrij worden gestart.

### Procedure

Ga als volgt te werk om opties voor het opbouwen van de index voor het systeem op te geven:

1. Klik op **Systeem** om de view Systeem te openen.
2. Klik op  **Wijzigen** om naar de view Systeem wijzigen te gaan.
3. Klik op de pagina Index op **Indexeringsopties configureren**.
4. Voer op de pagina Indexeringsopties op systeemniveau het aantal collecties in waarvoor de systeemresources gemeenschappelijk kunnen worden gebruikt en waarvoor de opbouwopdrachten voor de index parallel kunnen worden verwerkt.  
Het aantal collecties dat indexeringsresources gemeenschappelijk gebruikt, kan niet groter zijn dan het aantal collecties dat in het enterprise search-systeem aanwezig is. Als er bijvoorbeeld vijf collecties aanwezig zijn, moet u een getal invoeren dat lager is dan of gelijk is aan vijf.
5. Als er voor afzonderlijke collectie voldoende systeemresources aanwezig zijn voor het uitvoeren van meerdere gelijktijdige opbouwopdrachten voor indexen, kunt u de optie selecteren waarmee opbouwopdrachten voor delta-indexen op hetzelfde moment worden uitgevoerd als het moment waarop de hoofdindex wordt opgebouwd.
6. Klik op **OK**.

---

## Opties die van invloed zijn op de doorzoekbare indexweergave

Nadat documenten zijn geïndexeerd, kunt u opties opgeven waarmee wordt bepaald hoe gebruikers documenten kunnen zoeken en documenten kunnen bekijken in de zoekresultaten.

Om opties op te geven die van invloed zijn op de gebruikersweergave van de index, kunt u de volgende taken uitvoeren:

- Ondersteuning voor jokertekens in zoektermen configureren. U kunt ondersteuning inbouwen voor query's met jokertekens in de index of u kunt opties opgeven om de zoektermen tijdens de verwerking van query's uit te breiden.

- Scopes configureren om het aantal documenten te beperken dat gebruikers kunnen doorzoeken. Bij het doorzoeken van een collectie doorzoeken de gebruikers alleen de documenten die binnen de scope vallen, niet de volledige index.
- Documenten van dezelfde bron samenvouwen in de zoekresultaten. U kunt documenten groeperen waarvan de URI of het URI-patroon in de index overeenkomt en alleen de bovenste resultaatdocumenten in de zoekresultaten weergeven (gebruikers kunnen opties opgeven om de samengevouwen documenten te bekijken).
- URI's uit de index verwijderen. U kunt voorkomen dat gebruikers tijdelijk bepaalde documenten in de index kunnen doorzoeken.

## Geïndexeerde opties voor het doorzoeken van documenten

Tijdens het configureren van opties voor het doorzoeken van gecrawelde gegevens of tijdens het toewijzen van XML- en HTML-metagegevelementen aan zoekvelden kunt u aangeven hoe documenten kunnen worden doorzocht en hoe documenten in de zoekresultaten worden afgebeeld.

De zoekopties die u opgeeft, worden met de documenten in de index opgeslagen. Met de opties kunt u bepalen welke gegevens gebruikers kunnen opvragen en welke gegevens in de zoekresultaten worden afgebeeld.

### Crawleropties:

Wanneer u een crawler configureert voor het crawlen van gegevensbronnen die velden bevatten, kunt u de volgende opties opgeven om te bepalen of een veld kan worden doorzocht, hoe het veld kan worden doorzocht en of het veld in de zoekresultaten kan worden weergegeven:

- Vrije zoekopdracht
- Veldzoekopdracht
- Volledige overeenkomst
- Sorteerkbaar
- Zoeken met parameters
- Zoekresultaten
- Documentcontent

### XML- en HTML-veldtoewijzingsopties:

Wanneer u de parser configureert en opgeeft dat u de XML-elementen en HTML-metagegevelementen wilt toewijzen aan doorzoekbare velden in de index, geeft u de volgende opties op:

- Veldzoekopdracht
- Volledige overeenkomst
- Sorteerkbaar
- Zoekresultaten

Als u zoekopties configureert voor specifieke HTML-metagegevelementen (en niet voor alle elementen of voor de elementen in de Dublin Core-metagegevelementenset), kunt u ook opgeven dat velden die numerieke waarden bevatten, kunnen worden doorzocht met behulp van een parametrische zoekopdracht.

### Vrije zoekopdracht

De enterprise search-index is een volledige-tekstindex met content uit verschillende gegevensbronnen. U kunt de content doorzoeken door een eenvoudige query in

natuurlijke taal op te geven. Met de zoekprocessen wordt veld- en document-content doorzocht om te documenten te zoeken die relevant zijn voor de query.

Om velden in te schakelen zodat ze kunnen worden doorzocht met behulp van een vrije zoekopdracht, schakelt u het aankruisvakje **Vrije zoekopdracht** in tijdens het configureren van een crawler. Als u de titel-, trefwoord- en beschrijvingsvelden wilt doorzoeken, schakelt u dit aankruisvakje en het aankruisvakje **Veldzoekopdracht** in.

#### **Voorbeeld 1:**

Een vrije zoekopdracht kan zo eenvoudig zijn als de volgende query:

```
ketting fiets
```

Om aan te geven welke woorden in een document moeten worden weergegeven, kunt u speciale notaties gebruiken. Als er vóór een woord bijvoorbeeld een plusteken (+) staat, moet een document dat woord bevatten om te worden aangemerkt als een treffer. Staat er vóór een woord een minteken (-), dan worden documenten die dat woord bevatten, niet in de zoekresultaten weergegeven. Als u twee of meer woorden tussen aanhalingstekens plaatst ("), kunt u zoeken naar exacte zinnen.

#### **Voorbeeld 2:**

In de volgende vrije zoekopdracht is alleen sprake van een treffer als het document de exacte zin science fiction bevat en het woord robot niet voorkomt:

```
+"science fiction" -robot
```

### **Veldzoekopdracht**

Met een veldzoekopdracht kunt u ervoor zorgen dat alleen bepaalde gegevensvelden en metagegevensvelden van een document als resultaat worden gegeven. U kunt bijvoorbeeld aangeven dat bepaalde woorden moeten voorkomen in de titel van een document.

Om velden in te schakelen zodat ze kunnen worden doorzocht op basis van de veldnaam, schakelt u het aankruisvakje **Veldzoekopdracht** in tijdens het configureren van een crawler of tijdens het configureren van de veldtoewijzingsopties voor XML- en HTML-elementen. Als u de titel-, trefwoord- en beschrijvingsvelden wilt doorzoeken, schakelt u dit aankruisvakje en het aankruisvakje **Vrije zoekopdracht** in.

#### **Bijvoorbeeld:**

Om een veldzoekopdracht in enterprise search op te geven, neemt u veldnaam en het woord of de woordcombinaties die dat veld moeten bevatten in de query op.

Met de volgende query zoekt u naar documenten die het woord `ibm` en de woordcombinatie `enterprise search` in het titelveld bevatten:

```
title:ibm title:"enterprise search"
```

## Volledige overeenkomst

Met een volledige overeenkomst kunt u de kwaliteit van de zoekresultaten verbeteren doordat u nauwkeurige query's kunt opgeven. Aan de hand van een zoekopdracht met een volledige overeenkomst kunt u een query uitvoeren op velden en XML-elementen, en alleen die documenten ophalen waarin de volledige veldwaarde of XML-elementwaarde voldoet aan de zoektermen. Als de waarde van het veld of element minder of meer content bevat, levert de zoekopdracht niets op.

**Tip:** Bij het vaststellen of er een volledige overeenkomst bestaat, worden de zoektermen omgezet in kleine letters, worden extra spaties uit de zoekreeks verwijderd en wordt er patroonmatching op basis van jokertekens uitgevoerd. Er vindt echter geen lemmatisering plaats, er worden geen synoniemen gezocht en stopwoorden worden niet verwijderd. Voor een volledige overeenkomst voor XML-elementen is de elementnaam (zonder geneste elementen) en de volledige waarde van het element vereist, zodat een exacte overeenkomst met de zoektermen kan worden gevonden.

U kunt velden doorzoeken voor een volledige overeenkomst door de enterprise search-queriesyntaxis (SI-API) of de XMLFrag2-queriesyntaxis te gebruiken (XPath-query's worden niet ondersteund). Met het gelijkteken (=) vóór de zoektermen wordt aangegeven dat een zoekopdracht met volledige overeenkomst moet worden uitgevoerd.

Om velden in te schakelen zodat ze kunnen worden doorzocht met volledige overeenkomst, schakelt u het aankruisvakje **Volledige overeenkomst** in tijdens het configureren van een crawler of tijdens het configureren van de veldtoewijzingsopties voor XML- en HTML-elementen. Als u tijdens het configureren van de analyseopties voor een collectie instelt dat gebruikers XML-documenten kunnen doorzoeken met native XML-zoekopdrachten, kunnen alle XML-elementen worden doorzocht met volledige overeenkomst van de zoektermen.

### Voorbeeld 1:

Het veld `kleur` bevat de waarde `donkerblauw`.

- De volgende query met volledige overeenkomst levert een treffer op, omdat de query geen andere termen bevat:  
`kleur:"donkerblauw"`
- De volgende query met volledige overeenkomst levert geen treffer op, omdat de query naast het woord `donkerblauw` het woord `overhemd` bevat:  
`kleur:"donkerblauw overhemd"`
- De volgende query met volledige overeenkomst levert geen treffer op, omdat het veld `kleur` ook het woord `donker` bevat:  
`kleur:=blauw`

### Voorbeeld 2:

Zonder volledige overeenkomst kan de volgende XMLFrag2-query documenten als resultaat geven waarin `<diagnosis>intraductal carcinoma comedo type</diagnosis>` of andere termen worden aangegeven die niet exact aan de zoektermen voldoen:

```
@xmlf2::'<diagnosis>intraductal carcinoma</diagnosis>'
```

Met een volledige overeenkomst zorgt de volgende XMLFrag2-query ervoor dat alleen documenten als resultaat worden gegeven waarin de volledige content van de XML-elementwaarde overeenkomt met de zoektermen:

```
@xmlf2::'=<diagnosis>intraductal carcinoma</diagnosis>'
```

## Sorteerbaar

Als een gegevensbron velden bevat of als gebruikers XML- of HTML-documenten doorzoeken, wilt u de resultaten mogelijk sorteren op basis van de waarden in een bepaald veld. In het voorbeeldzoekprogramma voor enterprise search worden de namen weergegeven van alle velden die zijn geconfigureerd als sorteerbare velden. Gebruikers kunnen de resultaten sorteren op basis van een van de weergegeven velden in plaats van de velden te sorteren op relevantie of documentdatum. Daarnaast kunnen gebruikers kiezen of de documenten in oplopende of in aflopende volgorde moeten worden gesorteerd.

De resultaatdocumenten waarin het sorteerveld niet voorkomt, worden aan het einde van de zoekresultaten weergegeven. Resultaatdocumenten die het sorteerveld bevatten maar zijn geïndexeerd voordat het veld is geconfigureerd als sorteerbaar, worden aan het einde van de zoekresultaten weergegeven.

Om in te stellen dat gebruikers de zoekresultaten op basis van een veld kunnen sorteren, schakelt u het aankruisvakje **Sorteerbaar** in tijdens het configureren van de crawler of tijdens het configureren van de veldtoewijzingsopties voor XML- en HTML-elementen. Als het veld numerieke waarden bevat, schakelt u het aankruisvakje **Parametrische zoekopdracht** in om aan te geven dat het veld kan worden doorzocht met een parametrische zoekopdracht en kan worden gebruikt om de zoekresultaten te sorteren.

## Zoeken met parameters

Een parametrische zoekopdracht is een type veldzoekopdracht waarmee u vergelijkende of waardebepalende query's kunt uitvoeren op numerieke velden, datumvelden en metagegevens. U kunt bijvoorbeeld zoeken naar documenten met een bepaalde grootte of die na een bepaalde datum zijn geschreven. Daarnaast kunt u zoeken naar documenten met kenmerken die groter zijn dan, kleiner zijn dan of gelijk zijn aan een opgegeven waarde.

Om velden in te schakelen zodat ze kunnen worden doorzocht met een parametrische zoekopdracht, schakelt u het aankruisvakje **Parametrische zoekopdracht** in tijdens het configureren van een crawler of tijdens het configureren van de veldtoewijzingsopties voor specifieke HTML-metagegevenselementen.

### Voorbeeld 1:

Met de volgende query zoekt u naar artikelen die precies 50 euro kosten (of een andere valuta die voor het veld price is geïndexeerd):

```
#price::=50
```

### Voorbeeld 2:

Met de volgende query zoekt u naar documenten met een bestandsgrootte die groter is dan 1024, maar kleiner is dan 2048:

```
#filesize::>1024<=2048
```

## Zoekresultaten

Het kan voorkomen dat u bepaalde velden wilt doorzoeken terwijl u deze velden niet in de zoekresultaten wilt weergeven, of dat u een veld in de zoekresultaten wilt weergeven terwijl u geen query op het veld wilt uitvoeren. Dit kan het geval zijn als u een query moet uitvoeren op financiële gegevens om een belangrijk rapport te kunnen maken, terwijl u bij de salarissen van werknemers niet de namen van de werknemers wilt weergeven.

Om velden in te schakelen zodat ze worden weergegeven in de zoekresultaten, schakelt u het aankruisvakje **Zoekresultaten** in tijdens het configureren van een crawler of tijdens het configureren van de veldtoewijzingsopties voor XML- en HTML-elementen.

## Documentcontent

Voor bepaalde documenttypen, zoals webdocumenten, wordt het volledige document als content gezien. Voor andere documenttypen, zoals documenten met velden, kunt u aangeven welke velden nuttige content bevatten in tegenstelling tot metagegevens.

Om op te geven dat een veld de documentcontent vormt, schakelt u het aankruisvakje **Documentcontent** in tijdens het configureren van de crawler. Als de aankruisvakjes **Documentcontent** en **Vrije zoekopdracht** beide zijn ingeschakeld, wordt de waarde van het veld gebruikt voor het detecteren van duplicaatdocumenten en wordt deze waarde onderdeel van het gebied met dynamische documentsamenvattingen in de zoekresultaten.

### Verwante onderwerpen

Syntaxis voor zoekopdrachten

“Zoekprogramma’s voor enterprise search” op pagina 203

Met behulp van een zoekprogramma kunt u collecties en externe bronnen in het enterprise search-systeem doorzoeken. U kunt een onbeperkt aantal zoekprogramma’s maken en met elk programma kunt u een willekeurig aantal collecties en externe bronnen doorzoeken.

## Jokertekens in query’s

U kunt gebruikers in staat stellen jokertekens in zoektermen op te nemen en te zoeken naar woorden die overeenkomen met een opgegeven patroon.

Een zoekterm met een jokerteken is een term die een sterretje (\*) bevat. Als een gebruiker een query indient die een jokerteken bevat, worden in de zoekresultaten alle documenten uit de index opgenomen die voldoen aan de zoekterm, plus alle documenten uit de index die voldoen aan het patroon dat door het jokerteken wordt aangegeven. Het jokerteken achteraan de zoekterm sea\* komt overeen met search, season en seals.

Tijdens het configureren van jokertekenopties voor een index kunt u aangeven of gebruikers jokertekens in query’s kunnen opnemen en zo ja, op welke manier hiervoor ondersteuning moet worden geboden:

- U kunt instellen dat in alle delen van een document kan worden gezocht naar woorden die overeenkomen met het jokertekenpatroon of u kunt de patroonmatching beperken tot bepaalde velden.
- U kunt opgeven dat alle velden query’s met jokertekens ondersteunen of u kunt de patroonmatching beperken tot bepaalde velden.



- U kunt aangeven dat alleen het laatste teken in een zoekterm een jokerteken kan zijn of toestaan dat een jokerteken op elke gewenste plek in een zoekterm kan voorkomen. (Jokertekens kunnen niet voorkomen in veldnamen.)
- Afhankelijk van de positie waar de jokertekens moeten verschijnen, kunt u kiezen hoe de zoektermen moeten worden uitgebreid (zoektermen met jokertekens moeten worden uitgebreid tot alle indextermen die overeenkomen). In de index kunnen alle mogelijke uitbreidingen van termen worden opgeslagen of de zoekprocessen kunnen de termen tijdens de queryverwerking uitbreiden.

Wijzigingen die u in de instellingen voor jokertekens aanbrengt, worden van kracht als de hoofdindex opnieuw wordt opgebouwd.

## Indexuitbreiding

Om uitbreidingen van termen in de index op te nemen, geeft u op hoeveel voorafgaande tekens in een woord moeten overeenkomen met het jokertekenpatroon in een zoekterm wil er sprake zijn van een treffer. Alleen met de zoektermen die minimaal dit aantal tekens bevatten (exclusief het sterretje, \*) worden treffers gegeven. Als u bijvoorbeeld 4 opgeeft, is er pas sprake van een treffer als de zoekterm minimaal vier tekens bevat.

Als u 4 opgeeft, komt het woord `technologie` overeen met de zoektermen `tech*` en `techno*`, maar niet met de zoekterm `te*`.

Tijdens het opbouwen van een delta- of hoofdindex worden naast de oorspronkelijke termen alle mogelijke uitbreidingen van alle termen in een document geïndexeerd. Het voordeel hiervan is dat u tijdens de queryverwerking geen extra tijd kwijt bent met het uitbreiden van de termen. Als u deze optie selecteert, neemt de omvang van de index toe. U moet dan ook over voldoende systeemresources beschikken om die grotere index te ondersteunen.

Deze optie is met name nuttig als de collectie relatief klein is of als de ruimte en de tijd voor het opbouwen van de index minder belangrijk zijn dan de responstijd van de query. U kunt deze optie bijvoorbeeld gebruiken als u een catalogus of adressenlijst met werknemers wilt doorzoeken.

U kunt deze optie alleen kiezen als u instelt dat jokertekens het laatste teken in een zoekterm kunnen zijn. Als is ingesteld dat jokertekens op een willekeurige plek in zoektermen kunnen worden geplaatst, kunt u deze optie niet selecteren voor de uitbreiding van termen in de index.

## Query-uitbreiding

Als u query's wilt uitbreiden en regels voor patroonmatching wilt toepassen wanneer gebruikers query's met jokertekens indienen, geeft u op hoeveel variaties van een zoekterm een treffer opleveren. Als u bijvoorbeeld 50 opgeeft, kunnen er maximaal 50 variaties van een zoekterm als treffers van de zoekterm worden gegeven.

Stel dat de zoekterm `tech*` overeenkomt met de woorden `technisch`, `techniek` en `technologie` en maximaal 50 verschillende woorden met de tekens `tech` beginnen.

Hoewel de query-uitbreiding beperkt van invloed is op de grootte van de index, kunnen de prestaties van de zoekopdracht worden verslechterd. De zoekprocessen moeten alle mogelijke uitbreidingen van de zoekterm met jokertekens doorlopen, tot de limiet die u hebt opgegeven in de instellingen voor jokertekens.

Deze optie is met name nuttig als de collectie relatief groot is en het opbouwen van de index zo min mogelijk ruimte en tijd moet innemen. U kunt deze optie bijvoorbeeld kiezen voor e-mailrepository's, waarvoor de index altijd moet zijn bijgewerkt met de steeds veranderende documenten, terwijl de responstijd van de query's minder belangrijk is.

Deze optie is altijd beschikbaar, ongeacht het feit of u hebt ingesteld dat joker-tokens de laatste tekens in een zoekterm kunnen zijn of dat u hebt ingesteld dat jokertekens op een willekeurige plek in zoektermen kunnen worden geplaatst.

## Ondersteuning voor jokertekens in query's

De set uitbreidingen voor een zoekterm met jokertekens bevat alle termen in de index die kunnen worden opgehaald door het jokerteken te vervangen door een willekeurige tekenreeks. De set wordt als volgt vastgesteld:

- Als een collectie ondersteuning biedt voor jokertekens die op een willekeurige plek in een zoekterm kunnen worden geplaatst, wordt elke zoekterm met een sterretje (\*) geïnterpreteerd als een term met een jokerteken.
- De set bevat (maximaal) het maximum aantal uitbreidingen dat door de beheerder van enterprise search is geconfigureerd. Als de index meer uitbreidingen dan het opgegeven aantal uitbreidingen bevat, worden deze genegeerd. (In de zoekresultaten wordt aangegeven of er uitbreidingen zijn genegeerd.)
- Als de ondersteuning voor jokertekens is beperkt tot een set velden, bevat de set alleen de termen die in een van de opgegeven velden voorkomen. Een term hoeft in slechts een van de velden in minimaal één document in de index voor te komen.
- Als de zoekterm een term van een veldzoekopdracht is, moet het jokerteken achter de veldspecifier staan (bijvoorbeeld veldnaam:\*sphere). In de veldnaam mogen geen dubbele punten voorkomen (:).
- Als de ondersteuning voor jokertekens is beperkt tot een set velden, moet de veldnaam in de zoekterm met jokertekens een van de velden zijn die is opgegeven in de beheerconsole van enterprise search. Anders worden er geen uitbreidingen voor de term gevonden.
- Jokertekens worden alleen ondersteund in termen met platte tekst, niet in XML-elementnamen, kenmerknamen of kenmerkwaarden. Termen die uit alleen jokertekens bestaan, worden niet ondersteund.

## De invloed van jokertekens op de index

De ondersteuning voor jokertekens op basis van indexuitbreiding zorgt voor een grotere index en meer tijd voor het opbouwen van de index.

Als indexuitbreiding is ingeschakeld, worden elk voorvoegsel van een term en de term zelf geïndexeerd. De volgende termen worden bijvoorbeeld geïndexeerd voor de term support:

```
s su sup supp suppo suppor support
```

Het aantal termen dat in de index wordt opgeslagen, neemt toe op basis van de gemiddelde lengte van een woord. Met indexcompressie wordt de grootte van de index enigszins verminderd, maar niet drastisch. De tijd die nodig is om de index op te bouwen, neemt toe op basis van de gemiddelde lengte van een woord.

Een index met Engelse documenten neemt toe met een factor van ongeveer vier, omdat de gemiddelde lengte van een Engels woord vijf tot zes tekens is. Een index met n-gramtokenisatie wordt ongeveer twee keer zo groot, omdat elke n-gram twee tekens bevat.

Indexuitbreiding wordt in de volgende situaties aanbevolen voor jokertekenondersteuning:

- De collectie is klein genoeg zodat de ruimte en tijd die door de uitgebreide index worden ingenomen, niet voor prestatieproblemen zorgen.
- Alle mogelijke uitbreidingen van jokertekens worden in de zoekresultaten opgenomen om te voldoen aan de vereisten van de gebruiker (of het bedrijf).

U kunt de invloed op de index verminderen door een minimumlengte voor voorvoegsels op te geven voor jokertekenuitbreidingen. Als de minimumlengte voor voorvoegsels bijvoorbeeld drie is, worden de voorvoegsels *s* en *su* niet geïndexeerd voor het woord *support* en wordt een Engelse index verminderd met de factor drie (in plaats van vier).

Met query-uitbreiding voor jokertekens worden geen voorvoegsels naar de index geschreven. Termen worden uitgebreid wanneer de query wordt ingediend en de index groeit alleen op basis van een kleine gegevensstructuur die vereist is voor ondersteuning van die uitbreiding. Een index met query-uitbreiding is doorgaans tussen 10% en 20% groter dan een index zonder jokertekenondersteuning en het opbouwen van de index neemt minder dan 10% meer tijd in beslag. De configuratie van het maximum aantal uitbreidingen is niet van invloed op de grootte van de index en de tijd voor het opbouwen van de index.

## Opties configureren voor jokertekens in query's

Bij het configureren van indexeringsopties voor een enterprise search-collectie kunt u opgeven of gebruikers jokertekens in zoektermen kunnen opnemen.

### Voordat u begint

Om opties te configureren voor jokertekens, moet u de rol van beheerder van enterprise search hebben of beheerder zijn van de collectie waarvan de index deel uitmaakt.

### Over deze taak

De wijzigingen die u in de opties voor jokertekens opgeeft, worden van kracht zodra de hoofdindex de volgende keer wordt opgebouwd.

### Procedure

Ga als volgt te werk om ondersteuning voor jokertekens in query's te configureren:

1. Open een collectie, selecteer de pagina **Index** en klik op **Opties voor jokertekens configureren**.
2. Schakel op de pagina **Opties voor jokertekens** het aankruisvakje **Jokertekens in query's accepteren** in.
3. Optioneel: u kunt opgeven dat u jokertekens wilt toestaan in vrije zoekopdrachten. Met de vrije zoekopdracht *tech\** (waarmee niet wordt gezocht in benoemde velden) worden bijvoorbeeld alleen uitgebreide resultaten teruggestuurd (zoals technologie of techniek) als dit aankruisvakje is ingeschakeld.
4. Geef op welke velden jokertekens ondersteunen:

- Om op te geven dat jokertekens niet kunnen worden verwerkt in query's waarmee velden worden doorzocht, selecteert u **Geen velden**.
  - Om op te geven dat alle velden in een document query's met jokertekens ondersteunen, selecteert u **Alle velden**.
  - Om de ondersteuning voor jokertekens te beperken tot bepaalde velden, selecteert u **Bepaalde velden** en voert u vervolgens de veldnamen in. Alleen voor de velden die u opgeeft worden uitgebreide resultaten teruggestuurd. Met de query auteur:john\* worden bijvoorbeeld alleen uitgebreide resultaten teruggestuurd als u opgeeft dat het veld auteur ondersteuning biedt voor jokertekens.
5. Geef op of het jokerteken het laatste teken in een zoekterm moet zijn (een volgteken) of dat het op elke positie in een zoekterm kan worden gebruikt. Bij het selecteren van de positie en het type voor jokertekens moet u ook opgeven op welke manier u ondersteuning voor jokertekens wilt inschakelen. Klik voor meer informatie in de beheerconsole op **Help**.

## Scopes

Een scope configureren voor een beperkte collectieweergave

Een scope is een groep gerelateerde URI's in een index. Wanneer u een scope configureert, beperkt u het aantal documenten dat gebruikers in de collectie kunnen bekijken. Bij het doorzoeken van de collectie doorzoeken de gebruikers alleen de documenten in de scope, niet de volledige index. Als u deze functie wilt gebruiken, moeten uw zoekprogramma's ondersteuning bieden voor het doorzoeken van scopes.

Bij het maken van een scope geeft u een reeks URI's in de index op die gebruikers kunnen doorzoeken. Door beperkingen op te leggen aan de documenten die de gebruikers kunnen doorzoeken, draagt u ertoe bij dat de documenten in de zoekresultaten specifiek zijn voor de informatie waarnaar de gebruikers op zoek zijn.

U zou bijvoorbeeld een één scope kunnen definiëren voor de URI's van documenten van de Technische Dienst en een andere scope voor de URI's van documenten van de afdeling HR. Als uw zoekprogramma kan werken met scopes, zullen gebruikers van de afdeling Technische Dienst alleen documenten vinden die zich in de scope Technische Dienst bevinden, en gebruikers van de afdeling Human Resources alleen documenten in de scope Human Resources.

U kunt zo veel scopes definiëren als u wilt, maar een te groot aantal scopes heeft wel negatieve gevolgen voor de snelheid. Definieer liefst een zodanig aantal scopes dat er bij de meeste zoekopdrachten op slechts één of twee scopes hoeft te worden gefilterd. Omdat scopes hele URI's of URI-patronen kunnen bevatten, kan hetzelfde document tot meer dan één scope behoren.

Als u een scope hebt geconfigureerd, kan het nodig zijn om de hoofdindex tot twee keer toe op te bouwen om de wijzigingen actief te maken. Als u een scope configureert voordat de hoofdindex van de collectie voor het eerst wordt opgebouwd, zijn de gebruikers wel in staat de collectie te doorzoeken, maar kunnen de scopegegevens niet in de zoekresultaten worden weergegeven. Bouw de index opnieuw op om ervoor te zorgen dat het bereik van URI's in de scope, in de zoekresultaten wordt weergegeven.

Als u een scope configureert nadat de hoofdindex ten minste één keer is opgebouwd, worden de wijzigingen van kracht nadat de hoofdindex de volgende keer opnieuw is opgebouwd.

## Scopes configureren

Bij het configureren van een scope voor een enterprise search-collectie geeft u de URI's of URI-patronen op voor een reeks documenten in de index die door de gebruikers kunnen worden doorzocht.

### Voordat u begint

Om scopes te configureren, moet u de rol van beheerder van enterprise search hebben of beheerder zijn van de collecties waarvan de scopes deel uitmaken.

### Over deze taak

Als uw zoekprogramma's scopes ondersteunen, kunnen gebruikers tijdens het doorzoeken van de collectie alleen de documenten doorzoeken waarvan de URI's binnen de gedefinieerde scope vallen.

Als u een scope hebt geconfigureerd, kan het nodig zijn om de hoofdindex tot twee keer toe op te bouwen om de wijzigingen actief te maken. Als u een scope configureert voordat de index voor het eerst wordt opgebouwd, zijn de gebruikers wel in staat de collectie te doorzoeken, maar kunnen de scopegegevens niet in de zoekresultaten worden weergegeven. Bouw de index opnieuw op om ervoor te zorgen dat het bereik van URI's in de scope, in de zoekresultaten wordt weergegeven.

Als u een scope configureert nadat de hoofdindex ten minste één keer is opgebouwd, worden de wijzigingen van kracht nadat de hoofdindex de volgende keer opnieuw is opgebouwd.

### Procedure

Ga als volgt te werk om een scope te configureren:

1. Open een collectie, selecteer de pagina Index en klik op **Scopes configureren**.
2. Klik op de pagina Scopes op **Scope maken**.
3. Geef een naam op voor de scope en geef de URI's en URI-patronen op waarmee het bereik van de scope wordt aangegeven. U kunt ook opgeven dat bepaalde URI's of URI-patronen moeten worden uitgesloten van de scope.
4. Klik op **OK**.

De nieuwe scope wordt op de pagina Scopes weergegeven met de overige scopes die deel uitmaken van de collectie.

#### Verwante verwijzing

“URI-indelingen in een enterprise search-index” op pagina 111

Met de URI (Uniform Resource Identifier) van documenten in een enterprise search-index wordt het type crawler aangegeven waarmee het document aan de collectie is toegevoegd.

## Samengevouwen URI's

In enterprise search kunnen de zoekresultaten zodanig worden georganiseerd, dat de documenten van bronnen met dezelfde URI-prefix als groep worden weergegeven en in de zoekresultaten worden samengevouwen.

Wanneer de resultaten zijn samengevouwen, verschijnt het resultaat met de hoogste ranking linkslijnend. Een of meer resultaten met een lagere ranking worden gegroepeerd weergegeven onder het eerste resultaat.

Als u resultaatdocumenten met verschillende URI-prefixen wilt samenvouwen als één groep, kunt u de URI-prefixen koppelen aan een groepsnaam. Als u bijvoorbeeld drie servers hebt voor het beheren van financiële gegevens, kunt u de documenten van alle drie de servers in de zoekresultaten groeperen en de resultaten met een lagere ranking onder de hoogste resultaatdocumenten samenvouwen.

Zoekprogramma's kunnen de URI-prefix of de groepsnaam gebruiken om documenten in de zoekresultaten samen te vouwen. In het voorbeeldzoekprogramma voor enterprise search, worden de belangrijkste twee zoekresultaten weergegeven. Als meer dan twee resultaatdocumenten met dezelfde URI-prefix (of documenten die tot dezelfde URI-groep behoren) als resultaat worden gegeven, kunt u een optie selecteren om de samengevouwen resultaten weer te geven.

Gebruikers kunnen de enterprise search-queriesyntaxis (*samegroupas:URI-prefix*) gebruiken om alle documenten te doorzoeken die tot dezelfde groep behoren als de opgegeven URI-prefix die in de query is opgegeven.

## URI-prefixen en groepsnamen organiseren

Als u de beheerconsole gebruikt om regels voor het samenvouwen van zoekresultaten te configureren, geeft u de URI-prefixen op van de documenten die u wilt samenvouwen en koppelt u de URI-prefixen eventueel aan een groepsnaam.

De volgorde van de URI-prefixen die u configureert, is van belang. De indexserver gebruikt bij het berekenen van elke URI in een collectie de volgorde van de URI-prefixen. Voor elke URI geldt het volgende:

1. De URI-prefixen worden op de indexserver opeenvolgend gescand.
2. Wanneer de indexserver de eerste URI-prefix vindt die overeenkomt met een prefix van een document in de index, koppelt de indexserver de groepsnaam (of de URI-prefix, als er geen groepsnaam aanwezig is) als een extra zoekterm aan het document.

Als een webdocument niet overeenkomt met een URI-prefix, gebruikt de indexserver de hostnaam van de URL als de URI-prefix. Als een NNTP-document niet overeenkomt met een URI-prefix, gebruikt de indexserver het eerste bericht-ID in de waarde van de verwijzingsheader als de URI-prefix.

Nadat u een URI-prefix hebt toegevoegd aan de lijst van prefixen die in de zoekresultaten moeten worden samengevouwen, moet u de URI-prefix in de volgorde plaatsen waarin deze door de indexserver moet worden gescand en de prefix eventueel als extra zoekterm aan documenten in de index koppelen:

- Wanneer u een URI-prefix toevoegt en niet aan een groepsnaam koppelt, kunt u de afzonderlijke URI-prefix selecteren en omhoog of omlaag in de lijst verplaatsen.
- Wanneer u een URI-prefix toevoegt en aan een groepsnaam koppelt, wordt de volledige groep van URI-prefixen die tot dezelfde groep behoren, verplaatst wanneer u een URI-prefix omhoog of omlaag in de lijst verplaatst. De volgorde van URI-prefixen in een groep is niet van belang; als u een afzonderlijke URI-prefix selecteert, wordt automatisch de hele groep geselecteerd.

## URI's samenvouwen in de zoekresultaten

U kunt opties opgeven voor het groeperen en samenvouwen van de resultaatdocumenten van bronnen die dezelfde URI-prefix hebben. Bovendien kunt u een groepsnaam opgeven zodat resultaatdocumenten met andere URI-prefixen kunnen worden samengevouwen.

### Voordat u begint



Om opties op te geven voor het samenvouwen van zoekresultaten, moet u de rol van beheerder van enterprise search hebben of beheerder van de collectie zijn.

### Over deze taak

De wijzigingen die u aanbrengt voor het samenvouwen van de zoekresultaten, worden pas van kracht als u de hoofdindex opnieuw hebt opgebouwd.

### Procedure

Ga als volgt te werk om opties op te geven voor het samenvouwen van de zoekresultaten:

1. Open een collectie, selecteer de pagina Index en klik op **Zoekresultaten samenvouwen**.
2. Klik op de pagina Zoekresultaten samenvouwen op **URI-prefix toevoegen**.
3. Voer op de pagina URI-prefix voor samenvouwen van resultaten toevoegen de URI-prefix in voor de documenten die u in de zoekresultaten wilt samenvouwen. Bijvoorbeeld:  
`http://finance/ROI/`  
`http://server1.com/finance/db2://LOCALDB/SCHEMA1.TABLE1/exchange://exchangesvr.ibm.com/public/TeamRoom/Folder1/`
4. U kunt een beschrijvende groepsnaam invoeren en deze vervolgens aan de URI-prefix koppelen. Om de resultaatdocumenten van verschillende bronnen samen te vouwen als één groep, voert u de groepsnaam in tijdens het toevoegen van de verschillende URI-prefixen.
5. Klik op **OK**.
6. Ga naar de pagina Zoekresultaten samenvouwen plaats de nieuwe regel in de gewenste scanvolgorde voor de indexserver:
  - Als u een URI-prefix hebt toegevoegd en hieraan geen groepsnaam hebt gekoppeld, verschijnt de nieuwe URI-prefix onder in de lijst. Gebruik de cursortoetsen om de prefix op de juiste plek te plaatsen.
  - Als u een groepsnaam aan de nieuwe URI-prefix hebt gekoppeld, verschijnt de nieuwe URI-prefix onder aan de set met URI-prefixen die deel uitmaken van dezelfde groep. Gebruik de cursortoetsen om de hele groep met URI-prefixen op de juiste plek te plaatsen.
7. Om de URI-prefix of groepsnaam te wijzigen, selecteert u de URI-prefix en klikt u op  **Bewerken**.
8. Om een URI-prefix uit de lijst te verwijderen, selecteert u de URI-prefix en klikt u op  **Verwijderen**.

## URI's uit de index verwijderen

Om te voorkomen dat gebruikers bepaalde documenten in een collectie doorzoeken, kunt u de URI's van die documenten uit de index verwijderen.

### Voordat u begint

Om URI's uit de index te verwijderen, moet u de rol van beheerder van enterprise search hebben of beheerder van de collectie zijn.

### Over deze taak

Als u een volledige URI opgeeft, zien gebruikers de URI niet langer in de zoekresultaten staan. Als een gebruiker echter dezelfde query opgeeft en de resultaatdocumenten voor die query nog in de zoekcache staan, blijft de resultaatpagina voor de verwijderde URI in de zoekresultaten verschijnen. De zoekcache wordt pas vernieuwd en de URI wordt pas uit de index verwijderd, als de hoofd- of delta-index volledig opnieuw wordt opgebouwd.

Als u een URI-patroon opgeeft voor het verwijderen van meerdere URI's, blijven gebruikers de URI's die met dat patroon overeenkomen in de zoekresultaten zien totdat de hoofdindex opnieuw wordt opgebouwd.

Wanneer u een URI uit de index verwijdert, blijft deze in de crawlruimte staan. De volgende keer dat de crawler het document crawlt, wordt de URI in de index opgebouwd en kan deze weer worden doorzocht. Om een URI uit de crawlruimte te verwijderen, moet u de crawlingregels aanpassen om het document uit te sluiten en moet u vervolgens de crawler stoppen en opnieuw starten.

### Procedure

Ga als volgt te werk om URI's voor bepaalde documenten uit de index te verwijderen:

1. Open een collectie, selecteer de pagina Index en klik op **URI's uit de index verwijderen**.
2. Voer op de pagina URI's uit de index verwijderen de URI's (of URI-patronen) in die u uit de index wilt verwijderen.

Bijvoorbeeld:

```
http://domain.org/hr/*  
db2://knowledgeManagement/ROI*  
cm://enterprise/finance*
```

#### Verwante verwijzing

"URI-indelingen in een enterprise search-index" op pagina 111

Met de URI (Uniform Resource Identifier) van documenten in een enterprise search-index wordt het type crawler aangegeven waarmee het document aan de collectie is toegevoegd.



---

## Beheer van zoekserver voor enterprise search

Voorbeelden van opties die u voor zoekservers kunt opgeven, zijn het gebruik van cacheruimte voor de zoekresultaten, het beheer van de maximale weergavelengte van documentsamenvattingen in de zoekresultaten, het koppelen van aangepaste woordenboeken om de kwaliteit van zoekopdrachten te verbeteren en de weergave van vooraf gedefinieerde URI's in de zoekresultaten wanneer bepaalde termen in de query voorkomen.

Als een gebruiker een query opgeeft, gebruiken zoekservers de index om snel de relevante documenten te zoeken. Om de metagegevens voor de relevante documenten te kunnen ophalen, maken de zoekservers gebruik van de datastore in enterprise search, waarin de geanalyseerde en getokeniseerde gegevens zijn opgeslagen. Metagegevens bevatten, maar zijn niet beperkt tot de URI, de titel, de beschrijving, de datum, het gegevenstype, enzovoort van het document.

Als u de zoekservers configureert voor een collectie, kunt u opties opgeven voor de manier waarop query's worden verwerkt, waaronder opties die van invloed zijn op de prestaties van query's:

### **Een zoekcache configureren**

Om de prestaties van query's te verbeteren, kunt u aangeven dat de zoekresultaten (de reacties op query's) in een cache moeten worden opgeslagen en kunt u aangeven hoeveel ruimte aan de zoekresultaten in de cache moet worden toegewezen.

### **Maximale weergavelengte voor documentsamenvattingen configureren**

Voor de meeste resultaatdocumenten wordt een samenvatting van de documentcontent weergegeven, zodat gebruikers kunnen bepalen of het document de gewenste informatie bevat. U kunt aangeven hoeveel ruimte in de zoekresultaten moet worden gebruikt voor het weergeven van deze samenvattingsinformatie.

### **Een andere standaardtaal opgeven**

Tijdens het maken van de collectie wordt de standaardtaal voor het zoeken van documenten in de collectie opgegeven. Op elk gewenst moment kunt u een andere taal opgeven.

### **Aangepaste woordenboeken koppelen**

Als uw programmaontwikkelaars aangepaste woordenboeken voor synoniemen, stopwoorden of gewogen woorden hebben gemaakt, kunt u opgeven welke woordenboeken moeten worden gebruikt bij het doorzoeken van de collectie.

### **Snellinks configureren**

U kunt vooraf vaststellen welke URI's worden geretourneerd voor bepaalde trefwoorden en termen. Wanneer gebruikers de trefwoorden of termen in een query opgeven, wordt de vooraf gedefinieerde URI geretourneerd met de zoekresultaten. De snellink-URI's worden geretourneerd in aanvulling op de URI's die door de zoekservers worden geretourneerd door de index te doorzoeken.

### **Verwante onderwerpen**

"Ranking van documenten in enterprise search" op pagina 189

Wanneer een gebruiker een collectie zoekt, sturen de zoekprocessen de meest relevante resultaten terug voor de voorwaarden en bepalingen van de query.

“Aangepaste gewogen woordenboeken” op pagina 192

Om de kwaliteit van de zoekresultaten te verbeteren, kunt u invloed uitoefenen op de manier waarop documenten worden gerangschikt in het zoekresultaat. Een van de manieren om dat te doen, is een aangepast gewogen woordenboek te maken.

---

## Zoekcache

Als de werkbelasting van de zoekservers relatief hoog is, kunt u de prestaties verbeteren door de zoekresultaten op te slaan in een cache.

Als de zoekservers een zoekopdracht verwerken, kijken ze eerst of er nog geen resultaten voor dezelfde zoekopdracht aanwezig zijn in de cache. Als de juiste respons op de query worden gevonden, kunnen de resultaten snel worden teruggestuurd aan de gebruiker. Kunnen de zoekservers de juiste respons niet vinden in de cache, dan wordt de index doorzocht.

Als de zoekcache vol raakt, worden de oudste zoekresultaten en de resultaten van zoekopdrachten die minder vaak worden uitgevoerd, gewist om plaats te maken voor nieuwe zoekresultaten.

U kunt het gebruik van een zoekcache inschakelen vanuit de beheerconsole van enterprise search. Daar kunt u ook de capaciteit van de cache opgeven (het aantal queryresultaten die gelijktijdig in de cache kunnen worden opgeslagen).

Als u wijzigingen aanbrengt in de instellingen van de zoekcache, moet u de zoekservers opnieuw starten om die wijzigingen te activeren.

---

## Een zoekcache configureren

U kunt de zoekcache voor een collectie in- of uitschakelen. Daarnaast kunt u opties opgeven om de grootte van de zoekcache te bepalen.

### Voordat u begint

Om de zoekcache voor een collectie te configureren, moet u de rol van beheerder van enterprise search hebben of beheerder van een collectie zijn.

### Procedure

Ga als volgt te werk om de zoekcache te configureren:

1. Open een collectie, ga naar de pagina Zoeken en klik op **Opties voor zoekserver configureren**.
2. Schakel het aankruisvakje **Zoekcache gebruiken** op de pagina Opties voor zoekservers in.
3. Voer in het veld **Maximum aantal vermeldingen in de cache** het maximum aantal query-antwoorden in dat in de zoekcache kan worden opgeslagen.
4. Klik op **OK**.
5. Om de wijzigingen te activeren, moet u de zoekservers bewaken en de zoekprocessen opnieuw starten.

---

## Aangepaste synoniemenwoordenboeken

Om de kwaliteit van de zoekresultaten te verbeteren, kunt u opgeven dat gebruikers bij het doorzoeken van een collectie kunnen zoeken naar synoniemen van de zoektermen.

Als u een synoniemenwoordenboek maakt, dit toevoegt aan het enterprise search-systeem en koppelt aan een collectie, kunnen gebruikers tijdens het doorzoeken van de collectie naar documenten zoeken die synoniemen bevatten van de zoektermen. Als u query's op deze manier uitbreidt, is de kans groter dat gebruikers alle documenten van interesse vinden en niet alleen de documenten die exact aan de zoektermen voldoen. Omdat u tijdens het maken van het synoniemenwoordenboek aangeeft welke woorden synoniemen van elkaar zijn, zorgt u dat gebruikers de relevante documenten kunnen vinden zonder dat ze hiervoor verschillende varianties van de zoekterm hoeven op te geven.

Het kan bijvoorbeeld zijn dat uw bedrijf synoniemen en afkortingen gebruikt als verwijzing naar afdelingen, materiaal, enzovoort, of dat de documenten in uw collecties jargon bevat dat specifiek van toepassing is op uw bedrijfstak. Door een synoniemenwoordenboek te maken, kunt u ervoor zorgen dat een query die een acroniem bevat (zoals toegang) documenten retourneert waarin uitbreidingen van dat acroniem voorkomen (zoals toegangsbesturing, toegangsbesturingslijsten, enzovoort).

Gebruikers kunnen synoniemen gebruiken in de enterprise search-querytaal door een tilde vóór de zoekterm te plaatsen. Met de query ~WAS kunnen bijvoorbeeld documenten als resultaat worden gegeven waarin WebSphere Application Server wordt besproken. Programmaontwikkelaars kunnen de ondersteuning voor synoniemen ook inschakelen via de query-eigenschappen (hiervoor is geen speciale syntaxis vereist).

Synoniemenwoordenboeken bevatten varianten van woorden en hebben de volgende kenmerken:

- De woorden zijn niet specifiek van toepassing op een taal, maar kunnen in verschillende talen worden gebruikt. Voor elke collectie bestaat slechts één synoniemenwoordenboek.
- De woorden zijn niet verbogen. Alle mogelijke verbuigingen moeten aan de synoniemenlijst worden toegevoegd. Een verbuiging kan bijvoorbeeld het woord in enkelvoud en in meervoud zijn (toegangslijst en toegangslijsten).

De meeste termen die u aan een synoniemenwoordenboek toevoegt, zijn semantische equivalenten, wat betekent dat als term A een synoniem is van term B, B een synoniem is van A. Elke keer dat term A in een query wordt gebruikt, zou term B kunnen worden gebruikt (en omgekeerd).

U kunt echter ook termen toevoegen die overeenkomen met een ander gebruik van de term, zoals generieke of meer specifieke varianten van een term. U kunt bijvoorbeeld werken met een synoniemengroep die zowel de termen bouwen als huis bevat en met een andere groep die de termen bank, shore en kredietvereniging bevat.

Hoe minder strikt de relatie tussen de termen is, hoe meer zoekresultaten worden weergegeven, hoewel het kan voorkomen dat bepaalde zoekresultaten niet relevant zijn voor de query. De SI-API (Search and Index API) biedt methoden waarmee gebruikers de juiste synoniemen kunnen selecteren tijdens het indienen van een

zoekopdracht, en methoden waarmee gebruikers kunnen bekijken welke zoektermen aan welke synoniemen zijn toegevoegd.

Om een synoniemenwoordenboek te maken, moet een expert op het gebied van het onderwerp van de collectie een gewogen woordenlijst in XML-indeling opstellen (of dit doen in samenwerking met een programma-ontwikkelaar). Met behulp van een tool in enterprise search, *essyndictbuilder*, kan het XML-bestand worden geconverteerd naar een binair bestand (.dic).

De enterprise search-beheerder uploadt het binaire bestand naar het systeem en kent er een weergavenaam aan toe. De collectiebeheerders kunnen het synoniemenwoordenboek vervolgens selecteren voor gebruik bij het doorzoeken van de documenten in een collectie. Dit doen ze tijdens het configureren van de zoekserveropties voor die collectie.

**Beperking:** Als een aangepast synoniemenwoordenboek eenmaal in het systeem is opgenomen, kan het niet meer worden gewijzigd. Als u de synoniemen wilt wijzigen die beschikbaar zijn in een collectie, moet u de volgende stappen uitvoeren:

1. Het bron-XML-bestand bijwerken.
2. De XML-bron converteren naar een nieuw woordenboekbestand.
3. Het oude synoniemenwoordenboek verwijderen uit de collecties die er gebruik van maken.
4. Het oude synoniemenwoordenboek wissen van het systeem.
5. Het nieuwe synoniemenwoordenboek in het systeem opnemen.
6. Het nieuwe synoniemenwoordenboek koppelen aan de collecties die het moeten gaan gebruiken.

#### **Verwante onderwerpen**

Ondersteuning voor synoniemen in zoekprogramma's

#### **Verwante taken**

Een XML-bestand voor synoniemen maken

Een synoniemenwoordenboek maken

## **Synoniemenwoordenboeken aan het systeem toevoegen**

Als u aangepaste stopwoordenboeken maakt voor het doorzoeken van documenten in een collectie, moet u die woordenboeken koppelen aan het enterprise search-systeem. Later kunt u dan kiezen welk synoniemenwoordenboek u wilt gebruiken voor het doorzoeken van een bepaalde collectie.

### **Voordat u begint**

Om synoniemenwoordenboeken toe te voegen voor gebruik met enterprise search-query's, moet u de rol van beheerder van enterprise search hebben.


### **Beperkingen**

De maximumgrootte van een synoniemenwoordenboek is 8 MB.

### **Procedure**

Ga als volgt te werk om synoniemen aan het enterprise search-systeem te koppelen:

1. Klik op **Systeem** om de view Systeem te openen.

2. Klik op  **Wijzigen** om naar de view **Systeem wijzigen** te gaan.
3. Klik op de pagina **Zoeken** op **Synoniemenwoordenboeken configureren**.
4. Klik op de pagina **Synoniemenwoordenboeken configureren** op **Synoniemenwoordenboek toevoegen**.
5. Voer op de pagina **Een synoniemenwoordenboek toevoegen** een unieke weergavenaam in voor het synoniemenwoordenboek en voer indien gewenst een beschrijving in.
6. Geef de locatie van het dic-bestand op. Als het bestand is opgeslagen op het lokale systeem, kunt u op **Bladeren** klikken en het bestand opzoeken. Als het bestand is opgeslagen op de indexserver, moet u het volledige pad voor het bestand typen.
7. Klik op **OK**. Het aangepaste synoniemenwoordenboek wordt aan het enterprise search-systeem toegevoegd en wordt beschikbaar gesteld voor het doorzoeken van collecties.

## Een synoniemenwoordenboek aan een collectie koppelen

Als er synoniemenwoordenboeken in het enterprise search-systeem zijn opgenomen, kunt u een van die woordenboeken selecteren om te gebruiken bij het doorzoeken van een collectie. Als een zoekterm gelijk is aan een woord in het woordenboek, worden in de zoekresultaten ook resultaatdocumenten weergegeven die synoniemen van die term bevatten.

### Voordat u begint

Om een synoniemenwoordenboek voor een collectie te selecteren, moet u de rol van beheerder van enterprise search hebben of beheerder van een collectie zijn.

### Procedure

Ga als volgt te werk om een synoniemenwoordenboek aan een collectie te koppelen:

1. Open een collectie, ga naar de pagina **Zoeken** en klik op **Opties voor zoekserver configureren**.
2. Selecteer in het veld **Naam synoniemenwoordenboek** op de pagina **Opties voor zoekservers** het synoniemenwoordenboek dat moet worden gebruikt als gebruikers query's op deze collectie uitvoeren.

De lijst van beschikbare synoniemenwoordenboeken bevat alle synoniemenwoordenboeken die in het enterprise search-systeem zijn opgenomen.

3. Klik op **OK**.

---

## Aangepaste stopwoordenboeken

Om de kwaliteit van de zoekresultaten te verbeteren, kunt u opgeven dat bepaalde woorden automatisch uit de zoektermen moeten worden verwijderd tijdens de queryverwerking.

Een stopwoordenboek bevat termen die veel worden gebruikt in uw bedrijf en dus nuttig zijn om als zoektermen te worden gebruikt. Als u deze woorden uitsluit van query's, zorgt u dat gebruikers geen minder relevante resultaatdocumenten te zien krijgen (alleen documenten die overeenkomen met de overige termen in de query worden als resultaat gegeven). Tijdens de queryverwerking worden de stopwoorden door de zoekservers uit de query's verwijderd. Voorbeelden van woorden die

worden verwijderd zijn stopwoorden in het aangepaste woordenboek en stopwoorden die vooraf zijn gedefinieerd voor enterprise search (zoals veelvoorkomende voorzetsels en lidwoorden).

In enterprise search wordt standaard taalspecifieke herkenning van stopwoorden uitgevoerd. Met dit proces worden veelvoorkomende woorden zoals een en het uit een query verwijderd. U moet alleen een aangepast woordenboek definiëren voor bedrijfs- of domeinspecifieke stopwoorden.

Wanneer een query wordt verwerkt, worden de stopwoorden verwijderd voordat de spellingsuggesties worden gegeven. Als alle woorden in een query stopwoorden zijn, worden er geen stopwoorden verwijderd tijdens de queryverwerking. Om te voorkomen dat er geen zoekresultaten worden geretourneerd, wordt het verwijderen van stopwoorden uitgeschakeld als alle zoektermen stopwoorden zijn. Als het woordauto een stopwoord is en u zoekt naar auto, worden in de zoekresultaten documenten weergegeven waarin het woord auto voorkomt. Als u zoekt naar auto volvo worden in de zoekresultaten alleen documenten weergegeven waarin het woord volvo voorkomt.

Om een stopwoordenboek te maken, moet een expert op het gebied van het onderwerp van de collectie een stopwoordenlijst in XML-indeling opstellen (of dit doen in samenwerking met een programma-ontwikkelaar). Met behulp van een tool in enterprise search, esstopworddictbuilder, kan het XML-bestand worden geconverteerd naar een binair bestand (.dic).

De enterprise search-beheerder uploadt het binaire bestand naar het systeem en kent er een weergavenaam aan toe. De collectiebeheerders kunnen dat stopwoordenboek vervolgens selecteren voor gebruik bij het doorzoeken van de documenten in een collectie. Dit doen ze op het moment dat de de zoekserveropties voor die collectie configureren.

**Beperking:** Als een aangepast stopwoordenboek eenmaal in het systeem is opgenomen, kan dit niet meer worden gewijzigd. Als u wijzigingen wilt aanbrengen in de stopwoorden die beschikbaar zijn voor de verwerking van zoekopdrachten, moet u:

1. Het bron-XML-bestand bijwerken.
2. De XML-bron converteren naar een nieuw woordenboekbestand.
3. Het oude stopwoordenboek verwijderen uit de collecties die er gebruik van maken.
4. Het oude stopwoordenboek wissen van het systeem.
5. Het nieuwe stopwoordenboek in het systeem opnemen.
6. Het nieuwe stopwoordenboek koppelen aan de collecties die het moeten gaan gebruiken.

#### **Verwante onderwerpen**

Aangepaste stopwoordenboeken

#### **Verwante taken**

Een XML-bestand voor stopwoorden maken

Een stopwoordenboek maken

## **Stopwoordenboeken aan het systeem toevoegen**

Als u aangepaste stopwoordenboeken maakt zodat bepaalde woorden uit query's worden verwijderd, moet u die woordenboeken opnemen in het enterprise search-

systeem. Later kunt u dan kiezen welk stopwoordenboek u wilt gebruiken voor het doorzoeken van een bepaalde collectie.

### Voordat u begint


Om stopwoordenboeken toe te voegen aan het systeem, moet u de rol van beheerder van enterprise search hebben.

### Beperkingen

De maximumgrootte van een stopwoordenboek is 8 MB.

### Procedure

Ga als volgt te werk om aangepaste stopwoorden aan het enterprise search-systeem te koppelen:

1. Klik op **Systeem** om de view Systeem te openen.
2. Klik op  **Wijzigen** om naar de view Systeem wijzigen te gaan.
3. Ga naar de pagina Zoeken en klik op **Stopwoordenboeken configureren**.
4. Klik op de pagina Stopwoordenboeken configureren op **Stopwoordenboek toevoegen**.
5. Geef op de pagina Stopwoordenboek toevoegen een unieke weergavenaam voor het woordenboek op en typ desgewenst ook een beschrijving.
6. Geef de locatie van het dic-bestand op. Als het bestand is opgeslagen op het lokale systeem, kunt u op Bladeren klikken en het bestand opzoeken. Als het bestand is opgeslagen op de indexserver, moet u het volledige pad voor het bestand typen.
7. Klik op **OK**. Uw aangepaste stopwoordenboek wordt toegevoegd aan het enterprise search-systeem en wordt beschikbaar gesteld voor het doorzoeken van collecties.

## Een stopwoordenboek aan een collectie koppelen

Als er stopwoordenboeken in het enterprise search-systeem zijn opgenomen, kunt u een van die woordenboeken selecteren om te gebruiken bij het doorzoeken van een collectie. Als een zoekterm overeenkomt met een term in het woordenboek, wordt die term uit de query verwijderd voordat deze wordt verwerkt.

### Voordat u begint

Om een stopwoordenboek voor een collectie te selecteren, moet u de rol van beheerder van enterprise search hebben of beheerder van een collectie zijn.

### Procedure

Ga als volgt te werk om een stopwoordenboek aan een collectie te koppelen:

1. Open een collectie, ga naar de pagina Zoeken en klik op **Opties voor zoekserver configureren**.
2. Ga naar het veld **Naam stopwoordenboek** op de pagina Opties voor zoekserver en selecteer het stopwoordenboek dat moet worden gebruikt als gebruikers deze collectie doorzoeken.

De lijst van beschikbare woordenboeken bevat alle stopwoordenboeken die in het enterprise search-systeem zijn opgenomen.

3. Klik op **OK**.

---

## Dynamisch samenvatten

Dynamisch samenvatten is een methode waarmee wordt bepaald welke termen in een resultaatdocument de beste afspiegeling geven van de begrippen die de gebruiker zoekt.

Voor enterprise search probeert de functie voor dynamisch samenvatten zinnen in documenten vast te leggen die een grote verscheidenheid aan zoektermen bevatten. Enkele zinnen, of delen van zinnen, worden geselecteerd om te worden afgebeeld in de zoekresultaten. De zoektermen worden in de HTML-weergave geaccentueerd weergegeven.

Bij het configureren van zoekserveropties voor een collectie kunt u opgeven hoeveel ruimte in de zoekresultaten moet worden vrijgemaakt voor de document-samenvattingen. Omdat de samenvatting geaccentueerde tekens bevat, wordt de buffer die wordt teruggestuurd naar het zoekprogramma groter dan het maximum dat u hebt opgegeven. De weergegeven lengte wordt echter nooit groter dan de opgegeven waarde, maar de samenvatting kan natuurlijk wel korter zijn (afhankelijk van de samenvattingsgegevens die uit het brondocument zijn opgehaald).

## Documentsumenvattingen aanpassen in de beheerconsole

U kunt de hoeveelheid informatie aanpassen die in de documentsumenvattingen wordt weergegeven door opties voor de zoekserver op te geven in de beheerconsole van enterprise search.

### Voordat u begint

Om de weergavelengte van de samenvattingen voor een collectie vast te stellen, moet u de rol van beheerder van enterprise search hebben of beheerder van een collectie zijn.

### Over deze taak

De waarde die u opgeeft voor de weergavelengte van documentsumenvattingen, wordt gebruikt in combinatie met de waarde die u opgeeft voor het aantal zinnen dat elke samenvatting kan bevatten. De waarde die leidt tot de kortste document-samenvatting heeft voorrang.

Als u bijvoorbeeld een maximum van vier zinnen opgeeft, dan zal de document-samenvatting nooit meer dan vier zinnen bevatten, ook niet als het maximum aantal toegestane tekens veel groter is dan het aantal tekens in die vier zinnen. Maar als u een maximum van 10 zinnen opgeeft in combinatie met een maximum van 500 tekens, kan het gebeuren dat de documentsumenvatting minder dan 10 zinnen lang is.

### Procedure

Ga als volgt te werk om de weergavelengte voor documentsumenvattingen te configureren:

1. Open een collectie, ga naar de pagina Zoeken en klik op **Opties voor zoekserver configureren**.
2. Geef op de pagina Opties voor zoekservers een maximumweergavelengte op voor documentsumenvattingen. Wanneer gebruikers de zoekresultaten bekijken, zullen de documentsumenvattingen de opgegeven waarde niet overschrijven.



3. Geef aan hoeveel zinnen elke documentsamenvatting kan bevatten (samenvattingen kunnen maximaal tien zinnen bevatten).
4. Klik op **OK**.
5. Om de wijzigingen te activeren, moet u de zoekservers bewaken en de zoekprocessen opnieuw starten.

## Documentsamenvattingen aanpassen door de eigenschappen te bewerken

Elk resultaatdocument voor een enterprise search-query bevat een samenvatting. U kunt de hoeveelheid informatie aanpassen die elke samenvatting bevat door een eigenschappenbestand te bewerken.

### Over deze taak

U kunt de beschrijvingen van de zoekresultaten aanpassen door de waarden voor de volgende eigenschappen in het bestand `ES_NODE_ROOT/master_config/collectie_ID.runtime.node1/runtime-generic.properties` te wijzigen:

#### **MinWordsPerSentence**

Het minimum aantal woorden dat een zin met een beschrijving kan bevatten. De standaardwaarde is 4.

#### **MaxWordsPerSentence**

Het maximum aantal woorden dat een zin met een beschrijving kan bevatten. De standaardwaarde is 20.

#### **NumberOfReturnedSentences**

Het aantal zinnen dat een beschrijving vormt. De standaardwaarde is 5.

#### **MaxSentencesPerDocument**

Het maximum aantal zinnen in een document dat wordt beschouwd als kandidaat tijdens het maken van de beschrijving. De standaardwaarde is 1000.

### Procedure

Ga als volgt te werk om documentsamenvattingen in de zoekresultaten aan te passen:

1. Meld u bij de zoekservers aan als enterprise search-beheerder. Dit gebruikers-ID is opgegeven bij de installatie van WebSphere II OmniFind Edition.
2. Gebruik een teksteditor om het volgende bestand te bewerken, waarbij *coll\_ID* het ID is voor de collectie (of het ID dat door het systeem is toegewezen) dat tijdens het maken van de collectie is opgegeven:

```
ES_NODE_ROOT/master_config/coll_ID.runtime.node1/runtime-generic.properties
```

**Tip:** Om de toewijzing tussen een collectienaam en het ID vast te stellen, raadpleegt u het bestand `ES_NODE_ROOT/master_config/collections.ini`.

3. Wijzig de eigenschappen die u wilt aanpassen, sla het bestand op en sluit dit vervolgens af.
4. Stop de zoekservers en start deze opnieuw op om de wijzigingen toe te passen.

---

## Werken met snellinks

Snellinks zijn documenten die in de zoekresultaten worden afgebeeld wanneer een gebruiker een query opgeeft die bepaalde woorden en woordcombinaties bevat.

U kunt de beheerconsole van enterprise search gebruiken om snellinks voor een collectie te configureren.

## Snellinks

Met snellinks beschikken gebruikers over links naar documenten waarvan is vastgesteld dat ze relevant zijn voor de zoektermen.

Een snellink is een URI die enterprise search automatisch in de zoekresultaten opneemt als een query bepaalde woorden of woordcombinaties bevat. De snellinks verschijnen meestal boven aan de lijst van resultaten, zodat u zeker weet dat gebruikers de documenten die relevant zijn voor de query, niet over het hoofd zien.

Snellinks worden afgebeeld als aanvulling op de andere zoekresultaten. De zoekprocessen zoeken in de index naar documenten die voldoen aan de zoekcriteria en beelden de URI's van die documenten af, plus de URI's voor de snellinks.

Bij het configureren van een snellink kunt u een beschrijvende titel en een samenvatting voor de URI opgeven, zodat gebruikers het document eenvoudig kunnen herkennen en snel kunnen beslissen of ze het document willen ophalen.

Voor de URI <http://www.ibm.com/education/us/> kunt u bijvoorbeeld de titel IBM-training in de Verenigde Staten opgeven, met als samenvatting Oplossingen, producten en informatiebronnen voor professionals, docenten en studenten in te Verenigde Staten.

Om snellinks te kunnen gebruiken in enterprise search-collecties, moet de optie voor het afbeelden van snellinks beschikbaar zijn in het zoekprogramma. In sommige zoekprogramma's heeft de gebruiker de mogelijkheid om de weergave van snellinks in de zoekresultaten uit te schakelen.

## Snellinks configureren

Als u een snellink wilt maken voor een enterprise search-collectie, moet u de URI van een document koppelen aan de trefwoorden op basis waarvan het document in de zoekresultaten wordt opgenomen.

### Voordat u begint

Om snellinks te configureren, moet u de rol van beheerder van enterprise search hebben of beheerder zijn van de collectie waarvan de snellink deel uitmaakt.

### Over deze taak

Als u voorbeelden wilt van de manier waarop u trefwoorden en URI's voor snellinks kunt opgeven, klikt u tijdens het maken of bewerken van een snellink op **Help**.

U hoeft de zoekservers niet opnieuw te starten om de wijzigingen te activeren.

### Procedure

Ga als volgt te werk om een snellink te configureren:

1. Open een collectie, selecteer de pagina Zoeken en klik op **Snellinks configureren**.

2. Klik op de pagina Snellinks op **Snellink maken**.
3. Geef de trefwoorden en termen op aan de hand waarvan de snellink in de zoekresultaten wordt weergegeven, voer de URI in van het document dat u vooraf hebt ingesteld als relevant document voor de query en geef de overige opties voor de snellink op.

U kunt per regel één trefwoord, meerdere trefwoorden of één woordcombinatie (twee of meer woorden tussen aanhalingstekens) opgeven. U kunt de trefwoorden scheiden met een spatie (u kunt hiervoor geen komma gebruiken). Druk op de Enter-toets om naar een nieuwe regel te gaan.

4. Klik op **OK**.

De nieuwe snellink wordt op de pagina Snellinks weergegeven met de andere snellinks die bij de collectie horen.

#### **Verwante verwijzing**

“URI-indelingen in een enterprise search-index” op pagina 111

Met de URI (Uniform Resource Identifier) van documenten in een enterprise search-index wordt het type crawler aangegeven waarmee het document aan de collectie is toegevoegd.



---

## Ranking van documenten in enterprise search

Wanneer een gebruiker een collectie zoekt, sturen de zoekprocessen de meest relevante resultaten terug voor de voorwaarden en bepalingen van de query.

De zoekservers ondersteunen een uitgebreide querysyntaxis en gebruiken verschillende technieken om de meest relevante zoekresultaten te kunnen produceren, zoals op tekst gebaseerde scores en statische ranking. U kunt het standaardrankinggedrag uitbreiden door opties te configureren die van invloed zijn op het belang van documenten in de zoekresultaten:

- U kunt aangepaste gewogen woordenboeken maken om te manier te beïnvloeden waarop documenten met de opgegeven gewogen woorden, worden gerangschikt in de zoekresultaten.
- U kunt de scores beïnvloeden van documenten die aan het opgegeven URI-patroon voldoen.
- U kunt de scores beïnvloeden van documenten die velden bevatten die aan wegingsklassen zijn toegewezen.

### **Verwante onderwerpen**

“Documentranking op basis van wegingsklassen” op pagina 196

Als u velden koppelt aan wegingsklassen, kunt u op die manier invloed uitoefenen op de manier waarop documenten worden gerangschikt in het zoekresultaat.

“Documentranking op basis van URI-patronen” op pagina 194

U kunt het belang van documenten vergroten of verkleinen door wegingsfactoren aan URI-patronen toe te wijzen.

“Aangepaste gewogen woordenboeken” op pagina 192

Om de kwaliteit van de zoekresultaten te verbeteren, kunt u invloed uitoefenen op de manier waarop documenten worden gerangschikt in het zoekresultaat.

Een van de manieren om dat te doen, is een aangepast gewogen woordenboek te maken.

---

## Op tekst gebaseerde score

In enterprise search wordt de score voor elk document dat aan de woordcombinaties in een query voldoet, op dynamische wijze berekend.

Om de tekstscore te berekenen van elk document dat aan een query voldoet, neemt enterprise search verschillende factoren in acht, waaronder:

- De frequentie van elke zoekterm in de hele collectie. In het algemeen geldt dat zoektermen die in de meeste documenten voorkomen, minder bijdragen aan de score van een document dan zoektermen die in een selectievere set documenten voorkomen.
- Het aantal keer dat een zoekterm in het overeenkomende document voorkomt. In het algemeen geldt dat hoe vaker een zoekterm in een document voorkomt, hoe hoger de score is.
- Het feit of zoektermen bij elkaar in de buurt voorkomen in het overeenkomende document. In het algemeen geldt dat zoektermen die bij elkaar in de buurt voorkomen in een document, een hogere bijdrage leveren aan de score van het document dan wanneer dezelfde termen minder dicht bij elkaar in de buurt voorkomen.

- De context waarin zoektermen in elk overeenkomend document voorkomen. Een zoekterm in de titel van een document levert bijvoorbeeld een hogere bijdrage aan de score van het document dan wanneer dezelfde zoekterm in de lopende tekst van het document voorkomt.

De lengte van elk document en de uitgebreidheid van het woordenschat zijn ook factoren bij het bepalen van de score.

---

## Statische ranking

Aan bepaalde typen documenten kunt u een statische rankingfactor toewijzen waarmee het belang van deze documenten in de zoekresultaten wordt vergroot.

Als u een collectie maakt, geeft u op of u een statische rankingfactor aan de documenten in de collectie wilt koppelen. Voor webcontent kan het aantal in andere documenten aanwezige links naar een document en de oorsprong van deze links van invloed zijn op het belang van dat document in de zoekresultaten.

Voor documenten met datumvelden of datummetagegevens kunt u de datum van het document gebruiken om het belang te vergroten. Recente artikelen in NNTP-nieuwsgroepen kunnen bijvoorbeeld relevanter zijn dan oudere artikelen. Als een gegevensbron meerdere datumwaarden bevat, kunt u kiezen welke waarde het belangrijkste is voor het bepalen van het belang van documenten in de gegevensbron.

Als u statische ranking voor een collectie gebruikt, moet u ervoor zorgen dat u geen gegevensbronnen met verschillende rankingtypen in dezelfde collectie gebruikt. Als u bijvoorbeeld de links naar een document als statische rankingfactor wilt gebruiken, moet u zorgen dat de collectie alleen webdocumenten bevat. De kwaliteit van de zoekopdracht kan verslechteren als u bronnen met verschillende rankingmodellen in dezelfde collectie gebruikt.

Daarnaast moet u ervoor zorgen dat de documenten velden en waarden bevatten waarop statische ranking kan worden toegepast. Als u bijvoorbeeld opgeeft dat u de documentdatum als factor voor de ranking van documenten in de collectie wilt gebruiken en de documenten geen datumvelden of -attributen bevatten, kan de kwaliteit van de zoekopdracht worden verslechterd.

---

## Standaardwaarden voor statische documentranking herstellen

Als u een optie voor statische documentranking configureert tijdens het maken van een collectie, kunt u de standaardwaarden van de eigenschappen herstellen via de `runtime.properties`-bestanden voor de collectie.

### Voordat u begint

Om de standaardwaarden voor de documentranking van een document te herstellen, moet u enterprise search-beheerder zijn.

### Over deze taak

Als u de standaardwaarden voor de documentranking van een collectie wilt herstellen, moet u de `runtime.properties`-bestanden bijwerken voor de desbetreffende collectie en voor alle zoekservers in het enterprise search-systeem. In een configuratie met meerdere servers bevindt het bestand `runtime.properties` zich op de

indexserver in de directory `ES_NODE_ROOT/master_config/collectie_ID.runtime.node_ID`, waarbij *collectie\_ID* het ID is voor de collectie en *node\_ID* het ID is voor de zoekservers.

Als u bijvoorbeeld de collectie `coll` wilt bijwerken in een enterprise search-systeem met meerdere servers, werkt u het bestand `runtime.properties` bij voor de desbetreffende collectie en voor beide zoekservers (node3 en node4):

```
ES_NODE_ROOT/master_config/coll.runtime.node3/runtime.properties
ES_NODE_ROOT/master_config/coll.runtime.node4/runtime.properties
```

## Procedure

Ga als volgt te werk om de standaardwaarden van de documentranking te herstellen voor een collectie:

1. Meld u als enterprise search-beheerder aan bij de indexserver.
2. Geef het collectie-ID op voor de collectie waarvoor u de standaardrankingwaarden wilt herstellen. Het collectie-ID bevindt zich in het bestand `ES_NODE_ROOT/master_config/collections.ini`. Sorteert het bestand zodat dit beter leesbaar is. In het volgende voorbeeld is `coll` het collectie-ID:

```
% sort $ES_NODE_ROOT/master_config/collections.ini | more
collection1.configfile=coll_config.ini
collection1.datadir=/home/esearch/node/data/coll
collection1.description=
collection1.displayname=Collection1
collection1.flags=0
collection1.id=coll
collection1.sectiontype=collection
collection1.type=1
...
```

3. Open voor de collectie die u wilt herstellen het bestand `runtime.properties` in een editor en breng de volgende wijzigingen aan:
  - a. Wis de volgende eigenschappen:

```
trevi.autorank.dfthreshold1
trevi.autorank.dfthreshold2
trevi.autorank.dfthreshold3
trevi.autorank.rc0.*
trevi.autorank.rc1.*
```
  - b. Als het bestand `runtime.properties` de waarde `trevi.sourcetype=1` bevat, waarmee wordt aangegeven dat documenten worden gerangschikt op links, opent u het bestand `ES_INSTALL_ROOT/default_config/runtime.1/runtime.properties` in een editor en kopieert u de volgende standaard-eigenschappen naar het bestand `runtime.properties`:

```
trevi.autorank.dfthreshold1
trevi.autorank.dfthreshold2
trevi.autorank.dfthreshold3
trevi.autorank.rc0.*
trevi.autorank.rc1.*
```
  - c. Als het bestand `runtime.properties` de waarde `trevi.sourcetype=2` bevat, waarmee wordt aangegeven dat documenten worden gerangschikt op datum, opent u het bestand `ES_INSTALL_ROOT/default_config/runtime.2/runtime.properties` in een editor en kopieert u de volgende standaard-eigenschappen naar het bestand `runtime.properties`:

```
trevi.autorank.dfthreshold1
trevi.autorank.dfthreshold2
trevi.autorank.dfthreshold3
trevi.autorank.rc0.*
trevi.autorank.rc1.*
```

- d. Als het bestand `runtime.properties` de waarde `trevi.sourcetype=3` bevat, waarmee wordt aangegeven dat er geen statische rankingfactor wordt gebruikt voor de ranking van documenten in de collectie, opent u het bestand `ES_INSTALL_ROOT/default_config/runtime.0/runtime.properties` in een editor en kopieert u de volgende standaard eigenschappen naar het bestand `runtime.properties`:

```
trevi.autorank.dfthreshold1
trevi.autorank.dfthreshold2
trevi.autorank.dfthreshold3
trevi.autorank.rc0.*
trevi.autorank.rc1.*
```

4. Voor een configuratie met meerdere servers herhaalt u stap 3 op pagina 191 om het bestand `runtime.properties` bij te werken voor dezelfde collectie en de tweede zoekserver.
5. Open de beheerconsole, bewaak de pagina Zoeken en start de zoekprocessen voor deze collectie opnieuw.

Herhaal deze stappen indien nodig voor elke collectie waarvan u de standaardwaarden voor documentranking wilt herstellen.

---

## Aangepaste gewogen woordenboeken

Om de kwaliteit van de zoekresultaten te verbeteren, kunt u invloed uitoefenen op de manier waarop documenten worden gerangschikt in het zoekresultaat. Een van de manieren om dat te doen, is een aangepast gewogen woordenboek te maken.

Als in een zoekopdracht een woord voorkomt dat in een gewogen woordenboek staat, wordt het belang van documenten waarin dat woord voorkomt, verhoogd of verlaagd overeenkomstig de wegingsfactor die in het woordenboek voor dat woord is geconfigureerd.

De wegingsfactoren variëren van -10 tot 10. Tijdens de verwerking van de query verhogen de zoekservers het belang van documenten die woorden met een positieve wegingsfactor bevatten, en verlagen ze het belang van documenten waarin woorden met een negatieve wegingsfactor voorkomen.

Bijvoorbeeld: een document waarin zoektermen met een hoge wegingsfactor voorkomen, krijgt een hogere ranking dan wanneer die wegingsfactor niet zou worden toegepast. (De wegingsfactor is de enige factor die een rol speelt in de score van een document.)

Als u een woordenboek samenstelt, kunt u aan een willekeurig aantal woorden dezelfde wegingsfactor toekennen. Het woordenboek kan termen van één woord en van meerdere woorden bevatten. Termen van meerdere woorden worden opgevat als een woordcombinatie.

Als een woord waarvoor een wegingsfactor is opgegeven, wordt gebruikt in een query waarin de operator OR wordt gebruikt (bijvoorbeeld: `dit | dat`), wordt er voor de zoektermen een gewogen gemiddelde berekend. De resulterende totaalscore wordt gebruikt voor alle gevallen waarin de OR-operanden voorkomen. Er worden geen verschillende scores berekend voor verschillende OR-operanden.

Weging op basis van gewogen woordenboeken wordt niet ondersteund bij veldzoekopdrachten. Als de zoektermen door de parser worden geanalyseerd, wordt alleen de zoektekst (en dus niet de veldnaam) gebruikt voor het berekenen van de



score van het document. Om wegingsfactoren toe te passen op zoektermen die voorkomen in velden, wijst u de veldnamen toe aan wegingsklassen.

Om een gewogen woordenboek samen te stellen, moet een expert op het gebied van het onderwerp van de collectie een gewogen woordenlijst in XML-indeling opstellen (of dit doen in samenwerking met een programma-ontwikkelaar). Met behulp van een tool in enterprise search, esboosttermdictbuilder, kan het XML-bestand worden geconverteerd naar een binair bestand (.dic).

De enterprise search-beheerder uploadt het binaire bestand naar het systeem en kent er een weergavenaam aan toe. De collectiebeheerders kunnen dat gewogen woordenboek vervolgens selecteren voor gebruik bij het doorzoeken van de documenten in een collectie. Dit doen ze tijdens het configureren van de zoekserveropties voor die collectie.

**Beperking:** Als een aangepast gewogen woordenboek eenmaal in het systeem is opgenomen, kan het niet meer worden gewijzigd. Als u wijzigingen wilt aanbrengen in de gewogen woorden die beschikbaar zijn voor de verwerking van zoekopdrachten, moet u:

1. Het bron-XML-bestand bijwerken.
2. De XML-bron converteren naar een nieuw woordenboekbestand.
3. Het oude gewogen woordenboek verwijderen uit de collecties die er gebruik van maken.
4. Het oude gewogen woordenboek wissen van het systeem.
5. Het nieuwe gewogen woordenboek in het systeem opnemen.
6. Het nieuwe gewogen woordenboek koppelen aan de collecties die het moeten gaan gebruiken.

#### **Verwante onderwerpen**

“Ranking van documenten in enterprise search” op pagina 189

Wanneer een gebruiker een collectie zoekt, sturen de zoekprocessen de meest relevante resultaten terug voor de voorwaarden en bepalingen van de query.

Aangepaste gewogen woordenboeken

#### **Verwante taken**

Een XML-bestand voor gewogen woorden maken

Een gewogen woordenboek maken

## **Gewogen woordenboeken toevoegen aan het systeem**

Als u aangepaste gewogen woordenboeken samenstelt, moet u die woordenboeken opnemen in het enterprise search-systeem. Later kunt u dan kiezen welk gewogen woordenboek u wilt gebruiken voor het doorzoeken van een bepaalde collectie.

### **Voordat u begint**


Om gewogen woordenboeken toe te voegen aan het systeem, moet u de rol van beheerder van enterprise search hebben.

### **Beperkingen**

De maximumgrootte van een gewogen woordenboek is 8 MB.

### **Procedure**

U neemt als volgt een gewogen woordenboek op in het enterprise search-systeem:

1. Klik op **Systeem** om de view Systeem te openen.
2. Klik op  **Wijzigen** om naar de view Systeem wijzigen te gaan.
3. Ga naar de pagina Zoeken en klik op **Gewogen woordenboeken configureren**.
4. Klik op de pagina Gewogen woordenboeken configureren op **Gewogen woordenboek toevoegen**.
5. Geef op de pagina Gewogen woordenboek toevoegen een unieke weergavenaam voor het woordenboek op en typ desgewenst ook een beschrijving.
6. Geef de locatie van het dic-bestand op. Als het bestand is opgeslagen op het lokale systeem, kunt u op Bladeren klikken en het bestand opzoeken. Als het bestand is opgeslagen op de indexserver, moet u het volledige pad voor het bestand typen.
7. Klik op **OK**. Uw aangepaste gewogen woordenboek wordt toegevoegd aan het enterprise search-systeem en wordt beschikbaar gesteld voor het doorzoeken van collecties.

## Een gewogen woordenboek koppelen aan een collectie

Als er gewogen woordenboeken in het enterprise search-systeem zijn opgenomen, kunt u een van die woordenboeken selecteren om te gebruiken bij het doorzoeken van een collectie. Als een zoekterm gelijk is aan een term in het woordenboek, wordt het belang van documenten die die term bevatten, verhoogd of verlaagd overeenkomstig de wegingsfactor die in het woordenboek aan die term is toegewezen.

### Voordat u begint

Om een gewogen woordenboek voor een collectie te selecteren, moet u de rol van beheerder van enterprise search hebben of beheerder van een collectie zijn.

### Procedure

U kunt als volgt een gewogen woordenboek aan een collectie koppelen:

1. Open een collectie, ga naar de pagina Zoeken en klik op **Opties voor zoekserver configureren**.
2. Ga naar het veld **Naam gewogen woordenboek** op de pagina Opties voor zoekserver en selecteer het gewogen woordenboek dat moet worden gebruikt als gebruikers deze collectie doorzoeken.  
De lijst van beschikbare woordenboeken bevat alle gewogen woordenboeken die in het enterprise search-systeem zijn opgenomen.
3. Klik op **OK**.

---

## Documentranking op basis van URI-patronen

U kunt het belang van documenten vergroten of verkleinen door wegingsfactoren aan URI-patronen toe te wijzen.

Bij het toevoegen aan de index wordt aan alle documenten een standaard statische rankingscore toegewezen. De standaardscore is afhankelijk van de vraag of statische ranking voor de collectie is ingeschakeld en zo ja, welk type statische ranking is ingesteld (op documentdatum of, voor webdocumenten, het aantal andere documenten dat links naar dat document heeft).

U kunt het relatieve belang van een document beïnvloeden door wegingsfactoren toe te wijzen aan URI-patronen. De wegingsfactor wordt, samen met de standaard statische score en andere factoren, gebruikt om de uiteindelijke statische score van het document te bepalen.

De volgorde van de URI-patronen die u configureert, is van belang. Op de indexserver worden de URI-patronen geëvalueerd in de volgorde waarin ze worden weergegeven tijdens het berekenen van de waarde van elk document in een collectie. Voor elke URI geldt het volgende:

1. De URI-patronen worden op de indexserver opeenvolgend gescand.
2. Wanneer de indexserver het eerste URI-patroon vindt dat overeenkomt met een document in de index, wordt op dat document de wegingsfactor toegepast die voor dat URI-patroon is geconfigureerd.
3. Als voor een document geen overeenkomst met een URI-patroon wordt gevonden, wordt de standaard statische rangingscore gebruikt.

Nadat u een wegingsfactor voor een URI-patroon hebt geconfigureerd, moet u het URI-patroon in de gewenste scanvolgorde voor de indexserver plaatsen.

#### **Verwante onderwerpen**

“Ranking van documenten in enterprise search” op pagina 189

Wanneer een gebruiker een collectie zoekt, sturen de zoekprocessen de meest relevante resultaten terug voor de voorwaarden en bepalingen van de query.

## **De scores beïnvloeden van documenten die aan URI-patronen voldoen**

U kunt het belang van documenten die aan een URI-patroon voldoen, vergroten of verkleinen door een wegingsfactor aan de standaard statische rangingscore toe te wijzen.

#### **Voordat u begint**

Om het belang te beïnvloeden van documenten die voldoen aan een URI-patroon, moet u de rol van beheerder van enterprise search hebben of beheerder van een collectie zijn.

#### **Over deze taak**

De wegingsfactor die u configureert, wordt met de standaard statische rangingscore gebruikt om een nieuwe statische score te berekenen voor alle documenten die aan het opgegeven URI-patroon voldoen.

De wegingsfactor is alleen van invloed op statische scores en vormt slechts één van de onderdelen in de berekening van de uiteindelijke ranking van een document. Als er naar een bepaald document bijvoorbeeld heel veel links zijn (hetgeen resulteert in een hoge aanvangsscore), krijgt een document waarnaar helemaal geen links verwijzen altijd een lagere ranking.

#### **Procedure**

Ga als volgt te werk om de scores te beïnvloeden van documenten die aan een URI-patroon voldoen:

1. Open een collectie, selecteer de pagina Index en klik vervolgens op **Scores beïnvloeden door matching van URI-patronen**.



2. Klik op de pagina Scores beïnvloeden door matching van URI-patronen op **URI-patroon toevoegen**.
3. Geef een URI-patroon op voor documenten waarvoor u het belang in de zoekresultaten wilt vergroten of verkleinen. Bijvoorbeeld:

```
http://domain.org/hr/*
db2://*ROI*
*/afs/*
```

4. Voer een waarde tussen -10 en 10 in voor de wegingsfactor. De uiteindelijke statische score voor alle documenten die voldoen aan het URI-patroon wordt berekend op basis van deze wegingsfactor.
5. Klik op **OK**.
6. Ga naar de pagina Scores beïnvloeden door matching van URI-patronen en plaats het nieuwe URI-patroon in de volgorde waarin dit door de indexserver moet worden gescand.

De indexserver berekent de statische rankingscores in de volgorde waarin u de URI's plaatst. De beste resultaten verkrijgt u door de meer specifieke URI's bovenaan in de lijst te zetten. In het volgende voorbeeld voldoet de subdirectory /forms aan het URI-patroon `http://www.ibm.com/hr/*`. Om te zorgen dat de score van documenten in de subdirectory /forms correct wordt berekend, moet u het URI-patroon voor de subdirectory /forms subdirectory het eerst noemen:

```
http://www.ibm.com/hr/forms/* 8
http://www.ibm.com/hr/* -2
```

7. Als u het URI-patroon of de wegingsfactor wilt wijzigen, selecteert u het URI-patroon en klikt u op  **Bewerken**.
8. Als u een URI-patroon uit de lijst wilt verwijderen, selecteert u het URI-patroon en klikt u op  **Verwijderen**.
9. Als u de wegingsfactoren wilt toepassen op documenten die eerder zijn geïndexeerd, moet u de hoofdindex opnieuw opbouwen.

#### **Verwante onderwerpen**

"Ranking van documenten in enterprise search" op pagina 189

Wanneer een gebruiker een collectie zoekt, sturen de zoekprocessen de meest relevante resultaten terug voor de voorwaarden en bepalingen van de query.

#### **Verwante verwijzing**

"URI-indelingen in een enterprise search-index" op pagina 111

Met de URI (Uniform Resource Identifier) van documenten in een enterprise search-index wordt het type crawler aangegeven waarmee het document aan de collectie is toegevoegd.

---

## **Documentranking op basis van wegingsklassen**

Als u velden koppelt aan wegingsklassen, kunt u op die manier invloed uitoefenen op de manier waarop documenten worden gerangschikt in het zoekresultaat.

Bij het analyseren of "parseren" van documenten wijst de parser *wegingsklassen* toe aan documenttokens, op basis van de velden waartoe die tokens behoren. Deze wegingsklassen zijn opgenomen in de index en worden tijdens de evaluatie van de query gebruikt om scores te berekenen aan de hand waarvan documenten een ranking krijgen.

Om invloed uit te oefenen op de manier waarop die scores worden berekend, kunt u numerieke wegingsfactoren voor de wegingsklassen opgeven. Als een zoekterm overeenkomt met een token in een veld dat is toegewezen aan een dergelijke

wegingsklasse, wordt de totaalscore van het document beïnvloed door de bijdrage van het feit dat dit token aanwezig is. De score wordt berekend middels het toepassen van de wegingsfactor die voor de wegingsklasse is geconfigureerd.

U zou bijvoorbeeld de scores van titelvelden een hoger belang kunnen geven; als de term die u zoekt dan voorkomt in de titel van een document, krijgt dat document een hogere plaats in de zoekresultaten dan een document waarin de term niet in de titel staat.

Om de ranking van documenten te beïnvloeden, moet u in de beheerconsole van enterprise search wegingsfactoren voor wegingsklassen opgeven en velden koppelen of toewijzen aan wegingsklassen. Er zijn voor enterprise search zestien wegingsklassen vooraf geconfigureerd. Acht van die wegingsklassen zijn bedoeld voor contentvelden; de andere acht zijn bedoeld voor velden met metagegevens. U kunt de scores voor de standaardwegingsklassen aanpassen en u kunt die wegingsklassen toewijzen of koppelen aan andere of aanvullende velden.

Als u de veldtoewijzingen wijzigt, moet u de documenten opnieuw veldtoewijzingen en analyseren, zodat de wijzigingen kunnen worden toegepast op documenten die voorheen zijn geïndexeerd. Als u de voor een wegingsklasse opgegeven factoren wijzigt, moet u de zoekservers bewaken en de processen van de zoekservers stoppen en opnieuw starten om de wijzigingen te activeren.

## **Detectie van duplicaatdocumenten en documentsamenvattingen**

Wanneer u een veld aan een wegingsklasse toewijst, moet u opgeven of het veld wordt gebruikt om duplicaatdocumenten te detecteren en of de content van het veld in de documentsamenvattingen in de zoekresultaten kan worden opgenomen.

- Als een veld wordt gebruikt voor het detecteren van duplicaatdocumenten, wordt dat veld beschouwd als een contentveld en kunnen er alleen wegingsklassen worden geselecteerd die bedoeld zijn voor contentvelden. De content van deze veldtypen kan worden gebruikt in dynamische documentsamenvattingen in de zoekresultaten.
- Als een veld niet wordt gebruikt voor het detecteren van duplicaatdocumenten, wordt dat veld beschouwd als een veld voor metagegevens en kunnen er alleen wegingsklassen worden geselecteerd die bedoeld zijn voor metagegevensvelden. In dit geval worden twee documenten die in alle opzichten behalve het opgegeven veld aan elkaar gelijk zijn, beschouwd als duplicaten en wordt het veld niet gebruikt in dynamische documentsamenvattingen.

## **Hoge en lage terughaalwaarden**

Bij het evalueren van een query maakt het zoekproces een schatting van het aantal resultaatdocumenten dat zal worden geretourneerd. Aan de hand van drempelwaarden wordt vastgesteld of een query wordt geacht een lage of een hoge terughaalwaarde te hebben of een waarde die tussen de lage en hoge waarde ligt:

### **Lage terughaalwaarde**

Als het geschatte aantal documenten in het resultaat onder de lage drempelwaarde ligt, wordt de query beschouwd als een query met een lage terughaalwaarde.

### **Hoge terughaalwaarde**

Als het geschatte aantal documenten in het resultaat boven de hoge drempelwaarde ligt, wordt de query beschouwd als een query met een hoge terughaalwaarde.

### **Gecombineerde terughaalwaarde**

Als het geschatte aantal documenten tussen de twee drempelwaarden ligt, is de terughaalwaarde van de query een combinatie van de twee drempelwaarden.

In elke wegingsklasse zijn wegingsfactoren vastgelegd die tijdens de verwerking van de query worden gekoppeld aan query's met een lage en query's met een hoge terughaalwaarde. De lage wegingsfactor is van invloed op het relatieve belang van query's met een lage terughaalwaarde, de hoge wegingsfactor op query's met een hoge terughaalwaarde. Query's met een gemengde terughaalwaarde worden beïnvloed door een combinatie van deze twee wegingsfactoren.

De waarden van de wegingsfactoren bepalen het relatieve belang van het voorkomen van een zoekterm in een document. Elke keer dat een zoekterm in een document wordt gevonden, wordt dat geteld aan de hand van de overeenkomstige wegingsfactor.

Als u wegingsklassen voor een collectie opgeeft, kunt u de standaardwegingsfactoren wijzigen. U zou bijvoorbeeld wegingsfactoren kunnen opgeven om te zorgen dat zoektermen die in titelvelden voorkomen, vijf keer zo zwaar tellen als zoektermen die worden gevonden in de lopende tekst.

#### **Verwante onderwerpen**

“Ranking van documenten in enterprise search” op pagina 189

Wanneer een gebruiker een collectie zoekt, sturen de zoekprocessen de meest relevante resultaten terug voor de voorwaarden en bepalingen van de query.

## **Velden toewijzen aan wegingsklassen**

U kunt het relatieve belang van velden beïnvloeden door velden toe te wijzen aan wegingsklassen.

### **Voordat u begint**

Om velden toe te wijzen aan wegingsklassen, moet u de rol van beheerder van enterprise search hebben of beheerder van een collectie zijn.

### **Over deze taak**

Het systeem gebruikt de wegingsfactor om de ranking te beïnvloeden van documenten waarin zoektermen voorkomen in de velden die zijn toegewezen aan wegingsklassen.

In enterprise search zijn bepaalde toewijzingen gereserveerd voor interne velden en standaardtekst waarvoor geen andere kenmerken zijn gedefinieerd. U kunt wel andere velden toewijzen aan de wegingsklassen die door de gereserveerde velden worden gebruikt, maar u kunt de gereserveerde velden niet bewerken of verwijderen.

### **Procedure**

U kunt als volgt velden toewijzen aan wegingsklassen:

1. Open een collectie, ga naar de pagina Parser en klik op **Velden toewijzen aan wegingsklassen**.
2. Klik op de pagina Velden toewijzen aan wegingsklassen op **Veld toevoegen**.

3. Geef op de pagina Veld toevoegen aan een wegingsklasse de naam op van het veld dat u aan de wegingsklasse wilt toewijzen.  
U kunt de naam opgeven van een veld dat aanwezig is in een gecrawlde bron of in een externe bron, de naam van een veld dat is toegewezen vanuit een XML-element of vanuit een HTML metagegevens-element, of een van de vooraf gedefinieerde veldnamen.
4. Geef aan of het veld wordt gebruikt voor het detecteren van duplicaatdocumenten. Als u dit vakje selecteert, bevat de lijst van beschikbare wegingsklassen alleen nog klassen die van toepassing zijn op contentvelden. Als een document met dit veld in de zoekresultaten wordt geretourneerd, wordt de inhoud van dit veld afgebeeld in het gebied met de documentsamenvatting. Maakt u dit vakje leeg, dan bevat de lijst van beschikbare wegingsklassen alleen nog klassen die van toepassing zijn op metagegevensvelden. De content van het veld wordt niet afgebeeld in het gebied met de documentsamenvatting in de zoekresultaten.
5. Selecteer een wegingsklasse en klik op **OK**.  
Het veld dat u hebt toegevoegd, wordt afgebeeld op de pagina Velden toewijzen aan wegingsklassen. U kunt een optie selecteren om de wegingsklasse te wijzigen en andere wegingsfactoren te configureren voor het vaststellen van de score voor documenten die dit veld bevatten.
6. Om de instelling die aangeeft of een veld wordt gebruikt voor het detecteren van duplicaatdocumenten te wijzigen of om een veld toe te wijzen aan een andere wegingsklasse, klikt u op  **Wijzigen**. (Velden die zijn gereserveerd voor gebruik door enterprise search, kunnen niet worden gewijzigd.)
7. Om een veld uit een wegingsklasse te verwijderen, klikt u op  **Verwijderen**. (Velden die zijn gereserveerd voor gebruik door enterprise search, kunnen niet worden verwijderd.)
8. Om de wijzigingen door te voeren voor documenten die al eerder zijn geïndexeerd, moet u de documenten opnieuw crawlen en indexeren.

#### **Verwante onderwerpen**

“Ranking van documenten in enterprise search” op pagina 189

Wanneer een gebruiker een collectie zoekt, sturen de zoekprocessen de meest relevante resultaten terug voor de voorwaarden en bepalingen van de query.

## **Wegingsfactor voor wegingsklassen configureren**

De wegingsfactoren die u voor wegingsklassen configureert, geven aan hoe u het belang inschat van bepaalde velden in de documenten die bij een query worden gevonden. Wegingsklassen met een hoge wegingsfactor kunnen het belang verhogen van gevonden documenten die velden bevatten welke aan die wegingsklasse zijn toegewezen.

### **Voordat u begint**


Om wegingsfactoren voor wegingsklassen te configureren, moet u de rol van beheerder van enterprise search hebben of beheerder van een collectie zijn.

### **Over deze taak**

Het systeem maakt voor het berekenen van een nieuwe score voor gevonden documenten die velden bevatten welke aan een wegingsklasse zijn gekoppeld, gebruik van de wegingsfactoren die voor een wegingsklasse zijn geconfigureerd, de standaardscore voor statische ranking en andere factoren.

## Procedure

U kunt als volgt wegingsfactoren voor een wegingsklasse configureren:

1. Open een collectie, ga naar de pagina Parser en klik op **Velden toewijzen aan wegingsklassen**.
2. Klik op de pagina Velden toewijzen aan wegingsklassen op **Wegingsklassen wijzigen**.
3. Kijk op de pagina Wegingsklassen waar de wegingsklasse zich bevindt die u wilt wijzigen en klik op  **Bewerken**.
4. Geef op de pagina Wegingsklasse wijzigen de nieuwe waarden op voor de hoge en lage wegingsfactoren. U kunt ook voor beide factoren dezelfde waarde opgeven.
5. Klik op **OK**.
6. Om de wijzigingen te activeren, moet u de zoekservers bewaken en op de pictogrammen klikken voor het stoppen en opnieuw starten van de zoekprocessen. Als gebruikers query's indienen, wordt het relatieve belang van de gevonden documenten die velden bevatten welke zijn toegewezen aan wegingsklassen, nu bepaald door de nieuwe wegingsfactoren.

### Verwante onderwerpen

"Ranking van documenten in enterprise search" op pagina 189

Wanneer een gebruiker een collectie zoekt, sturen de zoekprocessen de meest relevante resultaten terug voor de voorwaarden en bepalingen van de query.

## Standaardwaarden voor wegingsklassen

Enterprise search bevat zestien wegingsklassen die u kunt gebruiken om de ranking van documenten in de zoekresultaten te beïnvloeden.

Om scores te berekenen voor velden en tekst die geen andere bepalende kenmerken hebben, zijn de volgende velden gereserveerd voor gebruikt door enterprise search:

```
es_special_field.regular_text  
es_special_field.default_fieldes_special_field.default_metadata_field
```

U kunt wel andere velden toewijzen aan de wegingsklassen die door de gereserveerde velden worden gebruikt, maar u kunt de gereserveerde velden niet bewerken of verwijderen.

Voor alle andere velden geldt dat u wijzigingen kunt aanbrengen in de wegingsfactor die door het systeem wordt gebruikt om de ranking van een document te berekenen. U kunt tevens een willekeurig aantal velden toewijzen aan elk van de wegingsklassen, inclusief de wegingsklassen die worden gebruikt voor gereserveerde velden.

In de volgende tabel ziet u een overzicht van de namen van de wegingsklassen, de standaardwegingsfactoren voor query's met een lage respectievelijk hoge terughaalwaarde en de namen van de vooraf gedefinieerde velden die in de standaardconfiguratie aan de wegingsklassen zijn toegewezen.

De standaardwegingsfactoren variëren afhankelijk van de statische rankingmethode die voor de collectie is geselecteerd toen de collectie werd gemaakt. De mogelijke opties zijn: geen statische ranking, statische ranking op basis van het aantal links naar een document (voor internetbronnen) en een ranking die wordt bepaald door de documentdatum.



Tabel 6. Standaardwaarden voor wegingsklassen

Standaard lage en hoge wegingsfactoren				
Naam wegings-klasse	Geen statische ranking	Documentlinks	Document-datum	Vooraf gedefinieerde veldtoewijzingen
Contentklasse A	Laag: 4 Hoog: 2	Laag: 6 Hoog: 1	Laag: 4 Hoog: 2	es_special_field.regular_text
Contentklasse B	Laag: 5 Hoog: 4	Laag: 7 Hoog: 3	Laag: 5 Hoog: 4	es_special_field.html_emphasized_text  Bevat de volgende HTML-elementen: b, big, caption, dfn, em, h4, h5, h6, strong
Contentklasse C	Laag: 7 Hoog: 4	Laag: 9 Hoog: 3	Laag: 7 Hoog: 4	es_special_field.html_headers  Bevat de volgende HTML-elementen: h1, h2, h3
Contentklasse D	Laag: 2 Hoog: 5	Laag: 1 Hoog: 5	Laag: 2 Hoog: 5	titel
Contentklasse E	Laag: 1 Hoog: 1	Laag: 5 Hoog: 10	Laag: 1 Hoog: 1	es_special_field.anchor
Contentklasse F	Laag: 1 Hoog: 1	Laag: 1 Hoog: 1	Laag: 1 Hoog: 1	es_special_field.anchor_same_dir
Contentklasse G	Laag: 1 Hoog: 1	Laag: 1 Hoog: 1	Laag: 1 Hoog: 1	es_special_field.anchor_same_host
Contentklasse H	Laag: 1 Hoog: 1	Laag: 1 Hoog: 1	Laag: 1 Hoog: 1	es_special_field.default_field
Metagegevensklasse A	Laag: 1 Hoog: 1	Laag: 1 Hoog: 1	Laag: 1 Hoog: 1	es_special_field.default_metadata_field
Metagegevensklasse B	Laag: 1 Hoog: 1	Laag: 1 Hoog: 1	Laag: 1 Hoog: 1	
Metagegevensklasse C	Laag: 1 Hoog: 1	Laag: 1 Hoog: 1	Laag: 1 Hoog: 1	
Metagegevensklasse D	Laag: 1 Hoog: 1	Laag: 1 Hoog: 1	Laag: 1 Hoog: 1	
Metagegevensklasse E	Laag: 1 Hoog: 1	Laag: 5 Hoog: 1	Laag: 1 Hoog: 1	trefwoorden
Metagegevensklasse F	Laag: 1 Hoog: 1	Laag: 3 Hoog: 1	Laag: 1 Hoog: 1	es_special_field.urlhost
Metagegevensklasse G	Laag: 1 Hoog: 1	Laag: 1 Hoog: 1	Laag: 1 Hoog: 1	es_special_field.urlpath
Metagegevensklasse H	Laag: 1 Hoog: 1	Laag: 1 Hoog: 1	Laag: 1 Hoog: 1	description

### Verwante onderwerpen

“Ranking van documenten in enterprise search” op pagina 189

Wanneer een gebruiker een collectie zoekt, sturen de zoekprocessen de meest relevante resultaten terug voor de voorwaarden en bepalingen van de query.



---

## Zoekprogramma's voor enterprise search

Met behulp van een zoekprogramma kunt u collecties en externe bronnen in het enterprise search-systeem doorzoeken. U kunt een onbeperkt aantal zoekprogramma's maken en met elk programma kunt u een willekeurig aantal collecties en externe bronnen doorzoeken.

### voorbeeldzoekprogramma

In het voorbeeldzoekprogramma vindt u een demonstratie van de meeste zoek- en ophaalfuncties die in enterprise search beschikbaar zijn. Daarnaast wordt in het voorbeeldprogramma geïllustreerd hoe u de IBM Search and Index API (SI-API) kunt gebruiken om interactieve, aangepaste zoekprogramma's te maken waarmee de doelen van uw onderneming worden aangegeven.

Tenzij er wijzigingen zijn aangebracht in het standaardconfiguratiebestand, kunt u met het standaardzoekprogramma alle collecties en externe bronnen in het enterprise search-systeem doorzoeken. U kunt dit programma gebruiken om nieuwe collecties en externe bronnen te testen voordat u ze beschikbaar stelt aan uw gebruikers.

Het voorbeeldzoekprogramma is automatisch aan alle collecties en externe bronnen gekoppeld. In een productieomgeving bepaalt de beheerder van enterprise search met welke zoekprogramma's een bepaalde collectie kan worden doorzocht.

### Aangepaste zoekprogramma's

U kunt de zoekprogramma's die u maakt als zelfstandige webprogramma's in een IBM WebSphere Application Server-omgeving uitvoeren of de programma's als portlets in een IBM WebSphere Portal-omgeving uitvoeren. Als u de Search and Index API gebruikt, kunt u zoekprogramma's ontwerpen die, net zoals het voorbeeldzoekprogramma, naadloos in beide omgevingen kunnen worden uitgevoerd.

Voor hulp bij het aanpassen van zoekprogramma's kunt u het Aanpassingstool Zoekprogramma gebruiken. Met dit tool kunt u opties selecteren via een grafische interface en wijzigingen bekijken terwijl u deze aanbrengt. Als u de wijzigingen opslaat, wordt het configuratiebestand voor het zoekprogramma bijgewerkt.

#### Verwante onderwerpen

"Geïndexeerde opties voor het doorzoeken van documenten" op pagina 164  
Tijdens het configureren van opties voor het doorzoeken van gecrawelde gegevens of tijdens het toewijzen van XML- en HTML-metagegeveenselementen aan zoekvelden kunt u aangeven hoe documenten kunnen worden doorzocht en hoe documenten in de zoekresultaten worden afgebeeld.

"Zoekprogramma-ID's" op pagina 236

Om in te stellen dat verschillende collecties kunnen worden doorzocht, wijst u de zoekprogramma's toe aan de collecties en externe bronnen die door de zoekprogramma's kunnen worden doorzocht. Met het programma Default kunt u het voorbeeldzoekprogramma gebruiken om alle collecties en externe bronnen te doorzoeken.

"Beveiliging op documentniveau" op pagina 237

Als bij het maken van de collectie de beveiliging is ingeschakeld, kunt u

beveiligingsopties op documentniveau configureren. Als u beveiligingsopties op documentniveau configureert, zorgt u dat in de zoekresultaten alleen de documenten worden afgebeeld waarvoor de gebruiker die de zoekopdracht heeft ingediend, weergavemachtigingen heeft.

Overzicht van de Search and Index API

Syntaxis voor zoekopdrachten

---

## Zoekprogramma's aan collecties koppelen

U kunt uw nieuwe zoekprogramma pas gebruiken als u dit hebt gekoppeld aan de collecties die met het programma kunnen worden doorzocht.

### Voordat u begint

Om zoekprogramma's te koppelen aan de collecties die met de programma's kunnen worden doorzocht, moet u de rol van beheerder van enterprise search hebben.

### Procedure

Ga als volgt te werk om een zoekprogramma aan een of meer collecties te koppelen:

1. Klik op **Beveiliging** in de werkbalk van de beheerconsole.
2. Klik op de pagina Zoekprogramma's op **Zoekprogramma's configureren**.
3. Klik op de pagina Zoekprogramma's configureren op **Zoekprogramma toevoegen**.
4. Typ de naam van het zoekprogramma.
5. Selecteer de collecties die met het programma kunnen worden doorzocht:
  - Klik op **Alle collecties en externe bronnen** als het zoekprogramma toegang moet krijgen tot alle collecties die u aan het systeem hebt toegevoegd.
  - Klik op **Bepaalde collecties en externe bronnen** als het zoekprogramma alleen toegang moet krijgen tot de collecties die u opgeeft.  
Als u deze optie selecteert, wordt er een lijst met collecties en externe bronnen afgebeeld. Schakel het aankruisvakje **Selecteren** in voor elke collectie die het zoekprogramma kan doorzoeken.
6. Klik op **OK**.

---

## Functies van het voorbeeldzoekprogramma

Het voorbeeldzoekprogramma voor enterprise search bevat een demonstratie van de meeste zoekfuncties die u in uw aangepaste zoekprogramma's kunt integreren.

U kunt het voorbeeldzoekprogramma gebruiken om alle collecties en externe bronnen tegelijkertijd te doorzoeken. Tenzij de standaardprogramma-eigenschappen zijn gewijzigd, kunt u dit programma gebruiken om alle collecties en externe bronnen in het enterprise search-systeem te doorzoeken.

### Queryfuncties

Met deze functies kunt u de volgende taken uitvoeren:

- Eenvoudige query's in de vorm van vrije tekst opgeven.
- Complexere query's opgeven om nauwkeurigere zoekresultaten weer te geven.  
U kunt bijvoorbeeld specifieke velden of XML-elementen zoeken of de

querysyntaxis gebruiken om te zoeken naar documenten waarin bepaalde woorden en woordcombinaties zijn opgenomen of hiervan zijn uitgesloten.

- Aangeven welke collecties en externe bronnen moeten worden doorzocht.
- Bepaalde brontypen of alle brontypen doorzoeken.
- Bepaalde typen documenten zoeken. U kunt bijvoorbeeld alleen zoeken naar Microsoft Word- of PDF-documenten (Portable Document Format).
- De taal voor de zoektermen opgeven. Daarnaast kunt u de talen opgeven van de documenten waarin u wilt zoeken.
- Bepaalde subsets van een collectie doorzoeken. U kunt met een zoekprogramma bijvoorbeeld alleen zoeken in een vooraf gedefinieerd bereik van documenten (een scope) of u kunt een query opgeven waarmee alleen de documenten worden doorzocht die deel uitmaken van een benoemde categorie.
- U kunt de query uitbreiden en synoniemen van de zoektermen opnemen. Als aan de collectie een synoniemenwoordenboek is gekoppeld, worden de documenten waarin de synoniemen van de zoektermen voorkomen, in de zoekresultaten weergegeven.

## **Functies voor zoekresultaten**

Met deze functies kunt u de volgende taken uitvoeren:

- De zoekresultaten bekijken die overeenkomen met de zoektermen.
- Bepalen hoeveel resultaatdocumenten op elke pagina worden afgebeeld, en vooruit en achteruit door de resultaatset bladeren.
- Details over de resultaatdocumenten verbergen en weergeven. U kunt bijvoorbeeld korte beschrijvingen of details van de documenten in de resultaatdocumenten weergeven, zoals de namen van velden.
- Documenten van dezelfde bron samenvouwen. Als een bron bijvoorbeeld 100 documenten als resultaat geeft, worden de twee meest relevante documenten samen in de resultaatset gegroepeerd. De overige 98 documenten kunt u bekijken door een optie te selecteren waarmee u meer documenten van dezelfde bron kunt weergeven.
- Documenten sorteren op relevantie, op documentdatum of op de waarden in een bepaald veld. Als u de documenten sorteert op datum of veld, kunt u opgeven of u de resultaten in oplopende volgorde of in aflopende volgorde wilt weergeven.
- Suggesties weergeven voor spellingcorrecties als er mogelijk onjuist gespelde woorden in de queryreeks worden aangetroffen.
- Informatie bekijken over de categorieën waarvan een resultaatdocument deel uitmaakt (als de collectie gebruikmaakt van categorieën) en alleen door de documenten bladeren die deel uitmaken van een bepaalde categorie.
- Extra zoektermen opgeven om binnen de zoekresultaten te zoeken.

## **Functies voor het ophalen van documenten**

Met deze functies kunt u de volgende taken uitvoeren:

- Documenten ophalen door op de URI van een document te klikken zodat het document in een webbrowser wordt geopend. Als een Notes- of Domino Document Manager-crawler is geconfigureerd voor het gebruik van het DIOP-protocol, kunnen de documenten die door deze crawlers worden gecrawld, worden bekeken in de Lotus Notes-clientviewer in plaats van een webbrowser.

Als beveiliging op documentniveau voor een crawler is geconfigureerd, kunnen alleen gebruikers die zijn gemachtigd om de beveiligde content te openen, documenten ophalen.

- Documenten ophalen met behulp van snellinks. Met een snellink worden bepaalde trefwoorden aan URI's gekoppeld. Als een query de opgegeven trefwoorden bevat, verschijnen de gekoppelde URI's (waarvan vooraf is vastgesteld dat ze relevant zijn voor de desbetreffende trefwoorden) boven aan de zoekresultaten.

---

## Eigenschappen van zoekprogramma

U kunt het configuratiebestand voor een zoekprogramma bewerken om opties op te geven voor uw omgeving, de weergave van het programma aan te passen en vast te stellen welke opties beschikbaar zijn voor gebruikers na het starten van het zoekprogramma.

Daarnaast kunt u eigenschappen bewerken met behulp van het Aanpassingstool Zoekprogramma. In dit programma kunt u opties selecteren, waarna het effect van de wijzigingen direct zichtbaar zijn. Als u tevreden bent met de opgegeven opties voor het doorzoeken van collecties en de weergave van de zoekresultaten, kunt u de opties opslaan zodat het configuratiebestand voor het zoekprogramma wordt bijgewerkt.

Het configuratiebestand voor het voorbeeldzoekprogramma voor enterprise search is het bestand `config.properties`. In dit onderwerp vindt u een beschrijving van de eigenschappen in dit bestand en worden de standaardeigenschappen beschreven. Als u configuratiebestanden voor uw aangepaste zoekprogramma's maakt, kunnen de eigenschappen in deze bestanden en de waarden die voor deze eigenschappen zijn opgegeven, afwijken.

### Omgevingsparameters

U kunt opties opgeven die van invloed zijn op de werking van het zoekprogramma.

#### **applicationName**

Hiermee wordt de naam van een geldig zoekprogramma aangegeven. De standaardwaarde is `Standaard`.

Verander de standaardwaarde als u een ander programma als zoekprogramma wilt gebruiken.

**Tip:** Als de naam van het zoekprogramma `Standaard` is, kunt u het voorbeeldzoekprogramma gebruiken om alle collecties en externe bronnen met een enkele query te doorzoeken.

#### **timeout**

Hiermee wordt het aantal seconden aangegeven dat moet worden gewacht op antwoord van de zoekserver voordat een timeout voor de zoekopdracht optreedt. Dit getal moet een geheel getal zijn (zoals 60, niet 60,5 of zestig). Als u geen waarde voor de timeout opgeeft, wordt er een standaardwaarde van 30 seconden gehanteerd.

#### **hostname**

Hiermee wordt de volledige hostnaam van de webserver aangegeven die is geconfigureerd voor ondersteuning van uw WebSphere Application Server-`instance`. De standaardwaarde is `localhost`.

Om te zorgen dat het zoekprogramma correct werkt, dient u de standaardwaarde te veranderen in de volledige hostnaam die WebSphere Application Server volgens de configuratie gebruikt. Bijvoorbeeld: als de hostnaam van de lokale computer MijnMachine is en de hostnaam van de webserver www.ibm.com, dan geeft u `www.ibm.com` op.

**protocol**

Hiermee wordt het protocol aangegeven voor communicatie met de webserver: `http` of `https`. Als deze eigenschap leeg blijft, wordt de standaardwaarde `http` gebruikt.

**port**

Hiermee wordt het poortnummer aangegeven van de webserver die is geconfigureerd voor ondersteuning van uw WebSphere Application Server-*instance*. De standaardwaarde is 80, hetgeen gebruikelijk is als het protocol HTTP wordt gebruikt. De poort die meestal voor het protocol HTTPS wordt gebruikt, is 443.

**trustStore**

Als u het protocol HTTPS gebruikt, geeft u het volledige pad op voor het sleutelruimtebestand (het databasebestand dat de openbare sleutels bevat). Dit bestand wordt ook wel *truststore* genoemd en maakt het mogelijk dat het SSL-protocol (Secure Sockets Layer) wordt gebruikt voor betrouwbare communicatie. Als u een Windows-pad wilt opgeven, plaatst u voor de schuine streep naar links een tweede schuine streep naar links. Bijvoorbeeld: `x:\Application Server\webserver.key`

**trustPassword**

Als u het protocol HTTPS gebruikt, geeft u een wachtwoord op voor het opgegeven sleutelruimtebestand.

**username**

Voor deze waarde wordt in het zoekprogramma automatisch de gebruikersnaam gebruikt die de gebruiker heeft opgegeven tijdens het aanmelden bij het zoekprogramma. Geef hier alleen een gebruikersnaam op als u het standaardgedrag voor de verificatie van gebruikers wilt wijzigen. Dit veld wordt alleen gebruikt als global security in WebSphere Application Server is ingeschakeld.

**password**

Voor deze waarde wordt in het zoekprogramma automatisch het wachtwoord gebruikt dat de gebruiker heeft opgegeven tijdens het aanmelden bij het zoekprogramma. Voer hier alleen een wachtwoord in als u een gebruikersnaam hebt opgegeven. Dit veld wordt alleen gebruikt als global security in WebSphere Application Server is ingeschakeld.

**ssoCookieName**

Hiermee wordt de naam aangegeven van de cookie die de SSO-tekenreeks (Single Sign-On) bevat. De standaardwaarde is `LtpaToken`.

**proxyHost**

Hiermee wordt de volledige hostnaam van een proxyserver aangegeven als een proxyserver vereist is voor toegang tot de zoekserver.

**proxyPort**

Hiermee wordt de poort voor de opgegeven proxyserverhost aangegeven.

**proxyUser**

Hiermee wordt de gebruikersnaam aangegeven die wordt gebruikt voor aanmelding bij de proxyserver als de proxyserver Basic Authentication vereist.

**proxyPassword**

Hiermee wordt het wachtwoord aangegeven voor de opgegeven gebruikersnaam voor de proxyserver.

**filter** Hiermee wordt de klasse aangegeven die wordt gebruikt om de documenten op te halen die in de zoekresultaten worden afgebeeld. De standaardklasse is `com.ibm.es.api.filters.SetDocumentURIFilterFetch`. Wijzig deze waarde alleen als u een aangepaste klasse hebt die u voor het ophalen van documenten wilt gebruiken.

**logging.level**

Hiermee wordt het detailniveau aangegeven voor het loggen van informatie:

**OFF** Er worden geen berichten gelogd.

**SEVERE**

Berichten die duiden op een ernstig probleem worden gelogd. Dit is de standaardwaarde.

**INFO** De informatieberichten worden gelogd.

**FINE** Er worden traceerberichten met weinig details gelogd. (Deze optie komt overeen met het logniveau FINE in de Java-klasse `java.util.logging.Level`.)

**ALL** Alle berichten worden gelogd.

## Pictogrammen voor brontypen

U kunt de afbeeldingen aanpassen die het type gegevensbron aangeven waarvan een resultaatdocument deel uitmaakt. De volgende pictogrammen voor brontypen (waarmee de crawlers en externe bronnen worden aangegeven die worden ondersteund wanneer WebSphere Information Integrator OmniFind Edition is geïnstalleerd) zijn vooraf gedefinieerd in het bestand `config.properties`.

**documentSource.vbr.icon**

Het pad en de naam van een afbeeldingsbestand waarmee wordt aangegeven dat het document is gecrawld door een Content Edition-crawler. Het standaardpictogram is `/images/sourceVBR.gif`.

**documentSource.db2.icon**

Het pad en de naam van een afbeeldingsbestand waarmee wordt aangegeven dat het document is gecrawld door een DB2-crawler. Het standaardpictogram is `/images/sourceDB2.gif`.

**documentSource.cm.icon**

Het pad en de naam van een afbeeldingsbestand waarmee wordt aangegeven dat het document is gecrawld door een DB2 Content Manager-crawler. Het standaardpictogram is `/images/sourceCM.gif`.

**documentSource.dominodoc.icon**

Het pad en de naam van een afbeeldingsbestand waarmee wordt aangegeven dat het document is gecrawld door een Domino Document Manager-crawler. Het standaardpictogram is `/images/sourceDominoDoc.gif`.

**documentSource.exchange.icon**

Het pad en de naam van een afbeeldingsbestand waarmee wordt aangege-



ven dat het document is gecrawld door een Exchange Server-crawler. Het standaardpictogram is /images/sourceExchange.gif.



**documentSource.database.icon**

Het pad en de naam van een afbeeldingsbestand waarmee wordt aangegeven dat het document is gecrawld door een JDBC database-crawler. Het standaardpictogram is /images/sourceJDBC.gif.



**documentSource.nntp.icon**

Het pad en de naam van een afbeeldingsbestand waarmee wordt aangegeven dat het document is gecrawld door een NNTP-crawler. Het standaardpictogram is /images/sourceNNTP.gif.



**documentSource.notes.icon**

Het pad en de naam van een afbeeldingsbestand waarmee wordt aangegeven dat het document is gecrawld door een Notes-crawler. Het standaardpictogram is /images/sourceNotes.gif.



**documentSource.quickplace.icon**

Het pad en de naam van een afbeeldingsbestand waarmee wordt aangegeven dat het document is gecrawld door een QuickPlace-crawler. Het standaardpictogram is /images/sourceWorkplace.gif.



**documentSource.unixfs.icon**

Het pad en de naam van een afbeeldingsbestand waarmee wordt aangegeven dat het document is gecrawld door een UNIX-bestandssysteem-crawler. Het standaardpictogram is /images/sourceUnixFS.gif.



**documentSource.web.icon**

Het pad en de naam van een afbeeldingsbestand waarmee wordt aangegeven dat het document is gecrawld door een webcrawler. Het standaardpictogram is /images/sourceWeb.gif.



**documentSource.wcm.icon**

Het pad en de naam van een afbeeldingsbestand waarmee wordt aangegeven dat het document is gecrawld door een Web Content Management-crawler. Het standaardpictogram is /images/sourceWorkplace.gif.



**documentSource.wps.icon**

Het pad en de naam van een afbeeldingsbestand waarmee wordt aangegeven dat het document is gecrawld door een WebSphere Portal-crawler. Het standaardpictogram is /images/sourceWPS.gif.



**documentSource.winfs.icon**

Het pad en de naam van een afbeeldingsbestand waarmee wordt aangegeven dat het document is gecrawld door een Windows-bestandssysteem-crawler. Het standaardpictogram is /images/sourceWindowsFS.gif.



**documentSource.ldap.icon**

Het pad en de naam van een afbeeldingsbestand waarmee wordt aangegeven dat het document deel uitmaakt van een externe bron die voor een LDAP-server is gemaakt. Het standaardpictogram is /images/sourceLDAP.gif.



### documentSource.jdbc.icon

Het pad en de naam van een afbeeldingsbestand waarmee wordt aangegeven dat het document deel uitmaakt van een externe bron die voor een JDBC-databasetabel (Java Database Connectivity) is gemaakt. Het standaardpictogram is /images/sourceJDBC.gif.

## Pictogrammen voor clientviewer

U kunt de resultaatdocumenten in de webbrowser bekijken. Documenten die zijn gecrawld door Notes- of Domino Document Manager-crawlers die zijn geconfigureerd voor gebruik van het DIIOP-protocol, kunnen ook worden bekeken met een Lotus Notes-clientviewerprogramma.

Om in te stellen dat documenten kunnen worden bekeken met een clientviewerprogramma, moet de volgende eigenschap zijn ingesteld op true:

```
clientViewer.show=true
```

U kunt de afbeeldingen aanpassen die het clientviewerprogramma aangeven. In het volgende voorbeeld wordt het pictogram Lotus Notes gebruikt om aan te geven dat het document kan worden bekeken in het viewerprogramma:

```
client.notes.icon=/images/notes.gif client.dominodoc.icon=/images/notes.gif
```

In de zoekresultaten worden het pictogram en de link naar het clientviewerprogramma als volgt weergegeven:



### Clientviewer

## Documentvelden

Voor gegevensbrontypen met velden kunt u bepalen welke velden in de resultaatdocumenten moeten worden weergegeven.

### fields.URI prefix=*door spaties gescheiden lijst van veldnamen*

In de URI-prefix moet voor de dubbele punt (:) een schuine streep naar links (\) worden geplaatst. Als u meer namen aan de lijst van veldnamen wilt toevoegen op de volgende regel, beëindigt u de voorgaande regel met een schuine streep naar links (\). Bijvoorbeeld:

```
fields.db2\://=databasenaam tabelnaam
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.vbr\://=itemname repositorytype revisionuser
fields.wcm\://=author owner modifier
fields.web\://=
fields.wp6\://=
fields.wps\://=
```

## Veldpictogrammen

Voor gegevensbrontypen en documenten met velden kunt u de afbeeldingen aanpassen waarmee de velden worden aangegeven. Alle velden boven de document-

samenvatting zijn voorzien van een afbeelding. De volgende veldpictogrammen zijn vooraf gedefinieerd in het bestand `config.properties`.



**field.icon.databasetitle**

Het pad en de naam van een afbeeldingsbestand waarmee wordt aangegeven dat het veld de documenttitel bevat. Het standaardpictogram is `/images/notesdb.gif`.



**field.icon.databasename**

Het pad en de naam van een afbeeldingsbestand waarmee wordt aangegeven dat het veld de naam bevat van de database waarvan het document deel uitmaakt. Het standaardpictogram is `/images/db2.gif`.



**field.icon.tablename**

Het pad en de naam van een afbeeldingsbestand waarmee wordt aangegeven dat het veld de naam bevat van de tabel waarvan het document deel uitmaakt. Het standaardpictogram is `/images/table.gif`.



**field.icon.directory**

Het pad en de naam van een afbeeldingsbestand waarmee wordt aangegeven dat het veld de naam bevat van de directory waarvan het document deel uitmaakt. Het standaardpictogram is `/images/closedFolder.gif`.



**field.icon.filename**

Het pad en de naam van een afbeeldingsbestand waarmee wordt aangegeven dat het veld de bestandsnaam van het document bevat. Het standaardpictogram is `/images/document.gif`.

**field.icon.documentID**

Het pad en de naam van een afbeeldingsbestand waarmee wordt aangegeven dat het veld het document-ID bevat. Mogelijk wilt u deze lege afbeelding gebruiken voor webdocumenten, om bijvoorbeeld een afbeelding voor de URL op te geven maar geen afbeelding voor de gebruiker weer te geven. Het standaardpictogram is `/images/dot.gif`.



**field.icon.group**

Het pad en de naam van een afbeeldingsbestand waarmee wordt aangegeven dat het veld het document-ID bevat. Mogelijk wilt u deze lege afbeelding gebruiken voor webdocumenten, om bijvoorbeeld een afbeelding voor de URL op te geven maar geen afbeelding voor de gebruiker weer te geven. Het standaardpictogram is `/images/document.gif`.



**field.icon.from**

Het pad en de naam van een afbeeldingsbestand waarmee wordt aangegeven dat het veld de persoon voorstelt die het document heeft verzonden. Het standaardpictogram is `/images/author.gif`.



**field.icon.creator**

Het pad en de naam van een afbeeldingsbestand waarmee wordt aangegeven dat het veld de maker van het document voorstelt. Het standaardpictogram is `/images/author.gif`.

**field.icon.author**

Het pad en de naam van een afbeeldingsbestand waarmee wordt aangegeven dat het veld de auteur van het document voorstelt. Het standaardpictogram is /images/author.gif.

**field.icon.revisionuser**

Het pad en de naam van een afbeeldingsbestand waarmee wordt aangegeven dat het veld de persoon voorstelt die het document heeft gereviseerd. Het standaardpictogram is /images/author.gif.

**field.icon.owner**

Het pad en de naam van een afbeeldingsbestand waarmee wordt aangegeven dat het veld de eigenaar van het document voorstelt. Het standaardpictogram is /images/author.gif.

**field.icon.modifier**

Het pad en de naam van een afbeeldingsbestand waarmee wordt aangegeven dat het veld de persoon voorstelt die het document heeft gewijzigd. Het standaardpictogram is /images/author.gif.

## Standaardveldpictogram

U kunt de afbeelding opgeven die moet worden gebruikt als er geen veldpictogrammen zijn geconfigureerd voor velden die in de zoekresultaten worden afgebeeld. Het volgende standaardveldpictogram is vooraf gedefinieerd in het bestand config.properties.

**field.defaultIcon**

Het pad en de naam van een afbeeldingsbestand dat het standaardpictogram is voor velden in de zoekresultaten. Het standaardpictogram is /images/database.gif.

## Datumvelden

U kunt opgeven welke velden datumvelden zijn. De veldnamen die u opgeeft, worden in de zoekresultaten weergegeven in de datumnotatie. Welke datumnotatie wordt gebruikt, is afhankelijk van de landinstellingen in de webbrowser.

**date.fields=door\_spaties\_gescheiden\_lijst\_van\_veldnamen**

Als u meer namen aan de lijst met veldnamen wilt toevoegen op de volgende regel, beëindigt u de voorgaande regel met een schuine streep naar links (\).

Voorbeeld:

```
date.fields=wijzigingsdatum aanmaakdatum
```

## Documenttitels

U kunt alternatieve titels voor documenten opgeven door de tekst van de titel te vervangen door meer begrijpelijke gegevens (dat wil zeggen: u kunt de titels *opschonen*). U kunt bijvoorbeeld opgeven dat documenttitels zonder informatieve waarde zoals Dia 1 niet in de zoekresultaten moeten worden weergegeven. (U kunt dan een veldnaam met meer betekenis, zoals de bestandsnaam, gebruiken om het resultaatdocument aan te geven.)

Daarnaast kunt u alternatieve titels voor documenten opgeven door woorden zonder informatieve waarde uit de documenttitels te verwijderen (dit wordt ook wel het *afkappen* van titels genoemd). Als een aantal resultaatdocumenten bijvoorbeeld met Microsoft Word - begint, kunt u de leesbaarheid van de zoekresultaten verbeteren door de terugkerende begintekst te onderdrukken.

**titles.clean=***door\_komma's\_gescheiden\_lijst\_titels*

**titles.truncatePrefix=***door\_komma's\_gescheiden\_lijst\_prefixen*

De door komma's gescheiden lijsten kunnen spaties en andere tekens bevatten, met uitzondering van de komma. Als u meer items aan de lijst wilt toevoegen op de volgende regel, beëindigt u de voorgaande regel met een schuine streep naar links (\).

Bijvoorbeeld:

```
titles.clean=Dia 1, Layout 1, IBM Software Group-presentatiesjabloon, \
naamloos, Naamloos document, PowerPoint-presentatie, \
(geen titel voor deze pagina)
```

```
titles.truncatePrefix=Microsoft Word -, Microsoft Powerpoint -
```

## Standaardwaarden voor gebruikersvoorkeuren

U kunt standaardwaarden opgeven voor de pagina Voorkeuren in het zoekprogramma. Als een gebruiker de voorkeuren wijzigt, zijn de nieuwe waarden alleen van kracht in de huidige sessie van de gebruiker. De volgende voorkeuren zijn vooraf gedefinieerd in het bestand config.properties.

### **preferences.resultsRange=10**

Hiermee wordt aangegeven dat elke pagina in de zoekresultaten tien resultaatdocumenten kan bevatten.

### **preferences.siteCollapsing=Yes**

Hiermee wordt aangegeven dat URI's die van dezelfde bron afkomstig zijn, in de zoekresultaten worden samengevouwen. Het samenvouwen van sites is alleen beschikbaar als de documenten zijn gesorteerd op relevantie. Voor web- en NNTP-gegevensbronnen worden URI's die dezelfde basis-URI hebben (zoals www.ibm.com) automatisch samengevouwen. Voor andere soorten gegevensbronnen en voor websites met een diepere hiërarchie van paden (zoals www.ibm.com/hr) moeten er in de beheerconsole van enterprise search regels voor het samenvouwen van sites worden opgegeven.

### **preferences.spellCorrections=Yes**

Hiermee wordt aangegeven dat de aanbevolen spellingscorrecties worden weergegeven wanneer een gebruiker een query indient die een mogelijk verkeerd gespeld woord bevat. Stopwoorden worden altijd verwijderd voordat de spellingsuggesties worden berekend.

### **preferences.extendedHighlighting=No**

Hiermee wordt aangegeven dat de zoektermen naast het document-samenvattingsveld niet gemarkeerd worden weergegeven in extra velden (zoals de documenttitel).

## Standaardcollecties en externe bronnen

U kunt opgeven welke collecties en externe bronnen vooraf moeten worden geselecteerd op de pagina's Voorkeuren en Uitgebreide zoekopdracht. Gebruikers kunnen de standaardset aanpassen zodat in minder collecties en externe bronnen wordt gezocht dan de collecties en bronnen die standaard beschikbaar zijn. Als u hier beperkingen instelt voor de set collecties en externe bronnen, kunnen gebrui-

kers bij het wijzigen van de voorkeuren of de opties voor uitgebreid zoeken elke gewenste collectie of externe bron selecteren die beschikbaar is.

**preferences.defaultCollections=\***

**preferences.defaultCollections=door\_spaties\_gescheiden\_lijst\_collectie\_ID's**

Geef een sterretje (\*) op om in te stellen dat alle collecties en externe bronnen kunnen worden doorzocht. (De collecties en externe bronnen moeten in de beheerconsole van enterprise search aan het zoekprogramma zijn gekoppeld.) Dit is de standaardinstelling in het bestand config.properties.

Als u beperkingen wilt instellen voor de items die voor gebruikers worden weergegeven als zij de voorkeuren of opties voor uitgebreid zoeken niet aanpassen, geeft u de collectie-ID's op voor de collecties en externe bronnen die standaard doorzoekbaar zijn voor gebruikers.

Bijvoorbeeld:

```
preferences.defaultCollections=*  
preferences.defaultCollections=coll_id1 coll_id2
```

## Extra informatie voor de zoekresultaten

U kunt de hoeveelheid informatie aanpassen die bij de zoekresultaten wordt weergegeven en bepalen of gebruikers de zoekresultaten kunnen filteren. De volgende instellingen zijn de standaardinstellingen in het bestand config.properties.

**refreshButton.show=false**

Hiermee wordt bepaald of de knop **Vernieuwen** op de basiszoekpagina wordt weergegeven. Voor geavanceerde zoekopdrachten is de knop **Vernieuwen** altijd beschikbaar. Als u deze optie instelt op true, kunnen gebruikers de lijst van collecties en externe bronnen vernieuwen die kunnen worden doorzocht.

Als u werkt met het Aanpassingstool Zoekprogramma, hebt u de knop **Vernieuwen** niet nodig.

Als u niet met dit tool werkt, wilt u de knop **Vernieuwen** mogelijk weergeven tijdens het testen van de wijzigingen die u aanbrengt in het configuratiebestand. Als u de wijzigingen hebt opgeslagen, kunt u op **Vernieuwen** klikken om te zien hoe de wijzigingen van invloed zijn op het zoekprogramma. Als u de knop **Vernieuwen** niet gebruikt, worden de wijzigingen pas van kracht nadat u het programma ESSearchServer in WebSphere Application Server opnieuw hebt gestart.

Als er geen collecties of externe bronnen beschikbaar zijn om te worden doorzocht (als bijvoorbeeld de verkeerde hostnaam is opgegeven, als de zoekservers niet actief zijn of als het programma ESSearchServer niet is gestart in WebSphere Application Server), wordt de knop **Vernieuwen** automatisch weergegeven om u te helpen het probleem op te lossen.

**builtQueryString.show=false**

Hiermee wordt bepaald of de volledige querysyntaxis moet worden weergegeven in een gebied dat vooraf gaat aan de lijst van resultaatdocumenten. U kunt deze optie instellen op true als u de werkelijke query wilt weergeven die is verwerkt.

**extraQueryData.show=false**

Hiermee wordt bepaald of extra informatie over de query wordt weergegeven. Stel deze optie in op true als u informatie wilt weergegeven over toegangslijstbeperkingen, de namen van de collecties en de externe bronnen die worden doorzocht, en de query language.

**refineResults.show=true**

Hiermee wordt bepaald of gebruikers de zoekresultaten kunnen verfijnen door extra zoektermen op te geven. Als u deze optie instelt op true, verschijnt onder aan de pagina met zoekresultaten een queryvak met het label **Zoeken binnen de resultaten**.

**sorting.show=true**

Hiermee wordt bepaald of de optie voor het sorteren van de zoekresultaten wordt weergegeven. Stel deze optie in op false om de opties **Sorteren op** en **Sorteervolgorde** voor het sorteren van de zoekresultaten te verbergen.

**sourceTypeFilter.show=true**

Hiermee wordt bepaald of de optie voor het filteren van de resultaten op brontype wordt weergegeven in de zoekresultaten. Stel deze optie in op false als u wilt voorkomen dat gebruikers de resultaten kunnen filteren op brontype.

U kunt dit voorkomen door de geselecteerde vermeldingen of alle vermeldingen van documenttypen te wissen in het configuratiebestand (`documentType.label=document_typen`).

**filter.showOnTwoLines=true**

Hiermee wordt bepaald of de opties voor het filteren van resultaten op basis van het brontype en het filteren van resultaten op basis van het bestandstype op een of twee regels in de zoekresultaten worden weergegeven. Bij het bekijken van de zoekresultaten kunnen gebruikers een brontype en een bestandstype selecteren om alleen de resultaatdocumenten weer te geven die overeenkomen met de geselecteerde filters.

U kunt deze optie instellen op false om de hoeveelheid ruimte te maximaliseren die beschikbaar is voor het weergeven van de zoekresultaten. Om de leesbaarheid van de filters te verbeteren, met name als de beschikbare filters meerdere regels in beslag nemen, kunt u deze eigenschap instellen op true zodat elk filter op een aparte regel wordt weergegeven.

**clientViewer.show=true**

Hiermee wordt bepaald of het Lotus Notes-clientviewerprogramma moet worden gebruikt om een resultaatdocument weer te geven. Stel deze optie in op false als u het viewerprogramma niet wilt gebruiken om Domino-documenten weer te geven.

**showDetails.show=true**

Hiermee wordt bepaald of de links Details afbeelden en Details verbergen in de zoekresultaten worden weergegeven. Stel deze optie in op false als u wilt voorkomen dat gebruikers details over de resultaatdocumenten kunnen bekijken.

**showDetailsImage.show=true**

Hiermee wordt bepaald of details over de resultaatdocumenten in een venster kunnen worden afgebeeld. Stel deze optie in op false als u wilt voorkomen dat gebruikers extra details over de resultaten kunnen bekijken door de cursor boven een document-URI te plaatsen.

**numberSearchResultsReturned.show=true**

Hiermee wordt bepaald of het totale aantal zoekresultaten wordt afgebeeld. Stel deze optie in op false als u wilt voorkomen dat gebruikers kunnen zien hoeveel documenten in de zoekresultaten worden weergegeven.

**showMessage.error=true**

Hiermee wordt de weergave van foutberichten bepaald. Stel deze optie in op false als u boven in het zoekprogramma geen foutberichten wilt weergeven.

**showMessage.warning=true**

Hiermee wordt de weergave van waarschuwingsberichten bepaald. Stel deze optie in op false als boven in het zoekprogramma geen waarschuwingsberichten wilt weergeven.

**showMessage.info=true**

Hiermee wordt de weergave van informatieve berichten bepaald. Stel deze optie in op false als u boven in het zoekprogramma geen informatieve berichten wilt weergeven.

**showMessage.success=true**

Hiermee wordt de weergave van succesberichten bepaald. Stel deze optie in op false als u boven aan de pagina in het zoekprogramma geen berichten wilt weergeven waarmee de succesvolle voltooiing van een actie wordt aangegeven.

## Aangepaste banners en logo's

U kunt bepalen welke afbeeldingen in het bannergebied boven in het zoekprogramma worden weergegeven. Zo kunt u de standaardafbeeldingen voor WebSphere II OmniFind Edition vervangen door afbeeldingen die van toepassing zijn op uw bedrijf. Als u geen banner wilt weergeven, kunt u een of beide regels instellen als commentaarregels. Met de eigenschap banner.icon wordt de afbeelding aangegeven die links van het bannergebied wordt weergegeven. Met de eigenschap banner2.icon wordt de afbeelding aangegeven die rechts van het bannergebied wordt weergegeven.

```
banner.icon=/images/WS_II_OFEdition.gif  
banner2.icon=/images/WS_II_mosaic.gif
```

## Aangepaste achtergrondafbeelding

U kunt de afbeeldingen aanpassen die op de achtergrond van de pagina's van het zoekprogramma worden weergegeven. Zo kunt u standaardafbeeldingen voor enterprise search vervangen door afbeeldingen die van toepassing zijn op uw bedrijf. Als u geen achtergrondafbeelding op de pagina's wilt weergeven, kunt u een of beide regels instellen als commentaarregels.

```
search.backgroundImage=/images/IIOF_search.gif  
preferences.backgroundImage=/images/IIOF_options.gif  
advanced.backgroundImage=/images/IIOF_advanced.gif  
browse.backgroundImage=/images/IIOF_tree.gif  
myProfile.backgroundImage=/images/IIOF_profile.gif  
logoff.backgroundImage=/images/IIOF_logout.gif
```

## Links

Met behulp van de eigenschappen in het gebied Links van het bestand config.properties kunt u de namen van de pagina's in het zoekprogramma op elke pagina weergeven als links (en niet op de werkbalk en op de pagina's met tabs). Dit is met name handig wanneer u het zoekprogramma uitvoert als een portlet en de hoeveelheid ruimte wilt minimaliseren die nodig is om het zoekprogramma op een portalpagina weer te geven.



Als u liever door het zoekprogramma navigeert door opties op de werkbalk en op pagina's met tabs te selecteren, kunt u deze regels uitschakelen.

## Zoektabs

Met de eigenschappen in het gebied Search tabs van het bestand config.properties worden de namen aangegeven van de JSP-pagina's (Java Server Pages) die voor de pagina's met tabs worden gebruikt in de view Zoekopdrachten van het zoekprogramma (Eenvoudige zoekopdracht, Uitgebreid zoeken en Categoriestructuur). U wordt aangeraden deze pagina's alleen te bewerken als u ervaring hebt met Java en JSP's.

In het onderstaande overzicht vindt u voorbeelden van de manieren waarop u dit gebied kunt aanpassen.

- U kunt in het zoekprogramma de aangepaste JSP's gebruiken voor een andere weergave van de pagina's met tabs.
- U kunt de items in de categoriestructuur uitschakelen. Als u bijvoorbeeld geen categorieën voor uw collecties hebt geconfigureerd, hoeft u de pagina Categoriestructuur niet op te nemen in het zoekprogramma.
- U kunt de items voor de pagina's met tabs kopiëren naar het gebied Toolbars van het bestand config.properties en deze regels uitschakelen. U kunt bijvoorbeeld alleen de werkbalk weergeven en geen pagina's met tabs.

## Werkbalken

Met de eigenschappen in het gebied Toolbars van het bestand config.properties worden de namen van de JSP-pagina's (Java Server Pages) aangegeven die voor de werkbalk in het zoekprogramma worden gebruikt. U wordt aangeraden deze pagina's alleen te bewerken als u ervaring hebt met Java en JSP's.

In het onderstaande overzicht vindt u voorbeelden van de manieren waarop u dit gebied kunt aanpassen.

- U kunt in het zoekprogramma de aangepaste JSP's gebruiken voor een andere weergave van de werkbalk.
- U kunt de werkbalkopties uitschakelen voor de items die u niet wilt weergeven. U kunt bijvoorbeeld instellen dat de link naar de pagina Info niet moet worden weergegeven op de werkbalk.
- U kunt de functie voor het weergeven van de pagina Uitgebreid zoeken vanuit het tabgebied in het bestand config.properties verplaatsen, zodat deze optie alleen beschikbaar is op de werkbalk.

## Betekenisvolle labels voor documenttypen

U kunt de leesbaarheid van het filter voor documenttypen verbeteren door aan de werkelijke namen van de documenttypen namen met meer betekenis toe te wijzen. Welke documenttypen beschikbaar zijn in zoekprogramma's, wordt gedefinieerd op basis van de AvailableDocumentTypes-klasse van de SI-API (Search and Index API). De beschikbare documenttypen worden voor het gemak ook weergegeven aan het einde van het bestand config.properties.

### **documentType.label=door\_spaties\_gescheiden\_lijst\_van\_documenttypen**

Hiermee wordt de naam aangegeven die op de regel voor het documenttypefilter wordt weergegeven in de zoekresultaten, en een lijst van de werkelijke documenttypen die moeten worden weergegeven wanneer een gebruiker het filter selecteert.

U kunt bijvoorbeeld het label `html` opgeven en de bestandsextensies en MIME-typen voor verschillende webdocumenten aan deze naam toewijzen. Wanneer een gebruiker op **html** klikt om de zoekresultaten te filteren, worden alleen de documenten met de opgegeven extensies en MIME-typen weergegeven.

De volgende documenttypetoewijzingen zijn vooraf gedefinieerd in het bestand `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
```

Als de waarde die u voor het documenttypelabel opgeeft gelijk is aan de naam van een eigenschap in het bestand `application.properties`, wordt de waarde van die eigenschap in het bestand `application.properties` afgebeeld, en dus niet de waarde die u hier opgeeft. Als u bijvoorbeeld `documentType.unixfs` opgeeft als label voor een bestandstypefilter, wordt de waarde van de eigenschap `unixfs` in het bestand `application.properties` (d.w.z. **UNIX file system**) afgebeeld als een selecteerbare naam voor een bestandstypefilter.

## Aangepaste filters

U kunt aangepaste query's opgeven om de weergave van resultaatdocumenten te filteren.

### **filterCustom.label=zoek\_termen**

Hiermee wordt de naam aangegeven die voor de aangepaste filterregel in de zoekresultaten wordt weergegeven, en een query waarmee de zoekresultaten worden verfijnd wanneer een gebruiker het filter selecteert. (Bij het bekijken van de zoekresultaten kunnen gebruikers een aangepast filter selecteren om alleen de resultaatdocumenten weer te geven die overeenkomen met de vooraf gedefinieerde query.)

In het volgende voorbeeld worden de zoekresultaten gefilterd zodat alleen de documenten worden weergegeven die deel uitmaken van de `human resources-database (hr-database)`:

```
filterCustom.HR_database_only=databasenaam:hr
```

Wanneer een gebruiker op **HR\_database\_only** klikt om de zoekresultaten te filteren, wordt de query `databasenaam:hr` verwerkt. Bij het weergeven van de zoekresultaten worden er dan alleen documenten uit de database `hr` afgebeeld.

Als de waarde die u voor het label van een aangepaste filter opgeeft gelijk is aan de naam van een eigenschap in het bestand `application.properties`, wordt de waarde van die eigenschap in het bestand `application.properties` afgebeeld, en dus niet de waarde die u hier opgeeft. Als u bijvoorbeeld `filterCustom.hostData=` als label voor een aangepaste filter, wordt de waarde van de eigenschap `hostData` in het bestand `application.properties` (d.w.z. **Server settings**) afgebeeld als een selecteerbare naam voor een aangepaste filter.

Het bestand config.properties bevat verschillende filters die zijn uitgeschakeld en die als voorbeeld dienen.

## Eigenschappen van het voorbeeldzoekprogramma wijzigen

Met het voorbeeldzoekprogramma voor enterprise search kunt u alle actieve collecties en externe bronnen in het systeem doorzoeken. U kunt een configuratiebestand bewerken om opties op te geven voor de webserveromgeving, een ander zoekprogramma als standaardprogramma gebruiken of bepalen welke opties worden weergegeven bij het starten van het zoekprogramma.

### Over deze taak

Met het installatieprogramma wordt op de zoekservers voor enterprise search een voorbeeldzoekprogramma voor enterprise search geïntegreerd in IBM WebSphere Application Server. Om dit zoekprogramma te configureren, moet u het configuratiebestand config.properties bewerken dat samen met het programma wordt geïnstalleerd.

Als u de wijzigingen wilt doorvoeren, moet u het programma ESSearchServer in WebSphere Application Server stoppen en opnieuw starten.

### Procedure

Ga als volgt te werk om de eigenschappen van het voorbeeldzoekprogramma te bewerken:

1. Meld u bij de zoekserver aan als de enterprise search-beheerder.
2. Open het bestand config.properties in een standaardteksteditor.  
Het bestand config.properties is op de volgende locatie geïnstalleerd, waarbij ES\_INSTALL\_ROOT de installatiedirectory van WebSphere II OmniFind Edition op de zoekserver is:  
`ES_INSTALL_ROOT/installedApps/ESSearchApplication.ear/ESSearchApplication.war/WEB-INF/config.properties`
3. Bewerk de eigenschappen om informatie over uw webserveromgeving en zoekvoorkeuren op te geven, sla het bestand vervolgens op en sluit het bestand. In het bestand worden commentaarregels aangegeven met het teken # (hekje).
4. Stop het programma ESSearchServer en start het opnieuw.

### AIX, Linux of Solaris

```
./stopServer.sh ESSearchServer  
./startServer.sh ESSearchServer
```

### Windows

```
stopServer ESSearchServer  
startServer ESSearchServer
```

Deze scripts zijn opgeslagen in de directory WAS\_INSTALL\_ROOT/AppServer/bin:

- Voor WebSphere Application Server versie 5 is het standaardinstallatiepad /usr/WebSphere (op AIX-systemen), /opt/WebSphere (op Linux- of Solaris-systemen) of C:\Program Files\WebSphere (op Windows-systemen).
- Voor WebSphere Application Server versie 6 is het standaardinstallatiepad /usr/IBM/WebSphere (op AIX-systemen), /opt/IBM/WebSphere (op Linux- of Solaris-systemen) of C:\Program Files\IBM\WebSphere (op Windows-systemen).

### Verwante taken

“De zoekservers configureren om uitsluitend veilige (SSL) opdrachten te accepteren” op pagina 224

U kunt de HTTP-interface op de zoekservers uitschakelen en de servers zodanig configureren dat ze uitsluitend zoekopdrachten via SSL en de veilige HTTPS-interface accepteren.

## Zoekprogramma's aanpassen

Het Aanpassingstool Zoekprogramma is een grafische interface waarmee u zoekprogramma's voor enterprise search of uw aangepaste zoekprogramma's kunt aanpassen.

### Beperkingen

Het Aanpassingstool Zoekprogramma is beschikbaar als een stand-alone programma. U kunt het Aanpassingstool Zoekprogramma niet starten als een portlet binnen WebSphere Portal.

### Over deze taak

Met het Aanpassingstool Zoekprogramma kunt u de wijzigingen visualiseren die u wilt aanbrengen en zoekprogramma's wijzigen zonder dat u hiervoor het configuratiebestand hoeft te bewerken. U kunt bijvoorbeeld de afbeelding voor de banner en de achtergrond wijzigen, de layout van de zoekinterface aanpassen en opties opgeven voor het werken met de zoekresultaten.

Wanneer u opties in het Aanpassingstool Zoekprogramma selecteert, worden de effecten van de selecties weergegeven. Bij het opslaan van de wijzigingen wordt het configuratiebestand voor het zoekprogramma bijgewerkt.

Als u de wijzigingen wilt doorvoeren, moet u het programma ESSearchServer in WebSphere Application Server stoppen en opnieuw starten.

### Procedure

Ga als volgt te werk om een zoekprogramma aan te passen:

1. Als u het voorbeeldzoekprogramma voor enterprise search wilt aanpassen, voert u de URL voor het Aanpassingstool Zoekprogramma in een webbrowser in. Bijvoorbeeld:

`http://ZoekServer.com/ESSearchApplication/palette.do`

*ZoekServer.com* is de hostnaam van de zoekserver.

Als de webserver niet is geconfigureerd voor gebruik van poort 80, moet u ook het juiste poortnummer opgeven. Bijvoorbeeld:

`http://ZoekServer.com:9080/ESSearchApplication/palette.do`

**Tip:** Als u enterprise search-beheerder bent, kunt u het Aanpassingstool Zoekprogramma ook openen door de optie **Aanpassingstool Zoekprogramma** in de beheerconsole van enterprise search te selecteren.

2. Om een aangepast zoekprogramma aan te passen, voert u de URL voor het Aanpassingstool Zoekprogramma in, gevolgd door de naam van het configuratiebestand van het zoekprogramma. Bijvoorbeeld:

`http://ZoekServer.com/ESSearchApplication/palette.do?configFile=/WEB-INF/myConfig.properties`

Als het bestand dat u opgeeft niet bestaat, worden de waarden in het bestand config.properties van het voorbeeldzoekprogramma weergegeven.

**Tip:** U kunt ook het configuratiebestand voor een zoekprogramma opgeven door op **Laden** te klikken nadat u het Aanpassingstool Zoekprogramma hebt gestart en de naam van het bestand hebt opgegeven.

3. Als algemene beveiliging is ingeschakeld in WebSphere Application Server, meldt u zich aan met een geldig gebruikers-ID en wachtwoord.
4. Selecteer de opties die u wilt aanpassen, zoals de informatie over de zoekserver, de typen berichten die moeten worden weergegeven, de opties voor query- en zoekresultaten en de namen van de afbeeldingen die u wilt gebruiken om de verschillende typen documenten in de zoekresultaten aan te duiden. Voor hulp bij het opgeven van opties klik u op **Help voor aanpassingsprogramma**. Als u het resultaat van bepaalde wijzigingen wilt bekijken, voert u een query in en klikt u op **Zoeken**.
5. Als u tevreden bent met de aangebrachte wijzigingen, klikt u op **Opslaan** om het configuratiebestand bij te werken. Als u op **Reset** klikt, worden voor de opties die in het Aanpassingstool Zoekprogramma worden weergegeven, de waarden in de laatst opgeslagen versie van het configuratiebestand gebruikt.
6. Meld u op de zoekserver aan als enterprise search-beheerder, stop het programma ESSearchServer en start het programma vervolgens opnieuw.

#### **AIX, Linux of Solaris**

```
./stopServer.sh ESSearchServer  
./startServer.sh ESSearchServer
```

#### **Windows**

```
stopServer ESSearchServer  
startServer ESSearchServer
```

Deze scripts zijn opgeslagen in de directory WAS\_INSTALL\_ROOT/AppServer/bin:

- Voor WebSphere Application Server versie 5 is het standaardinstallatiepad /usr/WebSphere (op AIX-systemen), /opt/WebSphere (op Linux- of Solaris-systemen) of C:\Program Files\WebSphere (op Windows-systemen).
- Voor WebSphere Application Server versie 6 is het standaardinstallatiepad /usr/IBM/WebSphere (op AIX-systemen), /opt/IBM/WebSphere (op Linux- of Solaris-systemen) of C:\Program Files\IBM\WebSphere (op Windows-systemen).

#### **Verwante taken**

“De zoekservers configureren om uitsluitend veilige (SSL) opdrachten te accepteren” op pagina 224

U kunt de HTTP-interface op de zoekservers uitschakelen en de servers zodanig configureren dat ze uitsluitend zoekopdrachten via SSL en de veilige HTTPS-interface accepteren.

## **Het voorbeeldzoekprogramma klonen**

Als u het voorbeeldzoekprogramma wilt gebruiken als basis voor uw eigen zoekprogramma's, kopieert u het bestand config.properties of gebruikt u het Aanpassingstool Zoekprogramma.

#### **Over deze taak**

U kunt het maken van aangepaste zoekprogramma's vereenvoudigen door de configuratieopties te klonen die u voor het voorbeeldzoekprogramma hebt opgegeven, waarna u de gewenste opties kunt aanpassen.

Nadat u het voorbeeldzoekprogramma hebt gekloond, geeft u de naam op van het configuratiebestand om het nieuwe zoekprogramma te starten. U kunt ook de naam van het configuratiebestand opgeven om het nieuwe zoekprogramma aan te passen met behulp van het Aanpassingstool Zoekprogramma.

Door het voorbeeldzoekprogramma te klonen, kunt u snel zoekprogramma's voor specifieke doelen of doelgroepen maken. U kunt bijvoorbeeld een zoekprogramma maken voor werknemers op de HR-afdeling en een ander zoekprogramma voor de verkopers.

Als u de wijzigingen wilt doorvoeren, moet u het programma ESSearchServer in WebSphere Application Server stoppen en opnieuw starten.

### Procedure

Ga als volgt te werk om het voorbeeldzoekprogramma te klonen:

1. Als u een configuratiebestand wilt bewerken:
  - a. Kopieer het bestand `config.properties` voor het voorbeeldzoekprogramma en wijzig de naam.

Het bestand `config.properties` is op de volgende locatie geïnstalleerd, waarbij `ES_INSTALL_ROOT` de installatiedirectory van WebSphere II OmniFind Edition op de zoekserver is:

```
ES_INSTALL_ROOT/installedApps/ESSearchApplication.ear/  
  ESSearchApplication.war/WEB-INF/config.properties
```

Het bestand dat u maakt, moet aanwezig zijn in de subdirectory `WEB-INF`.

- b. Bewerk de eigenschappen die u wilt gebruiken in het aangepaste zoekprogramma en sla het bestand op. U moet minimaal de eigenschap `applicationName` wijzigen en de naam van het zoekprogramma opgeven.
2. Als u het voorbeeldzoekprogramma wilt klonen met behulp van het Aanpassingstool Zoekprogramma:

- a. Start het Aanpassingstool Zoekprogramma door de naam op te geven van het configuratiebestand dat u wilt maken. In het volgende voorbeeld wordt het bestand `myNewFile.properties` gemaakt:

```
http://ESServer.com/ESSearchApplication/palette.do?configFile=/WEB-INF/myNewFile.properties
```

Omdat het bestand nog niet bestaat, worden de waarden gebruikt die zijn opgegeven in het configuratiebestand van het voorbeeldzoekprogramma (`config.properties`).

**Tip:** U kunt ook een configuratiebestand voor een zoekprogramma maken door het Aanpassingstool Zoekprogramma te starten, de naam van een bestand op te geven en vervolgens te klikken op **Laden**. Het bestand wordt gemaakt wanneer u op **Opslaan** klikt om de opties op te slaan.

- b. Als algemene beveiliging is ingeschakeld in WebSphere Application Server, meldt u zich aan met een geldig gebruikers-ID en wachtwoord.
  - c. Geef een naam op voor het zoekprogramma, geef de opties op die u wilt gebruiken voor het doorzoeken van collecties en klik vervolgens op

**Opslaan.** De wijzigingen die u aanbrengt, worden in het nieuwe configuratiebestand in de subdirectory WEB-INF opgeslagen.

Voor hulp bij het opgeven van opties klik u op **Help voor aanpassingsprogramma**. Als u het resultaat van bepaalde wijzigingen wilt bekijken, voert u een query in en klikt u op **Zoeken**.

3. Meld u op de zoekserver aan als enterprise search-beheerder, stop het programma ESSearchServer en start het programma vervolgens opnieuw.

#### **AIX, Linux of Solaris**

```
./stopServer.sh ESSearchServer  
./startServer.sh ESSearchServer
```

#### **Windows**

```
stopServer ESSearchServer  
startServer ESSearchServer
```

Deze scripts zijn opgeslagen in de directory WAS\_INSTALL\_ROOT/AppServer/bin:

- Voor WebSphere Application Server versie 5 is het standaardinstallatiepad /usr/WebSphere (op AIX-systemen), /opt/WebSphere (op Linux- of Solaris-systemen) of C:\Program Files\WebSphere (op Windows-systemen).
- Voor WebSphere Application Server versie 6 is het standaardinstallatiepad /usr/IBM/WebSphere (op AIX-systemen), /opt/IBM/WebSphere (op Linux- of Solaris-systemen) of C:\Program Files\IBM\WebSphere (op Windows-systemen).

---

## **Toegang krijgen tot zoekprogramma's**

U kunt een zoekprogramma openen door de URL in een webbrowser op te geven.

### **Voordat u begint**

U moet het zoekprogramma configureren voor uw webserveromgeving.

### **Over deze taak**

Het voorbeeldzoekprogramma wordt op de zoekservers voor enterprise search geïnstalleerd. U kunt dit programma gebruiken om collecties en externe bronnen te testen voordat u ze beschikbaar stelt aan uw gebruikers. Daarnaast kunt u het programma gebruiken als model voor uw eigen zoekprogramma's.

### **Procedure**

Ga als volgt te werk om een zoekprogramma te starten:

1. Voer de URL voor het zoekprogramma in een webbrowser in. Bijvoorbeeld:

```
http://ZoekServer.com/ESSearchApplication/
```

*ZoekServer.com* is de hostnaam van de zoekserver.

Als de webserver niet is geconfigureerd voor gebruik van poort 80, moet u ook het juiste poortnummer opgeven. Bijvoorbeeld:

```
http://ZoekServer.com:9080/ESSearchApplication/
```

2. Om een aangepast zoekprogramma te starten, voert u de URL voor het voorbeeldzoekprogramma in, gevolgd door de naam van het configuratiebestand van het zoekprogramma. Bijvoorbeeld:

<http://ZoekServer.com/ESSearchApplication/search.do?configFile=/WEB-INF/myConfig.properties>

Als het bestand dat u opgeeft niet bestaat, wordt het voorbeeldzoekprogramma voor enterprise search weergegeven.

3. Als algemene beveiliging is ingeschakeld in WebSphere Application Server, meldt u zich bij het programma aan met een geldig gebruikers-ID en wachtwoord.

Als voor een van de collecties die beschikbaar is in het zoekprogramma, beveiliging is ingeschakeld en als de beveiligde collectie crawlers bevat die zijn geconfigureerd voor het valideren van legitimatiegegevens van gebruikers tijdens de queryverwerking, kunt u een gebruikersprofiel configureren. Geef op de pagina Mijn profiel legitimatiegegevens op voor toegang tot de beveiligde domeinen. Vervolgens kunt u deze domeinen doorzoeken zonder dat u zich bij de domeinen hoeft aan te melden.

Als de crawler ondersteuning biedt voor SSO-beveiliging (Single Sign-On), kunt u beveiligde domeinen doorzoeken zonder dat u hiervoor een gebruikersprofiel hoeft te maken.

4. Dien een query in op de pagina Zoeken. Alle collecties en externe bronnen worden doorzocht die u op de pagina Voorkeuren hebt geselecteerd om te worden doorzocht.

---

## De zoekservers configureren om uitsluitend veilige (SSL) opdrachten te accepteren

U kunt de HTTP-interface op de zoekservers uitschakelen en de servers zodanig configureren dat ze uitsluitend zoekopdrachten via SSL en de veilige HTTPS-interface accepteren.

### Over deze taak

Om de zoekservers zodanig te configureren dat ze bij het verwerken van zoekopdrachten uitsluitend gebruik maken van het SSL-protocol (Secure Sockets Layer), moet u de HTTP-interface uitschakelen. Tevens moet u ervoor zorgen dat hetzelfde sleutelruimtebestand is opgeslagen op zowel zoekservers als alle clientcomputers, zoals de WebSphere Portal-server waarop de portlet Zoeken voor enterprise search is geïnstalleerd. Het sleutelruimtebestand, dat ook wel een *truststore* of betrouwbaarheidsbestand wordt genoemd, bevat de openbare sleutels die het mogelijk maken dat SSL wordt gebruikt voor betrouwbare communicatie.

### Procedure

U kunt de zoekservers als volgt configureren om uitsluitend veilige (SSL) opdrachten te accepteren:

1. Meld u aan als enterprise search-beheerder. Als u een installatie met meerdere servers hebt, voer de volgende procedure dan uit op de indexserver:
  - a. Open het bestand `ES_NODE_ROOT/master_config/nodes.ini` in een editor.
  - b. Wijzig de waarde van `knooppunt.searchserverport` van de HTTP-poort (meestal 80) in de HTTPS-poort (meestal 443) en sla het bestand op. Bij een installatie met meerdere servers wijzigt u beide waarde van `knooppunt.searchserverport` (één voor elke zoekserver).
2. Voer de volgende procedure uit om de zoekserver bij te werken: Als u een installatie met meerdere servers hebt, voer de volgende procedure dan uit op beide zoekservers:



- a. Meld u, voor een installatie met meerdere servers, bij de zoekserver aan als de enterprise search-beheerder.
- b. Roep het bestand `ES_NODE_ROOT/nodeinfo/es.cfg` op in een editor.
- c. Geef bij de eigenschap `TrustStore` het volledige pad van het SSL-sleutelruimtebestand op.
- d. Geef bij de eigenschap `HTTPProtocol` de waarde `HTTPS` op en sla het bestand op.
- e. Geef de volgende opdracht op, waarbij `trustStore_wachtwoord` het wachtwoord voor het sleutelruimtebestand is. Met deze opdracht wordt de waarde van het wachtwoord versleuteld en wordt de waarde van `TrustStorePassword` in het bestand `es.cfg` bijgewerkt.

#### AIX, Linux of Solaris

```
eschangetrustpw.sh -p trustStore_wachtwoord
```

#### Windows

```
eschangetrustpw -p trustStore_wachtwoord
```

- f. Zorg dat bij de eigenschappen `trustStore` en `trustPassword` in het bestand `config.properties` voor het zoekprogramma het juiste volledige pad en het juiste wachtwoord voor het sleutelruimtebestand worden genoemd. U kunt deze gegevens controleren of wijzigen door het bestand `config.properties` op te roepen in een editor of door gebruik te maken van het Aanpassingstool Zoekprogramma.
  - g. Als u gebruik maakt van de portlet Zoeken voor enterprise search, moet u ook controleren of met de portletparameters `trustStore` en `trustPassword` het juiste volledige pad en het juiste wachtwoord voor het sleutelruimtebestand worden opgegeven. U kunt deze gegevens controleren of wijzigen met behulp van de beheeropties in de beheerinterface van WebSphere Portal.
3. Geef de volgende opdrachten op om het enterprise search-systeem opnieuw te starten:

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

#### Verwante taken

“Eigenschappen van het voorbeeldzoekprogramma wijzigen” op pagina 219  
Met het voorbeeldzoekprogramma voor enterprise search kunt u alle actieve collecties en externe bronnen in het systeem doorzoeken. U kunt een configuratiebestand bewerken om opties op te geven voor de webserver-omgeving, een ander zoekprogramma als standaardprogramma gebruiken of bepalen welke opties worden weergegeven bij het starten van het zoekprogramma.

“Zoekprogramma’s aanpassen” op pagina 220

Het Aanpassingstool Zoekprogramma is een grafische interface waarmee u zoekprogramma’s voor enterprise search of uw aangepaste zoekprogramma’s kunt aanpassen.

“Enterprise search installeren in WebSphere Portal versie 5.1” op pagina 273

Om een enterprise search-systeem te integreren met WebSphere Portal versie 5.1.0 of later, gebruikt u het script `wp5_install`.

“Enterprise search installeren in WebSphere Portal versie 6” op pagina 279

Om een enterprise search-systeem te integreren met WebSphere Portal versie 6 of later, gebruikt u het script `wp6_install`.



---

## Externe bronnen voor enterprise search

Een *externe bron* is een gegevensbron die u inschakelt voor het uitvoeren van zoekopdrachten met een enterprise search-programma zonder dat u hiervoor de documenten in de gegevensbron hoeft te crawlen, analyseren of indexeren.

U kunt de volgende typen gegevensbronnen doorzoeken als externe bronnen:

- Databases die het JDBC-protocol (Java Database Connectivity) protocol. Alleen IBM DB2 Universal Database- (DB2 UDB), Oracle-, Microsoft SQL Server 2000- en Microsoft SQL Server 2005-databases worden ondersteund. Voor elke tabel die u geschikt maakt om te worden doorzocht, wordt er een afzonderlijk externe bron gecreëerd.

**Beperking:** De ondersteuning voor SQL Server 2005-databases is beperkt tot tabellen die NIET het gegevenstype Variant bevatten. De JDBC-driver voor SQL Server 2005 wordt niet ondersteund op AIX-systemen.

- LDAP-servers (Lightweight Directory Access Protocol). Voor elke LDAP-server wordt er één externe bron gemaakt.

Als u de gegevens over een externe bron hebt geconfigureerd, moet u de bron aan minimaal één zoekprogramma koppelen. Gebruikers kunnen de externe bron vervolgens doorzoeken op hetzelfde moment dat ze query's uitvoeren op de collecties die zijn gemaakt op basis van het crawlen, analyseren en indexeren van gegevens voor enterprise search.

### Verwante onderwerpen

Federators van de Search and Index API

---

## Externe bronnen aan het systeem toevoegen

Als u een externe bron aan het enterprise search-systeem wilt toevoegen, moet u het type bron opgeven dat u wilt toevoegen. Een wizard helpt u bij het opgeven van informatie over de gegevensbron en informatie over de manier waarop de bron kan worden doorzocht.

### Voordat u begint

Om een externe bron aan het systeem toe te voegen, moet u de rol van beheerder van enterprise search hebben.

### Beperkingen

Als u een Oracle-database als externe bron wilt doorzoeken, moet het Oracle-clientprogramma zijn geïnstalleerd op de zoekservers voor enterprise search.

De JDBC-driver voor Microsoft SQL Server 2005 wordt niet ondersteund op AIX-systemen.

### Over deze taak

Wanneer u informatie over een externe bron aan het systeem toevoegt, stelt u gebruikers in staat om met behulp van een enterprise search-programma een query

op de bron uit te voeren. U kunt LDAP-servers (Lightweight Directory Access Protocol) en JDBC-databasetabellen (Java Database Connectivity) doorzoekbaar maken.

Wanneer u een LDAP-server configureert, helpt een wizard u bij het opgeven van informatie waarmee het systeem verbinding met de server kan maken en bij het opgeven van opties voor de manier waarop de server kan worden doorzocht.

Wanneer u een JDBC-database configureert, helpt een wizard u bij het opgeven van informatie waarmee het systeem verbinding kan maken met de database en bij het opgeven van opties voor de manier waarop de gegevens in de tabellen kunnen worden doorzocht. Voor elke tabel die u aan het systeem toevoegt, wordt een aparte doorzoekbare externe bron gemaakt.

Voor informatie over SQL Server 2000-drivers raadpleegt u <http://www.microsoft.com/downloads/details.aspx?familyid=07287B11-0502-461A-B138-2AA54BFDC03A&displaylang=en>. Voor informatie over SQL Server 2005-drivers raadpleegt u <http://www.microsoft.com/downloads/details.aspx?familyid=e22bc83b-32ff-4474-a44a-22b6ae2c4e17&displaylang=en>.

## Procedure

Ga als volgt te werk om een externe bron aan het systeem toe te voegen:

1. Als u JDBC-databases in een enterprise search-systeem wilt opnemen, voert u de volgende stappen uit voordat u een externe bron toevoegt. U moet deze stap eenmaal uitvoeren. Met deze stap worden de juiste JDBC-drivers in het systeem gezocht.
  - a. Meld u bij de crawlerserver aan als de enterprise search-beheerder.
  - b. Open het bestand `ES_INSTALL_ROOT/configurations/interfaces/discovery__interface.ini` in een editor en geef de omgevingsvariabelen `CLASSPATH` en `LD_LIBPATH` op om het klassepad aan de JDBC-drivers en het pad aan de bibliotheekbestanden toe te voegen.
  - c. Open het bestand `ES_INSTALL_ROOT/configurations/interfaces/customcommunication__interface.ini` in een editor en geef de omgevingsvariabelen `CLASSPATH` en `LD_LIBPATH` op om het klassepad aan de JDBC-drivers en het pad aan de bibliotheekbestanden toe te voegen.
  - d. Optioneel: als u een Oracle JDBC-driver wilt gebruiken voor lokale of gecalogiseerde databases, voegt u het Oracle-bibliotheekpad toe aan de omgevingsvariabele `LD_LIBPATH` (bijvoorbeeld `LD_LIBPATH=.../home/oracle/OraHome1/lib32`) en opent u het bestand `escrset.sh` in een editor om het bibliotheekpad op te geven en de Oracle-installatiedirectoryvariabele te exporteren. Bijvoorbeeld:

```
ORACLE_HOME=/home/oracle/OraHome1
export ORACLE_HOME
```
  - e. Start het enterprise search-systeem opnieuw, inclusief de CCL (Common Communication Layer):

### AIX, Linux of Solaris

```
esadmin stop
stopccl.sh
startccl.sh
esadmin start
```

### Windows-opdrachtaanwijzing

```
esadmin stop
```

```
stopccl
startccl
esadmin start
```

### Systeembeheer van Windows-services

Ga als volgt te werk om CCL op de achtergrond te starten:

- 1) Geef esadmin stop op.
  - 2) Start Windows-services: **Start** → **Programma's** → **Systeembeheer** → **Services**.
  - 3) Klik met de rechtermuisknop op **WebSphere Information Integrator OmniFind Edition** en klik op **Stop**. Als de service is gestopt, klikt u op **Start**.
  - 4) Geef esadmin start op.
2. Klik op **Externe bronnen** om de view Externe bronnen te openen.
  3. Klik op **Externe bron toevoegen**.
  4. Selecteer het type externe bron dat u wilt toevoegen (LDAP-server of JDBC-database).
  5. Klik op **Volgende** om de configuratie van de externe bron te starten.

Er wordt een wizard geopend voor het type bron dat u wilt maken. Volg de instructies in de wizard om de externe bron te configureren. Als u meer informatie wilt over de opties die u kunt opgeven, klikt u op de verschillende pagina's in de wizard op **Help**.

De volgende standaard-JDBC-drivernamen en -locaties kunnen u helpen bij het configureren van verbidingsgegevens voor DB2 UDB- (DB2 Universal Database) en Oracle-databases:

#### DB2: Legacy JDBC-driver

Drivernaam: COM.ibm.db2.jdbc.app.DB2Driver  
Voorbeeldlocatie: *db2\_install\_root/java/db2java.zip*

#### DB2: Universal JDBC-driver

Drivernaam: com.ibm.db2.jcc.DB2Driver  
Voorbeeldlocaties:  
*db2\_install\_root/java/db2jcc.jar*  
*db2\_install\_root/java/db2jcc\_license\_cu.jar*

**Oracle** Drivernaam: oracle.jdbc.driver.OracleDriver  
Voorbeeldlocatie: *oracle\_home/jdbc/lib/ojdbc14.jar*

#### Microsoft SQL Server 2000

Drivernaam: com.microsoft.jdbc.sqlserver.SQLServerDriver  
Voorbeeldlocaties:  
*mssql\_jdbc\_home/lib/mssqlserver.jar*  
*mssql\_jdbc\_home/lib/msbase.jar*  
*mssql\_jdbc\_home/lib/msutil.jar*

#### Microsoft SQL Server 2005

Drivernaam: com.microsoft.sqlserver.jdbc.SQLServerDriver  
Voorbeeldlocatie: *install\_dir/sqljdbc\_1.0/locale/sqljdbc.jar*  
Bijvoorbeeld: *install\_dir/sqljdbc\_1.0/enu/sqljdbc.jar*

6. Als u opties hebt opgegeven voor het doorzoeken van de externe bron, klikt u op **Voltoeien**.

De nieuwe externe bron wordt in de view Externe bronnen weergegeven met de andere externe bronnen die aan het systeem zijn toegevoegd.

### Verwante onderwerpen

## Zoekprogramma's aan externe bronnen koppelen

Voordat u een externe bron kunt doorzoeken, moet u minimaal één zoekprogramma aan de bron koppelen.

### Voordat u begint

Om zoekprogramma's te koppelen aan de externe bronnen die met de programma's kunnen worden doorzocht, moet u de rol van beheerder van enterprise search hebben.

### Procedure

Ga als volgt te werk om een zoekprogramma aan een of meer externe bronnen te koppelen:

1. Klik op **Beveiliging** in de werkbalk van de beheerconsole.
2. Klik op de pagina Zoekprogramma's op **Zoekprogramma's configureren**.
3. Klik op de pagina Zoekprogramma's configureren op **Zoekprogramma toevoegen**.
4. Typ de naam van het zoekprogramma.
5. Selecteer de externe bronnen die met het programma kunnen worden doorzocht:
  - Klik op **Alle collecties en externe bronnen** als het zoekprogramma toegang moet krijgen tot alle externe bronnen die u aan het systeem hebt toegevoegd.
  - Klik op **Bepaalde collecties en externe bronnen** als het zoekprogramma alleen toegang moet krijgen tot de externe bronnen die u opgeeft.  
Als u deze optie selecteert, wordt er een lijst met collecties en externe bronnen afgebeeld. Schakel het aankruisvakje **Selecteren** in voor elke externe bron die het zoekprogramma kan doorzoeken.
6. Klik op **OK**.

### Verwante onderwerpen

Federators van de Search and Index API

---

## Beveiliging van enterprise search

Met behulp van de beveiligingsmethoden in enterprise search kunt u voorkomen dat onbevoegde toegang tot bronnen kan worden verkregen en kunt u beheerfuncties beperken tot bepaalde gebruikers.

Met enterprise search kunnen gebruikers een groot aantal gegevensbronnen doorzoeken. Om ervoor te zorgen dat alleen gebruikers met de juiste machtigingen toegang hebben tot bepaalde inhoud en tot de beheerconsole, kunt u beveiliging in enterprise search op verschillende niveaus instellen en afdwingen.

### Webserver

Het eerste beveiligingsniveau is het niveau van de webserver. Als u algemene beveiliging inschakelt in WebSphere Application Server, kunt u gebruikers toewijzen aan beheerdersrollen en kunt u de gebruikers verifiëren die het systeem beheren. Als een gebruiker zich bij de beheerconsole aanmeldt, zijn alleen de functies en collecties beschikbaar waarvoor de gebruiker beheerdersmachtigingen heeft.

In zoekprogramma's kunnen de beveiligingsmethoden van WebSphere Application Server ook worden gebruikt om te gebruikers te verifiëren die collecties doorzoeken.

### Beveiliging op collectieniveau

Wanneer u een collectie maakt, kunt u beveiliging op collectieniveau inschakelen. U kunt deze instelling niet wijzigen als de collectie is gemaakt. Als u geen beveiliging op collectieniveau inschakelt, kunt u later geen beveiligingsopties op documentniveau opgeven.

Als beveiliging op collectieniveau is ingeschakeld, hebt u de volgende mogelijkheden:

- Met de processen voor algemene analyses kunnen verschillende regels worden toegepast voor het indexeren van duplicaatdocumenten.
- U kunt opties configureren om beveiliging op documentniveau af te dwingen, zoals tijdens het crawlen beveiligingstokens koppelen aan de documenten, vereisen dat actuele legitimatiegegevens tijdens de queryverwerking moeten worden gevalideerd en aangeven of ankertekst in webdocumenten moet worden geïndexeerd.
- U kunt beveiliging afdwingen door zoekprogramma's (geen afzonderlijke gebruikers) toe te wijzen aan de collecties en externe bronnen die door de zoekprogramma's kunnen worden doorzocht. Vervolgens kunt u de standaardmethoden voor toegangsbesturing gebruiken om toegang tot de zoekprogramma's voor gebruikers te verlenen of weigeren.

Er is een verband tussen het inschakelen van de collectiebeveiliging en de kwaliteit van zoekopdrachten. Als u collectiebeveiliging inschakelt, worden voor alle documenten minder gegevens in de index opgenomen. Het gevolg hiervan is dat er minder resultaten voor bepaalde query's worden gevonden.

### Beveiliging op documentniveau

Bij het configureren van crawlers voor een collectie kunt u beveiliging op documentniveau inschakelen. U kunt bijvoorbeeld opties opgeven zodat tijdens het verzamelen van gegevens door crawlers, beveiligingstokens aan de gegevens worden toegewezen. De zoekprogramma's kunnen deze

tokens (die in de index bij de documenten worden bewaard) gebruiken om toegangsbesturing af te dwingen en om te garanderen dat alleen gebruikers met de juiste legitimatiegegevens in staat zijn query's op de gegevens uit te voeren en de resultaten te zien.

Voor bepaalde typen gegevensbronnen kunt u opties configureren om de aanmeldingsgegevens van een gebruiker tijdens de verwerking van query's te vergelijken met de actuele toegangsgegevens. Deze extra beveiligingslaag garandeert dat de bevoegdheden van een gebruiker in real-time worden gecontroleerd bij de native gegevensbron. Op die manier kunt u voorkomen dat er problemen ontstaan doordat de legitimatiegegevens van een gebruiker veranderen nadat een document en de bijbehorende beveiligingstokens zijn geïndexeerd.

Tijdens de fase voor de anker tekstverwerking in de algemene analyse wordt de tekst die in een document (het brondocument) verschijnt, normaal gesproken aan een ander document gekoppeld (het doeldocument) waarin de tekst niet noodzakelijkerwijs hoeft voor te komen. Bij het configureren van een webcrawler kunt u aangeven of u de anker tekst wilt uitsluiten van de index als de tekst links bevat naar een document die de webcrawler niet mag crawlen.

De beveiliging voor collecties gaat verder dan de verificatie- en toegangsbesturingsmethoden die in enterprise search kunnen worden gebruikt om de geïndexeerde inhoud te beveiligen. U kunt ook voorkomen dat kwaadwillende en niet-gemachtigde gebruikers toegang kunnen krijgen tot de gegevens die worden verzonden. Stel dat de zoekservers protocollen zoals SSL (Secure Sockets Layer), SSH (Secure Shell) en HTTPS (Secure Hypertext Transfer Protocol) gebruiken voor de communicatie tussen de indexserver en het zoekprogramma.

Via codering is aanvullende beveiliging nodig. Het wachtwoord voor de enterprise search-beheerder (dat tijdens de installatie van het product is opgegeven), wordt bijvoorbeeld in gecodeerde vorm opgeslagen. De wachtwoorden die de gebruikers in gebruikersprofielen opgeven, worden ook in gecodeerde vorm opgeslagen.

Voor een betere beveiliging moet u zorgen dat de serverhardware op de juiste wijze is geïsoleerd en beveiligd is tegen onrechtmatige toegang. Door een firewall te installeren, beveiligt u de enterprise search-servers tegen onrechtmatige toegang via een ander gedeelte van uw netwerk. Zorg ervoor dat er geen poorten openstaan op de enterprise search-servers. Configureer het systeem zodanig, dat alleen wordt geluisterd naar opdrachten op poorten die expliciet zijn toegewezen aan enterprise search-activiteiten en -programma's.

---

## Beveiliging tijdens de installatie

Via het installatieprogramma voor WebSphere Information Integrator OmniFind Edition wordt een omgeving geïmplementeerd waarmee de beveiliging kan worden afgedwongen wanneer gebruikers de enterprise search-collecties doorzoeken of beheren.

### Enterprise search-beheerders-ID

Tijdens het installatieproces wordt gevraagd om het gebruikers-ID en het wachtwoord van de enterprise search-beheerder. In het systeem worden de opgegeven legitimatiegegevens voor de volgende taken gebruikt:

- Het verifiëren van de enterprise search-beheerder tijdens het uitvoeren van beheertaken.



- Het maken van de interne database in enterprise search.
- Het starten van alle enterprise search-sessies of -processen.

Het gebruikers-ID dat tijdens de installatie wordt opgegeven, moet een geldig werkend systeemgebruikers-ID zijn met beheerdersbevoegdheden. Met het installatieprogramma worden de legitimatiegegevens (met de juiste codering) in de eigenschappenbestanden op de verschillende enterprise search-servers opgeslagen.

## Codering

Om vertrouwelijke gegevens te beveiligen, wordt er versleuteling ("encryption") gebruikt voor het coderen van de verificatiegegevens van alle berichten die via het enterprise search-systeem worden verzonden. Dit proces zorgt voor enige overhead omdat alleen de verificatie-ID's en -wachtwoorden worden gecodeerd. Ook de wachtwoorden die bij het systeem worden opgeslagen (in configuratiebestanden, de enterprise search-database, enzovoort) worden gecodeerd.

## Beveiliging in WebSphere Application Server

Als WebSphere Application Server niet eerder op de zoekserver is geïnstalleerd, wordt het product via het installatieprogramma geïnstalleerd waarbij de optie voor algemene beveiliging is uitgeschakeld. Als u de algemene beveiliging later inschakelt in WebSphere Application Server, is WebSphere Application Server verantwoordelijk voor de verificatie van de enterprise search-beheerder.

Als u algemene beveiliging inschakelt, moet u het ID en wachtwoord van de enterprise search-beheerder die tijdens de installatie zijn opgegeven, aan het gebruikersregister van WebSphere Application Server toevoegen (bijvoorbeeld een LDAP-adresboek).

Als u algemene beveiliging inschakelt nadat WebSphere II OmniFind Edition is geïnstalleerd, moet u de configuratiewaarden bijwerken en de opdracht eschangewaspw uitvoeren om de WebSphere Application Server-legitimatiegegevens in een enterprise search-eigenschappenbestand te coderen en op te slaan.

---

## Verificatie versus toegangsbesturing

Om te voorkomen dat niet-bevoegde gebruikers toegang hebben tot content en om de toegang tot beheerfuncties te besturen, biedt enterprise search ondersteuning voor gebruikersverificatie en -machtiging (toegangsbesturing).

### Gebruikersverificatie

Met het verificatieproces verifieert het systeem de identiteit van een gebruiker die toegang tot het systeem probeert te krijgen. Omdat toegangsbesturing meestal is gebaseerd op de identiteit van de gebruiker die toegang tot een resource aanvraagt, is verificatie van essentieel belang voor een effectieve beveiliging.

De verificatie van enterprise search-gebruikers wordt geïmplementeerd via legitimatiegegevens, die uit minimaal een gebruikers-ID en een wachtwoord bestaan.

Om de toegang van gebruikers tot de beheerconsole te verifiëren, wordt de verificatieondersteuning in WebSphere Application Server uitgebreid in enterprise search.

## Machtiging (toegangsbesturing)

Machtiging is de methode waarmee het recht voor toegang tot bepaalde gegevens of het uitvoeren van een bepaalde actie door het systeem wordt verleend of afgevoerd. Vaak moet een gebruiker zich bij het systeem aanmelden met een bepaalde manier van verificatie. Op basis van de methoden voor toegangsbesturing wordt bepaald welke bewerkingen de gebruiker kan uitvoeren door de identiteit van de gebruiker te vergelijken met een toegangslijst. Toegangsbesturing bestaat uit het volgende:

- Bestandsmachtigingen, zoals het recht om een bestand te maken, lezen, bewerken of wissen.
- Programmamachtigingen, zoals het recht om een programma uit te voeren.
- Gegevensmachtigingen, zoals het recht om gegevens in een database op te halen of bij te werken.

---

## Beheerdersrollen

Enterprise search gebruikt het concept van rollen om de toegang tot verschillende functies in de beheerconsole te besturen.

Als WebSphere Information Integrator OmniFind Edition (WebSphere II OmniFind Edition) is geïnstalleerd, wordt in het installatieprogramma een gebruikers-ID en wachtwoord voor de enterprise search-beheerder geconfigureerd. De eerste keer dat u de beheerconsole opent, moet u zich als deze gebruiker aanmelden. Als algemene beveiliging niet is ingeschakeld in WebSphere Application Server, is dit gebruikers-ID het enige gebruikers-ID dat u kunt gebruiken om toegang te krijgen tot de beheerconsole van enterprise search.

Als u algemene beveiliging inschakelt in WebSphere Application Server, kunt u extra gebruikers configureren als enterprise search-beheerders. Als u gebruikers toewijst aan rollen, kunt u de toegang tot bepaalde collecties beperken en controle uitoefenen over de functies die elke beheerder kan uitvoeren. De gebruikers-ID's die u aan de beheerdersrollen in enterprise search toewijst, moeten aanwezig zijn in een gebruikersregister in WebSphere Application Server.

Als een beheerder zich aanmeldt, wordt het gebruikers-ID in enterprise search geverifieerd. Alleen de collecties en functies die de gebruiker mag beheren, zijn beschikbaar in de console.

U kunt de volgende beheerdersrollen voor gebruikers configureren:

### **Enterprise search-beheerder**

Deze gebruikers maken collecties en hebben de machtiging om alle aspecten van uw enterprise search-systeem te beheren. Wanneer u WebSphere II OmniFind Edition installeert, heeft u het gebruikers-ID en wachtwoord op voor de eerste enterprise search-beheerder. Nadat u voor de eerste keer bent aangemeld, kunt u aan meerdere gebruikers de rol van enterprise search-beheerder toekennen.

### **Collectiebeheerder**

Deze gebruikers kunnen de werking van de collecties die ze kunnen beheren, bewerken, bewaken en besturen. Deze gebruikers kunnen geen collecties maken. Collectiebeheerders kunnen alleen activiteiten op systeemniveau bewaken en uitvoeren wanneer ze hiertoe zijn gemachtigd door een enterprise search-beheerder.

### Operator

Deze gebruikers kunnen de werking van de collecties die ze kunnen bewerken, bewaken en besturen. Deze gebruikers kunnen bijvoorbeeld collectieactiviteiten starten en stoppen, maar ze kunnen geen collectie maken of bewerken. Operators kunnen alleen activiteiten op systeemniveau bewaken en uitvoeren wanneer ze hiertoe zijn gemachtigd door een enterprise search-beheerder.

### Monitor

Deze gebruikers kunnen de collecties die ze kunnen beheren, bewaken. Deze gebruikers kunnen geen bewerkingen besturen (zoals het starten en stoppen van servers), collecties maken of collecties bewerken. Een monitor kan alleen activiteiten op systeemniveau bekijken (maar niet uitvoeren) wanneer hij hiertoe door een enterprise search-beheerder is gemachtigd.

#### Verwante taken

“Aanmelden bij de beheerconsole” op pagina 21

Om een enterprise search-systeem te beheren, geeft u een URL in een webbrowser op en meldt u zich vervolgens aan bij de beheerconsole.

“Een enterprise search-systeem starten” op pagina 291

Om in te stellen dat gebruikers een collectie kunnen doorzoeken, moet u de systeemprocessen starten en vervolgens de servers starten waarmee de collectie wordt gecrawld, geanalyseerd en geïndexeerd (de zoekservers worden automatisch gestart).

## Beheerders configureren

Door beheerdersrollen te configureren, kunt u de toegang tot collecties beperken en controle uitoefenen op de functies die elke beheerder kan uitvoeren.

### Voordat u begint

Voordat u een gebruiker aan een beheerdersrol toewijst, moet u zorgen dat de beveiliging is ingeschakeld in WebSphere Application Server. Daarnaast moet u zorgen dat het gebruikers-ID voorkomt in een gebruikersregister in WebSphere Application Server.

Om beheerders te configureren, moet u de rol van enterprise search-beheerder hebben.

### Procedure

Ga als volgt te werk om gebruikers aan beheerdersrollen toe te wijzen:

1. Klik op **Beveiliging** om de view Beveiliging te openen.
2. Klik op de pagina Beheerdersrollen op **Gebruiker toevoegen**.
3. Voer het gebruikers-ID in van de gebruiker die u als beheerder wilt instellen en selecteer de juiste beheerdersrol.
4. Als u de gebruiker niet instelt als enterprise search-beheerder, geeft u aan of de gebruiker toegang tot de pagina's kan krijgen via de werkbalk **Systeem**.  
U kunt bijvoorbeeld instellen dat bepaalde operators of collectiebeheerders logbestanden op systeemniveau kunnen bewaken.
5. Als u de gebruiker niet instelt als enterprise search-beheerder, selecteert u de collecties en externe bronnen die de gebruiker kan beheren.

U kunt de aankruisvakjes voor afzonderlijke collecties en externe bronnen inschakelen of u kunt instellen dat de gebruiker alle collecties en externe bronnen kan beheren.

---

## Beveiliging op collectieniveau

Om beveiliging op collectieniveau in te stellen, configureert u opties voor het indexeren van de inhoud en opties waarmee zoekprogramma's bepaalde collecties kunnen doorzoeken.

Wanneer u een collectie maakt, kunt u een optie kiezen waarmee collectiebeveiliging wordt ingeschakeld. Als u deze optie kiest, kunt u later beveiligingsinstellingen op documentniveau configureren. Als collectiebeveiliging is ingeschakeld, worden met de processen voor algemene analyses in enterprise search ook verschillende regels voor het indexeren van duplicaatdocumenten toegepast.

Als u een zoekprogramma hebt gestart, kunt u op basis van het zoekprogramma-ID aangeven welke collecties en externe bronnen het zoekprogramma kan doorzoeken en welke gebruikers toegang tot het zoekprogramma hebben.

## Analyse van duplicaatdocumenten

Als u collectiebeveiliging inschakelt, wordt met behulp van de processen voor algemene analyses niet gecontroleerd of de collectie duplicaatdocumenten bevat.

Tijdens de algemene analyse worden documenten geïdentificeerd die (bijna) gelijk zijn aan elkaar. Vervolgens wordt aan al deze documenten één canonieke voorstelling van de inhoud gekoppeld. Als is ingesteld dat duplicaatdocumenten moeten worden geïdentificeerd, zorgt u dat de zoekresultaten niet meerdere documenten met (bijna) dezelfde inhoud bevatten.

Als u tijdens het maken van een collectie collectiebeveiliging inschakelt, worden duplicaatdocumenten niet geïdentificeerd en wordt er dus geen algemene canonieke voorstelling aan de documenten gekoppeld. In plaats daarvan wordt elk document afzonderlijk geïndexeerd. Op deze manier wordt ervoor gezorgd dat de beveiligingsinstellingen voor elk document worden geëvalueerd, zodat gebruikers alleen de documenten kunnen doorzoeken als de beveiligingstokens overeenkomen met hun legitimatiegegevens. Zo kan de inhoud van twee documenten nagenoeg identiek zijn, terwijl voor het afdwingen van beveiliging verschillende toegangslijsten worden gebruikt.

Als u de analyse van duplicaatdocumenten uitschakelt, kan de beveiliging van documenten in een collectie worden verbeterd maar kan de kwaliteit van de zoekprestaties worden verslechterd omdat meerdere exemplaren van hetzelfde document in de zoekresultaten kunnen worden afgebeeld.

## Zoekprogramma-ID's

Om in te stellen dat verschillende collecties kunnen worden doorzocht, wijst u de zoekprogramma's toe aan de collecties en externe bronnen die door de zoekprogramma's kunnen worden doorzocht. Met het programma Default kunt u het voorbeeldzoekprogramma gebruiken om alle collecties en externe bronnen te doorzoeken.

Alle zoekprogramma's moeten een geldige programmanaam (APPID) doorgeven aan de API (Application Programming Interface) van enterprise search. Alleen de collecties en externe bronnen die aan deze APPID zijn gekoppeld, kunnen met het zoekprogramma worden doorzocht.

Voordat een zoekprogramma toegang tot een collectie of externe bron heeft, moet een enterprise search-beheerder het zoekprogramma koppelen aan de collecties en bronnen die met het programma kunnen worden doorzocht. Een zoekprogramma kan alle collecties en externe bronnen in een enterprise search-systeem doorzoeken of alleen de collecties en externe bronnen die u opgeeft.

Het voorbeeldzoekprogramma (ESSearchApplication) bevat een eigenschappenbestand waarin wordt aangegeven welke programmanaam wordt gebruikt. De standaardlocatie van dit eigenschappenbestand is `ES_INSTALL_ROOT\installedApps\ESSearchApplication.ear\ESSearchApplication.war\WEB-INF\config.properties`.

De beginwaarde voor de programmanaam is Default. Als u deze waarde wijzigt, wijzigt u de lijst van collecties en externe bronnen die het programma `ESSearchApplication` kan doorzoeken.

Om te bepalen welke gebruikers welke collecties kunnen doorzoeken, koppelt u gebruikers (of gebruikersgroepen) aan het clientprogramma met behulp van de standaardtoegangsbesturingsfuncties in WebSphere Application Server. U doet dit op dezelfde manier als de manier waarop u deze functies gebruikt voor het beperken van de toegang tot een URL. U kunt bijvoorbeeld de toegang beperken tot een URL waarmee het zoekprogramma wordt gestart.

Zie de SI-API (Search and Index API) voor enterprise search voor meer informatie over zoekprogramma's en de manier waarop u beveiligingsfuncties in de aangepaste zoekprogramma's kunt implementeren.

#### **Verwante onderwerpen**

“Zoekprogramma's voor enterprise search” op pagina 203

Met behulp van een zoekprogramma kunt u collecties en externe bronnen in het enterprise search-systeem doorzoeken. U kunt een onbeperkt aantal zoekprogramma's maken en met elk programma kunt u een willekeurig aantal collecties en externe bronnen doorzoeken.

Overzicht van de Search and Index API

Beveiliging van de Search and Index API

---

## **Beveiliging op documentniveau**

Als bij het maken van de collectie de beveiliging is ingeschakeld, kunt u beveiligingsopties op documentniveau configureren. Als u beveiligingsopties op documentniveau configureert, zorgt u dat in de zoekresultaten alleen de documenten worden afgebeeld waarvoor de gebruiker die de zoekopdracht heeft ingediend, weergavemachtigingen heeft.

In een enterprise search-systeem kunt u de beveiligingsopties op documentniveau op verschillende manieren configureren:

- U kunt documenten vooraf filteren en koppelen aan beveiligingstokens voordat de documenten aan de index worden toegevoegd.
- Voor bepaalde gegevenstypen kunnen de zoekresultaten achteraf worden gefilterd om de legitimatiegegevens van de gebruikers te valideren met de huidige

toegangsbesturingsgegevens. In de component Identiteitenbeheer in enterprise search kunnen de verschillende legitimatiegegevens worden gecodeerd die gebruikers nodig hebben voor toegang tot de verschillende repository's en worden de gecodeerde legitimatiegegevens in profielen opgeslagen. Als de bronnen die moeten worden doorzocht, zijn beveiligd met een product dat SSO-beveiliging (Single Sign-On) biedt, kan de component Identiteitenbeheer de toegang besturen tot documenten zonder dat gebruikers hiervoor profielen hoeven te maken.

- Voor de meeste crawlertypen kan een aangepaste Java-klasse (plugin) worden gebruikt om de beveiligingstokens aan de documenten in de index te koppelen.
- Voor documenten die door een webcrawler worden gecrawld, kan de ankertekst in documenten met links naar verboden documenten van de index worden uitgesloten.

#### Verwante onderwerpen

“Zoekprogramma's voor enterprise search” op pagina 203

Met behulp van een zoekprogramma kunt u collecties en externe bronnen in het enterprise search-systeem doorzoeken. U kunt een onbeperkt aantal zoekprogramma's maken en met elk programma kunt u een willekeurig aantal collecties en externe bronnen doorzoeken.

Beveiliging van de Search and Index API

## Zoekresultaten vooraf en achteraf filteren

U kunt documenten op twee manieren filteren om ervoor te zorgen dat in de zoekresultaten alleen documenten worden afgebeeld waarvoor de gebruiker die de zoekopdracht heeft ingediend, weergavemachtigingen heeft.

- Bij de eerste manier worden de native toegangslijsten van het document tijdens het crawlen in de index gerepliceerd en wordt het zoekprogramma gebruikt om de legitimatiegegevens van de gebruiker te vergelijken met de toegangslijsten van het gerepliceerde document. De beste prestaties worden verkregen door de documenten vooraf te filteren en te bepalen welke documenten aan de index worden toegevoegd. Het is echter niet eenvoudig één model te maken voor alle beveiligingsrichtlijnen van de verschillende back-endbronnen in de index en uniforme vergelijkingslogica te implementeren. Deze manier is ook minder geschikt als er veel wijzigingen worden aangebracht in de brontoegangslijsten.
- Met de tweede manier worden de documenten in de resultaatset achteraf gefilterd met behulp van de beveiligingsgegevens in de back-endbronnen. Op deze manier kunnen de back-endbronnen uiteindelijk bepalen welke documenten naar de gebruikers worden teruggezonden en ervoor zorgen dat de resultaatset de huidige toegangsbesturingfuncties bevat. Deze manier van filteren zorgt echter voor lagere zoekprestaties omdat verbinding moet worden gemaakt met alle back-endbronnen. Als een bron niet toegankelijk is, moeten de links naar de documenten uit de resultaatset worden gefilterd, samen met de documenten waarvoor de gebruiker geen weergavemachtigingen heeft.

**Belangrijk:** In een configuratie met meerdere servers wordt het achteraf filteren voor bepaalde brontypen op de crawlerserver uitgevoerd. Als de crawlerserver niet beschikbaar is vanwege onderhoud, worden er geen resultaten weergegeven als gebruikers query's uitvoeren op de enterprise search-collectie. Daarnaast worden er geen resultaten teruggezonden als de back-endservers die vereist zijn voor de toegangsbesturing, niet toegankelijk zijn.

Voor enterprise search bestaat de ondersteuning voor het afdwingen van toegangsbesturing uit een combinatie van deze twee methoden. Het ontwerp biedt optimale prestaties terwijl de beveiligingsrichtlijnen van de oorspronkelijke

documentrepository's exact behouden blijven. Door toegangsbesturingsgegevens op hoog niveau in de index op te slaan, kan het systeem een tijdelijke (mogelijk kleinere) resultaatset weergeven die vervolgens achteraf kan worden gefilterd om de huidige toegangsbesturing te verifiëren. De gedachte is dat, als de gebruiker toegang heeft tot de repository die eigenaar is van het document, die gebruiker ook toegang heeft tot het document.

Welke toegangsbesturingsgegevens in de index worden opgeslagen, is afhankelijk van het crawlertype. Met de Notes-crawler kunnen bijvoorbeeld toegangsbesturingsgegevens op database- en serverniveau worden opgeslagen en met de QuickPlace-crawler kunnen toegangsbesturingsgegevens worden opgeslagen voor servers, ruimten en rooms.

Alle gegevensbrontypen in een enterprise search-systeem bieden ondersteuning voor de mogelijkheid om native toegangslijsten tijdens het crawlen in de index op te slaan. Daarnaast bieden sommige gegevensbrontypen ondersteuning voor de mogelijkheid om de resultaatset achteraf te filteren en de actuele legitimatiegegevens van de gebruiker te verifiëren (dit type ondersteuning wordt geleverd via native beveiligingsmethoden of via de component Identiteitenbeheer in enterprise search).

Dit tweetakkige beveiligingsontwerp omvat de volgende taken:

- Het extraheren van native toegangslijstgegevens tijdens het crawlen.
- Het in de index opslaan van toegangslijstgegevens van de server en database.
- Het maken van de beveiligingscontext van de gebruiker tijdens het aanmelden van de gebruiker of tijdens de sessie-initialisatie. Met deze taak worden de verschillende ID's gecontroleerd die een gebruiker nodig heeft om toegang te kunnen krijgen tot de verschillende back-endbronnen.
- Het verwerken van de zoekopdracht met behulp van de beveiligingscontext van de gebruiker en het maken van een tijdelijke resultaatset met alleen die documenten waartoe de gebruiker toegang heeft op repositoryniveau.
- Het achteraf filteren van de tijdelijke resultaatset met behulp van de back-endbronnen op basis waarvan documenten aan de resultaatset zijn toegevoegd voor de huidige native toegangslijstgegevens.

## Validatie door middel van beveiligingstokens

Als bij het maken van een collectie de beveiliging is ingeschakeld, kunt u beveiligingsopties op documentniveau configureren door de beveiligingsgegevens in de index op te slaan.

Standaard wordt aan elk document een openbaar token toegewezen, zodat het document voor iedereen beschikbaar is. Als beveiliging voor de collectie is ingeschakeld, kan het openbare token worden vervangen door een waarde die door de beheerder wordt geleverd of door een waarde die uit een veld in het gecrawld document wordt gehaald. Tijdens het configureren van een crawler geeft u op dat u beveiligingstokens wilt gebruiken om beperkingen in te stellen voor de gebruikers die toegang kunnen krijgen tot de documenten die door de desbetreffende crawler zijn gecrawld.

Als een crawler door de collectiebeheerder wordt geconfigureerd, kan de beheerder beveiligingsopties opgeven voor afzonderlijke tabellen, bestandssystemen, enzovoort (voor verschillende gegevensbronnen in de crawlruimte kunnen meerdere beveiligingsregels worden geconfigureerd). De beheerder kan:

- Opgeven dat de documenten openbaar zijn (zodat alle gebruikers de documenten kunnen doorzoeken)
- Door de gebruiker gedefinieerde beveiligingstokens aan alle documenten toewijzen
- Beveiligingstokens ophalen uit een veld in de gecrawld gegevens en het token aan alle documenten toewijzen

Beveiligingstokens worden volledig door de gebruiker gedefinieerd (met uitzondering van het standaard openbare token). Een beveiligingstoken kan een gebruikers-ID, een groeps-ID, een gebruikersrol of een andere waarde voorstellen die u hebt ingesteld als geldige waarde voor de gegevensbron.

Een beheerder kan bijvoorbeeld opgeven dat het veld `hrDeptName` moet worden gebruikt voor de toegangsbesturing van documenten die door een Notes-crawler zijn gecrawld. Daarnaast kan de beheerder aangeven dat als het veld niet in een document voorkomt of als het veld geen beveiligingsgegevens bevat, de twee door de gebruiker gedefinieerde tokens `hrgroup1` en `hrgroup2` moeten worden gebruikt voor de toegangsbesturing tot documenten.

De beveiligingstokens worden in de crawler beschikbaar gemaakt via het configuratiebestand van de crawler. Voor elk document levert de crawler de waarde van het beveiligingstoken als metagegevens. De indexercomponent leest het beveiligingstoken en past dit toe op de plaatsingsgegevens voor het document in de index. Als de beheerder de toegangslijst bijwerkt voor de native gegevensbron, worden de bijgewerkte beveiligingsinstellingen beschikbaar wanneer de hoofd- of delta-index opnieuw wordt opgebouwd.

U kunt aangepaste bedrijfsregels toepassen om de waarde van de beveiligingstokens te bepalen door de regels in een Java-klasse te coderen. Wanneer u de crawlereigenschappen configureert, geeft u de naam op van de plugin die de crawler moet gebruiken bij het crawlen van documenten. De beveiligingstokens die door de plugin worden toegevoegd, worden in de index opgeslagen en kunnen worden gebruikt om de toegang tot documenten te besturen.

## Gebruik van beveiligingstokens in zoekprogramma's

Het is de verantwoordelijkheid van het clientzoekprogramma om de beveiligingstokens tijdens het zoeken te leveren, zodat de documenten op de juiste manier kunnen worden gefilterd. Als er geen beveiligingstoken wordt geleverd, wordt het standaard openbare token automatisch toegepast tijdens de verwerking van de zoekopdracht.

Het voorbeeldzoekprogramma voor enterprise search bevat een demonstratie van de manier waarop u beveiliging op documentniveau kunt implementeren. In dit voorbeeld wordt aangenomen dat de beheerder een waarde voor het beveiligingstoken heeft toegewezen aan een groep documenten (in plaats van het beveiligingstoken op te halen uit een veld in een gecrawld document). De zoekprogramma gebruikt het `aanmeld_ID` van de gebruiker om te bepalen tot welke documenten de gebruiker toegang heeft. In plaats van het actuele gebruikers-ID te gebruiken, gebruikt het zoekprogramma het ID van de groep waarvan de gebruiker deel uitmaakt. Door een groeps-ID als beveiligingstoken te gebruiken, kunnen gebruikers worden toegevoegd aan en worden verwijderd uit de groep zonder dat hiervoor de index opnieuw hoeft te worden opgebouwd.



Het beveiligingstoken dat door de beheerder aan een set documenten is toegewezen, is een geldig groeps-ID in het besturingssysteem. Aan de verschillende documenten in de crawlruimte worden andere groeps-ID toegewezen. Bijvoorbeeld:

Document1-5: beveiligingstoken = Groep1  
Document6-10: Beveiligingstoken = Groep2

## Validatie van actuele legitimatiegegevens tijdens queryverwerking

Als bij het maken van een collectie de beveiliging is ingeschakeld, kunt u voor bepaalde typen domeinen de actuele legitimatiegegevens van een gebruiker valideren wanneer de gebruiker een query indient.

Voordat op een query wordt gereageerd, maken de zoekservers een koppeling met de native repository's om de actuele machtigingen van de gebruiker te valideren, waarna alle documenten uit de zoekresultaten worden verwijderd waarvoor de gebruiker geen weergavemachtigingen heeft.

Tijdens het configureren van de volgende typen crawlers kunt u een optie selecteren waarmee de legitimatiegegevens van gebruikers worden gevalideerd door de legitimatiegegevens te vergelijken met de actuele toegangsbesturingsgegevens die in de native repository worden beheerd. Nadat documenten zijn gecrawld en geïndexeerd, wordt de component Identiteitenbeheer van enterprise search gebruikt om de gebruikers te valideren die de beveiligde collecties proberen te doorzoeken.

- Content Edition-crawler (Alleen voor de repositorytypen Documentum, FileNet Panagon Content Services, Hummingbird DM, Portal Document Manager en SharePoint)
- DB2 Content Manager-crawler
- Domino Document Manager-crawler
- Notes-crawler
- QuickPlace-crawler
- Windows-bestandssysteem-crawler

Voor de volgende typen crawlers kunnen de actuele legitimatiegegevens van gebruikers worden gevalideerd wanneer de gebruikers de zoekportlet in WebSphere Portal gebruiken om de enterprise search-collecties te doorzoeken.

- Web Content Management-crawler
- WebSphere Portal-crawler

### Verwante onderwerpen

“Beveiliging op documentniveau afdwingen voor documenten van het Windows-bestandssysteem” op pagina 263

Om in te stellen dat de actuele legitimatiegegevens moeten worden gevalideerd wanneer een gebruiker documenten doorzoekt die zijn gecrawld door een Windows-bestandssysteem-crawler, moet u de domeinaccountgegevens op zowel de crawlerserver als de Microsoft Windows-server configureren.

“Beveiliging op documentniveau afdwingen voor Lotus Domino-documenten” op pagina 259

Als de te crawlen Domino-server gebruikmaakt van het NRPC-protocol (Notes Remote Procedure Call), moet u de crawlerserver zodanig configureren, dat toegangsbesturing op documentniveau kan worden afgedwongen.

### Verwante taken

“Lotus Domino Trusted Servers configureren voor het controleren van legitimatiegegevens van gebruikers” op pagina 260  
Om beveiliging af te dwingen voor documenten die zijn gecrawld met een Notes-crawler die gebruikmaakt van het NRPC-protocol (Notes Remote Procedure Call), moeten de te crawlen Domino-servers worden geconfigureerd als Lotus Domino Trusted Servers.

## **Identiteitenbeheer voor enterprise search**

Het beheer van legitimatiegegevens voor meerdere gebruikers is een veelvoorkomend probleem in bedrijven. Dankzij de optionele component Identiteitenbeheer in enterprise search behoort dit probleem nu tot het verleden.

De gegevens in een bedrijf kunnen op verschillende manieren zijn ingedeeld. Zo kunnen de gegevens in het hele bedrijf worden gedistribueerd en kunnen de gegevens worden beheerd met behulp van de meest geschikte beheerssoftware die voorhanden is. Gebruikers kunnen bijvoorbeeld een SQL-programma gebruiken om toegang te krijgen tot relationele databases of kunnen een documentbeheersysteem gebruiken om toegang te krijgen tot relevante documenten.

Het besturen van de toegang tot gevoelige informatie in deze repository's wordt doorgaans afgehandeld door de beheerssoftware. Gebruikers identificeren zichzelf bij het hostsysteem door middel van een gebruikers-ID en een wachtwoord. Na de verificatie door het systeem wordt in de beheerssoftware op basis van de gedefinieerde toegangsrechten bepaald welke documenten de gebruiker mag bekijken en welke acties de gebruiker mag uitvoeren.

Het is gebruikelijk dat gebruikers over meerdere gebruikers-ID's en wachtwoorden beschikken voor de verschillende repository's. Op dezelfde manier als gebruikers worden gevraagd zichzelf te identificeren bij de oorspronkelijke bedrijfsrepository's, moeten gebruikers legitimatiegegevens opgeven voordat ze documenten kunnen bekijken in een enterprise search-collectie waarvoor de actuele legitimatiegegevens moeten worden gevalideerd. Gebruikers die over meerdere identiteiten beschikken, moeten voor elke identiteit de bijbehorende legitimatiegegevens opgeven.

Als u in de beheerconsole opgeeft dat u enterprise search wilt gebruiken voor Identiteitenbeheer, kunnen de zoekservers de actuele legitimatiegegevens van gebruikers op de volgende manieren valideren tijdens de queryverwerking:

- Het zoekprogramma kan de gebruiker vragen de legitimatiegegevens die nodig zijn om toegang te krijgen tot diverse domeinen, te registreren in een gebruikersprofiel. Dit profiel, dat in versleutelde vorm is opgeslagen in een beveiligde datastore, stelt gebruikers in staat om in de beveiligde domeinen te zoeken. Als er geen legitimatiegegevens zijn opgegeven voor een domein dat vereist dat de actuele legitimatiegegevens worden gevalideerd, worden documenten uit dat domein niet opgenomen in de zoekresultaten.
- Als de documenten in een collectie zijn gecrawld door een crawler die ondersteuning biedt voor SSO-beveiliging (Single Sign-On) en u opgeeft dat u SSO wilt gebruiken om de toegang tot documenten te besturen, worden de SSO-beveiligingsmethoden gebruikt om gebruikers te verifiëren voor de duur van een zoeksessie. De gebruiker hoeft geen profiel te maken waarin de legitimatiegegevens zijn vastgelegd, en hoeft ook geen gebruikers-ID en wachtwoord op te geven om beveiligde domeinen te doorzoeken.

Als gebruikers collecties doorzoeken die verlangen dat de actuele legitimatiegegevens tijdens het indienen van de query worden gevalideerd, kan het systeem gebruik maken van de profiel- of SSO-methode om al dan niet toegang te verlenen tot de documenten.

## De groepsgegevens van de gebruikers ophalen

Om de legitimatiegegevens van een gebruiker te valideren, moet de component Identiteitenbeheer de groepsgegevens van de gebruiker ophalen voor alle identiteiten van de gebruiker, waarna deze gegevens aan een USC-tekenreeks (User Security Context) moeten worden toegevoegd. Deze groepsgegevens worden gebruikt om de resultaten te filteren volgens de toegangsbesturingsgegevens die zijn opgeslagen in de enterprise search-index of volgens de SSO-verificatiegegevens. Het filteren gebeurt in de component Identiteitenbeheer met behulp van de SSO-tokens of met behulp van de legitimatiegegevens van gebruikers om een verbinding te maken met het back-endsysteem en de groepen aan te vragen waarvan de gebruiker deel uitmaakt.

Bij het configureren van de opties voor Identiteitenbeheer in de beheerconsole kunt u opgeven hoe vaak deze groepsgegevens moeten worden vernieuwd. U kunt de nieuwe groepsgegevens extraheren telkens wanneer de gebruiker zich bij het zoekprogramma aanmeldt of u kunt de groepsgegevens regelmatig extraheren (bijvoorbeeld elke drie dagen).

## Beveiliging zonder de component Identiteitenbeheer

Niet alle bedrijven willen meerdere identiteiten van de gebruikersgemeenschappen beheren met de enterprise search-component Identiteitenbeheer. Als u de component Identiteitenbeheer uitschakelt in de beheerconsole van enterprise search, is het genereren van de beveiligingscontexttekenreeks van de gebruiker de verantwoordelijkheid van het zoekprogramma. Nadat deze tekenreeks is gegenereerd, wordt de USC-tekenreeks gebruikt om de waarde van de toegangslijstbeperkingen voor elke query in te stellen. Bijvoorbeeld:

```
Query q = factory.createQuery("IBM");
q.setACLConstraints("User's Security Context in XML");
```

**Tip:** Om u te helpen bij het schrijven van een eigen functie voor Identiteitenbeheer, kunt u de SI-API-uitbreiding (Search and Index API) implementeren om besturing op programmeerniveau voor de database van Identiteitenbeheer in te stellen. Met deze API kunt u de USC genereren met Java-objecten, waarna de XML-tekenreeks automatisch wordt samengesteld.

De XML-querytekenreeks moet op de volgende manier zijn ingedeeld, waarbij ... de volledige XML-tekenreeks bevat:

```
@SecurityContext::'...'
```

De indeling van de XML-tekenreeks is als volgt:

```
<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>
  </identity>
</identities>
```

```

<properties>
  <property name="property_name">property_value</property>
  ...
</properties>
</identity>
...
</identities>

```

### **identities**

De waarde van het kenmerk `id` is het gebruikers-ID dat de gebruiker opgeeft tijdens het aanmelden bij het systeem.

### **ssoToken**

Optioneel: het LTPA-token (Lightweight Third-Party Authentication) dat voor de gebruiker wordt gemaakt voor de duur van de browsersessie. Deze parameter wordt alleen gebruikt als het doeldomein is ingeschakeld voor SSO en de crawler is geconfigureerd voor het gebruik van SSO-beveiliging.

### **identity**

De legitimatiegegevens van de gebruiker voor een bepaalde gegevensbron. De waarde van het kenmerk `id` is het domein dat de legitimatiegegevens van de gebruiker bevat (voor Domino is dit de naam van het Domino-domein).

### **type**

Het type gegevens. Deze waarde komt overeen met het crawlertype (Notes, DB2, Exchange Server, enzovoort).

### **username**

De gebruikersnaam die wordt gebruikt voor het doorzoeken van het domein.

### **password**

Hiermee wordt het wachtwoord voor de opgegeven gebruikersnaam aangegeven. Het kenmerk `encrypt` moet zijn ingesteld op `no` (enterprise search levert geen coderingsmethode buiten de component Identiteitenbeheer).

### **groups**

De namen van de groepen waarvan de gebruiker deel uitmaakt. Voor elke groepsnaam wordt een afzonderlijk groeps-element gebruikt.

### **properties**

Een lijst van verbindingsspecifieke eigenschappen, zoals het beheerders-ID en het gecodeerde wachtwoord die zijn gebruikt voor het maken van de crawler. Daarnaast wordt hiermee aangegeven of SSO is ingeschakeld voor de bron.

#### **property\_name**

De naam van de eigenschap.

#### **property\_value**

De waarde van de eigenschap.

## **Validatie van gebruikers met gebruikersprofielen**

Zoekprogramma's kunnen gebruikers vragen om de legitimatiegegevens die vereist zijn voor toegang tot de verschillende domeinen in een gebruikersprofiel vast te leggen.

Om een domein te doorzoeken waarvoor de legitimatiegegevens van gebruikers moeten worden gevalideerd wanneer een query wordt ingediend, moeten gebruikers de legitimatiegegevens waarmee ze zich bij het domein aanmelden, in het zoekprogramma opnemen. Met behulp van Identiteitenbeheer in enterprise search kunnen gebruikers de legitimatiegegevens in een gebruikersprofiel opslaan voor een willekeurig aantal domeinen. De legitimatiegegevens worden gecodeerd en beveiligd in het enterprise search-systeem opgeslagen.

Als er geen legitimatiegegevens zijn opgegeven voor een domein dat vereist dat de actuele legitimatiegegevens worden gevalideerd, worden documenten uit dat domein niet opgenomen in de zoekresultaten.

Gebruikers kunnen een gebruikersprofiel maken en hun legitimatiegegevens vastleggen tijdens het gebruik van een zoekprogramma. In het voorbeeldzoekprogramma voor enterprise search wordt deze mogelijkheid geboden door de optie **Mijn profiel**. In de aangepaste zoekprogramma's kan deze mogelijkheid op een andere wijze zijn geïmplementeerd.

Collecties kunnen documenten van verschillende gegevensbronnen bevatten. Een collectie kan bijvoorbeeld documenten bevatten die zijn gecrawld vanuit een Windows-bestandssysteem en verschillende Lotus Notes -databases. In de component Identiteitenbeheer wordt onderscheid gemaakt tussen de verschillende brontypen en wordt alleen om legitimatiegegevens gevraagd als deze nodig zijn voor toegang tot domeinen waarvoor validatie vereist is.

Standaard zijn legitimatiegegevens vereist voor zoekopdrachten en moeten gebruikers dus het gebruikers-ID en wachtwoord opgeven voor beveiligde domeinen. Als de gebruiker het gebruikers-ID of wachtwoord voor een bepaald domein is vergeten, kan het domein worden uitgeschakeld voor zoekopdrachten met behulp van het bijbehorende aankruisvakje. Als u een domein uitschakelt, voorkomt u dat de beveiligde documenten in het domein in een resultaatset worden afgebeeld.

Nadat een gebruiker een profiel heeft gemaakt, kan de gebruiker een zoekopdracht indienen. De component Identiteitenbeheer bevat de gegevens die nodig zijn om de USC-tekenreeks (User's Security Context) samen te stellen voor gebruik bij volgende zoekopdrachten. Als u geen gebruikmaakt van deze component, moet de USC-tekenreeks door het zoekprogramma worden geleverd wanneer de gebruiker een zoekopdracht uitvoert in domeinen waarvoor de actuele legitimatiegegevens moeten worden gevalideerd.

De volgende keer dat de gebruiker de enterprise search-collecties probeert te doorzoeken, herhaalt Identiteitenbeheer het verificatieproces voor de legitimatiegegevens, maar is dit keer in staat het gebruikersprofiel te raadplegen. Als er geen gegevens zijn gewijzigd, wordt automatisch de pagina geopend waarop de gebruiker zoekopdrachten kan indienen en wordt de gebruiker niet gevraagd een profiel te maken.

Als Identiteitenbeheer een wijziging in de legitimatiegegevens van de gebruiker aantreft, wordt automatisch de profielpagina afgebeeld wanneer het zoekprogramma wordt geopend. Deze situatie kan ontstaan als het wachtwoord bijvoorbeeld is gewijzigd voor een domein dat is ingeschakeld voor zoekopdrachten of als een domein waarvoor verificatie is vereist aan de collectie is toegevoegd.

Gebruikers kunnen het bericht negeren waarin het bijwerken van het profiel wordt aanbevolen. Het gevolg hiervan is echter dat de desbetreffende documenten niet in de zoekresultaten worden afgebeeld.

In het voorbeeldzoekprogramma bij enterprise search kunnen gebruikers hun profielen op elk gewenst moment bijwerken met **Mijn profiel** op de werkbalk.

## Validatie van gebruikers met SSO-beveiliging

Als documenten in een collectie zijn gecrawld met een crawler die SSO-beveiliging (Single Sign-On) ondersteunt, kunt u tijdens het configureren van de opties voor Identiteitenbeheer opgeven dat u SSO-beveiliging wilt gebruiken voor de toegangsbesturing van documenten.

### SSO inschakelen

Met behulp van SSO-verificatie (Single Sign-On) worden gebruikers één keer geverifieerd, waarna ze toegang kunnen krijgen tot verschillende resources, zonder dat gebruikers steeds opnieuw de legitimatiegegevens hoeven op te geven. In een enterprise search-systeem zorgt u er met SSO-verificatie voor dat u de verschillende gebruikersnamen en wachtwoorden niet hoeft te beheren die gebruikers nodig hebben voor toegang tot documenten in beveiligde collecties.

IBM WebSphere Application Server en Lotus Domino ondersteunen een SSO-methode die ook wel LTPA (Lightweight Third-Party Authentication) wordt genoemd. Wanneer een gebruiker toegang probeert te krijgen tot een product, wordt de gebruiker gevraagd een gebruikersnaam en wachtwoord voor verificatie op te geven. Deze gebruikersnaam en dit wachtwoord worden geverifieerd op basis van een LDAP-repository die door beide producten gemeenschappelijk wordt gebruikt. Nadat de gebruiker is geverifieerd, wordt er een sessiecookie gemaakt voor het LTPA-token. De gebruiker kan vervolgens toegang krijgen tot alle andere resources op verschillende servers waarvoor dezelfde verificatieconfiguratie wordt gebruikt, zonder dat de gebruiker hiervoor de legitimatiegegevens opnieuw hoeft op te geven. Dit token blijft aanwezig voor de duur van de browsersessie.

Ga als volgt te werk om SSO-ondersteuning in te schakelen voor gebruik in enterprise search-collecties:

- Zorg dat algemene beveiliging en een geldig LDAP-register in WebSphere Application Server zijn ingeschakeld op de zoekservers voor enterprise search. Het LDAP-register kan een willekeurig LDAP-product zijn dat wordt ondersteund in WebSphere Application Server.
- Zorg dat de WebSphere-verificatiemethode is geconfigureerd voor gebruik van een actieve LTPA-verificatiemethode. Als u LTPA configureert, geeft u een geldige maar flexibele domeinnaam op, zoals uw.server.com.
- Zorg dat de LTPA-sleutel vanuit WebSphere Application Server is geëxporteerd en is geïmporteerd in andere producten in hetzelfde domein als het domein waarvoor u LTPA-ondersteuning wilt inschakelen.

Nadat u via de browser hebt gecontroleerd of de bovenstaande beveiligingsconfiguratie op de juiste manier werkt, kunt u de beheerconsole van enterprise search gebruiken om crawlers te configureren die SSO-verificatie ondersteunen.

### SSO en Identiteitenbeheer

Als gebruikers collecties doorzoeken waarvoor de actuele legitimatiegegevens moeten worden gevalideerd, kan het systeem gebruikmaken van SSO-beveiligingsmethoden om al dan niet toegang te verlenen tot de documenten. Gebruikers worden niet om legitimatiegegevens gevraagd als ze bronnen doorzoeken die SSO-verificatie ondersteunen. De component Identiteitenbeheer is nuttig als aan al de volgende voorwaarden is voldaan:

- SSO is op de juiste wijze ingeschakeld in WebSphere Application Server en op de doeldomeinen.

- De beveiliging is ingeschakeld in minimaal één van de collecties die door het zoekprogramma kunnen worden doorzocht.
- De opties voor het gebruik van de component Identiteitenbeheer en SSO-beveiliging zijn ingeschakeld in de beheerconsole van enterprise search.
- De optie voor het gebruik van SSO-beveiliging en de opties voor het afdwingen van beveiliging op documentniveau (zoals opties voor indextoegangsbesturing of het valideren van de actuele legitimatiegegevens tijdens de queryverwerking) zijn geselecteerd als de volgende crawlertypen zijn geconfigureerd:
  - Content Edition (alleen beschikbaar voor Portal Document Manager-repository's)
  - Domino Document Manager (alleen beschikbaar voor crawlers die het protocol DIIOP gebruiken)
  - Notes (alleen beschikbaar voor crawlers die het protocol DIIOP gebruiken)
  - QuickPlace (alleen beschikbaar voor crawlers die het protocol DIIOP gebruiken)

Als gebruikers de zoekportlet voor enterprise search gebruiken voor het doorzoeken van collecties vanuit WebSphere Portal, is SSO-beveiliging tevens beschikbaar voor documenten die zijn gecrawld door de Web Content Management- en WebSphere Portal-crawlers.

## Identiteitenbeheer configureren

U kunt de component Identiteitenbeheer voor enterprise search gebruiken om aan te geven hoe de legitimatiegegevens van gebruikers moeten worden gevalideerd tijdens de queryverwerking.

### Voordat u begint

Om de opties voor Identiteitenbeheer te configureren, moet u de rol van beheerder van enterprise search hebben.

### Over deze taak

Als gebruikers collecties doorzoeken waarvoor de actuele legitimatiegegevens tijdens de queryverwerking moeten worden gevalideerd, kunt u in de component Identiteitenbeheer voor enterprise search een gebruikersprofiel gebruiken voor SSO-beveiligingsmethoden (Single Sign-On) om al dan niet toegang te verlenen tot de documenten.

### Procedure

Ga als volgt te werk om Identiteitenbeheer te configureren:

1. Klik op **Beveiliging** om de view Beveiliging te openen.
2. Klik op de pagina Zoekprogramma's op **Identiteitenbeheer configureren**.
3. Schakel op de pagina Identiteitenbeheer configureren het aankruisvakje in voor het gebruik van de component Identiteitenbeheer voor enterprise search om te bepalen hoe de legitimatiegegevens van gebruikers tijdens de queryverwerking moeten worden gevalideerd. Als dit vakje niet is aangekruist, moet het zoekprogramma de USC-tekenreeks (User Security Context) aanleveren op het moment dat gebruikers een zoekopdracht opgeven voor domeinen die verlangen dat de actuele legitimatiegegevens tijdens de verwerking van de query worden gecontroleerd.
4. Geef op hoe vaak de component Identiteitenbeheer de legitimatiegegevens van een gebruiker moet ophalen uit de groepsrecords in het gebruikersregister van

WebSphere Application Server. U kunt instellen dat de legitimatiegegevens worden vernieuwd elke keer dat de gebruikers het zoekprogramma benaderen of u kunt opgeven dat de gegevens na een bepaald aantal dagen moeten worden vernieuwd.

5. Voor de crawlertypen die SSO-verificatie ondersteunen, kunt u opgeven of de component Identiteitenbeheer bij het valideren van gebruikers gebruik moet maken van de SSO-beveiligingstokens in plaats van de gebruikersprofielen. U kunt het aankruisvakje inschakelen voor het gebruik van SSO-verificatie voor alle crawlertypen of u kunt de aankruisvakjes voor de afzonderlijke crawlertypen inschakelen.

**Belangrijk:** In de component Identiteitenbeheer worden de SSO-beveiligingsmethoden alleen gebruikt als SSO-beveiliging op de juiste wijze is geconfigureerd in WebSphere Application Server en op de doeldomeinen.

## Analyse van ankertekst

Als u collectiebeveiliging inschakelt, worden met behulp van de processen voor algemene analyses speciale regels toegepast waarmee de ankertekst wordt geïndexeerd in documenten die door webcrawlers worden gecrawld. Als u collectiebeveiliging niet inschakelt, kunt u tijdens het configureren van afzonderlijke websites opgeven of de ankertekst in links naar verboden documenten moet worden geïndexeerd.

Ankertekst is de informatie in een hyperlink waarmee de pagina wordt beschreven die op basis van de link wordt geopend. In de volgende link is de tekst `Querysyntax` de ankertekst in een link naar de pagina `syntax.htm`:

```
<a href=" ../doc/syntax.htm">Querysyntax</a>
```

Meestal volgt de webcrawler links in documenten om extra documenten te crawlen en worden deze gekoppelde pagina's in de index opgenomen. Tijdens de algemene analyse koppelen de indexprocessen de ankertekst niet alleen aan het document waarin de tekst is ingesloten (het brondocument), maar ook aan het doeldocument. In het bovenstaande voorbeeld is de ankertekst `Querysyntax` gekoppeld aan de doelpagina `syntax.htm` en aan de bronpagina die de ankerconstructie bevat. Op basis van deze koppeling kan het doeldocument worden opgehaald door query's waarin tekst wordt opgegeven die in het brondocument voorkomt. De koppeling levert echter ook een beveiligingsrisico op als gebruikers toestemming hebben het doeldocument te bekijken, maar ze niet het brondocument mogen bekijken.

Als u tijdens het maken van een collectie collectiebeveiliging inschakelt, wordt de verwerking van ankertekst uitgeschakeld. De ankertekst wordt niet meer geïndexeerd bij een document, tenzij de tekst daadwerkelijk voorkomt in het document of in de metagegevens van het document. Deze beveiligingsinstelling zorgt ervoor dat gebruikers geen gegevens in documenten te zien krijgen waarvoor ze geen toestemming hebben. Documenten worden alleen in de zoekresultaten afgebeeld als de eigen inhoud of de metagegevens aan de query voldoen.

Het inschakelen van collectiebeveiliging kan de beveiliging van webdocumenten verbeteren als u instelt dat gebruikers alleen de documenten kunnen doorzoeken als de beveiligingstokens overeenkomen met hun legitimatiegegevens. Doordat de ankertekst niet wordt verwerkt, kan het echter voorkomen dat de zoekresultaten niet alle documenten bevatten die mogelijk relevant zijn voor een query.



Als u collectiebeveiliging niet inschakelt, kunt u tijdens het configureren van geavanceerde webcrawlereigenschappen opgeven of de ankertekst in links naar verboden documenten moet worden geïndexeerd.

### **Ankertekst in links naar verboden documenten indexeren**

Als een document links bevat naar documenten die de webcrawler niet mag crawlen, kunt u tijdens het configureren van een webcrawler aangeven of u de ankertekst voor deze links in de index wilt opnemen.

#### **Voordat u begint**

Om opties te configureren voor het indexeren van ankertekst, moet u de rol van beheerder van enterprise search hebben of beheerder van de collectie zijn voor de webcrawler die u wilt configureren.


#### **Over deze taak**

Op basis van instructies in het bestand robots.txt of op basis van de metagegevens van webdocumenten kan worden aangegeven dat de webcrawler geen toegang heeft tot bepaalde documenten op een website. Als een document dat de webcrawler mag crawlen links bevat naar verboden documenten, kunt u aangeven op welke manier de ankertekst van deze links moet worden verwerkt.

Tijdens het configureren van de webcrawler kunt u aangeven of de ankertekst naar verboden documenten in de index moet worden opgenomen. Voor een optimale beveiliging neem u ankertekst in links naar verboden documenten niet in de index op. Als u de ankertekst niet in de index opneemt, kan het voorkomen dat de zoekresultaten niet alle documenten bevatten die mogelijk relevant zijn voor een query.

#### **Procedure**

Ga als volgt te werk om de ankertekst in links naar verboden documenten in de index op te nemen of van de index uit te sluiten:

1. Open een collectie en zoek op de pagina Crawler de webcrawler die u wilt configureren. Klik vervolgens op  **Crawlereigenschappen**.
2. Klik op **Geavanceerde eigenschappen van webcrawler wijzigen**.
3. Om de ankertekst van alle documenten die door de crawler kunnen worden gecrawld in de index op te nemen, schakelt u het aankruisvakje **Ankertekst in links naar verboden documenten indexeren** in. Gebruikers kunnen zien welke pagina's niet door de webcrawler kunnen worden gecrawld door te zoeken naar tekst die zich bevindt in de ankertekst van links naar deze pagina's.  
Om ankertekst in links naar verboden documenten van de index uit te sluiten, schakelt u dit aankruisvakje uit. Gebruikers kunnen de pagina's die niet door de webcrawler kunnen worden gecrawld, dan niet bekijken. De ankertekst wordt niet in de index opgenomen, net als de verboden documenten.
4. Klik op **OK** en klik op de pagina Eigenschappen van webcrawler nogmaals op **OK**.
5. Om de wijzigingen te activeren, moet u de crawler stoppen en opnieuw starten.

Om de wijzigingen door te voeren voor documenten die al eerder zijn geïndexeerd, moet u die documenten opnieuw crawlen zodat ze opnieuw kunnen worden geïndexeerd. Als tijdens een eerdere crawl gegevens over verboden documenten aan de index zijn toegevoegd, worden deze gegevens uit de index verwijderd.

---

## Beveiliging inschakelen voor enterprise search

Als u van plan bent beveiliging af te dwingen wanneer gebruikers een enterprise search-systeem beheren of doorzoeken, moet u algemene beveiliging inschakelen in WebSphere Application Server. Daarnaast moet u de beveiligingsgegevens in de enterprise search-configuratiebestanden, op de beheerconsole en in de zoekprogramma's configureren.

### Procedure

Ga als volgt te werk om de beveiliging voor een enterprise search-systeem in te schakelen:

1. Bepaal welk type gebruikersregister u wilt gebruiken voor de verificatie van gebruikers. Veel WebSphere Application Server-beheerders kiezen bijvoorbeeld voor het gebruikersregister LDAP (Lightweight Directory Access Protocol).
2. Selecteer in de beheerconsole van enterprise search **Beveiliging** en wijs minimaal één gebruiker in het WebSphere Application Server-gebruikersregister toe aan de beheerdersrol **Enterprise search-beheerder**.

U kunt ook het enterprise search-beheerders-ID dat tijdens de installatie van WebSphere Information Integrator OmniFind Edition is opgegeven aan het WebSphere Application Server-gebruikersregister toevoegen.

**Belangrijk:** Als de algemene beveiliging is ingeschakeld, hebben alleen de gebruikers-ID's die in het gebruikersregister voorkomen en waaraan de enterprise search-beheerdersrol is toegewezen, toegang tot de beheerconsole en kunnen alleen deze gebruikers enterprise search beheren.

3. Volg de instructies in "Algemene beveiliging en LDAP-gebruikersregister configureren in WebSphere Application Server" om de algemene beveiliging in te schakelen en het gebruikersregister te configureren.
4. Als u algemene beveiliging nadat WebSphere Information Integrator OmniFind Edition is geïnstalleerd, moet u het gebruikers-ID en wachtwoord voor WebSphere Application Server doorgeven aan het enterprise search-systeem. U geeft deze gegevens door met behulp van de opdracht `eschangewaspw`:
  - Als u WebSphere Information Integrator OmniFind Edition op één server hebt geïnstalleerd, volgt u de instructies in "Beveiliging inschakelen voor een enterprise search-systeem met één server" op pagina 252.
  - Als u WebSphere Information Integrator OmniFind Edition op meerdere servers hebt geïnstalleerd, volgt u de instructies in "Beveiliging inschakelen voor een enterprise search-systeem met meerdere servers" op pagina 253.
5. Voltooi de taken voor het type documenten dat u wilt crawlen en doorzoeken. Zie "Vereisten in de crawlerinstellingen voor de ondersteuning van beveiliging" op pagina 254 voor meer informatie.

## Algemene beveiliging en LDAP-gebruikersregister configureren in WebSphere Application Server

Om de beveiliging in te schakelen in WebSphere Information Integrator OmniFind Edition, moet u eerst algemene beveiliging inschakelen in WebSphere Application Server.

### Over deze taak

Als onderdeel van het inschakelen van algemene beveiliging moet u een gebruikersregister configureren voor de verificatie van gebruikers-ID's. In deze

taak wordt beschreven hoe u tijdens het inschakelen van algemene beveiliging een LDAP-gebruikersregister (Lightweight Directory Access Protocol) kunt configureren in WebSphere Application Server. Hoewel andere typen gebruikersregisters in WebSphere Application Server worden ondersteund, kunt u geen gebruikmaken van het lokale besturingssysteemregister om enterprise search-gebruikers te verifiëren. Om het lokale besturingssysteemregister te gebruiken, moeten voor elke gebruiker in het bedrijf besturingssysteemgebruikersaccounts op de zoekservers voor enterprise search bestaan.

Deze taak is gebaseerd op WebSphere Application Server versie 6. Als u werkt met een eerdere versie van WebSphere Application Server, kunnen er andere standaardlabels voor de paden en gebruikersinterface worden afgebeeld. In deze taak wordt ook gebruikgemaakt van IBM Tivoli Directory Server voor het LDAP-register. Als u een ander type register of een aangepast register gebruikt, moet u de gegevens voor het desbetreffende register opgeven.

In deze taak vindt u een samenvatting van de stappen die nodig zijn om algemene beveiliging te configureren voor gebruik in een enterprise search-systeem. Voor uitgebreide instructies raadpleegt u het Informatiecentrum voor *WebSphere Application Server, versie 6.0.x* op de volgende URL: <http://publib.boulder.ibm.com/infocenter/wasinfo/v6r0/index.jsp>

## Procedure

Ga als volgt te werk om algemene beveiliging in te schakelen in WebSphere Application Server:

1. Ga op de zoekserver voor enterprise search naar de volgende URL om de beheerconsole van WebSphere Application Server te openen, waarbij *lokalehost* de naam van de lokalehost of de naam van de server is, zoals `omnifind.search.xyz.com`.  
`http://lokalehost:9060/ibm/console`
2. Klik op **Security** en klik vervolgens op **Global Security**.
3. Stel WebSphere in voor het gebruik van een LDAP-register:
  - a. Klik onder **User registries** op **LDAP**.
  - b. Geef het gebruikers-ID en wachtwoord voor de server op waarmee de programmaserver wordt uitgevoerd.
  - c. Selecteer IBM Tivoli Directory Server als het registertype.
  - d. Geef de LDAP-serverhostnaam op (een IP-adres of een DNS-hostnaam). Het standaardpoortnummer is 389.
  - e. Geef de basis-DN (Distinguished Name) op die het startpunt vormt voor het doorzoeken van het register, zoals `ou=sales,o=ibm,c=us`.
  - f. Omdat bepaalde LDAP-servers tijdens het doorzoeken van het register geen ondersteuning bieden voor anonieme binding, geeft u de DN voor de programmaserver op, zoals `cn=searchuser,o=ibm,c=us`, en geeft u vervolgens het wachtwoord voor de programmaserver op. De programmaserver gebruikt deze DN en dit wachtwoord voor binding met het register.
  - g. Om SSL-communicatie (Secure Sockets Layer) tussen WebSphere en LDAP te gebruiken, schakelt u het aankruisvakje **SSL enabled** in.
  - h. Klik op **Apply** en klik vervolgens op **OK**.
4. Schakel onder **General Properties** de aankruisvakjes **Enable global security** en **Enforce Java 2 security** in.

5. Selecteer Simple WebSphere Authentication Mechanism (SWAM) voor de actieve verificatiemethode.
6. Selecteer voor het actieve gebruikersregister Lightweight Directory Access Protocol (LDAP).
7. Klik op **OK**.
8. Klik op de link **Save** boven aan de pagina. Als u wordt gevraagd uw wijzigingen op te slaan, klikt u op de knop **Save**.
9. Klik op de werkbalk op **Logout**.
10. Stop het programma ESSearchServer en start het opnieuw.

#### **AIX, Linux of Solaris**

```
./stopServer.sh ESSearchServer
./startServer.sh ESSearchServer
```

#### **Windows**

```
stopServer ESSearchServer
startServer ESSearchServer
```

Deze scripts zijn opgeslagen in de directory WAS\_INSTALL\_ROOT/AppServer/bin:

- Voor WebSphere Application Server versie 5 is het standaardinstallatiepad /usr/WebSphere (op AIX-systemen), /opt/WebSphere (op Linux- of Solaris-systemen) of C:\Program Files\WebSphere (op Windows-systemen).
- Voor WebSphere Application Server versie 6 is het standaardinstallatiepad /usr/IBM/WebSphere (op AIX-systemen), /opt/IBM/WebSphere (op Linux- of Solaris-systemen) of C:\Program Files\IBM\WebSphere (op Windows-systemen).

11. Start de beheerconsole van WebSphere Application Server opnieuw.
12. Omdat de server nu opnieuw wordt gestart in de beveiligde werkstand, voert u bij het aanmelden van de console het gebruikers-ID en wachtwoord voor de server in dat u hebt opgegeven tijdens het configureren van het LDAP-gebruikersregister (zie stap 3b op pagina 251).

## **Beveiliging inschakelen voor een enterprise search-systeem met één server**

Als u de algemene beveiliging (global security) in WebSphere Application Server inschakelt na de installatie van WebSphere Information Integrator OmniFind Edition, moet u de opdracht eschangewaspw gebruiken om het configuratiebestand van enterprise search (es.cfg) bij te werken met het wachtwoord voor de gebruiker van WebSphere Application Server.

### **Voordat u begint**

Zorg dat het bestand config.properties voor het programma ESSearchApplication een geldige gebruikersnaam en een geldig wachtwoord voor WebSphere Application Server bevat. De standaardlocatie van dit bestand is ES\_INSTALL\_ROOT/installedApps/ESSearchApplication.ear/ESSearchApplication.war/WEB-INF.

### **Over deze taak**

Met de opdracht eschangewaspw wordt het wachtwoord gecodeerd voordat dit in het bestand es.cfg wordt opgeslagen.

## Procedure

Ga als volgt te werk om algemene beveiliging in te schakelen voor een bestaand enterprise search-systeem met één server:

1. Meld u bij de enterprise search-server aan als de enterprise search-beheerder.
2. Zorg dat het item WASUser in het bestand ES\_NODE\_ROOT/nodeinfo/es.cfg een geldige gebruikersnaam voor WebSphere Application Server bevat.
3. Voer het volgende script uit, waarbij *WAS\_wachtwoord* het wachtwoord is voor de WebSphere Application Server-gebruikersnaam dat is opgegeven in het bestand ES\_NODE\_ROOT/nodeinfo/es.cfg (zie stap 2).

### AIX, Linux of Solaris

```
eschangewaspw.sh WAS_wachtwoord
```

### Windows

```
eschangewaspw WAS_wachtwoord
```

4. Onder Windows selecteert u **Configuratiescherm** → **Systeembeheer** → **Services** en voegt u de gebruikersnaam en het wachtwoord van WebSphere Application Server toe aan de services WebSphere Application Server en ESSearchServer.
5. Stop het enterprise search-systeem en start het opnieuw:

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

## Beveiliging inschakelen voor een enterprise search-systeem met meerdere servers

Als u de algemene beveiliging (global security) in WebSphere Application Server inschakelt na de installatie van WebSphere Information Integrator OmniFind Edition, moet u de opdracht `eschangewaspw` gebruiken om het configuratiebestand van enterprise search (`es.cfg`) bij te werken met het wachtwoord voor de gebruiker van WebSphere Application Server.

### Voordat u begint

Zorg dat het bestand `config.properties` voor het programma `ESSearchApplication` een geldige gebruikersnaam en een geldig wachtwoord voor WebSphere Application Server bevat. De standaardlocatie van dit bestand is `ES_INSTALL_ROOT/installedApps/ESSearchApplication.ear/ESSearchApplication.war/WEB-INF` op de zoekservers.

### Over deze taak

Met de opdracht `eschangewaspw` wordt het wachtwoord gecodeerd voordat dit in het bestand `es.cfg` wordt opgeslagen.

## Procedure

Ga als volgt te werk om algemene beveiliging in te schakelen voor een bestaand enterprise search-systeem met meerdere servers:

1. Voer de volgende stappen uit op de indexservers van enterprise search:
  - a. Meld u aan als enterprise search-beheerder.

- b. Zorg dat het item WASUser in het bestand ES\_NODE\_ROOT/nodeinfo/es.cfg een geldige gebruikersnaam voor WebSphere Application Server bevat.
- c. Voer het volgende script uit, waarbij *WAS\_wachtwoord* het wachtwoord is voor de WebSphere Application Server-gebruiker dat is opgegeven in het bestand ES\_NODE\_ROOT/nodeinfo/es.cfg (zie stap 1b).

**AIX, Linux of Solaris**

```
eschangewaspw.sh WAS_wachtwoord
```

**Windows**

```
eschangewaspw WAS_wachtwoord
```

2. Voer de volgende procedure uit op de tweede zoekserver (in een configuratie met twee servers), of op de crawlerserver en beide zoekservers (in een configuratie met vier servers):
  - a. Meld u aan als enterprise search-beheerder.
  - b. Voer het volgende script uit, waarbij *WAS\_wachtwoord* het wachtwoord is voor de WebSphere Application Server-gebruiker dat is opgegeven in het bestand ES\_NODE\_ROOT/nodeinfo/es.cfg (zie stap 1b).

**AIX, Linux of Solaris**

```
eschangewaspw.sh WAS_wachtwoord
```

**Windows-opdrachtaanwijzing**

```
eschangewaspw WAS_wachtwoord
```

3. Onder Windows selecteert u **Configuratiescherm** → **Systeembeheer** → **Services** en voegt u de gebruikersnaam en het wachtwoord van WebSphere Application Server toe aan de services WebSphere Application Server en ESSearchServer.
4. Ga naar de zoekserver van enterprise search en geef de volgende opdrachten op om het enterprise search-systeem te stoppen en opnieuw te starten:

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

---

## Vereisten in de crawlerinstellingen voor de ondersteuning van beveiliging

Om informatie te verzamelen waarmee beveiliging op documentniveau kan worden afgedwongen, moeten de crawlers gemachtigd zijn voor toegang tot de native beveiligingsgegevens. Voor bepaalde gegevenssoorten moeten aanvullende stappen worden uitgevoerd om een beveiligde omgeving te kunnen configureren.

*Tabel 7. Vereisten in de crawlerinstellingen voor de ondersteuning van beveiliging*

---

### Content Edition-crawlers

---

Tabel 7. Vereisten in de crawlerinstellingen voor de ondersteuning van beveiliging (vervolg)

---

<p>Voordat u een crawler maakt voor toegang tot de repository's in de directe werkstand, configureert u het WebSphere Information Integrator Content Edition-systeem voor uitvoering in de directe werkstand en configureert u een connector voor de crawlerserver.</p> <p>Voordat u een crawler maakt voor toegang tot de repository's in de serverwerkstand, voert u het script <code>escrvbr.sh</code> (onder AIX, Linux of Solaris) of het script <code>escrvbr.vbs</code> (onder Windows) uit om de crawlerserver te configureren.</p> <p>Bij het configureren van de crawler geeft u een gebruikers-ID en wachtwoord op waarmee de crawler toegang kan krijgen tot alle repository's die u wilt crawlen. Indien noodzakelijk kunt u voor elke repository in de crawlruimte een andere gebruikers-ID en wachtwoord opgeven.</p>	<p>Zie ook:</p> <ul style="list-style-type: none"><li>• "Directe werkstand-toegang tot WebSphere II Content Edition-repository's" op pagina 42</li><li>• "Serverwerkstand-toegang tot WebSphere II Content Edition-repository's" op pagina 43</li><li>• "De crawlerserver configureren op UNIX voor WebSphere II Content Edition" op pagina 43</li><li>• "De crawlerserver configureren op Windows voor WebSphere II Content Edition" op pagina 44</li></ul>
<hr/> <b>DB2-crawlers</b>	
<p>Voordat u de crawler maakt, voert u het script <code>escrdb2.sh</code> (onder AIX, Linux of Solaris) of het script <code>escrdb2.vbs</code> (onder Windows) uit om de crawlerserver te configureren.</p> <p>Bij het configureren van de crawler voor het crawlen van niet-gecatalogiseerde databases op afstand, geeft u een gebruikers-ID en wachtwoord op waarmee elke database op de doelserver kan worden gecrawld. U kunt voor elke database in de crawlruimte een andere gebruikers-ID en wachtwoord opgeven.</p>	<p>Zie ook:</p> <ul style="list-style-type: none"><li>• "De crawlerserver configureren op UNIX voor DB2-crawlers" op pagina 47</li><li>• "De crawlerserver configureren op Windows voor DB2-crawlers" op pagina 48</li></ul>
<hr/> <b>DB2 Content Manager-crawlers</b>	
<p>Voordat u de crawler maakt, voert u het script <code>escrcm.sh</code> (onder AIX, Linux of Solaris) of <code>escrcm.vbs</code> (onder Windows) uit om de crawlerserver te configureren.</p> <p>Bij het configureren van de crawler geeft u een gebruikers-ID en wachtwoord op waarmee de crawler toegang kan krijgen tot elke server die u wilt crawlen. Indien noodzakelijk kunt u voor elke server in de crawlruimte een andere gebruikers-ID en wachtwoord opgeven.</p>	<p>Zie ook:</p> <ul style="list-style-type: none"><li>• "De crawlerserver configureren op UNIX for DB2 Content Manager-crawlers" op pagina 55</li><li>• "De crawlerserver configureren op Windows for DB2 Content Manager-crawlers" op pagina 57</li></ul>
<hr/> <b>Domino Document Manager-, Notes- en QuickPlace-crawlers</b>	

---

Tabel 7. Vereisten in de crawlerinstellingen voor de ondersteuning van beveiliging (vervolg)

---

Ga als volgt te werk om Lotus Domino-servers te crawlen waarop het NRPC-protocol (Notes Remote Procedure Call) wordt gebruikt:

- Zorg er op een AIX-systeem voor dat de I/O Completion Port-module is geïnstalleerd en beschikbaar is op de crawlerserver.
- Voordat u de crawler maakt, voert u het script `escrnote.sh` (onder AIX, Linux of Solaris) of het script `escrnote.vbs` (onder Windows) uit om de crawlerserver te configureren.
- Op de enterprise search-crawlerserver moet een Domino-server zijn geïnstalleerd en deze server moet deel uitmaken van het Domino-domein dat moet worden gecrawld.
- Om de actuele legitimatiegegevens van de gebruiker te kunnen valideren wanneer de gebruiker een zoekopdracht indient, moet de te crawlen Domino-server als Lotus Domino Trusted Server zijn geconfigureerd.
- Geef tijdens het configureren van de crawler het pad voor het ID-bestand voor de Lotus Notes-gebruiker op waarmee toegang tot de server kan worden verkregen (zoals `c:\Program Files\lotus\notes\data\name.id` of `/local/notesdata/name.id`). Geef ook het wachtwoord voor dit ID-bestand op.

Ga als volgt te werk om Lotus Domino-servers te crawlen waarop het DIIOP-protocol (Domino Internet Inter-ORB Protocol) wordt gebruikt:

- Zorg er op een AIX-systeem voor dat de I/O Completion Port-module is geïnstalleerd en beschikbaar is op de crawlerserver.
- Configureer de crawlerserver voor gebruik van het protocol.
- Geef tijdens het configureren van de crawler het volledige ID op van de Lotus Notes-gebruiker die is gemachtigd voor toegang tot de server (zoals Gebruikersnaam/Plaats/Mijn bedrijf). Geef ook het wachtwoord voor dit gebruikers-ID op.

Om QuickPlace-servers te crawlen, moet u de QuickPlace-server configureren voor ondersteuning van beveiliging van lokale gebruikers of Directory Assistance, afhankelijk van het type beveiliging dat u wilt gebruiken.

Zie ook:

- “De I/O Completion Port op AIX configureren voor het crawlen van Lotus Domino-bronnen” op pagina 73
- “De crawlerserver configureren op UNIX voor het crawlen van Lotus Domino-bronnen” op pagina 68
- “De crawlerserver configureren op Windows voor het crawlen van Lotus Domino-bronnen” op pagina 71
- “Lotus Domino Trusted Servers configureren voor het controleren van legitimatiegegevens van gebruikers” op pagina 260
- “Servers configureren die het DIIOP-protocol gebruiken” op pagina 72
- “De QuickPlace-server configureren voor gebruik van de beveiligingsoptie Lokale gebruiker” op pagina 261
- “Directory Assistance op een QuickPlace-server configureren” op pagina 262

---

## Exchange Server-crawlers

---



Tabel 7. Vereisten in de crawlerinstellingen voor de ondersteuning van beveiliging (vervolg)

---

Geef tijdens het configureren van de crawler het ID op van de gebruiker die is gemachtigd voor toegang tot de openbare mappen op de te crawlen Exchange-server. Geef ook het wachtwoord voor dit gebruikers-ID op.

Zie ook:

- “Toegang beheren tot beveiligde Exchange Server-documenten” op pagina 258

Als u voor de crawler Exchange Server-sleutelbeheer en het SSL-protocol (Secure Sockets Layer) wilt instellen bij het crawlen van gegevens, geeft u ook het volledige pad naar het sleutelarchiefbestand op en voert u het wachtwoord in waarmee de crawler toegang tot dit bestand kan krijgen. Het sleutelarchiefbestand moet bestaan op de crawlerserver van enterprise search.

---

#### JDBC database-crawlers

---

Bij het configureren van de crawler geeft u een gebruikers-ID en wachtwoord op waarmee de tabellen in de doeldatabase kunnen worden gecrawld. U kunt voor elke database in de crawlruimte een andere gebruikers-ID en wachtwoord opgeven.

---

#### NNTP-crawlers

---

Voor de te crawlen NNTP-servers moet zijn ingesteld dat de crawlerserver gegevens moet kunnen lezen.

---

#### UNIX-bestandssysteem-crawlers

---

Voor de te crawlen AIX-, Linux- en Solaris-subdirectory's moet zijn ingesteld dat de crawler-server gegevens moet kunnen lezen.

---

#### Webcrawlers

---

De webcrawler voldoet aan het Robots Exclusion Protocol. Als een webserver het hoogste niveau van de serverdirectory een bestand robots.txt bevat, analyseert de crawler dit bestand en worden alleen de websites op die server gecrawld waarvoor toestemming bestaat. Informatie over dit protocol vindt u op <http://www.robotstxt.org/wc/exclusion.html>.

Zie ook:

- “Door HTTP Basic Authentication beveiligde websites” op pagina 91
- “Door Form-Based Authentication beveiligde websites” op pagina 92

Bij het configureren van de webcrawler geeft u het volgende op:

- U moet een gebruikersagentnaam voor de crawler opgeven. In de regels van het bestand robots.txt van de te crawlen servers kan deze naam zijn opgenomen voor het verlenen of weigeren van toegang.
  - Optioneel: als een webserver voor het beperken van de toegang tot websites gebruikmaakt van HTTP Basic Authentication, kunt u legitimatiegegevens opgeven aan de hand waarvan de webcrawler toegang kan krijgen tot pagina's die met een wachtwoord zijn beveiligd.
  - Optioneel: als een webserver voor het beperken van de toegang tot websites gebruikmaakt van HTML-formulieren, kunt u legitimatiegegevens opgeven aan de hand waarvan de webcrawler toegang kan krijgen tot pagina's die met een wachtwoord zijn beveiligd.
-

Tabel 7. Vereisten in de crawlerinstellingen voor de ondersteuning van beveiliging (vervolg)

---

### Web Content Management- en WebSphere Portal-crawlers

---

Voordat u een crawler maakt, moet u een installatiescript uitvoeren om een WebSphere Portal-server in enterprise search te integreren. Voor de verschillende versies van WebSphere Portal zijn verschillende scripts aanwezig.

Zie ook:

- “Installatiescripts voor integratie tussen enterprise search en WebSphere Portal” op pagina 272

Tijdens het configureren van de crawler geeft u de volledige DN (Distinguished Name) op waarmee de crawler pagina's kan ophalen van de te crawlen server (bijvoorbeeld uid=admin,cn=RegularEmployees,ou=Software Group,o=IBM,c=US). Geef ook het wachtwoord voor deze DN op. De DN moet overeenkomen met de DN die voor de portalsite-URL in WebSphere Portal is geconfigureerd.

Zorg dat de machtigingen die u voor de gebruikers-DN opgeeft, zijn gedefinieerd in de component PAC (Portal Access Control) van WebSphere Portal. De crawler gebruikt de PAC om gegevens voor toegangsbesturing te verkrijgen voor de documenten die worden gecrawld.

---

### Windows-bestandssysteem-crawlers

---

Voor de te crawlen subdirectory's moet zijn ingesteld dat de server gegevens moet kunnen lezen. Geef tijdens het configureren van de crawler voor het crawlen van bestandssystemen op afstand een gebruikers-ID op waarmee de crawler toegang kan krijgen tot de gegevens op afstand. Geef ook een wachtwoord voor dit gebruikers-ID op.

Zie ook:

- “Beveiliging op documentniveau afdwingen voor documenten van het Windows-bestandssysteem” op pagina 263

Om de actuele legitimatiegegevens van de gebruiker te kunnen valideren wanneer de gebruiker een zoekopdracht indient, moet u zorgen dat de domeinaccounts op de juiste wijze zijn geconfigureerd. De vereisten voor het instellen van domeinaccounts voor bestanden die op de lokale computer zijn gecrawld, zijn anders van de vereisten voor bestanden die op een Windows-server op afstand zijn gecrawld.

---

#### Verwante onderwerpen

“Beheer van enterprise search-crawler” op pagina 35

U moet crawlers configureren voor de verschillende gegevenstypen die u in een collectie wilt opnemen. Een collectie kan een willekeurig aantal crawlers bevatten.

## Toegang beheren tot beveiligde Exchange Server-documenten

Als u een Exchange Server-crawler wilt gebruiken om documenten te crawlen die door een firewall worden beveiligd, moet u controleren of de crawlerserver toegang kan krijgen tot de server met de openbare map van Microsoft Exchange Server.

#### Over deze taak

Als de crawlerserver geen toegang kan krijgen tot een beveiligde Exchange Server-server, ontvangt u HTTP-code 501 (Niet geïmplementeerd) van de server. Daarnaast kunnen berichten verschijnen waarin wordt aangegeven dat er onverwachte HTTP-respons is ontvangen.

### **Procedure**

Ga als volgt te werk om ervoor te zorgen dat de crawlerserver toegang kan krijgen tot de documenten achter de firewall:

1. Start een webbrowser op de crawlerserver.
2. Ga naar de URL voor de server met de openbare map van Exchange Server die u wilt crawlen. Bijvoorbeeld: <http://exchange.uwBedrijf.com/public/>
3. Controleer of u de Exchange Server-pagina kunt openen.

Als u geen toegang kunt krijgen tot de Exchange Server-server, neemt u contact op met de serverbeheerder in uw organisatie.

#### **Verwante onderwerpen**

“Exchange Server-crawlers” op pagina 61

Als u openbare mappen van Microsoft Exchange Server in een enterprise search-collectie wilt opnemen, moet u een Exchange Server-crawler configureren.

## **Beveiliging op documentniveau afdwingen voor Lotus Domino-documenten**

Als de te crawlen Domino-server gebruikmaakt van het NRPC-protocol (Notes Remote Procedure Call), moet u de crawlerserver zodanig configureren, dat toegangsbesturing op documentniveau kan worden afgedwongen.

Om beveiliging op documentniveau af te dwingen voor documenten op een Domino-server waarvoor het NRPC-protocol wordt gebruikt, moet u een Domino-server op de crawlerserver installeren. Deze Domino-server moet deel uitmaken van uw Domino-domein. Volg de instructies in de documentatie bij Lotus Domino om de Domino-server te installeren en configureren.

Daarnaast moet u de volgende taken uitvoeren, zodat de zoekservers kunnen controleren of de gebruiker die de beveiligde collectie doorzoekt, is gemachtigd om de documenten te bekijken die aan de zoekcriteria voldoen:

- “Lotus Domino Trusted Servers configureren voor het controleren van legitimatiegegevens van gebruikers” op pagina 260.
- “Algemene beveiliging en LDAP-gebruikersregister configureren in WebSphere Application Server” op pagina 250.

#### **Verwante onderwerpen**

“Domino Document Manager-crawlers” op pagina 58

Als u bibliotheken en kabinetten van Domino Document Manager wilt opnemen in een enterprise search-collectie, moet u een Domino Document Manager-crawler configureren.

“Notes-crawlers” op pagina 65

Als u IBM Lotus Notes-databases wilt opnemen in een enterprise search-collectie, moet u een Notes-crawler configureren.

“QuickPlace-crawlers” op pagina 75

Als u Lotus QuickPlace-ruimten en -rooms in een enterprise search-collectie wilt opnemen, moet u een QuickPlace-crawler configureren.

“Validatie van actuele legitimatiegegevens tijdens queryverwerking” op pagina 241

Als bij het maken van een collectie de beveiliging is ingeschakeld, kunt u voor bepaalde typen domeinen de actuele legitimatiegegevens van een gebruiker valideren wanneer de gebruiker een query indient.

## **Lotus Domino Trusted Servers configureren voor het controleren van legitimatiegegevens van gebruikers**

Om beveiliging af te dwingen voor documenten die zijn gecrawld met een Notes-crawler die gebruikmaakt van het NRPC-protocol (Notes Remote Procedure Call), moeten de te crawlen Domino-servers worden geconfigureerd als Lotus Domino Trusted Servers.

### **Voordat u begint**

Deze procedure is vereist als u beveiliging op documentniveau wilt afdwingen tijdens het doorzoeken van databases op afstand. Om databases te doorzoeken die zich op de lokale crawlerserver bevinden, is deze procedure niet vereist.

Om Trusted Servers te configureren, moet in de crawler een Domino-server zijn geïnstalleerd. Deze Domino-server moet deel uitmaken van uw Domino-domein.

### **Over deze taak**

Als u beveiligingsopties op documentniveau configureert voor een Notes-crawler, geeft u aan of u toegangsbesturing wilt afdwingen door de actuele legitimatiegegevens van de gebruiker te valideren wanneer de gebruiker een query indient. Om dit type beveiliging af te dwingen, moeten de te crawlen Domino-servers Lotus Domino Trusted Servers zijn.

Als gebruikers een domein doorzoeken waarvoor de actuele legitimatiegegevens moeten worden gevalideerd, stelt de Trusted Server het Domino-server-ID in staat te context over te schakelen naar het huidige gebruikers-ID. De Domino-database wordt geopend als de huidige gebruiker de database heeft geopend en alle databasetoegangslijstgegevens voor de gebruiker worden afgedwongen.

De mogelijkheid om op deze manier over te schakelen naar een context is alleen beschikbaar voor databases die zijn opgeslagen in de directory data voor de lokale Domino-server. In Lotus Domino versie 6.5.1 wordt deze optie geleverd via de Trusted Server. Om de Trusted Server te configureren, geeft de Domino-beheerder op welke Domino-servers Trusted Servers zijn voor het uitvoeren van vertrouwelijke bewerkingen (als een gebruiker zich bijvoorbeeld als een andere gebruiker gedraagt als een database wordt benaderd vanaf een computer op afstand).

### **Procedure**

Om een Trusted Server te configureren, voert u de volgende stappen uit op alle Domino-servers die door een Notes-crawler zijn gecrawld:

1. Gebruik op een Domino-server het ID-bestand van de Domino-domein-beheerder om de Lotus Domino Administrator-client te openen.
2. Klik op **File** en selecteer **Open server**.
3. Voer de naam in van de Domino-server waarvoor u de Trusted Server-opties wilt inschakelen.
4. Selecteer de tab **Configuration**.

5. Vouw het object **Server** uit, selecteer het document **Current Server** en klik op **Edit Server**.
6. Selecteer de tab **Security**, blader naar het einde van het document, zoek het item **Trusted Servers** en druk op de pijl omlaag.
7. Geef een van de volgende opties op:

#### **LocalDomainServers**

Selecteer deze optie als alle servers in het Domino-domein moeten worden beschouwd als Trusted Servers.

#### *server\_name*

Geef de naam op van de Domino-server die moet kunnen crawlen en worden doorzocht als Trusted Server.

Als de te crawlen Domino-server op een ander domein is geïnstalleerd, moet u de servernaam opgeven of de groep **OtherDomainServers** selecteren. Daarnaast moet u de Domino-procedures uitvoeren voor kruiscertificaten tussen het ID-bestand van de Domino-server in enterprise search en het andere Domino-domein. Raadpleeg de documentatie bij de Domino-server voor informatie over deze procedures.

8. Klik op **Save and Close** om de wijzigingen op te slaan.
9. Stop en start de Domino-servers op afstand opnieuw die u hebt ingesteld als Trusted Servers.

#### **Verwante onderwerpen**

“Domino Document Manager-crawlers” op pagina 58

Als u bibliotheken en kabinetten van Domino Document Manager wilt opnemen in een enterprise search-collectie, moet u een Domino Document Manager-crawler configureren.

“Notes-crawlers” op pagina 65

Als u IBM Lotus Notes-databases wilt opnemen in een enterprise search-collectie, moet u een Notes-crawler configureren.

“QuickPlace-crawlers” op pagina 75

Als u Lotus QuickPlace-ruimten en -rooms in een enterprise search-collectie wilt opnemen, moet u een QuickPlace-crawler configureren.

“Validatie van actuele legitimatiegegevens tijdens queryverwerking” op pagina 241

Als bij het maken van een collectie de beveiliging is ingeschakeld, kunt u voor bepaalde typen domeinen de actuele legitimatiegegevens van een gebruiker valideren wanneer de gebruiker een query indient.

## **De QuickPlace-server configureren voor gebruik van de beveiligingsoptie Lokale gebruiker**

Als u van plan bent een QuickPlace-crawler te configureren voor gebruik van de optie Lokale gebruiker om beveiliging te implementeren, moet u de Domino-directory op de Lotus QuickPlace-server configureren voordat u de crawler maakt.

### **Over deze taak**

Tijdens het configureren van een QuickPlace-crawler selecteert u een beveiligingswerkstand voor de crawler zodat de beveiliging op documentniveau kan worden afgedwongen. Als u de werkstand Lokale gebruiker selecteert, moet u ervoor zorgen dat alle lokale gebruikers-ID's en lokale groepen zijn geregistreerd in de Domino-directory (de hiërarchie van de Domino-directory moet overeenkomen met de QuickPlace-hiërarchie).

Daarnaast moet u zorgen dat het gebruikers-ID en het wachtwoord dat u voor de crawler opgeeft, in de Domino-directory zijn geregistreerd en leesmachtiging hebben voor de te crawlen database.

Om QuickPlace te kunnen gebruiken, is alleen de gebruikersnaam vereist. Om de QuickPlace-bronnen te kunnen crawlen, is het volledige gebruikers-ID vereist. Het volledige gebruikers-ID moet de volgende indeling hebben:

*gebruikersnaam/placenaam/QP/domeinnaam*

Gebruik deze procedure om te bepalen wat het volledige gebruikers-ID is, te zorgen dat het ID leesmachtiging voor de QuickPlace-database heeft en het gebruikers-ID aan de Domino-directory toe te voegen. De Domino-directory moet het gebruikers-ID bevatten dat moet worden gebruikt om de QuickPlace-databases en alle lokale gebruikers en lokale groepen van QuickPlace te crawlen (de hiërarchie van de Domino-directory moet overeenkomen met de QuickPlace-hiërarchie).

### Procedure

Ga als volgt te werk om de QuickPlace-server te configureren voor gebruik van de beveiligingsoptie Lokale gebruiker:

1. Controleer de machtigingen voor het gebruikers-ID:
  - a. Open het serverdocument op de QuickPlace-server.
  - b. Open de pagina Files en open vervolgens de toegangslijst (ACL) voor de database die u wilt crawlen.
  - c. Controleer of het ID van de lokale gebruiker dat voor de crawler wordt geconfigureerd in de toegangslijst voorkomt en het ID leesmachtiging voor de database heeft.  
U moet het volledige gebruikers-ID (zie stap 2) opgeven.
2. Voeg de gebruiker toe aan de Domino-directory:
  - a. Open het serverdocument op de QuickPlace-server.
  - b. Voeg op de pagina People and Groups in het structurelement People het volledige gebruikers-ID toe (zie stap 1).
  - c. Geef in het veld **Internet password** het wachtwoord voor het gebruikers-ID op.

#### Verwante onderwerpen

“QuickPlace-crawlers” op pagina 75

Als u Lotus QuickPlace-ruimten en -rooms in een enterprise search-collectie wilt opnemen, moet u een QuickPlace-crawler configureren.

### Directory Assistance op een QuickPlace-server configureren

Als u van plan bent een QuickPlace-crawler te configureren voor gebruik van een LDAP-directory om beveiliging te implementeren, moet u een Directory Assistance-database op de Lotus QuickPlace-server maken voordat u de crawler configureert.

### Beperkingen

Op de QuickPlace-server die u wilt crawlen, moeten de DIOP- en HTTP-taken worden uitgevoerd.

### Procedure

Ga als volgt te werk om LDAP Directory Assistance op een QuickPlace-server te configureren:

1. Maak een Directory Assistance-database:
  - a. Open het serverdocument op de QuickPlace-server.
  - b. Maak een database met behulp van het sjabloon **Directory Assistance(6)**. Dit sjabloon bevindt zich op de server.
  - c. Klik op **Add Directory Assistance** om een document in de database te maken.
  - d. Open de tab Basic en selecteer in het veld **DomainType** de waarde **LDAP**.
  - e. Open de tab Naming Contexts en zorg dat het aankruisvakje **Trusted for credentials** is ingeschakeld.
  - f. Open de tab LDAP en voer de gegevens over de LDAP-server in.
  - g. Sla het serverdocument op en sluit het document.
2. Configureer de QuickPlace-server voor gebruik van de Directory Assistance-database:
  - a. Open het serverdocument op de QuickPlace-server.
  - b. Open de tab Basic en voer in het veld **Directory assistance database name** de naam op van de database die u in stap 1 hebt gemaakt.
  - c. Sla het serverdocument op en sluit het document.

De QuickPlace-server kan de LDAP-server nu gebruiken als tweede Domino-directory.

#### **Verwante onderwerpen**

“QuickPlace-crawlers” op pagina 75

Als u Lotus QuickPlace-ruimten en -rooms in een enterprise search-collectie wilt opnemen, moet u een QuickPlace-crawler configureren.

## **Beveiliging op documentniveau afdwingen voor documenten van het Windows-bestandssysteem**

Om in te stellen dat de actuele legitimatiegegevens moeten worden gevalideerd wanneer een gebruiker documenten doorzoekt die zijn gecrawld door een Windows-bestandssysteem-crawler, moet u de domeinaccountgegevens op zowel de crawlerserver als de Microsoft Windows-server configureren.

Als u een Windows-bestandssysteem-crawler configureert, geeft u op of u de subdirectory's op de lokale computer of de subdirectory's op een computer op afstand wilt crawlen. Als de beveiliging is ingeschakeld voor de collectie, kunt u ook opties opgeven om de toegang tot de documenten in de gecrawelde subdirectory's te besturen.

Als u ervoor kiest toegangsbesturing af te dwingen door de actuele legitimatiegegevens van gebruikers te valideren wanneer de gebruiker een query indient, moet u zorgen dat de domeinaccounts op de juiste manier zijn geconfigureerd. De vereisten voor het instellen van domeinaccounts voor bestanden die op de lokale computer zijn gecrawld, zijn anders van de vereisten voor bestanden die op een Windows-server op afstand zijn gecrawld.

**Belangrijk:** De legitimatiegegevens van de gebruikers kunnen niet tijdens de queryverwerking worden gevalideerd als aan de volgende twee voorwaarden wordt voldaan:

- De te crawlen Windows-server maakt geen deel uit van een domein.
- De te crawlen directory is een directory op afstand, zoals \\servernaam\hostnaam.

## Validatie met lokale toegangsbesturingsgegevens

Om de actuele legitimatiegegevens van gebruikers te valideren, maakt het systeem gebruik van zowel de lokale gebruikersaccountgegevens als de domeinaccountgegevens (als de computer deel uitmaakt van een Windows-domein). Om de legitimatiegegevens tijdens de queryverwerking te valideren, moeten beide gebruikersnamen voorkomen in de beveiligingsgegevens voor de te doorzoeken documenten.

### Lokale accounts

Voor een lokale account wordt de gebruikersaccount in de volgende indeling afgebeeld:

*COMPUTERNAAM\GEBRUIKERSNAAM*

Bij het aanmelden geven gebruikers alleen de gebruikersnaam op, maar voor de op de juiste wijze opgegeven toewijzing van Windows-gebruikersrechten wordt de volledige naam gebruikt. Als de gebruikersnaam voor de lokale account bijvoorbeeld abcuser is, kan de volledige accountnaam WINSERVER1\abcuser zijn.

Als gebruikers een zoekprogramma gebruiken en een profiel configureren voor het doorzoeken van beveiligde documenten op een lokaal systeem, moet de gebruikersnaam worden opgegeven die wordt gebruikt voor aanmelding bij Windows (bijvoorbeeld abcuser).

### Domeinaccounts

Voor een domeinaccount wordt de gebruikersnaam in de volgende indeling afgebeeld:

*DOMEINNAAM\GEBRUIKERSNAAM*

Om aan te melden, geven gebruikers deze gegevens in de volgende indeling op:

*GEBRUIKERSNAAM@DOMEINNAAM*

Als u bijvoorbeeld gebruikersrechten configureert voor een bestand en het domein WIN1\abcuser selecteert, wordt de account afgebeeld als abcuser@win1.bedrijf.com.

Als gebruikers een zoekprogramma gebruiken en een profiel configureren waarmee ze documenten in een beveiligd domein kunnen doorzoeken, moet de gebruikersnaam worden opgegeven die wordt gebruikt voor aanmelding bij Windows (bijvoorbeeld abcuser@win1.bedrijf.com).

Om de validatie van actuele legitimatiegegevens af te dwingen op lokale computers, moeten de gebruikersaccounts die door de crawlerserver worden gebruikt over de volgende gebruikersrechten voor Windows beschikken. (Om gebruikersrechten toe te wijzen, gebruikt u Systeembeheer van Windows: **Systeembeheer** → **Lokaal beveiligingsbeleid** → **Lokaal beleid** → **Toewijzing van gebruikersrechten**.)

- Aan het gebruikers-ID dat op de crawlerserver wordt uitgevoerd, moet het recht **Werken als deel van het besturingssysteem** zijn toegewezen. (Dit recht wordt tijdens de installatie van WebSphere Information Integrator OmniFind Edition voor de enterprise search-beheerder op de crawlerserver geconfigureerd.)
- Aan gebruikers moet het gebruikersrecht **Lokaal aanmelden** zijn toegewezen.



## Validatie met domeintoegangsbesturingsgegevens op afstand

Voor het Windows-besturingssysteem wordt elke directory die begint met \\servernaam beschouwd als een directory op afstand. Bijvoorbeeld:

```
\\software\utilities\IBM
```

Om toegang te krijgen tot een directory op afstand moeten gebruikers de gebruikersnamen in de volgende indeling opgeven:

```
GEBRUIKERSNAAM@DOMEINNAAM
```

Als gebruikers een zoekprogramma gebruiken en een profiel configureren waarmee ze beveiligde documenten op een systeem op afstand kunnen doorzoeken, moet de gebruikersnaam worden opgegeven die wordt gebruikt voor aanmelding bij het Windows-systeem op afstand (bijvoorbeeld abcuser@win1.bedrijf.com).

Om de validatie van actuele legitimatiegegevens af te dwingen op computers op afstand, moeten de gebruikersaccounts over de volgende gebruikersrechten voor Windows beschikken. (Om gebruikersrechten toe te wijzen, gebruikt u Systeembeheer van Windows: **Systeembeheer** → **Beveiligingsbeleid voor domeinen**.)

- De te doorzoeken crawlerserver en Windows-server moeten deel uitmaken van hetzelfde domein.
- Aan het gebruikers-ID dat op de crawlerserver wordt uitgevoerd, moet het recht **Werken als deel van het besturingssysteem** zijn toegewezen. (Dit recht wordt tijdens de installatie van WebSphere Information Integrator OmniFind Edition voor de enterprise search-beheerder op de crawlerserver geconfigureerd.)
- Aan gebruikers moet het gebruikersrecht **Aanmelden als batchtaak** zijn toegewezen.

### Verwante onderwerpen

“Windows-bestandssysteemcrawler” op pagina 105

Als u documenten die zijn opgeslagen in een Microsoft Windows-bestandssysteem wilt opnemen in een enterprise search-collectie, moet u een Windows-bestandssysteem-crawler configureren.

“Validatie van actuele legitimatiegegevens tijdens queryverwerking” op pagina 241

Als bij het maken van een collectie de beveiliging is ingeschakeld, kunt u voor bepaalde typen domeinen de actuele legitimatiegegevens van een gebruiker valideren wanneer de gebruiker een query indient.

---

## Beveiliging uitschakelen voor enterprise search

U kunt de beveiliging uitschakelen voor een enterprise search-programma in WebSphere Application Server. Als u eerder beveiligingsinstellingen op document-niveau hebt geconfigureerd, kunt u opgeven dat de instellingen moeten worden genegeerd. De beveiligingsinstellingen zijn ook van invloed op de manier waarop samengevouwen resultaten in de zoekresultaten worden weergegeven.

## Beveiliging uitschakelen voor een bedrijfstoepassing in WebSphere Application Server

Om te bepalen voor welke enterprise search-activiteiten verificatie nodig is, kunt u de algemene beveiliging voor afzonderlijke bedrijfstoepassingen uitschakelen in WebSphere Application Server.

Over deze taak

Het installatieprogramma van WebSphere Information Integrator OmniFind Edition installeert drie bedrijfstoepassingen in WebSphere Application Server:

- Het programma ESAdmin bevat de interface voor de beheerconsole van enterprise search.
- Het programma ESSearchApplication bevat de interface voor het voorbeeldzoekprogramma.
- Het programma ESSearchServer levert alle communicatie voor de SI-API-implementatie in enterprise search en schakelt de SI-API-interfaces in voor communicatie met de zoekservers.

Standaard ondersteunen alle drie de bedrijfstoepassingen algemene beveiliging van WebSphere Application Server global security. Als in deze programma's wordt ontdekt dat algemene beveiliging is ingeschakeld, wordt begonnen met de verificatie van alle opdrachten die in de programma's worden ontvangen.

Bepaalde organisaties kunnen de beveiliging voor bepaalde bedrijfstoepassingen in- of uitschakelen. U kunt bijvoorbeeld instellen dat alle gebruikers moeten worden geverifieerd die de beheerconsole van enterprise search openen, terwijl de gebruikers die de SI-API-interfaces of het voorbeeldzoekprogramma gebruiken, niet worden geverifieerd.

### Procedure

Ga als volgt te werk om de beveiliging uit te schakelen voor een bedrijfstoepassing:

1. Start op de zoekserver de beheerconsole van WebSphere Application Server. U kunt de beheerconsole op de volgende manieren openen:
  - Selecteer het programma via het menu **Start** van Windows.
  - Voor WebSphere Application Server versie 5 opent u een webbrowser en gaat u naar `http://hostnaam:poort/admin`, waarbij *hostnaam* de hostnaam van de zoekserver is en *poort* het poortnummer van de beheerconsole van WebSphere Application Server is. Het poortnummer van de beheerconsole is meestal 9090.
  - Voor WebSphere Application Server versie 6 opent u een webbrowser en gaat u naar `http://hostnaam:poort/ibm/console`, waarbij *hostnaam* de hostnaam is van de zoekserver en *poort* het poortnummer is voor de beheerconsole van WebSphere Application Server. Het poortnummer van de beheerconsole is meestal 9060.
2. Als u wordt gevraagd het gebruikers-ID en wachtwoord in te voeren, geeft u het gebruikers-ID en wachtwoord op dat is opgegeven tijdens het inschakelen van algemene beveiliging in WebSphere Application Server.
3. Nadat u zich hebt aangemeld bij de beheerconsole klikt u op **Programma's** en klikt u vervolgens op **Bedrijfstoepassingen**.
4. Schakel het aankruisvakje in naast de naam van de bedrijfstoepassing waarvoor u de beveiliging wilt uitschakelen.
5. Blader omlaag en klik op de link **Map security roles to users/groups**.
6. Zoek de rol **AllAuthenticated** en schakel het aankruisvakje in onder de kolom **Everyone?**.
7. Klik op **OK**.
8. Klik op de link **Save** om de wijzigingen op te slaan.
9. Als u werkt met WebSphere Network Deployment schakelt u het aankruisvakje **Synchronize changes with Nodes** in.

10. Klik op **Save**.
11. Stop het programma ESSearchServer en start het opnieuw.

#### **AIX, Linux of Solaris**

```
./stopServer.sh ESSearchServer  
./startServer.sh ESSearchServer
```

#### **Windows**

```
stopServer ESSearchServer  
startServer ESSearchServer
```

Deze scripts zijn opgeslagen in de directory WAS\_INSTALL\_ROOT/AppServer/bin:

- Voor WebSphere Application Server versie 5 is het standaardinstallatiepad /usr/WebSphere (op AIX-systemen), /opt/WebSphere (op Linux- of Solaris-systemen) of C:\Program Files\WebSphere (op Windows-systemen).
- Voor WebSphere Application Server versie 6 is het standaardinstallatiepad /usr/IBM/WebSphere (op AIX-systemen), /opt/IBM/WebSphere (op Linux- of Solaris-systemen) of C:\Program Files\IBM\WebSphere (op Windows-systemen).

## **Beveiliging op documentniveau uitschakelen**

U kunt gebruikers in staat stellen een collectie te doorzoeken ongeacht of er toeganginstellingen voor de documenten in de index zijn geconfigureerd. Voor crawlers die ondersteuning bieden voor validatie van actuele legitimatiegegevens kunt u ook instellen dat gebruikers een collectie kunnen doorzoeken zonder dat de actuele toeganginstellingen tijdens de queryverwerking worden gevalideerd.

### **Voordat u begint**

Om beveiliging op documentniveau in of uit te schakelen voor alle documenten in een collectie, moet u de rol van beheerder van enterprise search hebben of beheerder van de collectie zijn.

### **Beperkingen**

U kunt alleen beveiligingsopties op documentniveau opgeven als tijdens het maken van de collectie beveiliging voor de collectie is ingeschakeld.

### **Over deze taak**

U kunt crawlers configureren om ervoor te zorgen dat tijdens het crawlen van documenten beveiligingstokens aan de documenten worden gekoppeld. De zoekprogramma's kunnen deze tokens (die in de index worden bewaard) gebruiken om toegangsbesturing af te dwingen wanneer gebruikers de collectie doorzoeken. Voor bepaalde crawlers kunt u ook opgeven dat u tijdens het indienen van query's actuele toeganginstellingen wilt afdwingen die in de native repository's van gebruikers aan de documenten zijn gekoppeld.

Om deze beveiligingsbeperkingen op te heffen, kunt u opgeven dat de zoekservers de beveiligingstoken moeten negeren die met een query worden doorgegeven. Daarnaast kunt u instellen dat gebruikers query's kunnen uitvoeren op documenten zonder dat hiervoor de legitimatiegegevens hoeven te worden vergeleken met de actuele toeganginstellingen.

U kunt de beveiliging op documentniveau tijdelijk uitschakelen als u bijvoorbeeld een nieuwe collectie wilt testen of een probleem met een zoekprogramma wilt oplossen.

### Procedure

Ga als volgt te werk om toegangsinstellingen op documentniveau uit te schakelen:

1. Open een collectie, selecteer de pagina Algemeen en klik op **Beveiliging op documentniveau in- of uitschakelen**.
2. Schakel op de pagina Beveiliging op documentniveau voor alle documenten het aankruisvakje **Toegangsbesturing op documentniveau in de index negeren** in als u wilt instellen dat de beveiligingstokens die via de crawlers aan documenten zijn toegewezen, niet worden gebruikt wanneer gebruikers query's uitvoeren op de collectie.

De crawlers gaan door met het toevoegen van beveiligingstokens aan documenten, maar de zoekservers negeren de tokens zodat gebruikers de eerder beveiligde documenten kunnen doorzoeken.

3. Schakel het aankruisvakje **Actuele legitimatiegegevens niet controleren tijdens verwerking van query** in als u wilt instellen dat de actuele toegangsinstellingen die in de native repository's aan documenten zijn gekoppeld, niet wilt valideren wanneer gebruikers query's indienen. Dit aankruisvakje is alleen beschikbaar voor documenten die zijn gecrawld door een crawler die deze functie ondersteunt.

Als u dit vakje selecteert, blijven andere beveiligingsinstellingen op documentniveau van kracht. Als u bijvoorbeeld tijdens het configureren van de crawler opties hebt opgegeven voor het opslaan van toegangsinstellingen in de index, wordt doorgedaan met het toepassen van deze beveiligingsinstellingen tenzij u ook het aankruisvakje **Toegangsbesturing op documentniveau in de index negeren** inschakelt.

## Beveiliging uitschakelen voor samengevouwen zoekresultaten

Als collectiebeveiliging is ingeschakeld, kunnen de zoekresultaten van dezelfde site niet worden samengevouwen in de zoekresultaten tenzij u opgeeft dat u de legitimatiegegevens van gebruikers niet wilt valideren tijdens de queryverwerking.

### Voordat u begint

Om validatie van de actuele legitimatiegegevens in of uit te schakelen voor alle documenten in een collectie, moet u de rol van beheerder van enterprise search hebben of beheerder van de collectie zijn.

### Beperkingen

U kunt alleen beveiligingsopties op documentniveau opgeven als tijdens het maken van de collectie beveiliging voor de collectie is ingeschakeld.

### Over deze taak

Als de actuele legitimatiegegevens zijn gevalideerd, wordt de bron van elk document gecontroleerd en mogelijk doorgestuurd voor validatie. U moet de validatie van actuele legitimatiegegevens uitschakelen als u documenten met dezelfde URI-prefix of documenten die deel uitmaken van een eerder geconfigureerde samengevouwen URI-groep, in de zoekresultaten wilt samenvouwen.

## Procedure

Ga als volgt te werk om de validatie van actuele legitimatiegegevens uit te schakelen zodat documenten in de zoekresultaten kunnen worden samengevouwen:

1. Open een collectie, selecteer de pagina Algemeen en klik op **Beveiliging op documentniveau in- of uitschakelen**.
2. Schakel op de pagina Beveiliging op documentniveau voor alle documenten het aankruisvakje **Actuele legitimatiegegevens niet controleren tijdens verwerking van query** in.
3. Bewaak de collectie, selecteer de pagina Zoeken, stop de zoekserverprocessen en start deze vervolgens opnieuw.

Als gebruikers een query uitvoeren op de collectie, worden de documenten met dezelfde URI-prefix of documenten die deel uitmaken van sites die zijn geconfigureerd om te worden samengevouwen, in de zoekresultaten samengevouwen. In het voorbeeldzoekprogramma kunnen gebruikers de samengevouwen resultaten bekijken door op de link **Meer resultaten uit dezelfde bron** te klikken.



---

## Integratie van enterprise search met WebSphere Portal

U kunt de zoekmogelijkheden van IBM WebSphere Portal uitbreiden door de zoekportlet voor enterprise search-portlets in WebSphere Portal te implementeren en WebSphere Portal zodanig te configureren, dat de zoekportlet als standaardzoekmachine wordt gebruikt.

### Integratiepunten

Het installatieprogramma van WebSphere Information Integrator OmniFind Edition bevat installatiescripts voor de integratie tussen enterprise search en WebSphere Portal. Als u deze scripts hebt uitgevoerd, kunt u het enterprise search-systeem op verschillende manieren met WebSphere Portal integreren:

#### Zoekportlet voor enterprise search

WebSphere Portal biedt gebruikers één toegangspunt om interactief te kunnen werken met programma's, content, processen en personen. Met behulp van het WebSphere Portal-framework kunnen nieuwe programma's, ook wel portlets genoemd, worden geïntegreerd en geïmplementeerd zonder dat dit van invloed is op de andere programma's in de portal.

Als u de zoekportlet voor enterprise search in WebSphere Portal implementeert, kunt u de WebSphere Portal-interface gebruiken om enterprise search-collecties te doorzoeken en met de zoekresultaten te werken. Via de configuratie-instellingen van WebSphere Portal kunt u ervoor zorgen dat de enterprise search-portlet er hetzelfde uitziet en op dezelfde manier werkt als de andere portlets in uw WebSphere Portal-omgeving.

#### Zoekcentrum van WebSphere Portal

Het Zoekcentrum van WebSphere Portal biedt een centraal beginpunt voor het doorzoeken van alle bronnen die via WebSphere Portal beschikbaar zijn gemaakt om te worden doorzocht. Met het Zoekcentrum en de universele zoekportlet kunt u WebSphere Portal-content en andere collecties doorzoeken die bij het Zoekcentrum zijn geregistreerd.

Als u het installatiescript hebt uitgevoerd voor de integratie tussen enterprise search en WebSphere Portal versie 5.1, wordt een enterprise search-pagina aan een pagina in de interface van het Zoekcentrum toegevoegd. U kunt deze pagina selecteren als u alleen enterprise search-collecties wilt doorzoeken of u kunt een query invoeren waarmee de enterprise search-collecties en alle andere collecties worden doorzocht die beschikbaar zijn in het Zoekcentrum.

Als u de installatiescripts hebt uitgevoerd voor de integratie tussen enterprise search en WebSphere Portal versie 6, wordt de enterprise search-functionaliteit als federatieve service geïntegreerd die u kunt gebruiken om de enterprise search-collecties en de andere beschikbare collecties in het Zoekcentrum te doorzoeken.

#### Zoekbalk in WebSphere Portal

In de rechterbovenhoek van de WebSphere Portal-interfacethema's vindt u een zoekbalk. Standaard worden via deze zoekbalk alle zoekopdrachten naar de standaardzoekmachine in het Zoekcentrum doorgestuurd. Als u krachtigere enterprise search-functies wilt gebruiken voor de

queryverwerking, kunt u dit standaardgedrag wijzigen zodat alle zoekopdrachten naar de zoekportlet voor enterprise search worden doorgestuurd.

### **WebSphere Portal- en Web Content Management-crawlers**

Als u WebSphere Portal-sites en IBM Workplace Web Content Management-sites in een enterprise search-index wilt opnemen, kunt u de beheerconsole van enterprise search gebruiken om WebSphere Portal- en Web Content Management-crawlers te configureren. Vervolgens kunt u de enterprise search-portlet of een zoekprogramma gebruiken om de geïndexeerde content te doorzoeken.

De WebSphere Portal-crawler kan sites met WebSphere Portal versie 5.1 en WebSphere Portal versie 6 crawlen. De Web Content Management-crawler kan sites met WebSphere Portal versie 6 crawlen.

### **Voordelen van integratie**

Met enterprise search kunt u de zoekomgeving in WebSphere Portal verbeteren omdat enterprise search ondersteuning biedt voor het doorzoeken van meer gegevensbrontypen. Via de zoekportlet voor enterprise search kunt u websites plus all andere gegevensbrontypen doorzoeken doe worden ondersteund door een enterprise search-systeem.

Daarnaast biedt enterprise search voordelen op het gebied van schaalbaarheid. De portalzoekmachine is geschikt voor kleine tot middelgrote bedrijven waarin de werkbelasting voor zoek- en ophaalbewerkingen met één server kan worden afgehandeld. Voor ondersteuning van mogelijkheden op bedrijfsniveau kan de werkbelasting in enterprise search over verschillende servers worden verdeeld, waarbij twee servers ondersteuning bieden voor de verwerking van zoek- en ophaaltaken.

#### **Verwante onderwerpen**

“WebSphere Portal-crawlers” op pagina 102

Als u pagina’s van een IBM WebSphere Portal-site wilt opnemen in een enterprise search-collectie, moet u een WebSphere Portal-crawler configureren.

“Web Content Management-crawlers” op pagina 100

Als u IBM Workplace Web Content Management-documenten in een enterprise search-collectie wilt opnemen, moet u een Web Content Management-crawler configureren.

---

## **Installatiescripts voor integratie tussen enterprise search en WebSphere Portal**

Om enterprise search te integreren met IBM WebSphere Portal, kunt u de installatiescripts uitvoeren die bij het installatieprogramma van WebSphere Information Integrator OmniFind Edition zijn geleverd.

U moet het JAR-bestand dat de installatiescripts voor uw versie van WebSphere Portal bevat, vanaf de enterprise search-server kopiëren naar de server waarop WebSphere Portal is geïnstalleerd. Met de installatiescripts worden de volgende taken uitgevoerd:

- De EAR-bestanden worden geïmplementeerd, zodat u enterprise search in WebSphere Portal kunt gebruiken en crawlers kunt maken om WebSphere Portal- en IBM Workplace Web Content Management-content aan de enterprise search-collecties toe te voegen.



- De WAR-bestanden worden geïmplementeerd die vereist zijn voor de enterprise search-portlet.
- Er worden pagina's in WebSphere Portal gemaakt, waarna de enterprise search-portletbestanden aan deze pagina's worden toegewezen.
- Alle vereiste JAR-bestanden worden naar de installatiedirectory's van WebSphere Portal gekopieerd (van de JAR-bestanden die al in de installatiedirectory's zijn opgeslagen, worden backups gemaakt voordat de JAR-bestanden voor enterprise search worden gekopieerd).
- Er wordt een integratiepunt geleverd zodat WebSphere Information Integrator Content Edition de Portal Document Manager-documenten kan doorzoeken.

Als u de scripts hebt uitgevoerd, moet u de beheerinterface van WebSphere Portal gebruiken om de eigenschappen van de zoekportlet bij te werken en informatie op te geven over de zoekserver voor enterprise search.

### Richtlijnen voor de syntaxis

- Met de scripts worden alle integratiepunten tussen enterprise search en WebSphere Portal ingesteld. U kunt er bijvoorbeeld niet voor kiezen alleen de portlet te installeren, terwijl u de EAR-bestanden die ondersteuning bieden voor de WebSphere Portal- en Web Content Management-crawlers, niet installeert.
- Als u WebSphere Information Integrator Content Edition niet instelt en later besluit dat u een portlet wilt gebruiken om de Portal Document Manager-documenten te doorzoeken, moet u een script uitvoeren waarmee enterprise search uit WebSphere Portal wordt verwijderd. Vervolgens kunt u het installatiescript opnieuw uitvoeren en het installatiepad van WebSphere Information Integrator Content Edition opgeven.
- Met behulp van de scripts wordt WebSphere Portal gestopt en opnieuw gestart. Mogelijk wilt u de scripts na werktijd uitvoeren, zodat de portalservices gewoon beschikbaar blijven voor gebruikers.
- Als tijdens het uitvoeren van de scripts een fout optreedt, moet u het installatiescript nogmaals uitvoeren. Het kan voorkomen dat bepaalde taken tijdens de eerste poging zonder problemen worden voltooid, terwijl er wel fouten door de taken zijn vastgelegd. Het installatieproces wordt in dit geval voortgezet en de resterende taken worden uitgevoerd.
- De eerste keer dat u de enterprise search-portletpagina opent nadat het installatiescript is uitgevoerd, kan de pagina langzaam worden weergegeven omdat de JSP-bestanden (Java Server Pages) voor de portlet moeten worden gecompileerd.

---

## Enterprise search installeren in WebSphere Portal versie 5.1

Om een enterprise search-systeem te integreren met WebSphere Portal versie 5.1.0 of later, gebruikt u het script `wp5_install`.

### Over deze taak

De bestanden die nodig zijn voor de integratie tussen enterprise search en WebSphere Portal vindt u in het bestand `es.wp5.install.jar`. Wanneer u dit bestand uitpakt, worden de volgende bestanden geëxtraheerd:

- `ESSearchPortlet.war`
- `ESSearchAdapterPortlet.war`
- `ESSearchAdapter.ear`
- `ESPACServer.ear`

- esapi.jar
- siapi.jar
- es.security.jar
- Script-, batch-, XML- en JACL-bestanden voor de installatie

## Procedure

Ga als volgt te werk om enterprise search te integreren met een WebSphere Portal versie 5.1-systeem:

1. Kopieer het bestand es.wp5.install.jar vanaf de enterprise search-server naar de WebSphere Portal-server en gebruik de opdracht Java JAR (of de opdracht TAR) om het bestand uit te pakken.
2. Optioneel: als u integratie met WebSphere Portal Document Manager (PDM) wilt instellen, voert u een van de volgende stappen uit:
  - Voer het installatieprogramma van WebSphere Information Integrator Content Edition uit, selecteer de optie waarmee alleen de connector wordt geïnstalleerd en installeer de PDM-connector op de WebSphere Portal-server.
  - Maak de directorystructuur voor WebSphere Information Integrator Content Edition op de WebSphere Portal-server en kopieer de volgende bestanden vanuit een bestaande installatiedirectory voor WebSphere Information Integrator Content Edition naar de WebSphere Portal-server, waarbij CE\_ROOT de hoofdinstantiatiedirectory voor WebSphere Information Integrator Content Edition aangeeft:

```
CE_ROOT/lib/vbr.jar
CE_ROOT/ejb/vbr_pdm.jar
CE_ROOT/war/services.war
CE_ROOT/vbr_services.properties
```

3. Geef bij de opdrachtaanwijzing de opdracht wp5\_install.bat (onder Windows) of de opdracht wp5\_install.sh op (onder AIX, Linux of Solaris). In het volgende voorbeeld worden de parameters op aparte regels weergegeven zodat ze beter leesbaar zijn. U moet de parameters bij de opdracht opgeven:

```
wp5_install.bat
-WASDir "C:\\Program Files\\WebSphere\\AppServer"
-WASUser wpsbind -WASPassword wpsbind
-WPSDir "C:\\Program Files\\WebSphere\\PortalServer"
-WPSUser wpsadmin -WSPassword wpsadmin
-WPSHost "portalserver.ibm.com:9081" -IICEDir "C:\\IICE"
```

### **WASDir**

Het volledige pad voor de installatiedirectory van WebSphere Application Server.

### **WASUser**

De gebruikersnaam voor de beheerder van WebSphere Application Server. Deze waarde is alleen vereist als algemene beveiliging is ingeschakeld in WebSphere Application Server.

### **WASPassword**

Het wachtwoord voor de WebSphere Application Server-beheerder (indien opgegeven).

### **WPSDir**

Het volledige pad voor de installatiedirectory van WebSphere Portal.

### **WPSUser**

De gebruikersnaam voor de WebSphere Portal-beheerder.

**WSPassword**

Het wachtwoord voor de opgegeven WebSphere Portal-beheerder.

**WPSHost**

De hostnaam en het poortnummer van de WebSphere Portal-server.

**IICEDir**

Het volledige pad voor de installatiedirectory van WebSphere Information Integrator Content Edition. Deze waarde is alleen vereist als u eerder de Portal Document Manager-connector op de WebSphere Portal-server hebt ingesteld.

4. Als u het script hebt uitgevoerd (en WebSphere Portal is gestopt en opnieuw is gestart), werkt u de enterprise search-portlet bij om de zoekserver aan te geven:
  - a. Meld u bij WebSphere Portal aan met het Portal-beheerders-ID en -wachtwoord.
  - b. Klik op **Administration** in de rechterbovenhoek.
  - c. Klik op **Portlet Management** in het navigatiegebied links en klik vervolgens op **Portlets**.
  - d. Wijzig de optie **Search by** in **Title contains**.
  - e. Voer in het veld **Search** de waarde enterprise search in en klik vervolgens op de knop **Search**.
  - f. Als de nieuwe pictogrammen rechts worden weergegeven, klikt u op het pictogram met de moersleutel om de zoekportlet voor enterprise search te configureren.
  - g. Wijzig de volgende parameters in de lijst van portletparameters:

**hostname**

Geef de volledige hostnaam op van de zoekserver voor enterprise search.

**port**

Geef het poortnummer op dat door WebSphere Application Server op de zoekserver voor enterprise search moet worden gebruikt. De standaardwaarde is 80 (de standaardwaarde voor SSL-communicatie is 443).

**username**

Als algemene beveiliging in WebSphere Application Server is ingeschakeld op de zoekserver, geeft u een gebruikersnaam op die is toegestaan in het gebruikersregister van WebSphere Application Server.

**password**

Als u een WebSphere Application Server-gebruikersnaam hebt opgegeven, geeft u het bijbehorende wachtwoord op.

**protocol**

Geef het protocol op dat voor de communicatie tussen WebSphere Portal en de zoekserver moet worden gebruikt. De standaardwaarde is HTTP. Als u SSL gebruikt, geeft u HTTPS op.

**trustStore**

Als u SSL gebruikt, geeft u het volledige pad op (met de bestandsnaam) voor het SSL-certificaatbestand.

**trustPassword**

Als u SSL gebruikt, geeft u het wachtwoord voor het opgegeven betrouwbaarheidsbestand op.

**ssoCookieName**

Geef de naam op van de cookie die de SSO-tekenreeks (Single Sign-On) bevat. De standaardwaarde is LtpaToken.

**proxyHost**

Als een proxyserver vereist is voor toegang tot de zoekserver voor enterprise search, geeft u de volledige hostnaam van een proxyserver op.

**proxyPort**

Als u een proxyserver hebt opgegeven, geeft u het poortnummer voor de proxyserver op.

**proxyUser**

Als voor de proxyserver Basic Authentication vereist is, geeft u de gebruikersnaam op die moet worden gebruikt voor de aanmelding bij de proxyserver.

**proxyPassword**

Als u een gebruikersnaam voor de proxyserver hebt opgegeven, geeft u het bijbehorende wachtwoord op.

h. Klik op **OK** om de wijzigingen op te slaan.

**Verwante onderwerpen**

“WebSphere Portal-crawlers” op pagina 102

Als u pagina’s van een IBM WebSphere Portal-site wilt opnemen in een enterprise search-collectie, moet u een WebSphere Portal-crawler configureren.

**Verwante taken**

“De zoekservers configureren om uitsluitend veilige (SSL) opdrachten te accepteren” op pagina 224

U kunt de HTTP-interface op de zoekservers uitschakelen en de servers zodanig configureren dat ze uitsluitend zoekopdrachten via SSL en de veilige HTTPS-interface accepteren.

## De zoekbalk van WebSphere Portal versie 5.1 configureren voor gebruik van enterprise search

U kunt WebSphere Portal versie 5.1.0 of later configureren voor gebruik van enterprise search wanneer gebruikers query’s indienen via de zoekbalk in plaats van de standaardzoekmachine in WebSphere Portal.

**Voordat u begint**

U kunt zoekopdrachten pas doorsturen naar enterprise search als u het installatiescript wp5\_install hebt uitgevoerd om enterprise search te integreren met WebSphere Portal. Daarnaast moet u de enterprise search-portletparameters bijwerken voor de hostnaam, de poort en de andere informatie over de zoekserver voor enterprise search.

**Over deze taak**

In de rechterbovenhoek van de WebSphere Portal-interfacethema’s vindt u een zoekbalk. Standaard worden alle zoekopdrachten naar de portlet voor het Zoekcentrum doorgestuurd. Als u krachtigere enterprise search-functies wilt gebruiken voor de queryverwerking, kunt u dit standaardgedrag wijzigen zodat alle zoekopdrachten naar de zoekportlet voor enterprise search worden doorgestuurd.

Deze wijziging is van invloed op pagina's die hetzelfde WebSphere Portal-thema gebruiken als de zoekportlet voor enterprise search. Daarnaast moeten deze pagina's het bestand SearchBarInclude.jsp aanroepen. Pagina's die een ander thema gebruiken of die het bestand SearchBarInclude.jsp niet aanroepen, blijven de standaardportlet voor het Zoekcentrum gebruiken.

Als u deze taak hebt uitgevoerd, kunt u het Zoekcentrum pas gebruiken als u de wijzigingen ongedaan hebt gemaakt (u kunt bijvoorbeeld het oorspronkelijke bestand SearchBarInclude.jsp herstellen).

## Procedure

Ga als volgt te werk om in te stellen dat gebruikers de enterprise search-portlet kunnen gebruiken tijdens het indienen van query's op de WebSphere Portal-zoekbalk:

1. Ga op de WebSphere Portal-server naar de directory `WAS_INSTALL_ROOT/AppServer/installedApps/eswin1/wps.ear/wps.war/themes/html`.
2. Maak een backup van het bestand `SearchBarInclude.jsp` door dit bestand te kopiëren en de naam ervan te wijzigen (bijvoorbeeld `SearchBarInclude.jsp.BACKUP`).
3. Open het bestand `SearchBarInclude.jsp` in een editor, vervang de inhoud tussen de instructies `<wps:if loggedIn="yes">` en `</wps:if>` door de volgende tekst en sla het bestand vervolgens op.

```
<wps:if loggedIn="yes">
<wps-internal:adminlinkinfo name="<%=AdminUniqueNamesMappingService.SEARCH_CENTER%>"
  <wps:urlGeneration contentNode="wps.My Portal.OmniFindSearch"
    portletWindowState="Normal" pacCheck="NoCheck">

  <td nowrap="nowrap" align="right" valign="middle" class="wpsLinkBar">
    <form style="margin: 0px;" method="post" action="<%=wpsURL%>"
      <table border="0" cellpadding="0" cellspacing="0">
        <tr>
          <td valign="middle" nowrap>
            <td valign="middle" nowrap>
              <label for="q"><wps:text key="search.theme.searchbox"
                bundle="nls.engine"/></label>
            </td>
            <td valign="middle" style="padding: 0px 4px 0px 4px;">
              <input class="wpsEditField" name="q" type="text"/>
            </td>
            <td valign="middle"><input valign="middle"
              alt='<wps:text key="search.theme.searchbox.alttext"
                bundle="nls.engine"/>'
              src="<%= wpsBaseUrl %>/images/icons/Search.gif" type="image"/>
            </td>
          </tr>
        </table>
      </form>
    </td>

  </wps:urlGeneration>

</wps-internal:adminlinkinfo>

</wps:if>
```

**Belangrijk:** Als u de enterprise search-portletcode wijzigt en de aangepaste portlet implementeert op de pagina van uw keuze, moet u het kenmerk `contentNode="wps.My Portal.OmniFindSearch"` vervangen door het unieke pagina-ID voor de pagina waarop u de aangepaste portlet wilt implementeren.

4. Open het bestand `Default.jsp` en sla het bestand op. Deze stap, waarmee de wijzigingsdatum van het bestand wordt bijgewerkt om ervoor te zorgen dat het

bestand opnieuw wordt gecompileerd, is optioneel als u een eigen thema gebruikt in plaats van het standaard-WebSphere Portal-thema.

5. Open het bestand AdminLinkBarInclude.jsp en sla het bestand op. Deze stap, waarmee de wijzigingsdatum van het bestand wordt bijgewerkt om ervoor te zorgen dat het bestand opnieuw wordt gecompileerd, is optioneel als u een eigen thema gebruikt in plaats van het standaard-WebSphere Portal-thema.
6. Stop de instance van de WebSphere Portal-toepassingsserver en start deze vervolgens opnieuw op.

## Enterprise search verwijderen uit WebSphere Portal versie 5.1

Als u enterprise search wilt verwijderen uit een systeem met WebSphere Portal versie 5.1.0 of later, gebruikt u het script wp5\_uninstall.

### Over deze taak

Wanneer u enterprise search uit WebSphere Portal verwijdert, worden de portlet-parameters die u tijdens het installatieproces voor de enterprise search-portlet hebt opgegeven, niet opgeslagen.

Wanneer u het script start, wordt de WebSphere Portal-server gestopt. Als de enterprise search-software is verwijderd, wordt de WebSphere Portal-server opnieuw gestart met behulp van het script.

### Procedure

Ga als volgt te werk om enterprise search te verwijderen uit een WebSphere Portal-systeem met versie 5.1:

Geef bij de opdrachtaanwijzing de opdracht wp5\_uninstall.bat (onder Windows) of de opdracht wp5\_uninstall.sh op (onder AIX, Linux of Solaris). In het volgende voorbeeld worden de parameters op aparte regels weergegeven zodat ze beter leesbaar zijn. U moet de parameters bij de opdracht opgeven:

```
wp5_uninstall.bat
-WASDir "C:\\Program Files\\WebSphere\\AppServer"
-WASUser wpsbind -WASPassword wpsbind
-WPSDir "C:\\Program Files\\WebSphere\\PortalServer"
-WPSUser wpsadmin -WSPassword wpsadmin
-WPSHost "portalserver.ibm.com:9081"
```

#### WASDir

Het volledige pad voor de installatiedirectory van WebSphere Application Server.

#### WASUser

De gebruikersnaam voor de beheerder van WebSphere Application Server. Deze waarde is alleen vereist als algemene beveiliging is ingeschakeld in WebSphere Application Server.

#### WASPassword

Het wachtwoord voor de WebSphere Application Server-beheerder (indien opgegeven).

#### WPSDir

Het volledige pad voor de installatiedirectory van WebSphere Portal.

#### WPSUser

De gebruikersnaam voor de WebSphere Portal-beheerder.

**WSPassword**

Het wachtwoord voor de opgegeven WebSphere Portal-beheerder.

**WPSHost**

De hostnaam en het poortnummer van de WebSphere Portal-server.

---

## Enterprise search installeren in WebSphere Portal versie 6

Om een enterprise search-systeem te integreren met WebSphere Portal versie 6 of later, gebruikt u het script wp6\_install.

**Over deze taak**

De bestanden die nodig zijn voor de integratie tussen enterprise search en WebSphere Portal vindt u in het bestand es.wp6.install.jar. Wanneer u dit bestand uitpakt, worden de volgende bestanden geëxtraheerd:

- ESSearchPortlet.war
- ESPACServer.ear
- esapi.jar
- es.search.provider.jar
- es.security.jar
- Pictogrammen voor brontypen in zoekprogramma voor gebruik op de resultatenpagina van de zoekprovider
- Script-, batch-, XML- en JACL-bestanden voor de installatie

**Procedure**

Ga als volgt te werk om enterprise search te integreren met een WebSphere Portal versie 6-systeem:

1. Kopieer het bestand es.wp6.install.jar vanaf de enterprise search-server naar de WebSphere Portal-server en gebruik de opdracht Java JAR (of de opdracht TAR) om het bestand uit te pakken.
2. Optioneel: als u integratie met WebSphere Portal Document Manager (PDM) wilt instellen, voert u een van de volgende stappen uit:
  - Voer het installatieprogramma van WebSphere Information Integrator Content Edition uit, selecteer de optie waarmee alleen de connector wordt geïnstalleerd en installeer de PDM-connector op de WebSphere Portal-server.
  - Maak de directorystructuur voor WebSphere Information Integrator Content Edition op de WebSphere Portal-server en kopieer de volgende bestanden vanuit een bestaande installatiedirectory voor WebSphere Information Integrator Content Edition naar de WebSphere Portal-server, waarbij CE\_ROOT de hoofdinstantiatiedirectory voor WebSphere Information Integrator Content Edition aangeeft:  
  
CE\_ROOT/lib/vbr.jar  
CE\_ROOT/ejb/vbr\_pdm.jar  
CE\_ROOT/war/services.war  
CE\_ROOT/vbr\_services.properties
3. Geef bij de opdrachtaanwijzing de opdracht wp6\_install.bat (onder Windows) of de opdracht wp6\_install.sh op (onder AIX, Linux of Solaris). In het volgende voorbeeld worden de opties op aparte regels weergegeven zodat ze beter leesbaar zijn. U moet de opties bij de opdracht opgeven:

```

wp6_install.bat
-WSPProfileDir "C:\\Program Files\\IBM\\WebSphere\\profiles\\wp_profile"
-WASDir "C:\\Program Files\\IBM\\WebSphere\\AppServer"
-WASUser wpsbind -WASPassword wpsbind
-WPSDir "C:\\Program Files\\IBM\\WebSphere\\PortalServer"
-WPSUser wpsadmin -WSPassword wpsadmin
-WPSHost "portalserver.ibm.com:9081" -IICEDir "C:\\IICE"

```

#### **WSPProfileDir**

Het volledige pad voor de profieldirectory van WebSphere Portal. Het standaardpad is /usr/IBM/WebSphere/AppServer/profiles/wp\_profile op AIX-systemen, /opt/IBM/WebSphere/AppServer/profiles/wp\_profile op Linux- en Solaris-systemen en C:\Program Files\IBM\WebSphere\profiles\wp\_profile op Windows-systemen.

#### **WASDir**

Het volledige pad van de rootdirectory van WebSphere Application Server; alleen vereist voor AIX-, Linux- en Solaris-systemen. De standaard rootdirectory is /usr/IBM/WebSphere/AppServer op AIX-systemen, /opt/IBM/WebSphere/AppServer op Linux- en Solaris-systemen en C:\Program Files\IBM\WebSphere\AppServer op Windows-systemen.

#### **WASUser**

De gebruikersnaam voor de beheerder van WebSphere Application Server. Deze waarde is alleen vereist als algemene beveiliging is ingeschakeld in WebSphere Application Server.

#### **WASPassword**

Het wachtwoord voor de WebSphere Application Server-beheerder (indien opgegeven).

#### **WPSDir**

Het volledige pad voor de installatiedirectory van WebSphere Portal.

#### **WPSUser**

De gebruikersnaam voor de WebSphere Portal-beheerder.

#### **WSPassword**

Het wachtwoord voor de opgegeven WebSphere Portal-beheerder.

#### **WPSHost**

De hostnaam en het poortnummer van de WebSphere Portal-server.

#### **IICEDir**

Het volledige pad voor de installatiedirectory van WebSphere Information Integrator Content Edition. Deze waarde is alleen vereist als u eerder de Portal Document Manager-connector op de WebSphere Portal-server hebt ingesteld.

4. Als u het script hebt uitgevoerd (en WebSphere Portal is gestopt en opnieuw is gestart), werkt u de enterprise search-portlet bij om de zoekserver aan te geven:
  - a. Meld u bij WebSphere Portal aan met het Portal-beheerders-ID en -wachtwoord.
  - b. Klik op **Administration** in de linkerbenedenhoek.
  - c. Klik op **Portlet Management** in het navigatiegebied links en klik vervolgens op **Portlets**.
  - d. Wijzig de optie **Search by** in **Title contains**.
  - e. Voer in het veld **Search** de waarde enterprise search in en klik vervolgens op de knop **Search**.



f. Als de nieuwe pictogrammen rechts worden weergegeven, klikt u op het pictogram met de moersleutel om de zoekportlet voor enterprise search te configureren.

g. Wijzig de volgende parameters in de lijst van portletparameters:

**hostname**

Geef de volledige hostnaam op van de zoekserver voor enterprise search.

**port**

Geef het poortnummer op dat door WebSphere Application Server op de zoekserver voor enterprise search moet worden gebruikt. De standaardwaarde is 80 (de standaardwaarde voor SSL-communicatie is 443).

**username**

Als algemene beveiliging in WebSphere Application Server is ingeschakeld op de zoekserver, geeft u een gebruikersnaam op die is toegestaan in het gebruikersregister van WebSphere Application Server.

**password**

Als u een WebSphere Application Server-gebruikersnaam hebt opgegeven, geeft u het bijbehorende wachtwoord op.

**protocol**

Geef het protocol op dat voor de communicatie tussen WebSphere Portal en de zoekserver moet worden gebruikt. De standaardwaarde is HTTP. Als u SSL gebruikt, geeft u HTTPS op.

**trustStore**

Als u SSL gebruikt, geeft u het volledige pad op (met de bestandsnaam) voor het SSL-certificaatbestand.

**trustPassword**

Als u SSL gebruikt, geeft u het wachtwoord voor het opgegeven betrouwbaarheidsbestand op.

**ssoCookieName**

Geef de naam op van de cookie die de SSO-tekenreeks (Single Sign-On) bevat. De standaardwaarde is LtpaToken.

**proxyHost**

Als een proxyserver vereist is voor toegang tot de zoekserver voor enterprise search, geeft u de volledige hostnaam van een proxyserver op.

**proxyPort**

Als u een proxyserver hebt opgegeven, geeft u het poortnummer voor de proxyserver op.

**proxyUser**

Als voor de proxyserver Basic Authentication vereist is, geeft u de gebruikersnaam op die moet worden gebruikt voor de aanmelding bij de proxyserver.

**proxyPassword**

Als u een gebruikersnaam voor de proxyserver hebt opgegeven, geeft u het bijbehorende wachtwoord op.

h. Klik op **OK** om de wijzigingen op te slaan.

**Verwante onderwerpen**

“WebSphere Portal-crawlers” op pagina 102

Als u pagina’s van een IBM WebSphere Portal-site wilt opnemen in een enterprise search-collectie, moet u een WebSphere Portal-crawler configureren.

“Web Content Management-crawlers” op pagina 100

Als u IBM Workplace Web Content Management-documenten in een enterprise search-collectie wilt opnemen, moet u een Web Content Management-crawler configureren.

#### **Verwante taken**

“De zoekservers configureren om uitsluitend veilige (SSL) opdrachten te accepteren” op pagina 224

U kunt de HTTP-interface op de zoekservers uitschakelen en de servers zodanig configureren dat ze uitsluitend zoekopdrachten via SSL en de veilige HTTPS-interface accepteren.

## **Het Zoekcentrum van WebSphere Portal versie 6 configureren voor enterprise search**

U kunt WebSphere Portal versie 6 configureren voor het doorzoeken van enterprise search-collecties wanneer gebruikers query’s indienen via het Zoekcentrum van WebSphere Portal.

### **Beperkingen**

Als de enterprise search-collecties die u wilt doorzoeken, zijn beveiligd, moeten gebruikers de zoekportlet voor enterprise search uitvoeren en een gebruikersprofiel configureren. Het profiel is gecodeerd en is opgeslagen op een beveiligde opslagplek in enterprise search. Pas als het profiel aanwezig is, kunnen gebruikers vanuit het Zoekcentrum van WebSphere Portal query’s indienen voor het doorzoeken van beveiligde collecties.

### **Over deze taak**

Het Zoekcentrum van WebSphere Portal versie 6 biedt ondersteuning voor federatieve zoekmogelijkheden in meerdere collecties. De collecties kunnen verschillende typen content bevatten, zoals Portal Document-bibliotheken en portalcontent (pagina’s en portlets). Als u de installatiescripts hebt uitgevoerd voor de integratie tussen enterprise search en WebSphere Portal, kunt u het Zoekcentrum configureren zodat u met het Zoekcentrum ook de enterprise search-collecties kunt doorzoeken.

### **Procedure**

Ga als volgt te werk om in te stellen dat met het Zoekcentrum enterprise search-collecties kunnen worden doorzocht:

1. Meld u bij WebSphere Portal aan met het Portal-beheerders-ID en -wachtwoord.
2. Klik op **Beheer** in de linkerbenedenhoek.
3. Klik op **Zoekbeheer** in het navigatiegebied links en klik vervolgens op **Zoeken beheren**.
4. Klik op **Zoekservices** en klik vervolgens op **Nieuwe zoekservice**.
5. Selecteer in het veld **Zoekservice-implementatie** de optie Enterprise Search zoekservice en voer de naam die u voor de service wilt gebruiken in in het tekstvak **Servicenaam**.
6. Wijzig de volgende parameters in de lijst van parameters:

**hostname**

Geef de volledige hostnaam op van de zoekserver voor enterprise search.

**port**

Geef het poortnummer op dat door WebSphere Application Server op de zoekserver voor enterprise search moet worden gebruikt. De standaardwaarde is 80 (de standaardwaarde voor SSL-communicatie is 443).

**username**

Als algemene beveiliging in WebSphere Application Server is ingeschakeld op de zoekserver, geeft u een gebruikersnaam op die is toegestaan in het gebruikersregister van WebSphere Application Server.

**password**

Als u een WebSphere Application Server-gebruikersnaam hebt opgegeven, geeft u het bijbehorende wachtwoord op.

**protocol**

Geef het protocol op dat voor de communicatie tussen WebSphere Portal en de zoekserver moet worden gebruikt. De standaardwaarde is HTTP. Als u SSL gebruikt, geeft u HTTPS op.

**trustStore**

Als u SSL gebruikt, geeft u het volledige pad op (met de bestandsnaam) voor het SSL-certificaatbestand.

**trustPassword**

Als u SSL gebruikt, geeft u het wachtwoord voor het opgegeven betrouwbaarheidsbestand op.

**ssoCookieName**

Geef de naam op van de cookie die de SSO-tekenreeks (Single Sign-On) bevat. De standaardwaarde is LtpaToken.

**proxyHost**

Als een proxyserver vereist is voor toegang tot de zoekserver voor enterprise search, geeft u de volledige hostnaam van een proxyserver op.

**proxyPort**

Als u een proxyserver hebt opgegeven, geeft u het poortnummer voor de proxyserver op.

**proxyUser**

Als voor de proxyserver Basic Authentication vereist is, geeft u de gebruikersnaam op die moet worden gebruikt voor de aanmelding bij de proxyserver.

**proxyPassword**

Als u een gebruikersnaam voor de proxyserver hebt opgegeven, geeft u het bijbehorende wachtwoord op.

7. Klik op **OK** om de wijzigingen op te slaan.

## De zoekbalk van WebSphere Portal versie 6 configureren voor gebruik van enterprise search

U kunt WebSphere Portal versie 6 of later configureren voor gebruik van enterprise search wanneer gebruikers query's indienen via de zoekbalk in plaats van de standaardzoekmachine in WebSphere Portal.

## Voordat u begint

U kunt zoekopdrachten pas doorsturen naar enterprise search als u het installatiescript wp6\_install hebt uitgevoerd om enterprise search te integreren met WebSphere Portal. Daarnaast moet u de enterprise search-portletparameters bijwerken voor de hostnaam, de poort en de andere informatie over de zoekserver voor enterprise search.

## Over deze taak

In de rechterbovenhoek van de WebSphere Portal-interfacethema's vindt u een zoekbalk. Standaard worden alle zoekopdrachten naar de portlet voor het Zoekcentrum doorgestuurd. Als u krachtigere enterprise search-functies wilt gebruiken voor de queryverwerking, kunt u dit standaardgedrag wijzigen zodat alle zoekopdrachten naar de zoekportlet voor enterprise search worden doorgestuurd.

Deze wijziging is van invloed op pagina's die hetzelfde WebSphere Portal-thema gebruiken als de zoekportlet voor enterprise search. Daarnaast moeten deze pagina's het bestand banner\_searchControl.jspf aanroepen. Pagina's die een ander thema gebruiken of die het bestand banner\_searchControl.jspf niet aanroepen, blijven de standaardportlet voor het Zoekcentrum gebruiken.

Als u deze taak hebt uitgevoerd, kunt u het Zoekcentrum pas gebruiken als u de wijzigingen ongedaan hebt gemaakt (u kunt bijvoorbeeld het oorspronkelijke bestand banner\_searchControl.jspf herstellen).

## Procedure

Ga als volgt te werk om in te stellen dat gebruikers de enterprise search-portlet kunnen gebruiken tijdens het indienen van query's op de WebSphere Portal-zoekbalk:

1. Stop de instance van de WebSphere Portal-toepassingenserver.
2. Ga op de WebSphere Portal-server naar de directory WPS\_PROFILE\_ROOT/installedApps/knooppunt/wps.ear/wps.war/themes/html/huidige\_thema, waarbij *knooppunt* de naam is van het knooppunt voor de WebSphere Portal-server en *huidige\_thema* de naam is van het thema dat momenteel in gebruik is voor de WebSphere Portal-server. De standaardnaam voor het thema van een WebSphere Portal-server is IBM.
3. Maak een backup van het bestand banner\_searchControl.jspf door dit bestand te kopiëren en de naam ervan te wijzigen (bijvoorbeeld banner\_searchControl.jspf.BACKUP).
4. Open het bestand banner\_searchControl.jspf in een editor, vervang de inhoud door de volgende tekst en sla het bestand weer op.

```
<!-- Gelicentieerd materiaal - Eigendom van IBM, 5724-E76,  
(C) Copyright IBM Corp. 2001, 2004, 2006  
- Alle rechten voorbehouden.
```

Zie de opmerkingen in Default.jsp voor informatie over het wijzigen van thema-JSP-fragmenten.

```
--%>  
<%@ taglib uri="/WEB-INF/tld/SearchMenuControl.tld" prefix="searchmenu" %>  
<!-- search box --%>  
<%String ic = (bidiImageRTL == null) ? "icons/scope_search_submit.gif" :  
"icons/scope_search_submit"+bidiImageRTL+".gif";%>  
  
<searchmenu:adminlinkinfo name="SEARCH_CENTER">
```

```

<portal-navigation:urlGeneration contentNode="ibm.portal.OmniFindSearch"
portletWindowState="Normal"
portletMode="View"
pacCheck="NoCheck">

<div class="searchControl">
  <form name="searchFromThemeForm" style="margin: 0px;" method="POST"
    action="<%= wpsURL.write(escapeXmlWriter); %>">

    <table border="0" cellpadding="0" cellspacing="0">
    <tr>
    <td><span class="wpsInstructionText">
      <portal-fmt:text key='search.theme.control.label' bundle='nls.engine' />
    </span></td>
    <td <!-- nowrap is deprecated, use css -->
      valign="middle" style="padding: 0px 4px 0px 4px;">
      <input type="hidden" name="OCN" value="<%= wpsContentNodeID %>" />
      <input type="hidden" name="clearifblank" value="1" />
      <input type="hidden" name="srchproc" value="" />
      <searchmenu:menu
        scopeFieldName="scope"
        searchFieldName="q"
        output="all"
        uniqueId="searchTheme"
        tabIndex="3"/>
    </td>
    <td valign="middle">
      <input tabIndex="4" valign="middle"
        title="<portal-fmt:text key='search.theme.searchresultsicon.alttext'
          bundle='nls.engine' />"
        alt="<portal-fmt:text key='search.theme.searchresultsicon.alttext'
          bundle='nls.engine' />"
        src="<portal-logic:urlFindInTheme file="<%= ic %>" />" type="image">
    </td>
    </tr>
    </table>
  </form>
</div>
</portal-navigation:urlGeneration>
</searchmenu:adminlinkinfo>

```

**Belangrijk:** Als u de enterprise search-portletcode wijzigt en de aangepaste portlet implementeert op de pagina van uw keuze, moet u het kenmerk contentNode="ibm.portal.OmniFindSearch" vervangen door het unieke pagina-ID voor de pagina waarop u de aangepaste portlet wilt implementeren.

5. Open het bestand banner.jspf en sla het op. Deze stap, waarmee de wijzigingsdatum van het bestand wordt bijgewerkt om ervoor te zorgen dat het bestand opnieuw wordt gecompileerd, is optioneel als u een eigen thema gebruikt in plaats van het standaard-WebSphere Portal-thema.
6. Open het bestand Default.jsp en sla het bestand op.
7. Start de instance van de WebSphere Portal-toepassingenserver opnieuw.

## Enterprise search verwijderen uit WebSphere Portal versie 6

Als u enterprise search wilt verwijderen uit een systeem met WebSphere Portal versie 6, gebruikt u het script wp6\_uninstall.

### Over deze taak

Wanneer u enterprise search uit WebSphere Portal verwijdert, worden de portlet-parameters die u tijdens het installatieproces voor de enterprise search-portlet hebt opgegeven, niet opgeslagen.

Wanneer u het script start, wordt de WebSphere Portal-server gestopt. Als de enterprise search-software is verwijderd, wordt de WebSphere Portal-server opnieuw gestart met behulp van het script.

### **Procedure**

Ga als volgt te werk om enterprise search te verwijderen uit een WebSphere Portal-systeem met versie 6:

Geef bij de opdrachtaanwijzing de opdracht wp6\_uninstall.bat (onder Windows) of de opdracht wp6\_uninstall.sh op (onder AIX, Linux of Solaris). In het volgende voorbeeld worden de parameters op aparte regels weergegeven zodat ze beter leesbaar zijn. U moet de parameters bij de opdracht opgeven:

```
wp6_uninstall.bat
-WSPProfileDir "C:\\Program Files\\IBM\\WebSphere\\AppServer\\profiles\\wp_profile"
-WASDir "C:\\Program Files\\IBM\\WebSphere\\AppServer"
-WASUser wpsbind -WASPassword wpsbind
-WPSDir "C:\\Program Files\\IBM\\WebSphere\\PortalServer"
-WPSUser wpsadmin -WSPassword wpsadmin
-WPSHost "portalserver.ibm.com:9081"
```

#### **WSPProfileDir**

Het volledige pad voor de profieldirectory van WebSphere Portal.

#### **WASDir**

Het volledige pad van de rootdirectory van WebSphere Application Server; alleen vereist voor AIX-, Linux- en Solaris-systemen.

#### **WASUser**

De gebruikersnaam voor de beheerder van WebSphere Application Server. Deze waarde is alleen vereist als algemene beveiliging is ingeschakeld in WebSphere Application Server.

#### **WASPassword**

Het wachtwoord voor de WebSphere Application Server-beheerder (indien opgegeven).

#### **WPSDir**

Het volledige pad voor de installatiedirectory van WebSphere Portal.

#### **WPSUser**

De gebruikersnaam voor de WebSphere Portal-beheerder.

#### **WSPassword**

Het wachtwoord voor de opgegeven WebSphere Portal-beheerder.

#### **WPSHost**

De hostnaam en het poortnummer van de WebSphere Portal-server.

---

## Migratie vanuit WebSphere Portal naar enterprise search

Enterprise search bevat een migratiewizard die u kunt gebruiken als u collecties en regelgebaseerde taxonomieën vanuit IBM WebSphere Portal wilt migreren naar enterprise search.

In enterprise search wordt een taxonomie een *categoriestructuur* genoemd. Als u een taxonomie hebt gemigreerd, gebruikt u de beheerconsole van enterprise search om de categoriestructuur en de categorieregels te bewerken. Nadat u een collectie hebt gemigreerd, gebruikt u de beheerconsole om de collectie te beheren.

Als u taxonomieën en collecties wilt migreren, voert u de Migratiewizard uit op de enterprise search-indexserver.

---

### Een collectie migreren vanuit WebSphere Portal

Als u collecties en regelgebaseerde taxonomieën vanuit WebSphere Portal naar enterprise search wilt migreren, moet u de collecties in WebSphere Portal voorbereiden en vervolgens de Migratiewizard gebruiken om de collecties te migreren.

#### Voordat u begint

Als u van plan bent taxonomieën en collecties te migreren, moet u de taxonomiebestanden migreren voordat u de procedure voor het migreren van collecties uitvoert. Op deze manier zorgt u ervoor dat de gemigreerde categoriseringsregels kunnen worden gebruikt in combinatie met de gemigreerde collecties.

#### Procedure

Ga als volgt te werk om een collectie (en optioneel de taxonomie) vanuit WebSphere Portal naar enterprise search te migreren:

1. Stop in de WebSphere Portal-zoekmachine alle crawlerprocessen in de collecties die u wilt migreren en keur de wachtende documenten goed of wijs deze af. (Enterprise search biedt geen ondersteuning voor het concept van wachtende documenten.)
2. Gebruik voor elke collectie die u wilt migreren de portalzoekmachineportlets om de instellingen te exporteren naar XML-bestanden.
3. Als de enterprise search-indexserver op een aparte server is geïnstalleerd, kopieert u de geëxporteerde XML-bestanden naar de indexserver.
4. Meld u bij de enterprise search-indexserver aan als enterprise search-beheerder. Dit gebruikers-ID is opgegeven bij de installatie van WebSphere Information Integrator OmniFind Edition.
5. Ga naar de installatiedirectory van enterprise search:

UNIX: `cd $ES_INSTALL_ROOT/bin`

Windows: `cd %ES_INSTALL_ROOT%\bin`

6. Als u collecties wilt migreren waarvoor beveiliging is ingeschakeld, geeft u de volgende opdracht op om de Migratiewizard te starten en klikt u vervolgens op **Volgende**.

UNIX: ./eswpsmigrate.sh  
Windows: eswpsmigrate.bat

7. Als u de beveiliging op collectieniveau wilt uitschakelen voor de collecties die u wilt migreren, geeft u de volgende opdracht op om de Migratiewizard te starten en klikt u vervolgens op **Volgende**.

UNIX: ./eswpsmigrate.sh disable.security  
Windows: eswpsmigrate.bat disable.security

8. Selecteer **Zoekinstellingen migreren vanuit de Portal Zoekfunctie in WebSphere Portal** en klik vervolgens op **Volgende**.
9. Ga naar de directory waarin de geëxporteerde configuratiebestanden van de portalzoekmachine zijn opgeslagen, selecteer de bestanden die u wilt migreren en klik op **Volgende**. De geselecteerde configuratiebestanden worden geanalyseerd en gevalideerd.
10. Voer de volgende informatie in voor elke collectie en klik op **Volgende** om de collecties naar enterprise search te migreren:
  - De naam van de collectie zoals deze in enterprise search moet worden weergegeven.
  - Het criterium op basis waarvan het documentbelang voor de collectie wordt vastgesteld. De statische rankingfactor kan zijn gebaseerd op documentdatum, op het aantal links naar webdocumenten vanuit andere webdocumenten of u kunt statische rankingfactoren uitschakelen.
  - Het type categorisering dat u voor deze collectie wilt gebruiken. Als u Geen opgeeft, worden er geen taxonomiegegevens naar enterprise search gemigreerd. Als u regelgebaseerde categorieën selecteert, wordt de taxonomie samen met de collectie naar enterprise search gemigreerd.

Als tijdens de migratie fouten optreden, raadpleegt u het bestand MigrationWizard.log. Dit bestand is opgeslagen in de directory waarin de Migratiewizard is geïnstalleerd.

Nu kunt u de beheerconsole van enterprise search gebruiken om extra instellingen voor de gemigreerde collecties te configureren.

**Vereiste:** Als u webcrawlereigenschappen configureert voor de collectie die u hebt gemigreerd, moet u een e-mailadres (voor het ontvangen van opmerkingen over de crawler) en een gebruikersagentnaam opgeven. Voor hulp klikt u tijdens het configureren van de webcrawlereigenschappen op **Help**.

11. Start het crawlen, analyseren en indexeren voor de gemigreerde collectie in de beheerconsole van enterprise search.
12. Nadat u hebt vastgesteld dat de gemigreerde collectie doorzoekbaar is in enterprise search, wist u de oorspronkelijke collectie in de portalzoekmachine.
13. Optioneel: als WebSphere Portal-beheerder kunt u de volgende stappen uitvoeren als u wilt instellen dat gebruikers de gemigreerde collectie kunnen doorzoeken vanuit een portal in WebSphere Portal.
  - a. Implementeer de enterprise search-portlet in de WebSphere Portal-installatie.

In een WebSphere Portal-servercluster moet deze stap worden uitgevoerd op de server waarop de implementatiebeheerfunctie van WebSphere Application Server is geïnstalleerd. Met deze functie wordt de enterprise search-portlet naar de andere servers in de WebSphere Portal-servercluster gedistribueerd.

- b. Voeg de enterprise search-portlet toe aan de gewenste portalaanpak's.



In WebSphere Portal is de toegangsbesturing van de zoekportlet gebaseerd op de toegankelijkheid tot bepaalde pagina's en portlets. Hoewel de collectie-instellingen zijn gemigreerd, moet de portlet handmatig worden geplaatst door de WebSphere Portal-serverbeheerder.

---

## Instellingen van gemigreerde collecties

Als u collecties vanuit IBM WebSphere Portal migreert, maakt de Migratiewizard standaardinstellingen voor collecties en crawlers.

Als voor de collecties van de portalzoekmachine en de collecties van enterprise search dezelfde instellingen bestaan, gebruikt de wizard de instellingen in de portalzoekmachine tijdens het migreren van de collectie naar enterprise search. Voor instellingen die alleen aanwezig zijn in enterprise search gebruikt de wizard de instellingen die u opgeeft bij het migreren van de collectie of de standaardinstellingen voor collecties in enterprise search.

### Instellingen die aanwezig zijn in de portalzoekmachine en enterprise search

Voor elke collectie die u migreert, migreert de Migratiewizard de volgende instellingen:

- De sites van de portalzoekmachines binnen de collectie van de portalzoekmachine.
- De taal van de collectie.
- De taxonomie (of categoriestructuur) en de regels voor de regelgebaseerde categorieën als de enterprise search-collectie regelgebaseerde categorisering gebruikt.

Elke site van de portalzoekmachine in een collectie wordt samengevoegd in een enterprise search-webcrawler. De Migratiewizard migreert de volgende crawlerinstellingen:

- De start-URL's.
- Het aantal parallelle crawlprocessen.
- De crawldiepte.
- De timeout (in seconden) voor het ophalen van documenten.
- De standaardtekenset.
- Regels voor het crawlen van websites (opnemen of uitsluiten).

### Instellingen die alleen in enterprise search aanwezig zijn

Bij het migreren van een collectie geeft u informatie over de collectie op. De Migratiewizard migreert deze instellingen en gebruikt de standaardinstellingen voor collecties in enterprise search om elke collectie te configureren die u migreert.

U kunt de collectie- en webcrawlerconfiguratie wijzigen met behulp van de beheerconsole van enterprise search. De waarden die tussen haakjes worden weergegeven, ( ), zijn de standaardinstellingen voor de gemigreerde gegevens.

- De collectienaam.
- De strategie voor de statische ranking van het document.
- Het type categorisering dat wordt gebruikt (Regelgebaseerd of Geen).
- Of de zoekcache moet worden gebruikt en hoeveel query-antwoorden de zoekcache kan bevatten (Ja en 5000).

- Of de zoekresponstijd moet worden bewaakt en of een alertsignaal moet verschijnen als een limiet wordt overschreden (Ja en 5 seconden).
- Of gebruik moet worden gemaakt van toegangsbesturing (Nee).
- Een rooster voor het opbouwen van delta-indexen.
- Een rooster voor het opbouwen van de hoofdindex.
- Het detailniveau voor logbestanden (Alle berichten).

Voor elke webcrawler worden in de Migratiewizard de volgende instellingen gemaakt:

- De crawlernaam.
- De crawlerbeschrijving.
- De maximumpaginalengte.
- De beveiligingsinstellingen van het document.
- De MIME-typen (Multipurpose Internet Mail Extensions) van het document die moeten worden gecrawld (indien van toepassing op het gegevensbrontype).

Voordat u de webcrawler start die zojuist is gemigreerd, moet u alle crawler-eigenschappen en crawlruimte-instellingen controleren en ervoor zorgen dat de vereiste waarden zijn opgegeven (vereiste velden worden aangegeven met een rood sterretje). Zorg er met name voor dat u een e-mailadres voor het ontvangen van opmerkingen over de crawler en een gebruikersagentnaam voor de crawler hebt opgegeven. Voor hulp klikt u tijdens het configureren van de webcrawler-eigenschappen op **Help**.

---

## Logbestand van de Migratiewizard

Met de Migratiewizard worden alle berichten naar het bestand WpsMigratorLog.log geschreven. Dit bestand is opgeslagen in de directory waarin de wizard is geïnstalleerd.

Voor elke gemigreerde collectie bevat het bestand WpsMigratorLog.log de waarden van alle instellingen die zijn gelezen met de zoekmachine van WebSphere Portal. Daarnaast wordt in dit bestand aangegeven waar deze instellingen zijn geïmporteerd in de enterprise search-collecties.

---

## Een enterprise search-systeem starten en stoppen

Nadat u een collectie hebt gemaakt, moet u de servers starten zodat de gegevens kunnen worden gecrawld, geanalyseerd en geïndexeerd (de zoekservers worden automatisch gestart). Stop de servers en start deze opnieuw nadat u wijzigingen in de collectie hebt aangebracht.

De meeste enterprise search-servers kunnen doorlopend worden uitgevoerd of op basis van het rooster dat u opgeeft. U kunt bijvoorbeeld een rooster opgeven voor het opbouwen van de hoofdindex en de delta-index. Nadat u het enterprise search-systeem hebt gestart, hoeft u de serverprocessen meestal alleen te stoppen en opnieuw te starten als u de configuratie-instellingen hebt gewijzigd (als u bijvoorbeeld de categorieën hebt bijgewerkt of de zoekcache hebt vergroot).

Als u wijzigingen aanbrengt in de content van een collectie of als u de regels wijzigt voor de manier waarop crawlers gegevens moeten verzamelen van de bronnen in uw bedrijf, moet u de crawlers stoppen en opnieuw starten voordat de wijzigingen actief worden. Als u de crawlregels niet wijzigt, wordt de webcrawler doorlopend uitgevoerd en worden de overige crawlers uitgevoerd op basis van de roosters die u opgeeft.

Om de beschikbaarheid van de zoekservers te verbeteren op momenten dat de indexserver en de beheerconsole niet beschikbaar zijn, kunt u opdrachten opgeven om de zoekservers voor een collectie te starten in stand-alone werkstand.

---

## Een enterprise search-systeem starten

Om in te stellen dat gebruikers een collectie kunnen doorzoeken, moet u de systeemprocessen starten en vervolgens de servers starten waarmee de collectie wordt gecrawld, geanalyseerd en geïndexeerd (de zoekservers worden automatisch gestart).

### Voordat u begint

Configureer de gegevensbronnen die u wilt crawlen en geef opties op voor de manier waarop de gegevens moeten worden geanalyseerd, geïndexeerd en doorzocht. Als u bijvoorbeeld wilt dat gebruikers in de zoekresultaten de categorie-details kunnen bekijken, moet u categorieën configureren voordat u de parser start.

Om de enterprise search-servers te starten, moet u de rol van beheerder van enterprise search hebben, beheerder zijn van de collectie of operator zijn met machtigingen om de collectie te beheren.

U moet de enterprise search-servers voor een collectie in de juiste volgorde starten. U moet bijvoorbeeld een crawler starten en de gegevens crawlen voordat u de gecrawelde gegevens kunt indexeren.

### Beperkingen

Om een enterprise search-systeem te starten, moet u werken met een gebruikersaccount die kan worden gecontroleerd met lokale verificatie. Als u het systeem probeert te starten met een AFS-account (Andrew File System), treden er fouten op.

## Procedure

Ga als volgt te werk om een enterprise search-systeem te starten:

1. Als u enterprise search gebruikt in een configuratie met twee of vier servers, meldt u zich aan als enterprise search-beheerder en start u de CCL (Common Communication Layer) op elke server:

### AIX, Linux of Solaris



```
startccl.sh -bg
```

### Windows-opdrachtaanwijzing

```
startccl
```

### Systeembeheer van Windows-services

Ga als volgt te werk om CCL op de achtergrond te starten:

- a. Start Windows-services: **Start** → **Programma's** → **Systeembeheer** → **Services**.
  - b. Klik met de rechtermuisknop op **WebSphere Information Integrator OmniFind Edition** en klik op **Start**.
2. Start de enterprise search-systeemcomponenten:
    - a. Meld u bij elke enterprise search-server aan als enterprise search-beheerder.
    - b. Start alle systeemcomponenten: `esadmin system startall`  
Met deze opdracht worden de volgende processen en programma's gestart:
      - De webservice (in een configuratie met meerdere servers wordt de webservice op beide zoekservers gestart)
      - De programma's ESSearchServer en ESAdmin in WebSphere Application Server (in een configuratie met meerdere servers worden de programma's op beide zoekservers gestart)
      - De sessie ESAdmin op de indexserver
      - De CCL op de computer waarop u de opdracht hebt uitgevoerd (als de CCL nog niet actief is)
      - De databasenetwerkserver voor enterprise search
      - Het Informatiecentrum van enterprise search (in een configuratie met meerdere servers wordt het Informatiecentrum op beide zoekservers gestart)
  3. Start de beheerconsole van enterprise search en meld u aan als enterprise search-beheerder. Als u werkt met beheerdersrollen, kunt u zich aanmelden als collectiebeheerder of operator met machtigingen voor de collectie die moet worden gestart.
  4. Zoek in de view Collecties de collectie die u wilt beheren en klik op  **Monitor**.
  5. Klik op de pagina Crawler voor elke crawler die u wilt starten op  **Starten**.
    - Als u een webcrawler start, begint de crawler direct met het crawlen van gegevens. Deze typen crawlers kunnen doorlopend worden uitgevoerd om webdocumenten (opnieuw) te crawlen.
    - Als u een van de overige crawlertypen start, wordt de crawlersessie gestart. De crawler begint met crawlen op de geplande datum en tijd. Als u geen rooster voor de crawler hebt opgegeven of als u de crawler eerder wilt starten, bewaakt u de crawler en klikt u voor elke gegevensbron die u wilt crawlen op het startpictogram.  
Als de crawler is gestart, kunt u deze doorlopend uitvoeren. Als u een rooster voor de crawler hebt opgegeven, wordt de crawler op elke geplande datum en tijd uitgevoerd.

6. Als de gegevens zijn gecrawld, opent u de pagina Parser en klikt u op  **Starten** om de parser te starten.  
U kunt de parser doorlopend uitvoeren. Het is meestal niet nodig om de parser te stoppen. Dit hoeft alleen als u wijzigingen hebt aangebracht in de manier waarop documenten worden geanalyseerd (als u bijvoorbeeld de categorieën of XML-veldtoewijzingen hebt bijgewerkt).
7. Optioneel: als u de indexerend handmatig wilt starten en niet wilt wachten tot de geplande datum en tijd, opent u de pagina Index en klikt u in het gebied **Hoofdindex** op  **Starten**.  
U kunt het indexeringsproces doorlopend uitvoeren. De index wordt opgebouwd op de geplande datums en tijden.

**Tip:** De zoekservers worden automatisch gestart en u kunt ze doorlopend uitvoeren. Het is meestal niet nodig om de zoekservers te stoppen. Dit hoeft alleen als u wijzigingen in de instellingen van de zoekcache of documentsamenvattingen hebt aangebracht. Om de zoekservers te starten, opent u de pagina Zoeken. Klik op  **Stoppen** en klik vervolgens op  **Starten**.

Om de beschikbaarheid van de zoekservers te verbeteren op momenten dat de beheerconsole niet beschikbaar is, kunt u opdrachten opgeven om de zoekservers voor een collectie te starten in stand-alone werkstand. Als de indexserver niet actief is, is de beheerconsole niet beschikbaar.

#### **Verwante onderwerpen**

“Enterprise search - Systeembeheer” op pagina 17

U kunt de beheerconsole van enterprise search gebruiken om collecties en externe bronnen te maken en beheren, componenten te starten en stoppen, systeemactiviteit en logbestanden te bewaken, beheerders te configureren, zoekprogramma's te koppelen aan collecties en externe bronnen en informatie op te geven om de beveiliging te verbeteren.

“Beheerdersrollen” op pagina 234

Enterprise search gebruikt het concept van rollen om de toegang tot verschillende functies in de beheerconsole te besturen.

#### **Verwante taken**

“Aanmelden bij de beheerconsole” op pagina 21

Om een enterprise search-systeem te beheren, geeft u een URL in een webbrowser op en meldt u zich vervolgens aan bij de beheerconsole.

“Besturen welke componenten worden gestart of gestopt” op pagina 295

U kunt zelf opgeven welke componenten moeten worden gestart of gestopt met de opdrachten esadmin system startall en esadmin system stopall.

“Zoekservers in de stand-alone werkstand beheren” op pagina 296

Om te zorgen dat de zoekservers altijd beschikbaar zijn, kunt u de zoekservers voor afzonderlijke collecties starten zelfs als de indexserver niet actief is.

#### **Verwante verwijzing**

“Opdrachten, retourcodes en sessie-ID's in enterprise search” op pagina 337

U kunt opdrachten gebruiken om een diagnose te maken van problemen, de status vast te stellen van de verschillende delen van het systeem, sessies te starten en te stoppen of het systeem te starten en te stoppen.

---

## sEen enterprise search-systeem stoppen

Als u wijzigingen hebt aangebracht in een enterprise search-server of als u problemen moet oplossen, moet u de server mogelijk stoppen en opnieuw starten.

### Voordat u begint


Om de enterprise search-servers te stoppen, moet u de rol van beheerder van enterprise search hebben, beheerder zijn van de collectie of operator zijn met machtigingen om de collectie te beheren.

### Over deze taak

U kunt de enterprise search-servers los van elkaar stoppen. Als u bijvoorbeeld een crawler stopt en opnieuw start om de wijzigingen te activeren die u in de crawler-configuratie hebt aangebracht, hoeft u de zoekservers niet te stoppen en opnieuw te starten.





### Procedure

Ga als volgt te werk om de enterprise search-servers te stoppen:

1. Zoek in de view Collecties de collectie die u wilt beheren en klik op  **Monitor**.
2. Zoek op de pagina Crawler de crawler die u wilt beheren en stop of onderbreek deze.

Als u de crawlruimte of crawlereigenschappen wijzigt, moet u de crawler stoppen en opnieuw starten om de wijzigingen door te voeren. Als u de crawlruimte wijzigt en de wijzigingen wilt toepassen op de documenten die al zijn geïndexeerd, moet u ook de documenten opnieuw crawlen.

**Tip:** Mogelijk ziet u een bericht over een timeout van de aangevraagde bewerking hoewel het proces nog op de achtergrond wordt uitgevoerd. Om te bepalen of de taak is voltooid, klikt u op **Vernieuwen** in de beheerconsole (klik niet op **Vernieuwen** in de webbrowser). Het proces is voltooid wanneer het statuspictogram voor de crawler aangeeft dat het is gestopt.

3. Klik op de pagina Parser op  **Stoppen** om de parser te stoppen.  
Als u de regels voor het analyseren van gegevens wijzigt, moet u de parser stoppen en opnieuw starten om de wijzigingen door te voeren. De wijzigingen zijn alleen van toepassing op de documenten die zojuist zijn gecrawld. Als u de wijzigingen wilt toepassen op documenten die al in de index zijn opgenomen, moet u een volledige crawl uitvoeren om alle documenten opnieuw te crawlen, zodat deze opnieuw kunnen worden geanalyseerd en geïndexeerd.
4. Klik op de pagina Index op  **Stoppen** om de index te stoppen die worden opgebouwd.  
U kunt het opbouwen van een index ook stoppen tijdens het bewaken van de indexwachtrij. Hiertoe selecteert u **Systeem** op de werkbalk, opent u de pagina Index en klikt u vervolgens op  **Stoppen** voor de index waarvan u het opbouwen wilt stoppen.
5. Klik op de pagina Zoeken op  **Stoppen** om de zoekservers te stoppen.  
Meestal hoeft u de zoekservers alleen te stoppen en opnieuw te starten wanneer u de instellingen voor de zoekcache of de documentsamenvattingen wijzigt.
6. Ga als volgt te werk om de enterprise search-systeemcomponenten te stoppen in plaats van de afzonderlijke servers:

- a. Meld u bij elke enterprise search-server aan als enterprise search-beheerder.
- b. Stop alle systeemcomponenten: `esadmin system stopall`

Met deze opdracht worden de volgende processen en programma's gestopt:

- De webserver (in een configuratie met meerdere servers wordt de webserver op beide zoekservers gestopt)
- De programma's ESSearchServer en ESAdmin in WebSphere Application Server (in een configuratie met meerdere servers worden de programma's op beide zoekservers gestopt)
- De sessie ESAdmin op de indexserver
- De CCL (Common Communication Layer) voor enterprise search op de computer waarop u de opdracht hebt uitgevoerd
- De databasenetwerkserver voor enterprise search
- Het Informatiecentrum van enterprise search (in een configuratie met meerdere servers wordt het Informatiecentrum op beide zoekservers gestopt)

#### **Verwante onderwerpen**

"Enterprise search - Systeembeheer" op pagina 17

U kunt de beheerconsole van enterprise search gebruiken om collecties en externe bronnen te maken en beheren, componenten te starten en stoppen, systeemactiviteit en logbestanden te bewaken, beheerders te configureren, zoekprogramma's te koppelen aan collecties en externe bronnen en informatie op te geven om de beveiliging te verbeteren.

#### **Verwante taken**

"Aanmelden bij de beheerconsole" op pagina 21

Om een enterprise search-systeem te beheren, geeft u een URL in een webbrowser op en meldt u zich vervolgens aan bij de beheerconsole.

"Besturen welke componenten worden gestart of gestopt"

U kunt zelf opgeven welke componenten moeten worden gestart of gestopt met de opdrachten `esadmin system startall` en `esadmin system stopall`.

"Zoekservers in de stand-alone werkstand beheren" op pagina 296

Om te zorgen dat de zoekservers altijd beschikbaar zijn, kunt u de zoekservers voor afzonderlijke collecties starten zelfs als de indexserver niet actief is.

#### **Verwante verwijzing**

"Opdrachten, retourcodes en sessie-ID's in enterprise search" op pagina 337

U kunt opdrachten gebruiken om een diagnose te maken van problemen, de status vast te stellen van de verschillende delen van het systeem, sessies te starten en te stoppen of het systeem te starten en te stoppen.

---

## **Besturen welke componenten worden gestart of gestopt**

U kunt zelf opgeven welke componenten moeten worden gestart of gestopt met de opdrachten `esadmin system startall` en `esadmin system stopall`.

### **Over deze taak**

Het bestand `ES_INSTALL_ROOT/default_config/AutoRunComponents.properties` bevat een lijst van de enterprise search-componenten die met de opdrachten `esadmin system startall` en `esadmin system stopall` kunnen worden gestart en gestopt. Standaard worden alle opgesomde componenten gestart en gestopt met deze opdrachten.

Als u wilt dat bepaalde componenten niet worden gestart of gestopt, kunt u het eigenschappenbestand aanpassen.

### Procedure

U kunt als volgt opgeven welke componenten moeten worden gestart of gestopt wanneer u het enterprise search-systeem start of stopt:

1. Meld u aan als enterprise search-beheerder. Doe dit op de server waarop u de opdracht `esadmin system startall` of `esadmin system stopall` wilt geven.
2. Roep het bestand `ES_INSTALL_ROOT/default_config/AutoRunComponents.properties` op in een editor.
3. Om te voorkomen dat een bepaalde component wordt gestart, voegt u een veld toe met de naam `Component.startable.component-ID=false`, waarbij `component-ID` het ID is van de component die niet moet worden gestart.
4. Om te voorkomen dat een bepaalde component wordt gestopt, voegt u een veld toe met de naam `Component.stopable.component-ID=false`, waarbij `component-ID` het ID is van de component die niet moet worden gestopt.
5. Sla het bestand op en sluit het af.

De volgende keer dat u de opdracht `esadmin system startall` of `esadmin system stopall` gebruikt, wordt de component die u hebt toegevoegd niet gestart of gestopt, afhankelijk van de wijzigingen die u hebt aangebracht in het eigenschappenbestand.

**Voorbeeld:** In dit voorbeeld wordt de HTTP-server wel gestart met de opdracht `esadmin system startall` (de standaardinstelling), maar niet gestopt met de opdracht `esadmin system stopall` (zoals aangegeven met de vet afgebeelde regel in het voorbeeld):

```
#####  
# Details of component 3.  
#####  
Component.name.3=IBM HTTP server  
Component.impl.class.3=com.ibm.es.control.util.component.impl.HTTPControlImpl  
Component.nodes.3=search  
# By default all components are startable  
Component.stopable.3=false
```

#### Verwante taken

“Een enterprise search-systeem starten” op pagina 291

Om in te stellen dat gebruikers een collectie kunnen doorzoeken, moet u de systeemprocessen starten en vervolgens de servers starten waarmee de collectie wordt gecrawld, geanalyseerd en geïndexeerd (de zoekservers worden automatisch gestart).

“sEen enterprise search-systeem stoppen” op pagina 294

Als u wijzigingen hebt aangebracht in een enterprise search-server of als u problemen moet oplossen, moet u de server mogelijk stoppen en opnieuw starten.

---

## Zoekservers in de stand-alone werkstand beheren

Om te zorgen dat de zoekservers altijd beschikbaar zijn, kunt u de zoekservers voor afzonderlijke collecties starten zelfs als de indexserver niet actief is.

### Beperkingen

Om de zoekservers te starten en te stoppen, moet u een enterprise search-beheerder zijn.



De mogelijkheid om de zoekservers te starten en te stoppen in de stand-alone werkstand is niet beschikbaar vanaf de beheerconsole van enterprise search. Als de indexserver niet actief is, is de beheerconsole niet toegankelijk.

Voordat u de zoekservers in stand-alone werkstand kunt starten en stoppen, moet u controleren of de crawler, de parser, de indexserver en de zoekservers voor de collectie allemaal minimaal één keer zijn gestart. Dit is nodig om te zorgen dat de vereiste bestanden op de zoekservers zijn gesynchroniseerd.

Als de beveiliging op documentniveau voor de collectie is ingeschakeld, moet u zorgen dat de crawlerserver eveneens is gestart. Dit is nodig om te garanderen dat beveiligingsfuncties op documentniveau kunnen worden afgedwongen. Als de crawlerserver niet actief is, worden er in de zoekresultaten alleen documenten opgenomen waarvoor geen gebruikersverificatie vereist is.

### Over deze taak

Als de indexserver niet beschikbaar is, kunt u zorgen dat de gebruikers het systeem kunnen blijven doorzoeken door de zoekservers te starten in de stand-alone werkstand. In een installatie met meerdere servers kunt u de opdrachten voor het starten en stoppen van de zoekservers opgeven vanaf elke enterprise search-server. De opdrachten proberen de zoekservers te starten of te stoppen voor de opgegeven collectie op alle beschikbare zoekservers.

### Procedure

U kunt de zoekservers als volgt starten en stoppen in de stand-alone werkstand:

1. Om de zoekservers te starten voor een collectie op een moment dat de indexserver niet actief is:

- a. Meld u op de crawlerserver en de zoekservers aan als enterprise search-beheerder en start de CCL-service (common communication layer):

#### AIX, Linux of Solaris

```
startccl.sh -bg
```

#### Windows-opdrachtaanwijzing

```
startccl
```

#### Systeembeheer van Windows-services

Ga als volgt te werk om CCL op de achtergrond te starten:

- 1) Start Windows-services: **Start** → **Programma's** → **Systeembeheer** → **Services**.
  - 2) Klik met de rechtermuisknop op **WebSphere Information Integrator OmniFind Edition** en klik op **Start**.
- b. Geef de volgende opdracht op, waarbij *collectie-ID* de collectie aangeeft die de eigenaar is van de zoekservers welke u wilt starten:  

```
esadmin startSearch -cid collectie-ID
```
2. Om de zoekservers te stoppen voor een collectie op een moment dat de indexserver niet actief is:
    - a. Meld u bij elke enterprise search-server aan als enterprise search-beheerder.
    - b. Geef de volgende opdracht op, waarbij *collectie-ID* de collectie aangeeft die de eigenaar is van de zoekservers welke u wilt stoppen:  

```
esadmin stopSearch -cid collectie-ID
```

### Verwante onderwerpen

“Enterprise search - Systeembeheer” op pagina 17

U kunt de beheerconsole van enterprise search gebruiken om collecties en externe bronnen te maken en beheren, componenten te starten en stoppen, systeemactiviteit en logbestanden te bewaken, beheerders te configureren, zoekprogramma's te koppelen aan collecties en externe bronnen en informatie op te geven om de beveiliging te verbeteren.

#### **Verwante taken**

“Een enterprise search-systeem starten” op pagina 291

Om in te stellen dat gebruikers een collectie kunnen doorzoeken, moet u de systeemprocessen starten en vervolgens de servers starten waarmee de collectie wordt gecrawld, geanalyseerd en geïndexeerd (de zoekservers worden automatisch gestart).

“sEen enterprise search-systeem stoppen” op pagina 294

Als u wijzigingen hebt aangebracht in een enterprise search-server of als u problemen moet oplossen, moet u de server mogelijk stoppen en opnieuw starten.

#### **Verwante verwijzing**

“Opdrachten, retourcodes en sessie-ID's in enterprise search” op pagina 337

U kunt opdrachten gebruiken om een diagnose te maken van problemen, de status vast te stellen van de verschillende delen van het systeem, sessies te starten en te stoppen of het systeem te starten en te stoppen.

---

## Activiteit van enterprise search beheren

Tijdens het bewaken van systeem- en collectieactiviteiten kunt u de status van verschillende processen bekijken, mogelijke problemen controleren of de configuratie-instellingen aanpassen om de prestaties te verbeteren.

Met de beheerconsole van enterprise search kunt u het systeem bewaken en indien noodzakelijk bepaalde bewerkingen aanpassen. U kunt gedetailleerde gegevens oproepen voor elke hoofdactiviteit (crawlen, analyseren, indexeren en zoeken). Voorbeelden van deze gegevens zijn gemiddelde responstijden en voortgangsgegevens, zoals hoeveel documenten tijdens een sessie zijn gecrawld of geïndexeerd.

De meeste activiteiten kunt u starten en stoppen door op de bijbehorende pictogrammen te klikken. U kunt een activiteit bijvoorbeeld ook onderbreken (pauze), de configuratie ervan wijzigen of een probleem oplossen, en de werking hervatten zodat de activiteit kan worden voortgezet.

### Verwante onderwerpen

“Enterprise search - Systeembeheer” op pagina 17

U kunt de beheerconsole van enterprise search gebruiken om collecties en externe bronnen te maken en beheren, componenten te starten en stoppen, systeemactiviteit en logbestanden te bewaken, beheerders te configureren, zoekprogramma's te koppelen aan collecties en externe bronnen en informatie op te geven om de beveiliging te verbeteren.

### Verwante taken

“Een enterprise search-systeem starten” op pagina 291

Om in te stellen dat gebruikers een collectie kunnen doorzoeken, moet u de systeemprocessen starten en vervolgens de servers starten waarmee de collectie wordt gecrawld, geanalyseerd en geïndexeerd (de zoekservers worden automatisch gestart).

“sEen enterprise search-systeem stoppen” op pagina 294

Als u wijzigingen hebt aangebracht in een enterprise search-server of als u problemen moet oplossen, moet u de server mogelijk stoppen en opnieuw starten.

### Verwante verwijzing

“Opdrachten, retourcodes en sessie-ID's in enterprise search” op pagina 337

U kunt opdrachten gebruiken om een diagnose te maken van problemen, de status vast te stellen van de verschillende delen van het systeem, sessies te starten en te stoppen of het systeem te starten en te stoppen.

---

## Schatting maken van het aantal documenten in een collectie

Tijdens het maken of bewerken van een enterprise search-collectie geeft u een schatting van het aantal documenten dat in de collectie zal worden opgenomen. Dit aantal wordt in het systeem gebruikt om een schatting te maken van de geheugen- en schijfresources die voor de collectie nodig zijn (en niet om de grootte van de collectie te beperken).

### Voordat u begint

Om de geschatte grootte van een collectie te wijzigen, moet u de rol van beheerder van enterprise search hebben of beheerder van de collectie zijn.

## Over deze taak

Als de collectie groter wordt dan de geschatte omvang, stopt het systeem NIET met het toevoegen van documenten aan de index. Als u alertsignalen voor de collectie configureert en aangeeft dat u een melding wilt ontvangen wanneer het limiet voor het aantal toegestane documenten in de index wordt overschreden, komt het standaardlimiet overeen met de waarde die u voor het geschatte aantal documenten in de collectie hebt opgegeven. Deze schatting en het drempelpercentage voor het alertsignaal dat u opgeeft, worden bewaakt. Wanneer het maximum aantal documenten dat voor de collectie is geconfigureerd, bijna is bereikt, wordt er een e-mail verzonden.

## Procedure

Ga als volgt te werk om een schatting op te geven voor de mogelijke grootte van een collectie:

1. Open een collectie, selecteer de pagina Algemeen en klik op **Algemene opties configureren**.
2. Voer in het veld **Geschat aantal documenten** een aantal in waarmee wordt aangegeven hoe zeer de collectie naar verwachting zal groeien. De standaardwaarde is 1 miljoen documenten.

---

## Een collectie bewaken


U kunt algemene gegevens bekijken over de status van elke component in de collectie of opties selecteren om gedetailleerde informatie te bekijken over de activiteiten van elke component of URI.


### Voordat u begint


Alle enterprise search-beheerders kunnen collecties bewaken. Om componenten te starten of te stoppen, of roosters in of uit te schakelen moet u de rol van beheerder van enterprise search hebben, of beheerder of operator van de collectie zijn.

## Procedure

Ga als volgt te werk om een collectie te bewaken:

1. Zoek in de view Collecties de collectie die u wilt bewaken en klik op  **Monitor**. Er wordt informatie over de huidige status van elke collectiecomponent weergegeven.

**Tip:** Als u een collectie bewerkt en de pagina Algemeen is geopend, kunt u op  **Monitor** klikken om de view voor het bewaken van de collectie weer te geven.

2. Als u de gegevens over een URI wilt bekijken, klikt u op  **Gegevens URI**.  
U kunt bijvoorbeeld bekijken of een bepaalde URI in de index is opgenomen en bekijken of de index waarin de URI is opgenomen, naar de zoekservers is gekopieerd.
3. Als u een afzonderlijke component wilt bewaken en gedetailleerde gegevens over de activiteit van de component wilt bekijken, klikt u op het pictogram **Status**.

### Verwante onderwerpen

“Collecties van enterprise search” op pagina 29

Een enterprise search-collectie bevat de volledige set bronnen die door de

gebruikers met een enkele query kan worden doorzocht. Met behulp van federatie kunnen gebruikers met een enkele query meerdere collecties doorzoeken.

---

## Gegevens over een URI bekijken

U kunt gedetailleerde gegevens over een URI bekijken. U ziet actuele en historische informatie over de manier waarop het document dat door deze URI wordt vertegenwoordigd, is gecrawld, geïndexeerd en doorzocht.

### Voordat u begint

Voordat u een URI-rapport opvraagt of een rapport naar een e-mailadres verzendt, moet u zorgen dat de component actief is waarvan u gegevens wilt ontvangen. Als u bijvoorbeeld gegevens wilt bekijken over de manier waarop een document is gecrawld, geïndexeerd en doorzocht, zorgt u dat de webcrawler, indexserver en zoekservers actief zijn. Als u een gewist document wilt traceren, zorgt u dat de logopties voor documenttracering zijn geconfigureerd.

### Over deze taak

Het verzamelen van informatie over een URI neemt veel tijd in beslag. U kunt ervoor kiezen de informatie te bekijken en te wachten totdat de informatie wordt afgebeeld. U bespaart tijd door het rapport naar een e-mailadres van uw keuze te verzenden.


Voordat u een rapport kunt ontvangen, controleert u of de informatie over uw mailserver is geconfigureerd voor enterprise search. Deze informatie geeft u op tijdens het configureren van de e-mailopties op de pagina Loggen van de view Systeem.

De indexserver en zoekservers kunnen informatie verschaffen over alle URI's (bijvoorbeeld of een URI is opgenomen in de index en of de URI naar de zoekservers is gekopieerd). Om informatie te bekijken over de manier waarop een document is gecrawld, moet u de URI opgeven voor het document dat door de webcrawler is gecrawld.

### Procedure

Ga als volgt te werk om gegevens over een URI te bekijken:

1. Zoek in de view Collecties de collectie die u wilt bewaken en klik op  **Monitor**.

**Tip:** Als u een collectie bewerkt en de pagina Algemeen is geopend, kunt u op  **Monitor** klikken om de view voor het bewaken van de collectie weer te geven.

2. Klik op  **Gegevens URI**.
3. Voer op de pagina Gegevens URI de URI in waarvoor u de gegevens wilt bekijken.
4. Schakel de aankruisvakjes in voor het type gegevens dat u wilt bekijken:

#### Crawlergegevens (alleen beschikbaar voor webcrawlers)

Schakel dit aankruisvakje in als u informatie wilt bekijken over de manier waarop een document door de webcrawler is gecrawld en informatie wilt bekijken over de huidige status in de crawlruimte.

### **Indexgegevens**

Schakel dit aankruisvakje in als u wilt zien of een document is geïndexeerd en naar de zoekservers is gekopieerd.

### **Zoekgegevens**

Schakel dit aankruisvakje in als u informatie wilt bekijken over de manier waarop het document kan worden doorzocht en of het document beschikbaar is om te worden doorzocht.

### **Door de parser gewiste documenten**

Schakel dit aankruisvakje in als u wilt zien of het document van het enterprise search-systeem is gewist tijdens het analyseren, en zo ja, wat de reden daarvan was.

### **Uit de index gewiste documenten**

Schakel dit aankruisvakje in als u wilt zien of het document van het enterprise search-systeem is gewist tijdens het indexeren of analyseren, en zo ja, wat de reden daarvan was.

5. Als u wilt wachten totdat het rapport wordt weergegeven, klikt u op **Rapport bekijken**.
6. Als u een rapport naar een e-mailadres wilt verzenden zodat u het rapport later kunt bekijken, klikt u op **Rapport verzenden**.
  - a. Voer in het veld **E-mailadres om te waarschuwen** op de pagina Gedetailleerd URI-rapport verzenden het e-mailadres in waarnaar het rapport moet worden verzonden.
  - b. Klik op **Rapport verzenden**.

### **Verwante onderwerpen**

“Documenttracering” op pagina 317

Documenten kunnen tijdens verschillende verwerkingsfasen uit het systeem worden gewist. U kunt opties opgeven zodat u precies weet wanneer een document is gewist en wat de oorzaken hiervan zijn.

### **Verwante taken**

“Rapporten over gewiste documenten bekijken” op pagina 318

U kunt gedetailleerde informatie bekijken over documenten die uit een enterprise search-systeem zijn gewist. Deze informatie is alleen beschikbaar als documenttracering voor de collectie is ingeschakeld.

### **Verwante verwijzing**

“URI-indelingen in een enterprise search-index” op pagina 111

Met de URI (Uniform Resource Identifier) van documenten in een enterprise search-index wordt het type crawler aangegeven waarmee het document aan de collectie is toegevoegd.

---

## **Crawlers bewaken**


U kunt algemene gegevens bekijken over de status van elke crawler in een collectie of opties selecteren om gedetailleerde informatie te bekijken over de crawler-activiteiten.


### **Voordat u begint**

Als voor uw beheerdersrol beperkingen zijn ingesteld voor het bewaken van collecties, kunt u de gegevens van de crawler wel bekijken, maar kunt u het gedrag van de crawler niet aanpassen (zoals het starten en stoppen van de crawler).

### **Procedure**

Ga als volgt te werk om een crawler te bewaken:

1. Zoek in de view Collecties de collectie die u wilt bewaken en klik op  **Monitor**.
2. Open de pagina Crawler.

**Tip:** Als u een collectie bewerkt en de pagina Crawler is geopend, kunt u op  **Monitor** klikken om de view voor het bewaken van crawlers weer te geven.

3. Als de crawler actief is of is onderbroken en u gedetailleerde statusinformatie over de crawler wilt bekijken, klikt u op  **Details**. Welke typen informatie worden afgebeeld, is afhankelijk van het crawlertype.

Als voor uw beheerdersrol is ingesteld dat u de processen voor een collectie kunt beheren, kunt u tijdens het bekijken van de gegevens over de crawler-activiteiten de crawler starten, stoppen en onderbreken. Als voor de crawler een rooster kan worden ingesteld, kunt u het crawlrooster ook in- en uitschakelen.

4. Als de crawler is gestopt of onderbroken en u een crawlersessie wilt starten, klikt u op  **Starten** of **Hervatten**.

#### Voor webcrawlers:



Als de crawler is gestopt, begint de crawler opnieuw met crawlen en wordt de volledige crawlruimte gecrawld. Als de crawler is onderbroken, gaat de crawler verder met crawlen aan het begin van de doellocatie waar het crawlen is onderbroken.

Als u direct een volledige crawl wilt uitvoeren, klikt u op het pictogram **Details** en klikt u vervolgens op het pictogram **Volledig opnieuw crawlen starten**. De crawler begint met het crawlen van de volledige crawlruimte, inclusief pagina's die niet zijn gewijzigd sinds de laatste keer dat ze zijn gecrawld. U zou bijvoorbeeld het volledig opnieuw crawlen van alle documenten kunnen starten als u wijzigingen hebt aangebracht in de regels voor het analyseren van documenten en u die regels wilt toepassen op documenten die eerder zijn geïndexeerd.

#### Voor alle andere crawlertypen:

Als de crawler is gestopt, begint de crawler op de geplande datum en tijd met crawlen. De eerste keer dat de crawler een gegevensbron crawlt, wordt een volledige crawl uitgevoerd. Wanneer een geplande crawl wordt herhaald, worden of alle updates van de gegevensbron gecrawld (toevoegingen, verwijderingen en wijzigingen van documenten) of alleen toevoegingen en wijzigingen van documenten. U geeft het type crawl op in het crawlerrooster.

Als u geen rooster voor de crawler hebt opgesteld of als u de crawler eerder wilt starten, klikt u op het pictogram **Gegevens**. Klik vervolgens in het gebied met de gegevens van de crawlruimte op het pictogram van het type uit te voeren crawl: een volledige crawl, alle updates, of alleen nieuwe en gewijzigde documenten. U moet voor elke gegevensbron die u wilt crawlen (zoals een server, database of submap) op het bijbehorende startpictogram klikken.

5. Als de crawler actief is en u de crawler wilt stoppen, klikt u op  **Stoppen** of  **Onderbreken**. De crawler stopt met het crawlen van gegevens totdat u deze opnieuw start of hervat.

Als de crawler is onderbroken, gaat de crawler verder met crawlen aan het begin van de doellocatie waar het crawlen is onderbroken. De DB2-crawler gaat bijvoorbeeld verder met crawlen in de eerste rij van de tabel die werd gecrawld op het moment dat de crawler werd onderbroken.

### Verwante onderwerpen

“Beheer van enterprise search-crawler” op pagina 35

U moet crawlers configureren voor de verschillende gegevenstypen die u in een collectie wilt opnemen. Een collectie kan een willekeurig aantal crawlers bevatten.

### Verwante verwijzing

“Opdrachten, retourcodes en sessie-ID's in enterprise search” op pagina 337

U kunt opdrachten gebruiken om een diagnose te maken van problemen, de status vast te stellen van de verschillende delen van het systeem, sessies te starten en te stoppen of het systeem te starten en te stoppen.

---

## Gegevens over webcrawleractiviteit bekijken


Door de gegevens over de webcrawleractiviteiten te bekijken, kunt u de systeemprestaties beoordelen en indien nodig de webcrawlereigenschappen en crawlerruimtedefinities aanpassen.


### Voordat u begint


Alle enterprise search-beheerders kunnen crawleractiviteiten bewaken. Om een crawler te starten of te stoppen, moet u de rol van beheerder van enterprise search hebben, of beheerder of operator van de collectie zijn.

### Procedure

Ga als volgt te werk om gegevens over een webcrawleractiviteit te bekijken:

1. Zoek in de view **Collecties** de collectie die eigenaar is van de webcrawler die u wilt bewaken en klik op  **Monitor**.
2. Open de pagina **Crawler**.

**Tip:** Als u een collectie bewerkt en de pagina **Crawler** is geopend, kunt u op  **Monitor** klikken om de view voor het bewaken van crawlers weer te geven.

3. Als de webcrawler die u wilt bewaken, actief is of is onderbroken, klikt u op  **Details**.
4. Op de detailpagina voor de webcrawler kunt u de volgende opties bekijken of selecteren om meer informatie weer te geven over de huidige activiteiten en de activiteiten uit het verleden van de crawler.
  - Klik op **Threadgegevens** om te zien hoeveel threads er actief bezig zijn met het crawlen van websites en hoeveel threads er niet actief zijn.
  - Klik op **Actieve sites** om informatie te bekijken over de websites die de crawler actief aan het crawlen is.
  - Klik op **Recentelijk gecrawelde URL's**. Er wordt informatie afgebeeld over de URL's die recentelijk door de crawler zijn gecrawld. Als de items in de lijst niet worden gewijzigd nadat u de view hebt vernieuwd, vinden er geen crawlactiviteiten plaats.
  - Klik op **Crawlerhistorie** om rapporten met crawleractiviteiten in het verleden weer te geven.
  - Voer in het gebied **URL-status** de URL in waarvoor u informatie wilt afbeelden.
    - a. Klik op **URL-gegevens** om statusinformatie voor de URL weer te geven. U kunt alleen URL-gegevens aanvragen voor URL's die eerder zijn gecrawld.



- b. Klik op **Sitegegevens** om de informatie op te geven die u wilt opnemen in het rapport over de website waarvan de URL deel uitmaakt. U kunt sitegegevens opvragen voor eerder gecrawlde websites of voor websites die nog niet zijn gecrawld.

U kunt deze optie bijvoorbeeld gebruiken om te zien of een URL in de crawlruimte is opgenomen, of een URL is gecrawld of alleen is ontdekt, wanneer de URL opnieuw moet worden gecrawld en om informatie te bekijken over de laatste crawl poging voor de website. Daarnaast kunt u de inhoud van het bestand robots.txt voor de website bekijken, dat mogelijk meer duidelijkheid verschaft over de reden waarom de site niet is gecrawld.

## threadgegevens webcrawler

U kunt de webcrawler bewaken om te zien hoeveel threads actief websites aan het crawlen zijn en hoeveel threads niet actief zijn.

Als u tijdens het bewaken van een collectie de gegevens over een webcrawler bekijkt, kunt u de status van de crawlerthreads weergeven. De volgende statuswaarden kunnen worden weergegeven:

### Wachten

Hiermee wordt aangegeven dat voor de thread geen URL aanwezig is die kan worden gecrawld. Deze situatie kan ontstaan als een thread klaar is met crawlen en de crawler niet snel genoeg meer URL's kan vinden om te crawlen. Als voor de crawlereigenschap waarmee wordt bepaald hoe lang de crawler moet wachten voordat de volgende pagina van dezelfde website kan worden opgehaald, een te hoge waarde is ingesteld, kan het zijn dat er niet snel genoeg URL's worden aangeleverd.

### Ophalen

Hiermee wordt aangegeven dat de thread bezig is een pagina van een website te downloaden.

### Voltooid

Hiermee wordt aangegeven dat de thread bezig is de gecrawlde pagina's naar de rest van de crawler te verzenden, maar nog niet gereed is met het crawlen van een andere URL.

### Aangehouden

Hiermee wordt aangegeven dat de crawler is onderbroken.

In de meest ideale situatie halen de threads de pagina's altijd op. Als de status van threads vaak Voltooid is, kan het zijn dat de database te maken heeft met doorvoerproblemen.

Als de status van threads vaak Wachten is, moet u de waarde bekijken die is opgegeven in het veld **Maximum aantal actieve hosts** in de crawlereigenschappen. Als de waarde hoog is, zijn er mogelijk niet genoeg sites in de crawlruimte om de threads bezig te houden of komen niet genoeg URL's in aanmerking om te worden gecrawld. Situaties die kunnen zorgen voor een lage activiteit zijn fouten in DNS- en robotzoekacties.

## Actieve sites van de webcrawler

U kunt de webcrawler bewaken om informatie te bekijken over de websites die de crawler actief aan het crawlen is.

Als u tijdens het bewaken van een collectie de gegevens over een webcrawler bekijkt, kunt u de cijfers voor de actieve sites weergeven. De cijfers geven de volgende informatie aan:

- Hoeveel URL's de crawler op dit moment vanuit de interne database naar het geheugen heeft verplaatst om te crawlen.
- Hoeveel URL's de crawler tot nu toe heeft geprobeerd te crawlen.
- Hoe lang het duurt voordat een site wordt gedeactiveerd en uit het geheugen wordt verwijderd voor de iteratie van de crawler.
- Hoe lang een site tot nu toe in het geheugen is geplaatst.

Deze informatie wijzigt tijdens het doorlopen van de crawlregels die voor de crawler zijn geconfigureerd. In het meest ideale geval komt het aantal geactiveerde URL's zo veel mogelijk overeen met de waarde die is geconfigureerd voor het veld **Maximum aantal actieve hosts** in de crawlergeheugeneigenschappen.

Als het aantal geactiveerde URL's rond de nul is, kan de crawler geen in aanmerking komende URL's vinden. Dit probleem kan worden veroorzaakt door fouten in DNS-zoekacties, problemen met de netwerkconnectiviteit, databasefouten en problemen in de crawlruimtedefinitie. Bijvoorbeeld:

- Als veel sites te lang in het geheugen aanwezig zijn en er weinig URL's zijn gecrawld, zijn er mogelijk netwerkconnectiviteitsproblemen.
- Als er onvoldoende sites in de lijst worden afgebeeld, kunnen er problemen zijn met de crawlruimtedefinitie of DNS-zoekactie.
- Als sites met een redelijke snelheid worden gecrawld en uit het geheugen worden verwijderd terwijl veel URL's niet zijn gecrawld, bewerkt u de crawlergeheugeneigenschappen en past u de timeoutwaarde in het veld **Hoeveelheid tijd dat elke host actief kan blijven** aan zodat sites langer in het geheugen aanwezig zijn.

## crawlsnelheid webcrawler

U kunt de webcrawler bewaken om informatie te bekijken over de snelheid waarmee de crawler pagina's van websites downloadt.

Als u tijdens het bewaken van een collectie de gegevens over een webcrawler bekijkt, kunt u de cijfers weergeven over de snelheid waarmee de crawler gegevens crawlt (de crawlsnelheid). Daarnaast kunt u zien hoeveel URL's de crawler heeft gecrawld sinds de huidige sessie is gestart.

De crawlsnelheid is het aantal pagina's dat per seconde is gecrawld. Dit aantal staat in verband met verschillende eigenschappen die u voor de webcrawler kunt configureren:

- Het aantal crawlerthreads
- Het aantal actieve sites
- De hoeveelheid tijd die de crawler moet wachten voordat de volgende pagina van dezelfde webserver kan worden opgehaald

Als de crawler één actieve site per crawlerthread bevat en de crawler twee seconden moet wachten voordat de volgende pagina van dezelfde website kan worden opgehaald, kan de crawler niet sneller dan één pagina per thread per twee seconden crawlen. Als de crawler bijvoorbeeld het standaard aantal threads (200) gebruikt, kan de crawler 100 pagina's per seconde voor 200 threads crawlen.

Als er twee keer zo veel actieve sites als crawlerthreads zijn en de crawler twee seconden moet wachten totdat de volgende pagina van dezelfde webserver kan worden opgehaald, kan de crawler één pagina per thread per seconde ophalen. De downloadsnelheid van het netwerk en de databasedoorvoercapaciteit zijn dan echter beperkende factoren. Er is sprake van krachtige crawlerprestaties als de crawlsnelheid is aangepast aan het aantal crawlerthreads, het aantal actieve sites en de wachttijd van de crawler.

Een andere factor die u kunt bekijken tijdens het bewaken van de webcrawlerprestaties is het aantal URL's dat de crawler heeft gecrawld sinds de start van de huidige crawlersessie. Als u dit getal deelt door de totale hoeveelheid tijd die de crawler is uitgevoerd, krijgt u het gemiddelde van de doorvoercapaciteit op de lange termijn. Als dit getal niet toeneemt, is de crawler gereed of kan de crawler niet doorgaan. Als er bijvoorbeeld fouten zijn opgetreden in de netwerkconnectiviteit, database of DNS-zoekactie, kan de voortgang van de crawler worden geblokkeerd.

## Webcrawlerrapporten maken

Door de rapporten over de webcrawleractiviteiten te bekijken, kunt u de systeemprestaties beoordelen en indien nodig de webcrawlereigenschappen en crawleruimte-definities aanpassen.

### Voordat u begint

Als voor uw beheerdersrol beperkingen zijn ingesteld voor het bewaken van collecties, kunt u wel de cijfers van de crawler bekijken en de rapporten over de crawleractiviteiten maken, maar kunt u het gedrag van de crawler niet wijzigen (zoals het starten of stoppen van de crawler).


### Over deze taak


De gegevens over de webcrawleractiviteiten kunnen in verschillende typen rapporten worden afgebeeld. Voor bepaalde typen worden de gegevens teruggezonden zodra de gegevens van de interne database van de crawler zijn verzameld. Het maken van de rapporten Site en HTTP-statuscodes neemt enige tijd in beslag. Als u deze typen rapporten maakt, kunt u een e-mailadres opgeven voor de ontvangst van het rapport, zodat u niet hoeft te wachten totdat de resultaten naar de beheerconsole van enterprise search zijn teruggestuurd.

Klik tijdens het bewaken van de webcrawler en het maken van de rapporten op **Help** voor meer informatie over de interpretatie van de cijfers in de rapporten.

### Procedure

Ga als volgt te werk om webcrawlerrapporten te maken:

1. Zoek in de view Collecties de collectie die eigenaar is van de webcrawler die u wilt bewaken en klik op  **Monitor**.
2. Open de pagina Crawler.

**Tip:** Als u een collectie bewerkt en de pagina Crawler is geopend, kunt u op  **Monitor** klikken om de view voor het bewaken van crawlers weer te geven.

3. Als de webcrawler waarvoor u rapporten wilt maken, actief is of is onderbroken, klikt u op  **Details**.

4. Op de detailpagina voor de webcrawler kunt u de volgende opties selecteren voor het type rapport dat u wilt maken:
  - Klik in het gebied **Statusoverzicht crawler** op **Crawlerhistorie** als u rapporten wilt maken over de crawler en alle sites die door de crawler zijn ontdekt of gecrawld.
  - Geef in het gebied **URL-status** de URL op van de site waarvoor u een rapport wilt maken en klik vervolgens op **Sitegegevens**.
5. Voor elk cijfer dat u in het rapport wilt afbeelden, schakelt u het aankruisvakje in (voor zowel crawlerhistorie- als siterapporten). Klik vervolgens op **Rapport bekijken**.

Voor dergelijke cijfers stuurt de crawler een rapport naar de beheerconsole terug zodra de gegevens uit de interne database zijn opgehaald.

6. Als u een crawlerhistorierapport maakt, kunt u opties opgeven voor het maken van een siterapport. Klik als u klaar bent op **Rapport uitvoeren**.

Het rapport wordt gemaakt met de cijfers die u hebt geselecteerd en wordt opgeslagen in het bestand dat u opgeeft (de bestandsnaam moet absoluut zijn). U kunt aangeven dat u een e-mail wilt ontvangen zodra het rapport is gemaakt.

7. Als u een crawlerhistorierapport maakt, kunt u opties opgeven voor het maken van een HTTP-statuscoderapport. Klik als u klaar bent op **Rapport uitvoeren**.

Dit rapport bevat informatie over het aantal HTTP-statuscodes dat per site is gedistribueerd. Het rapport wordt opgeslagen in een bestand dat u opgeeft (de bestandsnaam moet absoluut zijn). U kunt aangeven dat u een e-mail wilt ontvangen zodra het rapport is gemaakt.

Gebruik dit rapport als u wilt zien welke sites een groot aantal 4xx-statuscodes (waarmee wordt aangegeven dat bepaalde pagina's niet zijn gevonden), 5xx-statuscodes (waarmee serverproblemen worden aangegeven), 6xx-statuscodes (waarmee connectiviteitsproblemen worden aangegeven) en andere codes als resultaat worden gegeven.

Dit rapport is met name nuttig als de crawler al enige tijd actief is (bijvoorbeeld enkele weken). Op basis van het rapport kunt u verwijderde sites identificeren, alsmede nieuwe sites, sites met veel URL's (wat kan duiden op overbodig crawlen van een Lotus Notes-database) en sites met een recursief bestandsstelsel dat wordt verzorgd door de HTTP-server. Als de sites met veel HTTP-statuscodes geen bijdrage leveren aan de index, kunt u de prestaties van de crawler verbeteren door de sites uit de crawlruimte te verwijderen.

## HTTP-statuscodes die naar de webcrawler worden verzonden

Tijdens het bewaken van een webcrawler kunt u informatie bekijken over de HTTP-statuscodes die de crawler ontvangt van de pagina's die de crawler probeert te crawlen.

### Tabeloverzicht

Bij het bewaken van de webcrawlerhistorie of het bewaken van de status van een bepaalde URL, kunt u informatie bekijken over de HTTP-statuscodes die naar de crawler worden verzonden. U kunt deze informatie gebruiken voor het beheren van de crawlruimte en het optimaliseren van de crawlerprestaties. Als de crawler bijvoorbeeld een groot aantal HTTP-statuscodes voor een URL ontvangt en de statuscodes aangeven dat de pagina's op die locatie niet kunnen worden gecrawld, kunt u de prestaties verbeteren door de desbetreffende URL uit de crawlruimte te verwijderen.

In de volgende tabel vindt u een overzicht van de HTTP-statuscodes en de manier waarop deze door de webcrawler worden geïnterpreteerd. Waarden tussen 100 en 505 zijn standaard-HTTP-statuscodes (zie voor meer informatie <http://www.w3.org/Protocols/rfc2616/rfc2616.html>). De overige HTTP-statuscodes zijn afkomstig van enterprise search en de webcrawler.

Tabel 8. HTTP-statuscodes van de webcrawler

Code	Beschrijving	Code	Beschrijving	Code	Beschrijving	Code	Beschrijving
NULL	Niet gecrawld	400	Ongeldige opdracht	500	Interne serverfout	693	Selecteren mislukt (URLFetcher)
100	Voortzetten	401	Niet gemachtigd	501	Niet geïmplementeerd	694	Schrijffout (URLFetcher)
101	Protocollen wisselen	402	Betaling vereist	502	Foutieve gateway	695	Onvolledige blokheader (URLFetcher)
200	Gelukt	403	Verboden	503	Service niet beschikbaar	699	Onverwachte fout (URLFetcher)
201	Gemaakt	404	Niet gevonden	504	Gatewaytimeout	700	Parsingfout (geen header-einde)
202	Geaccepteerd	405	Methode niet toegestaan	505	HTTP-versie niet ondersteund	710	Parsingfout (header)
203	Niet-gezaghebbende informatie	406	Niet aanvaardbaar	611	Leesfout	720	Parsingfout (geen HTTP-code)
204	Geen content	407	Proxyverificatie vereist	612	Verbindingsfout	730	Parsingfout (lopende tekst)
205	Content gereset	408	Opdrachttimeout	613	Leestimeout	740 of 4044	Uitgesloten op basis van bestand robots.txt
206	Gedeeltelijke content	409	Conflict	614	SSL-handshake mislukt	741	Robots tijdelijk niet beschikbaar
300	Meerdere keuzen	410	Verdwenen	615	Andere leesfout	760	Uitgesloten door definitie van crawlruimte
301	Permanent verplaatst	411	Lengte vereist	616	FBA-afwijking	761	Niet toegestaan door lokale crawlruimte; toegestaan door global
302	Gevonden	412	Niet voldaan aan voorwaarde	617	Coderingsfout	770	Onjuist protocol of niet-standaard systeem-poort
303	Zie andere	413	Opdrachtentiteit is te groot	618	Doorsturen zonder doorstuur-URL	780	Uitgesloten i.v.m. bestandstype

Tabel 8. HTTP-statuscodes van de webcrawler (vervolg)

Code	Beschrijving	Code	Beschrijving	Code	Beschrijving	Code	Beschrijving
304	Niet gewijzigd	414	Opdracht-URI is te lang	680	DNS-zoek-opdracht mislukt	786	Ongeldige URL
305	Proxy gebruikt	415	Niet-ondersteund medium	690	Onjuist samengestelde URL	2004	Geen index META-tag
306	(Ongebruikt)	416	Aangevraagd bereik is ongelidig	691	Verbinding verbroken (URLFetcher)	3020	Zachte redirect
307	Tijdelijk doorsturen	417	Niet voldaan aan verwachting	692	Timeout voor schrijven (URLFetcher)		

## Opmerkingen bij de tabel

### 4xx-statuscodes

400-codes (ongeldige opdracht) worden zelden weergegeven. Op basis van de HTTP-statuscodestandaard moeten 4xx-codes aangegeven dat de client (de crawler) is mislukt. Meestal wordt het probleem echter veroorzaakt door de server of de URL die de crawler als link heeft ontvangen. Op sommige webserverns kunnen bijvoorbeeld geen URL's worden gebruikt die vanaf de siteroot omhoog proberen te navigeren (zoals `http://xyz.ibm.com/../../sales`). Voor andere webserverns is deze opwaartse navigatie geen probleem en wordt de operator voor de bovenliggende directory genegeerd (..) als de crawler zich al in de root bevindt.

Op sommige serverns worden opdrachten voor de siteroot als fout beschouwd en worden met bepaalde verouderde links bewerkingen aangevraagd die niet meer worden herkend of niet meer zijn geïmplementeerd. Als een pagina wordt aangevraagd die niet meer bestaat, wordt door de server een uitzondering gegeven zodat de webserver HTTP-statuscode 400 als resultaat geeft omdat de opdracht niet meer geldig is.

- 615** Hiermee wordt aangegeven dat op de crawlerserver die gegevens van websites downloadt, een onverwachte uitzondering is opgetreden. Vaak wordt met dit type statuscode aangegeven dat er een probleem in de crawler is ontstaan.

### 61x-statuscodes

Met uitzondering van 615 geeft de statuscode 61x problemen aan die tijdens het crawlen kunnen optreden, zoals een timeout. Voor de volgende statuscodes kunnen herstelprocedures nodig zijn:

#### 611, 612 en 613

Trage sites of slechte netwerkprestaties kunnen de oorzaak van deze problemen zijn.

**611** Hiermee wordt aangegeven dat in de crawler een fout is opgetreden bij het ophalen van een document.

**612** Hiermee wordt aangegeven dat in de crawler een fout is opgetreden bij het maken van een verbinding met een webserver.

**613** Hiermee wordt aangegeven dat in de crawler een timeout is opgetreden bij het ophalen van een document.

- 614** Hiermee wordt aangegeven dat de crawler geen beveiligde (HTTPS-)sites kan crawlen. Als u denkt dat deze sites toegankelijk zouden moeten zijn, controleert u of de certificaten op de juiste manier zijn ingesteld op de crawlerserver en op de doelwebserver. Als voor een site een certificaat aanwezig is van een bekende certificaatgever, kunt u nieuwe certificaatgevers aan het trustarchief van de crawler toevoegen.

Controleer ook hoe certificaten met eigen handtekening zijn geconfigureerd op de sites die u wilt crawlen. De crawler is geconfigureerd om certificaten met eigen handtekening te accepteren. Sommige sites maken certificaten met eigen handtekening voor een root-URL (zoals `http://sales.ibm.com/`) en proberen vervolgens dat certificaat te gebruiken voor subdomeinen (zoals `http://internal.sales.ibm.com/`). De crawler kan geen certificaten accepteren die op deze manier worden gebruikt. Certificaten met eigen handtekening worden alleen geaccepteerd als de domeinnaam van het onderwerp (`sales.ibm.com`) en de ondertekenaar van het certificaat overeenkomen met de domeinnaam van de pagina die wordt aangevraagd.

- 616** Hiermee wordt aangegeven dat het aanmeldformulier voor FBS (Form-based Authentication) nog steeds in de download wordt afgebeeld nadat de verificatie is uitgevoerd.

Als de crawler niet kan worden geverifieerd op basis van de informatie die in het FBA-configuratiebestand is opgegeven (aanmeldformulier plus verificatiegegevens zoals de gebruikersnaam, het wachtwoord, enzovoort), wordt statuscode 616 toegewezen aan alle pagina's die afhankelijk zijn van de FBA. De beheerder moet uitvinden waarom de FBA-configuratie niet werkt.

- 617** Hiermee wordt aangegeven dat er geen tekenreeks kan worden gemaakt op basis van de bytecontent van een document omdat de coderingstekensreeks (tekenset) ongeldig is of omdat het document ongeldige bytes bevat.

- 618** Hiermee wordt aangegeven dat de doorstuur-URL ongeldig is wanneer de crawler de volgende HTTP-statuscodes ontvangt. Mogelijk is de locatie van de HTTP-responsheader ongeldig.

301 Permanent verplaatst  
302 Gevonden

- 680** Hiermee wordt aangegeven dat de crawler geen IP-adressen kan ophalen voor de hosts in de crawlruimte, mogelijk als gevolg netwerktoegangsproblemen. Dit type fout houdt in dat de crawler geen volledige sites kan crawlen, niet dat slechts enkele URL's niet kunnen worden gecrawld. Als u veel statuscodes van dit type ontvangt, wordt de doorvoer sterk vermindert.

#### **69x-statuscodes**

Statuscodes 690 t/m 699 worden nooit vastgelegd in de permanente database van de crawler. Met deze codes worden resultaten aangegeven die niet overeenkomen met het werkelijke resultaat van een downloadactie van een host op afstand. In plaats daarvan wordt een tijdelijke toestand in de crawler aangegeven, zoals een component die wordt afgesloten terwijl een andere component wacht op een resultaat of bezig is met het verzenden

van een resultaat. Deze statuscodes verschijnen in bepaalde logbestanden, maar niet in het permanente record en moeten dus niet worden gebruikt als waarden voor de selectieset.

### **7xx-statuscodes**

De 7xx-codes verschijnen meestal als gevolg van regels in de crawlruimte:

#### **710 - 730**

Hiermee wordt aangegeven dat de crawler als gevolg van problemen geen volledige download heeft uitgevoerd of dat de crawler ongeldige HTML-gegevens op een site is tegengekomen. Als veel van deze typen statuscodes worden afgebeeld, neemt u voor hulp contact op met de enterprise search-vertegenwoordiger.

#### **740 of 4044**

Hiermee wordt aangegeven dat de inhoud van een bestand niet kan worden geïndexeerd omdat het document is uitgesloten op basis van beperkingen in het bestand robots.txt van de site.

**740** Hiermee wordt aangegeven dat ankercodelinks die naar het uitgesloten document verwijzen, in de index kunnen worden opgenomen.

**4044** Hiermee wordt aangegeven dat de ankercodelinks in documenten die naar het uitgesloten document verwijzen, ook worden uitgesloten van de index.

**741** Hiermee wordt aangegeven dat voor de site het bestand robots.txt aanwezig is, dat regels bevat op basis waarvan de site kan worden gecrawld, maar niet kan worden gedownload. Als de URL herhaaldelijk niet kan worden gecrawld, wordt de URL uit de crawlruimte verwijderd. Als dit type statuscode vaak voorkomt, controleert u of de doelsite tijdelijk of permanent niet toegankelijk is. Als de doelsite niet meer beschikbaar is, verwijdert u de site uit de crawlruimte.

De resterende 7xx-statuscodes verschijnen meestal als u wijzigingen in de crawlruimte aanbrengt nadat de crawler enige tijd is uitgevoerd. Met deze statuscodes worden meestal geen problemen aangegeven die u moet oplossen.

**3020** Hiermee wordt aangegeven dat een document met statuscode een locatieheader bevat waarmee de gebruikersagent naar een andere URL wordt verwezen.

---

## **De parser bewaken**

U bewaakt de parser als u gegevens wilt bekijken over documenten die door de parser zijn geanalyseerd voordat deze aan de enterprise search-index worden toegevoegd. Met behulp van een aantal opties kunt gegevens bekijken en de activiteiten van de parser beheren.

### **Voordat u begint**

Als voor uw beheerdersrol beperkingen zijn ingesteld voor het bewaken van collecties, kunt u de status van de parser wel bekijken, maar kunt u de parser niet starten of stoppen.

### **Over deze taak**




Bij het bewaken van de parsergegevens wordt een momentopname van de parseractiviteit afgebeeld. Deze momentopname toont de cijfers van de analyseactiviteiten op een bepaald moment. In de cijfers vindt u informatie over het aantal documenten dat is gecrawld en geanalyseerd of dat wacht op analyse, en het aantal documenten dat is geanalyseerd en wacht op opslag in de index.


Als de parser actief is, worden berichten met aanvullende informatie over de status van de parser afgebeeld. Bijvoorbeeld:

- De parser is mogelijk bezig met het analyseren van documenten.
- De parser is mogelijk niet actief. De parser is inactief totdat er meer documenten beschikbaar zijn voor analyse. Als er fouten optreden, wordt gewacht totdat de parser opnieuw wordt gestart. De parser wordt automatisch opnieuw gestart als er geen parserservices beschikbaar zijn (bijvoorbeeld als er geen verbinding met de parserservice kan worden gemaakt of als alle virtuele Java-machines voor de parser bezig zijn met andere collecties).
- De parser is mogelijk onderbroken (de parser kan bijvoorbeeld wachten totdat het opbouwen van de index is voltooid).

### Procedure

Ga als volgt te werk om de parser voor een collectie te bewaken:

1. Zoek in de view Collecties de collectie die u wilt bewaken en klik op  **Monitor**.
2. Open de pagina Parser.


**Tip:** Als u een collectie bewerkt en de pagina Parser is geopend, kunt u op  **Monitor** klikken om de view voor het bewaken van de collectie weer te geven.

3. Als de parser actief is en u gedetailleerde statusinformatie over de analyseactiviteiten wilt bekijken, klikt u op  **Details**.

Als voor uw beheerdersrol is ingesteld dat u de processen voor een collectie kunt beheren, kunt u tijdens het bekijken van de gegevens over de analyseactiviteiten de parser starten en stoppen.

4. Als de parser is gestopt en u deze wilt starten, klikt u op  **Starten**.

Denk er bij het maken van een collectie aan dat u de parser pas start nadat de crawler is begonnen met het crawlen van gegevens. U weet dan zeker dat de parser gegevens heeft om te analyseren en te categoriseren. Tenzij u wijzigingen in de analyseregels aanbrengt, kunt u de parser doorlopend uitvoeren.

5. Als de parser actief is en u deze wilt stoppen, klikt u op  **Stoppen**.

U moet de parser stoppen en opnieuw starten als u wijzigingen aanbrengt in de parsingregels. Als u bijvoorbeeld de parserconfiguratie hebt gewijzigd, moet u de parser stoppen en opnieuw starten voordat de wijzigingen van kracht worden.

---

## De indexactiviteit voor een collectie bewaken


U kunt de index voor een collectie bewaken als u de voortgang van het opbouwen van een index wilt bekijken, het indexrooster wilt in- of uitschakelen of indexactiviteiten wilt starten en stoppen.


### Voordat u begint






Alle enterprise search-beheerders kunnen indexactiviteiten bewaken. Om het opbouwen van een index te starten of stoppen of het indexrooster in of uit te schakelen, moet u de rol van beheerder van enterprise search hebben of beheerder of operator van de collectie zijn.

### Procedure

Ga als volgt te werk om de index voor een collectie te bewaken:

1. Zoek in de view Collecties de collectie die u wilt bewaken en klik op  **Monitor**.
2. Open de pagina Index.

**Tip:** Als u een collectie bewerkt en de pagina Index al is geopend, kunt u op  **Monitor** klikken om de view voor het bewaken van de collectie te weer te geven.

3. Als een index is gepland en u wilt voorkomen dat de index op de geplande datum en tijd wordt opgebouwd, klikt u op  **Rooster uitschakelen**. De index wordt dan pas weer opgebouwd als u het rooster weer inschakelt of als u het opbouwen handmatig start.
4. Als een index is gepland maar het rooster voor het opbouwen is uitgeschakeld, klikt u op  **Rooster inschakelen**. De index komt dan in de wachtrij om te worden opgebouwd op de datum en tijd die in het rooster zijn ingesteld.
5. Als een index is gestopt en u deze wilt starten, klikt u op  **Starten**.  
Gewoonlijk vindt het indexeren op regelmatige basis plaats, via een rooster. Als u een index hebt gestopt tijdens het opbouwen of als u het rooster voor een index uitschakelt tijdens het opbouwen, kunt u op **Starten** klikken om het opbouwen van de index handmatig te starten.
6. Als een index wordt opgebouwd en u het opbouwen wilt stoppen, klikt u op  **Stoppen**.  
Het kan voorkomen dat u het opbouwen van een delta-index moet stoppen, bijvoorbeeld als u het opbouwen van de hoofdindex handmatig wilt starten nadat u het type categorisering hebt gewijzigd dat in de collectie wordt gebruikt.
7. Als tijdens het opbouwen van een index fouten optreden, klikt u op  **Fout**.  
De pagina Inhoud logbestand wordt afgebeeld, zodat u extra gegevens over de indexeerfouten kunt bekijken. Op die pagina kunt u ook afzonderlijke foutberichten selecteren om meer over een probleem te weten te komen.

#### Verwante onderwerpen

“Beheer van enterprise search-index” op pagina 159

Om ervoor te zorgen dat gebruikers altijd toegang hebben tot de laatste informatie, maakt enterprise voor elke collectie een index. Deze index wordt beheerd door op regelmatige basis de content bij te werken.

---

## De indexeerwachtrij van enterprise search bewaken

U kunt de status van het opbouwen van een index in de indexeerwachtrij bekijken, het opbouwen van een index stoppen of een index uit de wachtrij wissen.

### Voordat u begint

Om de indexeerwachtrij te beheren, moet u de rol van beheerder van enterprise search hebben.

### Over deze taak

U kunt meerdere indexen tegelijkertijd opbouwen, maar er kan slechts één index per collectie tegelijk in de wachtrij worden geplaatst. Bij het configureren van de indexopties voor het systeem kunt u opgeven hoeveel indexen gelijktijdig gemeenschappelijk gebruik kunnen maken van de wachtrij en de indexeerresources.

### Procedure

Ga als volgt te werk om de indexeerwachtrij te bewaken:

1. Klik op **Systeem** om de view Systeem te openen.

2. Selecteer de pagina Index.


Er verschijnt een lijst van de collecties waarvoor indexen in de indexeerwachtrij zijn opgenomen. Voor elke index kunt u het type index bekijken dat wordt afgebeeld (delta of volledig), de tijd weergegeven waarop de index in de indexeerwachtrij is geplaatst en de tijd weergegeven dat het opbouwen van de index is gestart (als de index momenteel wordt opgebouwd).


3. Om een afzonderlijke index te beheren, klikt u op het pictogram **Status**.

U kunt bijvoorbeeld de voortgang van het opbouwen van een index bekijken, zien hoeveel documenten in de index zijn geplaatst of het indexrooster uitschakelen.

4. Als u het opbouwen van een index wilt stoppen, klikt u op  **Stoppen**.

Als u bijvoorbeeld categorieregels hebt gewijzigd, is het verstandig om het opbouwen van een delta-index te stoppen zodat u opdracht kunt geven voor het opbouwen van een hoofdindex.

Om het opbouwen van een index te starten nadat u dit hebt gestopt, moet u wachten op de volgende geplande starttijd van de index of klikt u op het pictogram **Status** om de index te bewaken en klikt u vervolgens op  **Starten** om het opbouwen van de index te starten.

5. Om een index uit de indexeerwachtrij te verwijderen, klikt u op  **Verwijderen**.

#### Verwante onderwerpen

“Beheer van enterprise search-index” op pagina 159

Om ervoor te zorgen dat gebruikers altijd toegang hebben tot de laatste informatie, maakt enterprise voor elke collectie een index. Deze index wordt beheerd door op regelmatige basis de content bij te werken.

---

## De zoekservers bewaken

U kunt gedetailleerde statusinformatie bekijken over alle zoekserveractiviteiten voor een bepaalde collectie, of u kunt gedetailleerde statusinformatie bekijken voor de zoekservers in het hele enterprise search-systeem.


### Voordat u begint


Alle beheerders van enterprise search kunnen zoekservers bewaken voor de collecties waarvan ze beheerder zijn. Om alle zoekservers in het enterprise search-systeem te bewaken, moet u de rol van beheerder van enterprise search hebben.

Om een zoekserver te starten of te stoppen, moet u de rol van beheerder van enterprise search hebben, of beheerder of operator van de collectie zijn.


### Procedure

1. Ga als volgt te werk om de zoekservers voor één collectie te bewaken:

- a. Zoek in de view Collecties de collectie die u wilt bewaken en klik op  **Monitor**.
- b. Open de pagina Zoeken.

**Tip:** Als u een collectie bewerkt en de pagina Zoeken al is geopend, kunt u op  **Monitor** klikken om de view voor het bewaken van de collectie weer te geven.

2. Ga als volgt te werk om alle zoekservers in het enterprise search-systeem te bewaken:
  - a. Klik op **Systeem** om de view Systeem te openen.
  - b. Selecteer de pagina Zoeken.

3. Als een zoekserver is gestopt en u deze wilt starten, klikt u op  **Starten**.

4. Als een zoekserver actief is en u deze wilt stoppen, klikt u op  **Stoppen**.

Als u de zoekcache in- of uitschakelt, wijzigingen aanbrengt in de grootte van de zoekcache of wijzigingen aanbrengt in de snellinks, moet u de zoekservers stoppen en opnieuw starten om de wijzigingen te activeren.

5. Als u een overzicht wilt bekijken van de hoeveelheid tijd die een zoekserver besteedt aan het verwerken van zoekopdrachten, klikt u op **Historieresponstijd**.

In dit rapport ziet u hoeveel tijd (in milliseconden) de zoekserver op een bepaalde datum gemiddeld nodig heeft om te reageren op zoekopdrachten.

De gemiddelde responstijd is een indicator voor de systeemprestaties en laat de servicekwaliteit zien. En hoge responstijd kan betekenen dat het systeem zwaar wordt belast. Zo kan het zijn dat het aantal collecties dat wordt doorzocht en de grootte van de collectie het systeem te zwaar belasten.

6. Om een lijst af te beelden van de query's die het meest frequent worden uitgevoerd, klikt u op **Populaire query's**.

Het rapport bevat de trefwoorden van de 50 meest frequent uitgevoerde query's en laat zien hoe vaak een bepaalde query is uitgevoerd.

Op basis van dit rapport kunt u bepalen welke query's in aanmerking komen voor snellinks. Met snellinks kunt u de kwaliteit van zoekopdrachten voor veel gebruikers verbeteren. U kunt ervoor zorgen dat zeer relevante documenten altijd in de zoekresultaten worden weergegeven.

Mogelijk wilt u ook links maken naar de resources die de query's vanuit de enterpriseportal beantwoorden. Als gebruikers bijvoorbeeld regelmatig informatie over de onkostenprocedure opvragen, kunt u op de homepage van uw intranet een link opnemen naar de pagina waarop de onkostenprocedure wordt beschreven.

7. Om een lijst af te beelden van de meest recent uitgevoerde query's, klikt u op **Recente query's**.

Het rapport bevat de trefwoorden van de 50 meest recent uitgevoerde query's.

Op basis van de meest recent uitgevoerde query's kunt u de huidige trends en urgente situaties in uw organisatie vaststellen. Zo kunt u een piek ontdekken in de interesse voor een bepaald onderwerp. Deze piek kan aangeven dat een snellink voor dat onderwerp moet worden gemaakt zodat gebruikers op andere manieren toegang hebben tot dat onderwerp (bijvoorbeeld via een link in de enterpriseportal).

---

## De Data Listener bewaken



U kunt de Data Listener bewaken om de status ervan te bekijken en om gegevens over de activiteiten van de Data Listener-clientprogramma's weer te geven.

### Voordat u begint

Om de Data Listener te bewaken, moet u de rol van beheerder van enterprise search hebben.

### Procedure

Ga als volgt te werk om de Data Listener te bewaken:

1. Klik op **Systeem** om de view Systeem te openen.
2. Bekijk de statuspictogrammen op de pagina Data Listener om te zien of de Data Listener actief is of is gestopt.
3. Als de Data Listener actief is en u gedetailleerde statusinformatie over de activiteiten van het clientprogramma wilt bekijken, klikt u op  **Details**.  
Met de statuspictogrammen op de pagina Data Listener-gegevens wordt aangegeven of de Data Listener actief is of is gestopt. In de cijfers vindt u informatie over het aantal opdrachten dat wacht op verwerking, de huidige status van elke thread die werkt met opdrachten van clientprogramma's en het aantal threads dat actief is voor een bepaalde threadstatus.
4. Als u het poortnummer voor de Data Listener wijzigt of het maximum aantal documenten dat op een tijdelijke opslagplek kan worden bewaard, klikt u op  **Opnieuw starten**.

De Data Listener wordt automatisch gestart bij het starten van het enterprise search-systeem. Gewoonlijk is het niet nodig de Data Listener opnieuw te starten, tenzij u een of beide bovenstaande configuratieopties wijzigt.

### Verwante taken

"Ondersteuning voor Data Listener-programma's configureren" op pagina 107  
U kunt enterprise search uitbreiden door een externe crawler te maken met behulp van de Data Listener-API. Met de aangepaste Data Listener-programma's kunnen gegevens aan collecties worden toegevoegd, gegevens uit collecties worden verwijderd of webcrawlers opdracht worden gegeven URL's (opnieuw) te bezoeken.

---

## Documenttracering

Documenten kunnen tijdens verschillende verwerkingsfasen uit het systeem worden gewist. U kunt opties opgegeven zodat u precies weet wanneer een document is gewist en wat de oorzaken hiervan zijn.

Als in de parser een fout optreedt als gevolg waarvan het document niet kan worden geanalyseerd, wordt er een bericht met een redencode voor het gewiste document gelogd. (Op basis van dit fouttype worden oudere documentversies overigens niet uit de index verwijderd.)

Documenten kunnen tijdens de indexeringsfase worden gewist, de bijbehorende gegevens worden dan ook gelogd. URI's en URI-patronen kunnen expliciet worden gewist. Een document kan door een crawler zijn gecrawld die later is gewist. Het brondocument is mogelijk niet meer aanwezig (als aan het document een negatieve HTTP-code is gekoppeld) of de aan het document gekoppelde HTTP-code is onbe-

kend. Documenten kunnen ook worden gewist als de ranggegevens ontbreken voor een document waarvoor algemene analyse moet worden uitgevoerd.

Als u weet dat een document is gecrawld maar het document niet in de index wordt afgebeeld, kunt u de beheerconsole van enterprise search gebruiken om de stroom van het document in het systeem te traceren. In gedetailleerde rapporten wordt aangegeven wanneer, waar en waarom het document is gewist. In het rapport kan bijvoorbeeld worden aangegeven dat het document onverwacht is gewist tijdens de algemene analyse of dat een beheerder de URI uit de index heeft verwijderd.

#### **Verwante taken**

“Gegevens over een URI bekijken” op pagina 301

U kunt gedetailleerde gegevens over een URI bekijken. U ziet actuele en historische informatie over de manier waarop het document dat door deze URI wordt vertegenwoordigd, is gecrawld, geïndexeerd en doorzocht.

## **Logbestanden configureren voor documenttracering**

Om te bepalen wanneer, waar en waarom een document uit het systeem is gewist, kunt u logbestanden configureren waarin informatie over gewiste documenten wordt gelogd.

#### **Voordat u begint**

Om opties te configureren voor het traceren van gewiste documenten, moet u de rol van beheerder van enterprise search hebben of beheerder van de collectie zijn.

#### **Over deze taak**

Om te voorkomen dat logbestanden te veel schijfruimte innemen, worden de logbestanden automatisch overschreven en wordt er altijd een nieuw logbestand gestart wanneer de huidige datum verandert. Als een logbestand groeit tot de maximaal toegestane grootte en de datum niet is veranderd, maakt het systeem een nieuw logbestand. Wanneer het maximum aantal logbestanden is gemaakt, wordt het oudste logbestand gewist zodat er een nieuw bestand kan worden gemaakt.

#### **Procedure**

Ga als volgt te werk om logbestanden te configureren voor documenttracering:

1. Open een collectie, selecteer de pagina Loggen en klik op **Documenttracering configureren**.
2. Schakel het aankruisvakje voor het traceren van documenten in op de pagina Documenttracering.
3. Geef aan hoeveel logbestanden moeten worden gebruikt voor het loggen van informatie over documenten die uit het systeem zijn gewist. Deze logbestanden worden gemeenschappelijk gebruikt door alle sessies waarin documenten kunnen worden gewist.

## **Rapporten over gewiste documenten bekijken**

U kunt gedetailleerde informatie bekijken over documenten die uit een enterprise search-systeem zijn gewist. Deze informatie is alleen beschikbaar als documenttracering voor de collectie is ingeschakeld.

#### **Voordat u begint**

Voordat u een rapport over gewiste documenten opvraagt of een rapport naar een e-mailadres verzendt, moet u zorgen dat de sessies actief zijn waarvan u informatie wilt ontvangen. Als u bijvoorbeeld meer informatie wilt over documenten die zijn gewist tijdens het analyseren of indexeren, zorgt u dat de parser- en indexsessies voor de collectie zijn gestart.

Voordat u een rapport kunt ontvangen, controleert u of de informatie over uw mailserver is geconfigureerd voor enterprise search. Deze informatie geeft u op tijdens het configureren van de e-mailopties op de pagina Loggen van de view Systeem.


### Over deze taak


Het verzamelen van informatie over gewiste documenten kan veel tijd in beslag nemen. U kunt ervoor kiezen de informatie te bekijken en te wachten totdat de informatie wordt afgebeeld. U bespaart tijd door het rapport naar een e-mailadres van uw keuze te verzenden.


Als een bepaald document is gewist, vermeldt het rapport de datum en tijd waarop dat is gebeurd, het severityniveau van de fout, de component en de sessie waarin het probleem zich voordeed en het foutbericht.

### Procedure

Ga als volgt te werk om gegevens over gewiste documenten te bekijken:

1. Zoek in de view Collecties de collectie die u wilt bewaken en klik op  **Monitor**.

**Tip:** Als u een collectie bewerkt en de pagina Algemeen is geopend, kunt u op  **Monitor** klikken om de view voor het bewaken van de collectie weer te geven.

2. Klik op  **Gegevens URI**.
3. Voer op de pagina Gegevens URI de URI in waarvoor u de gegevens wilt bekijken.
4. Schakel de aankruisvakjes in voor het type gegevens dat u wilt bekijken:

#### Door de parser gewiste documenten

Schakel dit aankruisvakje in als u wilt zien of het document is gewist tijdens het analyseren, en zo ja, wat de reden daarvan was.

#### Uit de index gewiste documenten

Schakel dit aankruisvakje in als u wilt zien of het document is gewist tijdens het indexeren of analyseren, en zo ja, wat de reden daarvan was.

5. Geef aan hoe u het rapport wilt bekijken:
  - Als u wilt wachten totdat het rapport wordt weergegeven, klikt u op **Rapport bekijken**.
  - Als u een rapport naar een e-mailadres wilt verzenden zodat u het rapport later kunt bekijken, klikt u op **Rapport verzenden**.

Voer in het veld **E-mailadres om te waarschuwen** op de pagina Gedetailleerd URI-rapport verzenden het e-mailadres in waarnaar het rapport moet worden verzonden en klik op **Rapport verzenden**.

### Verwante taken

“Gegevens over een URI bekijken” op pagina 301

U kunt gedetailleerde gegevens over een URI bekijken. U ziet actuele en historische informatie over de manier waarop het document dat door deze URI wordt vertegenwoordigd, is gecrawld, geïndexeerd en doorzocht.

#### Verwante verwijzing

“URI-indelingen in een enterprise search-index” op pagina 111

Met de URI (Uniform Resource Identifier) van documenten in een enterprise search-index wordt het type crawler aangegeven waarmee het document aan de collectie is toegevoegd.

## Logbestanden over gewiste documenten bekijken

U kunt de gelogde berichten bekijken over documenten die uit een enterprise search-systeem zijn gewist. Deze informatie is alleen beschikbaar als document-tracering voor de collectie is ingeschakeld.


### Over deze taak


Om een rapport te bekijken over een gewist document, moet u de URI van het document weten. In de logbestanden van het gewiste document vindt u de datum en tijd waarop het document is gewist, het severityniveau van de fout, de component en de sessie waarin het probleem zich voordeed en het foutbericht.

### Procedure

Ga als volgt te werk om logbestanden voor gewiste documenten te bekijken:

1. Zoek in de view Collecties de collectie die u wilt bewaken en klik op  **Monitor**.

**Tip:** Als u een collectie bewerkt en de pagina Algemeen is geopend, kunt u op  **Monitor** klikken om de view voor het bewaken van de collectie weer te geven.

2. Klik op  **Gewiste documentlogbestanden**.
3. Selecteer op de pagina Gewiste documentlogbestanden het logbestand dat u wilt bekijken. De naam van elk logbestand geeft aan of het document is gewist door de parser (pd) of tijdens het indexeren (in); tevens geeft de naam aan wanneer het bestand is gemaakt. Als er op dezelfde datum meerdere logbestanden van hetzelfde type worden gemaakt, geeft een volgnummer de volgorde aan waarin de bestanden op die datum zijn gemaakt. Bijvoorbeeld:

```
dropped_doc_in_20060525.log  
dropped_doc_pd_20060524.log (bevat de meest recente vermeldingen op deze datum)  
dropped_doc_pd_20060524.log.1  
dropped_doc_pd_20060524.log.2 (bevat de oudste vermeldingen op deze datum)
```

4. Klik op **Logboek bekijken**.

Voor elk bericht op de pagina Inhoud logbestand vindt u de datum en tijd waarop het bericht is gegenereerd, het severityniveau van het bericht, de naam van de sessie waarin het bericht is gegenereerd, het ID van het bericht en de tekst van de fout.

U kunt op de knoppen klikken om naar de eerste pagina, de laatste pagina, de vorige pagina of de volgende pagina van het logbestand te gaan. U kunt ook een paginanummer opgeven om rechtstreeks naar die pagina te gaan.

5. Voor meer informatie over een bericht klikt u op  **Details**.

Op de pagina Gegevens logbericht ziet u de hostnaam van de enterprise search-server waarvan het bericht afkomstig is, de naam van het bestand



waarin de fout is opgetreden, de functienaam en het regelnummer waar de fout is opgetreden, het proces-ID en het thread-ID.

Klik op de knoppen om naar de volgende en vorige berichten in het logbestand te gaan.



---

## Logbestanden en alerts van enterprise search

U kunt de typen berichten kiezen die u voor een bepaalde collectie en voor het systeem wilt loggen, opties opgeven voor het maken en bekijken van logbestanden, alertsignalen ontvangen en e-mails over berichten ontvangen.

Tijdens de normale werking schrijven de enterprise search-componenten logberichten naar een gemeenschappelijk logbestand. Dit logbestand is opgeslagen in de directory `ES_NODE_ROOT/logs` op de indexserver. Met behulp van de beheerconsole kunt u deze gemeenschappelijke loggegevens bekijken.

Als er een probleem optreedt, bijvoorbeeld in de netwerkcommunicatie, schrijven de componenten de logberichten naar de directory logs op de server waarop de component is geïnstalleerd. U kunt deze lokale logbestanden bekijken met een bestandsviewer op die computer, bijvoorbeeld het hulpprogramma `tail` op UNIX-systemen. Deze typen logbestanden kunt u niet bekijken met de beheerconsole.

Tijdens het configureren van logbestanden kunt u de typen berichten kiezen die u wilt loggen (bijvoorbeeld foutberichten of waarschuwingsberichten), aangeven hoe vaak verouderde logbestanden moeten worden gewist om ruimte te maken voor nieuwe logbestanden, de maximumgrootte aangeven voor de logbestanden en de taal voor de berichten selecteren. Daarnaast kunt u opties opgeven voor het ontvangen van e-mail wanneer bepaalde events optreden of wanneer bepaalde berichten of berichttypen worden gelogd.

Bij het bewaken van logbestanden kunt u kiezen welk logbestand u wilt openen. U kunt de inhoud van het logbestand filteren, zodat alleen de berichten met een bepaald severityniveau (bijvoorbeeld alleen foutberichten) of berichten die door een bepaalde enterprise search-sessie zijn gegenereerd, worden weergegeven. Tijdens het bekijken van een logbestand kunt u de details van afzonderlijke berichten weergeven. U kunt bijvoorbeeld de naam bekijken van de functie op basis waarvan het bericht is gegenereerd en andere gegevens die u helpen bij het uitvoeren van de herstelprocedure (indien nodig).

### Verwante onderwerpen

Berichten voor enterprise search

---

## Alerts

U kunt enterprise search configureren zodat berichten naar het logbestand worden geschreven als bepaalde events hebben plaatsgevonden.

Berichten die worden getriggerd door events (ook wel alertsignalen genoemd), informeren u over situaties waaraan u mogelijk iets wilt doen, bijvoorbeeld wanneer er te weinig beschikbare ruimte is voor een resource. Bij het configureren van alertsignalen voor enterprise search geeft u de situaties op die door het systeem moeten worden bewaakt. Als de situatie zich voordoet, schrijft het systeem automatisch een bericht naar het logbestand.

Als u direct op de hoogte wilt worden gesteld van een situatie, kunt u opties opgeven voor het ontvangen van een e-mail als een van de bewaakte berichten in het logbestand wordt vastgelegd.

U kunt alertsignalen configureren voor events op collectieniveau en voor events die op systeemniveau optreden. Op collectieniveau kan het systeem:

- Het aantal documenten bewaken dat door elke crawler wordt gecrawld en een alertbericht versturen als het maximum aantal toegestane documenten bijna is bereikt.
- Het aantal documenten bewaken dat aan de index kan worden toegevoegd voor uw collecties en een alertbericht versturen als het maximum aantal documenten bijna is bereikt.
- U informeren wanneer de tijd die nodig is om op zoekopdrachten te reageren, de opgegeven limiet overschrijdt.

Op systeemniveau kan het systeem de schijfruimte voor elke enterprise search-server bewaken en een alertbericht versturen als er nog weinig beschikbare ruimte is.

## Alertsignalen op collectieniveau configureren

Door het configureren van alertsignalen zorgt u dat berichten naar het logbestand worden geschreven wanneer bepaalde events op collectieniveau optreden. Daarnaast kunt u e-mail ontvangen wanneer berichten over deze events worden gelogd.

### Voordat u begint

Om alertsignalen voor een collectie te configureren, moet u de rol van enterprise search-beheerder hebben of beheerder van de collectie zijn.

### Procedure

Ga als volgt te werk om alertsignalen op collectieniveau te configureren:

1. Open een collectie, selecteer de pagina Loggen en klik op **Waarschuwingen configureren**.
2. Als u het aantal documenten wilt bewaken dat door elke crawler wordt gecrawld, gaat u als volgt te werk:
  - a. Schakel het aankruisvakje **Als het aantal documenten dat door enige crawler is gecrawld een percentage van het toegestane maximum bereikt in**.
  - b. Geef in het veld **Percentage** een percentage op voor het loggen van berichten. Geef deze waarde op als percentage van het maximum aantal documenten dat de crawlers kunnen crawlen (de waarde voor **Maximum aantal te crawlen documenten** geeft u tijdens het configureren van de crawler-eigenschappen op). De standaardwaarde 90 procent.

Omdat u voor verschillende crawlers verschillende maxima kunt opgeven, worden er voor elke crawler andere berichten gelogd. Als u bijvoorbeeld als drempelwaarde voor de alert de standaardwaarde kiest en de ene DB2-crawler toestaat om 2 miljoen documenten te crawlen terwijl u de andere Notes-crawler toestaat om 1 miljoen documenten te crawlen, wordt er een bericht gelogd wanneer de DB2-crawler 1,8 miljoen documenten heeft gecrawld en wordt er een ander bericht gelogd wanneer de Notes-crawler 900.000 documenten heeft gecrawld.
3. Als u het aantal documenten wilt bewaken dat aan de index kan worden toegevoegd, gaat u als volgt te werk:
  - a. Schakel het aankruisvakje **Als het aantal documenten in de collectie een bepaald percentage van de geschatte grootte wordt in**.

- b. Geef in het veld **Percentage** een percentage op voor het loggen van berichten. Geef deze waarde op als een percentage van het aantal documenten dat volgens u in de collectie wordt opgenomen. De standaardwaarde is 85 procent.

In het veld **Limiet** ziet u de actuele geschatte grootte van de collectie. Om deze waarde te wijzigen, opent u de pagina Algemeen van de collectie, selecteert u de optie voor het configureren van algemene opties en typt u een nieuwe waarde in het veld **Geschatte aantal documenten**.

**Waarschuwing:** Deze limiet, en het geschatte aantal documenten dat u voor een collectie configureert, worden alleen gebruikt voor het bewaken van de groei van de collectie. Deze opties stellen geen absoluut limiet aan de groei van de index.

4. Als u wilt dat het systeem u laat weten wanneer de responstijd van zoekopdrachten de limiet overschrijdt, gaat u als volgt te werk:
  - a. Schakel het aankruisvakje **Als de responstijd van zoekopdrachten een limiet overschrijdt** in.
  - b. Voer in het veld **Limiet** het aantal seconden in dat u als maximale responstijd wilt instellen.

Wanneer deze limiet wordt overschreden, schrijft het systeem een logbericht. Als u de standaardwaarde behoudt, maakt het systeem een logbericht wanneer een zoekserver gemiddeld vijf seconden of langer nodig heeft om op zoekopdrachten te reageren.

Gangbare responstijden zijn minder dan een halve seconde. Gemiddelden groter dan 1 seconden kunnen betekenen dat uw besturingssysteem moet worden afgesteld of dat er een probleem in de configuratie-instelling van de zoekserver bestaat. U kunt bijvoorbeeld meer ruimte aan de zoekcache toewijzen.

5. Klik op **OK**.

Als u een e-mail wilt ontvangen wanneer het systeem berichten over deze events logt, opent u de pagina Loggen. Klik vervolgens op **E-mailopties voor berichten configureren** zodat u een e-mailadres kunt opgeven. De bericht-ID's voor de alertsignalen die u hebt ingeschakeld, worden automatisch toegevoegd aan de lijst van bericht-ID's waarvoor de e-mail wordt verzonden.

Voordat u e-mails kunt ontvangen, moet u zorgen dat de gegevens over uw mailserver zijn geconfigureerd. Hiervoor moet de enterprise search-beheerder **Systeem** selecteren op de werkbalk, de pagina Loggen openen en vervolgens op **E-mailopties voor berichten configureren** klikken.

#### **Verwante taken**

"E-mail over gelogde berichten ontvangen" op pagina 328

U kunt opties opgeven voor de ontvangst van e-mail wanneer bepaalde berichten of bepaalde typen berichten worden gelogd.

## **Alerts op systeemniveau configureren**


Door het configureren van alertsignalen zorgt u dat berichten naar het logbestand worden geschreven wanneer bepaalde events op systeemniveau optreden. Daarnaast kunt u e-mail ontvangen wanneer berichten over deze events worden gelogd.

#### **Voordat u begint**

Om alertsignalen op systeemniveau te configureren, moet u beheerder van enterprise search zijn.

## Procedure

Ga als volgt te werk om alertsignalen op systeemniveau te configureren:

1. Klik op **Systeem** om de view Systeem te openen.
2. Klik op  **Wijzigen** om naar de view Systeem wijzigen te gaan.
3. Klik op de pagina Loggen op **Waarschuwingen configureren**.
4. Als het systeem de hoeveelheid beschikbare ruimte op de enterprise search-servers moet bewaken, schakelt u het aankruisvakje **Als de hoeveelheid beschikbare ruimte voor het bestandssysteem een bepaald percentage van de totale ruimte wordt in**.
5. Geef in het veld **Percentage** op bij welke waarde u een bericht wilt ontvangen als de hoeveelheid beschikbare ruimte op een server gering is. Het getal dat u hier opgeeft, vertegenwoordigt het gewenste percentage van de totale ruimte op het bestandssysteem. De standaardwaarde is 80 procent.  
Als het enterprise search-systeem is geïnstalleerd op meerdere servers, wordt voor elke server een apart logbericht gegenereerd. U wordt bijvoorbeeld gewaarschuwd als op de crawlerserver weinig beschikbare ruimte is, en een andere bericht als er ruimtegebrek is op de index- en zoekservers.
6. Klik op **OK**.

Als u een e-mail wilt ontvangen wanneer het systeem berichten over deze event logt, opent u de pagina Loggen. Klik vervolgens op **E-mailopties voor berichten configureren** zodat u een e-mailadres en gegevens over uw mailserver kunt opgeven.

### Verwante taken

“E-mail over gelogde berichten ontvangen” op pagina 328

U kunt opties opgeven voor de ontvangst van e-mail wanneer bepaalde berichten of bepaalde typen berichten worden gelogd.

---

## Logbestanden configureren

U kunt de typen berichten opgeven die u wilt loggen en opties opgeven voor het maken van logbestanden.

### Voordat u begint

Om logbestanden op collectieniveau te configureren, moet u de rol van beheerder van enterprise search hebben of de beheerder van de collectie zijn. Om logbestanden op systeemniveau te configureren, moet u beheerder van enterprise search zijn.



### Over deze taak

Om te voorkomen dat logbestanden te veel schijfruimte innemen, worden de logbestanden automatisch overschreven en wordt er altijd een nieuw logbestand gestart wanneer de huidige datum verandert. Als een logbestand groeit tot de maximaal toegestane grootte en de datum niet is veranderd, maakt het systeem een nieuw logbestand. Wanneer het maximum aantal logbestanden is gemaakt, wordt het oudste logbestand gewist zodat er een nieuw bestand kan worden gemaakt.

Als u e-mail over gelogde bestanden wilt ontvangen, moet u eerst opgeven hoe de e-mail moet worden bezorgd. Vervolgens geeft u op voor welke berichten u e-mails wilt ontvangen.

## Procedure

Ga als volgt te werk om logbestanden voor enterprise search te configureren:

1. Als u opties wilt configureren voor het maken en overschrijven van logbestanden op systeemniveau, gaat u als volgt te werk:
  - a. Klik op **Systeem** om de view Systeem te openen.
  - b. Klik op  **Wijzigen** om naar de view Systeem wijzigen te gaan.
  - c. Klik op de pagina Loggen op **Opties voor logbestanden configureren**. De pagina Opties voor logbestanden op systeemniveau wordt afgebeeld.
2. Als u opties wilt configureren voor het maken en overschrijven van logbestanden op collectieniveau, gaat u als volgt te werk:
  - a. Zoek in de view Collecties de collectie waarvoor u opties wilt opgeven en klik op  **Bewerken**.
  - b. Klik op de pagina Loggen op **Opties voor logbestanden configureren**. De pagina Opties voor logbestanden op collectieniveau wordt afgebeeld.
3. Selecteer in het veld **Type te loggen informatie** de typen berichten die u wilt loggen:

### Alleen foutberichten

Foutberichten geven aan dat er een ongewenste situatie is opgetreden en dat het proces niet kan worden voortgezet. U moet actie ondernemen om het probleem op te lossen.

### Fout- en waarschuwingsberichten

Waarschuwingsberichten geven aan dat er mogelijk een conflict of inconsistentie is ontstaan, maar ze zorgen er niet voor dat het proces stopt. Dit is de standaardoptie.

### Alle berichten

Informatieberichten bieden algemene informatie over het systeem of de huidige taak en roepen u niet op tot actie.

4. Voer in het veld **Maximumgrootte van elk logbestanden** het maximum aantal megabytes voor elk logbestand. De standaardwaarde is 5 MB.

Als het logbestand deze grootte bereikt, wordt er een nieuw logbestand gemaakt tot het maximum aantal toegestane logbestanden is bereikt. Als u de logbestanden relatief klein houdt, kunt u ze gemakkelijker bekijken.
5. Voer in het veld **Maximum aantal logbestanden** het maximum aantal logbestanden in dat u wilt maken. De standaardwaarde is 10.

Als u wilt dat er ook oudere logbestanden beschikbaar zijn, verhoogt u deze waarde. Als u vooral geïnteresseerd bent in recente berichten en geen historisch overzicht van de activiteiten nodig hebt, verlaagt u deze waarde.
6. Selecteer in het veld **Standaardlocale** de taal die u voor het loggen van berichten wilt gebruiken. De standaardwaarde is Engels.
7. Klik op **OK**.

---

## SMTP-servergegevens configureren

Voordat u e-mails kunt ontvangen van enterprise search-activiteiten, moet u gegevens configureren over de SMTP-server (Simple Mail Transfer Protocol).

### Voordat u begint

Om gegevens over uw SMTP-server te configureren, moet u de rol van beheerder van enterprise search hebben.


## Over deze taak

Met behulp van verschillende beheerfuncties in enterprise search kunt u e-mail ontvangen. Voordat u e-mail kunt ontvangen van deze functies, moet u gegevens over uw SMTP-server opgeven:

- Als u alertsignalen op collectieniveau of alertsignalen op systeemniveau configureert, kunt u e-mail ontvangen wanneer deze berichten worden gelogd. Daarnaast kunt u e-mail ontvangen als andere berichten worden gelogd, niet alleen berichten die op basis van bewaakte events worden getriggerd.
- Als u gedetailleerde informatie wilt bekijken over een URI in de index of over een document dat uit het enterprise search-systeem is verwijderd, kunt u het rapport per e-mail ontvangen.
- Als u een webcrawler bewaakt en opgeeft dat u historierapporten voor de webcrawler wilt maken, kunt u per e-mail op de hoogte worden gesteld zodra een rapport is gemaakt.

## Procedure

Ga als volgt te werk om gegevens over uw SMTP-server te configureren:

1. Klik op **Systeem** om de view Systeem te openen.
2. Klik op  **Wijzigen** om naar de view Systeem wijzigen te gaan.
3. Klik op de pagina Loggen op **E-mailopties voor berichten configureren**.
4. Voer in het veld **SMTP-mailserver voor het afleveren van e-mail** op de pagina E-mailopties voor systeemberichten de volledige hostnaam of het volledige IP-adres in van de SMTP-server die u wilt gebruiken.  
Deze server wordt gebruikt voor het versturen van e-mail naar de adressen die u opgeeft.
5. Geef in het veld **Frequentie om te checken of er e-mail is** op hoe vaak er op berichten moet worden gecontroleerd en e-mails over de berichten moeten worden verstuurd.  
Het systeem voegt alle berichten voor een specifiek e-mailadres samen in één bericht en verzendt dat bericht met de door u opgegeven frequentie.
6. Klik op **OK**.

---

## E-mail over gelogde berichten ontvangen

U kunt opties opgeven voor de ontvangst van e-mail wanneer bepaalde berichten of bepaalde typen berichten worden gelogd.

### Voordat u begint

Om e-mailopties voor berichten op systeemniveau te configureren, moet u de rol van enterprise search-beheerder hebben. Om e-mailopties te configureren voor berichten op collectieniveau, moet u de rol van beheerder van enterprise search hebben of beheerder van de collectie zijn.

Voordat u e-mails kunt ontvangen, moet u gegevens configureren over de SMTP-server (Simple Mail Transfer Protocol) zodat de e-mails kunnen worden bezorgd.

### Over deze taak



Bij het configureren van alertsignalen kunt u een optie kiezen waarmee berichten worden gelogd wanneer bepaalde events optreden. Als u deze opties inschakelt,



kunt u vervolgens opties opgeven voor het automatisch ontvangen van e-mail wanneer er berichten van dit type worden gelogd. Daarnaast kunt u opties opgeven voor het ontvangen van e-mail wanneer andere berichten worden gelogd, niet alleen berichten die op basis van events worden getriggerd.

## Procedure

Ga als volgt te werk om e-mailopties voor berichten te configureren:

1. Als u e-mail wilt ontvangen over systeemberichten:
  - a. Klik op **Systeem** om de view Systeem te openen.
  - b. Klik op  **Wijzigen** om naar de view Systeem wijzigen te gaan.
  - c. Klik op de pagina Loggen op **E-mailopties voor berichten configureren**.
  - d. Schakel op de pagina E-mailopties voor systeemberichten het aankruisvakje **E-mail versturen over berichten op systeemniveau** in.
  - e. Voer in het veld **E-mailadres voor het ontvangen van e-mail** een of meer e-mailadressen in. Het is gebruikelijk dat de enterprise search-beheerder informatie over systeemberichten ontvangt.  
U kunt de adressen scheiden met een komma. Bijvoorbeeld:  
steinbeck@us.ibm.com, yeats@ireland.ibm.com, dante@it.ibm.com.
  - f. Als u een e-mail wilt ontvangen over alle foutberichten die worden gelogd, schakelt u het aankruisvakje **E-mail versturen over alle foutberichten** in.
  - g. Als u alleen e-mail wilt ontvangen wanneer er bepaalde berichten op systeemniveau worden gelogd, voert u de bericht-ID's voor deze berichten in het gebied **E-mail verzenden over berichten van een bepaald type** in. Typ één bericht-ID per regel. Bijvoorbeeld:  
FFQC4819E  
FFQ00005E  
Standaard worden verschillende bericht-ID's afgebeeld (klik op **Help** voor een beschrijving van deze berichten).
  - h. Klik op **OK**.
2. Als u e-mail wilt ontvangen over berichten voor een collectie:
  - a. Klik op **Collecties** om de view Collecties te openen.
  - b. Zoek in de lijst van collecties de collectie die u wilt configureren en klik op  **Bewerken**.
  - c. Klik op de pagina Loggen op **E-mailopties voor berichten configureren**.
  - d. Schakel op de pagina E-mailopties voor collectieberichten het aankruisvakje **E-mail versturen over berichten op collectieniveau** in.
  - e. Voer in het veld **E-mailadres voor het ontvangen van e-mail** een of meer e-mailadressen in. Het is gebruikelijk dat de collectiebeheerder informatie over berichten op collectieniveau ontvangt.  
U kunt de adressen scheiden met een komma. Bijvoorbeeld:  
steinbeck@us.ibm.com, yeats@ireland.ibm.com, dante@it.ibm.com.
  - f. Als u een e-mail wilt ontvangen over alle foutberichten die worden gelogd, schakelt u het aankruisvakje **E-mail versturen over alle foutberichten** in.
  - g. Als u alleen e-mail wilt ontvangen wanneer er bepaalde berichten op collectieniveau worden gelogd, voert u de bericht-ID's voor deze berichten in het gebied **E-mail verzenden over berichten van een bepaald type** in. Typ één bericht-ID per regel. Bijvoorbeeld:  
FFQC4819E  
FFQ00005E

Standaard worden verschillende bericht-ID's afgebeeld (klik op **Help** voor een beschrijving van deze berichten).

h. Klik op **OK**.

#### Verwante onderwerpen

Berichten voor enterprise search

#### Verwante taken

“Alertsignalen op collectieniveau configureren” op pagina 324

Door het configureren van alertsignalen zorgt u dat berichten naar het logbestand worden geschreven wanneer bepaalde events op collectieniveau optreden. Daarnaast kunt u e-mail ontvangen wanneer berichten over deze events worden gelogd.

“Alerts op systeemniveau configureren” op pagina 325

Door het configureren van alertsignalen zorgt u dat berichten naar het logbestand worden geschreven wanneer bepaalde events op systeemniveau optreden. Daarnaast kunt u e-mail ontvangen wanneer berichten over deze events worden gelogd.

---

## Logbestanden bekijken


U kunt de logbestanden bekijken die door de systeem- en collectiecomponenten naar een gemeenschappelijk logbestand worden geschreven. U kunt ook filters opgeven om berichten met een bepaald severityniveau en berichten van bepaalde zoeksessies te bekijken.

### Voordat u begint

Alle beheerders van enterprise search kunnen logbestanden bekijken voor de collecties waarvan ze beheerder zijn. Om logbestanden op systeemniveau te bekijken, moet u de rol van beheerder van enterprise search hebben of toegangsmachtiging voor de werkbalk **Systeem** hebben.

### Procedure

1. Ga als volgt te werk om de logbestanden voor één collectie te bekijken:
  - a. Klik op **Collecties** om de view Collecties te openen.
  - b. Zoek in de lijst van collecties de collectie die u wilt bekijken, klik op  **Monitor** en open de pagina Loggen.

**Tip:** Als u een collectie bewerkt en de pagina Loggen al is geopend, kunt u op  **Monitor** klikken om de view voor het bewaken van de collectie weer te geven.

2. Ga als volgt te werk om logbestanden op systeemniveau te bekijken:
  - a. Klik op **Systeem** om de view Systeem te openen.
  - b. Selecteer de pagina Loggen.
3. Selecteer in het veld **Logbestand** het logbestand dat u wilt bekijken.

De naam van elk logbestand bevat het type logbestand (zoals de systeem- of collectienaam) en de datum waarop het bestand is gemaakt. Als er op dezelfde datum meerdere logbestanden van hetzelfde type zijn gemaakt, geeft een volgnummer de volgorde aan waarin de bestanden zijn gemaakt. Bijvoorbeeld:


```
log_file_type_20060526.log (bevat de meest recente vermeldingen op deze datum)
log_file_type_20060526.log.1
log_file_type_20060526.log.2 (bevat de oudste vermeldingen op deze datum)
log_file_type_20060525.log (bevat de meest recente vermeldingen op deze datum)
```

*log\_file\_type\_20060525.log.1*  
*log\_file\_type\_20060525.log.2*  
*log\_file\_type\_20060525.log.3 (bevat de oudste vermeldingen op deze datum)*

4. Als u alleen de berichten met een bepaald severityniveau wilt bekijken, schakelt u het desbetreffende aankruisvakje in het veld **Severity** in.
5. Als u alleen berichten uit bepaalde sessies wilt bekijken, schakelt u het desbetreffende aankruisvakje in het veld **Sessie** in.
6. Klik op **Logboek bekijken**.

Voor elk bericht op de pagina Inhoud logbestand vindt u de datum en tijd waarop het bericht is gegenereerd, het severityniveau van het bericht, de naam van de sessie waarin het bericht is gegenereerd, het ID van het bericht en de tekst van de fout.

U kunt op de knoppen klikken om naar de eerste pagina, de laatste pagina, de vorige pagina of de volgende pagina van het logbestand te gaan. U kunt ook een paginanummer opgeven om rechtstreeks naar die pagina te gaan.

7. Voor meer informatie over een bericht klikt u op  **Details**.  
Op de pagina Gegevens logbericht ziet u de hostnaam van de enterprise search-server waarvan het bericht afkomstig is, de naam van het bestand waarin de fout is opgetreden, de functienaam en het regelnummer waar de fout is opgetreden, het proces-ID en het thread-ID.

Klik op de knoppen om naar de volgende en vorige berichten in het logbestand te gaan.



---

## Backup en herstel van een enterprise search-systeem

Met scripts voor backup en herstel kunt u een backup van het enterprise search-systeem maken en het systeem later weer herstellen.

### Gegevens waarvan backups worden gemaakt met scripts

De scripts hebben betrekking op backup en herstel van de volgende bestanden:

- Configuratiebestanden in de directory `ES_NODE_ROOT/master_config`
- Databasebestanden voor de crawlers, waaronder alle crawlermetagegevens zoals de datum waarop de gegevensbronnen voor het laatst zijn gecrawld
- Alle bestanden in de directory `ES_NODE_ROOT/data`
- Indexbestanden voor collecties die zijn geconfigureerd met niet-standaard gegevensdirectory's

### Backupdirectorystructuur

Het backupscript maakt de volgende subdirectory's onder de directory die u opgeeft op het moment dan u het script uitvoert. Het beheerders-ID van enterprise search moet gemachtigd zijn om te schrijven naar de directory die u opgeeft.

#### **master\_config**

Bevat de configuratiebestanden uit de directory `ES_NODE_ROOT/master_config`

#### **database**

Bevat de databasebestanden van de crawlerserver

#### **data**

Bevat de indexbestanden van de indexserver

### Richtlijnen voor de syntaxis

- U kunt een backup maken van de gegevens op de ene computer en de gegevens terugzetten naar een andere computer. Hiervoor gelden echter de volgende beperkingen:
  - Bestanden waarvan een backup is gemaakt met een bepaalde versie van WebSphere Information Integrator OmniFind Edition, kunnen niet worden teruggezet naar een systeem waarop een andere versie van WebSphere Information Integrator OmniFind Edition wordt uitgevoerd.
  - U moet de gegevens terugzetten naar een systeem met een gelijk of groter aantal enterprise search-servers. Als u bijvoorbeeld een backup maakt van een enterprise search-systeem dat op één server wordt uitgevoerd, kunt u de gegevens terugzetten naar een systeem waarop twee of vier enterprise search-servers worden gebruikt. Als u een backup hebt gemaakt van gegevens op een systeem met vier servers, kunt u de gegevens niet terugzetten naar een systeem met één of twee servers.
- U moet de hoofdindex opbouwen voordat u het backupproces start, zodat een backup wordt gemaakt van de laatst geïndexeerde gegevens.
- Alle instellingen voor de installatiedirectory (`ES_INSTALL_ROOT`), de gegevensdirectory (`ES_NODE_ROOT`) en het ID en wachtwoord voor de enterprise search-beheerder moeten gelijk zijn op het systeem waarop de backup wordt gemaakt en het systeem waarnaar de gegevens worden teruggezet.

- In een configuratie met meerdere servers moet u een backup maken van het systeem en de backup terugzetten vanaf de enterprise search-indexserver. Omdat alle crawlergegevens zijn opgeslagen in databases op de crawlerserver, worden de scriptopdrachten op afstand uitgevoerd om een backup te maken van de crawlergegevens en deze terug te zetten.
- Er moet uiteraard voldoende schijfruimte beschikbaar zijn om een backup van het enterprise search-systeem in een andere directory te maken. De backup- en herstelscripts voeren geen controle uit op de bestanden.
- Wanneer de backup- en herstelscripts actief zijn, worden alle systeemsessies gestopt. Om te voorkomen dat er onjuiste of inconsistente systeemgegevens worden afgebeeld, kunt u het beste ook de beheerconsole van enterprise niet gebruiken zolang deze scripts worden uitgevoerd.
- Als het systeem vastloopt nadat er een onherstelbare fout is opgetreden, moet u WebSphere II OmniFind Edition opnieuw installeren en vervolgens het herstelscript uitvoeren.

#### Verwante verwijzing

“Opdrachten, retourcodes en sessie-ID’s in enterprise search” op pagina 337  
 U kunt opdrachten gebruiken om een diagnose te maken van problemen, de status vast te stellen van de verschillende delen van het systeem, sessies te starten en te stoppen of het systeem te starten en te stoppen.

---

## Een backup van het enterprise search-systeem maken

U kunt een backup van het enterprise search-systeem maken met behulp van het script `esbackup.sh` (voor AIX, Linux of Solaris) of met het script `esbackup.bat` (voor Microsoft Windows).

### Beperkingen

Het beheerders-ID van enterprise search moet gemachtigd zijn om te schrijven naar de directory die u opgeeft wanneer u het backupscript uitvoert.

Wanneer de backup- en herstelscripts actief zijn, worden alle systeemsessies gestopt. Om te voorkomen dat er onjuiste of inconsistente systeemgegevens worden afgebeeld, kunt u het beste ook de beheerconsole van enterprise niet gebruiken zolang deze scripts worden uitgevoerd.

**Waarschuwing:** Als u op Ctrl+C drukt om het backupscript te onderbreken, wordt het systeem instabiel. U moet de volgende opdracht opgeven om alle servicesessies en de actieve sessies die tijdens het backupproces zijn gestopt, te starten:

```
esadmin system startall
```

### Procedure

U maakt als volgt een backup van het enterprise search-systeem:

1. Meld u bij de indexserver aan als enterprise search-beheerder. Dit gebruikers-ID is opgegeven bij de installatie van WebSphere II OmniFind Edition.
2. Als de CCL (Common Communication Layer) voor enterprise search niet actief is, start u deze:

**AIX, Linux of Solaris**  
`startccl.sh -bg`

### Windows-opdrachtaanwijzing

```
startccl
```

### Systeembeheer van Windows-services

Ga als volgt te werk om CCL op de achtergrond te starten:

- a. Start Windows-services: **Start** → **Programma's** → **Systeembeheer** → **Services**.
  - b. Klik met de rechtermuisknop op **WebSphere Information Integrator OmniFind Edition** en klik op **Start**.
3. Voor een configuratie met meerdere servers zorgt u dat de CCL op elke server is gestart. Herhaal indien nodig de vorige stappen.
  4. Start het backupproces, waarbij *backup\_directory* de directory is waar de backup van de gegevens is opgeslagen:

### AIX, Linux of Solaris

```
esbackup.sh backup_directory
```

### Windows-opdrachtaanwijzing

```
esbackup.bat backup_directory
```

### Verwante verwijzing

“Opdrachten, retourcodes en sessie-ID's in enterprise search” op pagina 337

U kunt opdrachten gebruiken om een diagnose te maken van problemen, de status vast te stellen van de verschillende delen van het systeem, sessies te starten en te stoppen of het systeem te starten en te stoppen.

---

## Het enterprise search-systeem herstellen

Als u WebSphere Information Integrator OmniFind Edition opnieuw hebt geïnstalleerd, kunt u het script *esrestore.sh* (voor AIX, Linux of Solaris) of het script *esrestore.bat* (voor Microsoft Windows) gebruiken om een enterprise search-systeem te herstellen.

### Beperkingen

Wanneer de backup- en herstelscripts actief zijn, worden alle systeemsessies gestopt. Om te voorkomen dat er onjuiste of inconsistente systeemgegevens worden afgebeeld, kunt u het beste ook de beheerconsole van enterprise niet gebruiken zolang deze scripts worden uitgevoerd.

Bestanden waarvan een backup is gemaakt met een bepaalde versie van WebSphere II OmniFind Edition, kunnen niet worden teruggezet naar een systeem waarop een andere versie van WebSphere II OmniFind Edition wordt uitgevoerd. Daarnaast moeten op het systeem waarnaar u de gegevens terugzet, hetzelfde aantal of meer enterprise search-servers zijn geïnstalleerd dan het systeem waarop de backup is uitgevoerd.

### Procedure

Ga als volgt te werk om het enterprise search-systeem te herstellen:

1. Meld u bij de indexserver aan als enterprise search-beheerder. Dit gebruikers-ID is opgegeven bij de installatie van WebSphere II OmniFind Edition.
2. Als de CCL (Common Communication Layer) voor enterprise search niet actief is, start u deze:

### AIX, Linux of Solaris

```
startccl.sh -bg
```

### Windows-opdrachtaanwijzing

```
startccl
```

### Systeembeheer van Windows-services

Ga als volgt te werk om CCL op de achtergrond te starten:

- a. Start Windows-services: **Start** → **Programma's** → **Systeembeheer** → **Services**.
  - b. Klik met de rechtermuisknop op **WebSphere Information Integrator OmniFind Edition** en klik op **Start**.
3. Voor een configuratie met meerdere servers zorgt u dat de CCL op elke server is gestart. Herhaal indien nodig de vorige stappen.
  4. Ga als volgt te werk om de controller te stoppen:  

```
esadmin stop
```
  5. Zet de enterprise search-gegevens terug met de volgende opdracht, waarbij *backup\_directory* de directory is waar u de backup van de bestanden hebt opgeslagen:

### AIX, Linux of Solaris

```
esrestore.sh backup_directory
```

### Windows-opdrachtaanwijzing

```
esrestore.bat backup_directory
```

### Verwante verwijzing

“Opdrachten, retourcodes en sessie-ID's in enterprise search” op pagina 337  
U kunt opdrachten gebruiken om een diagnose te maken van problemen, de status vast te stellen van de verschillende delen van het systeem, sessies te starten en te stoppen of het systeem te starten en te stoppen.



---

## Opdrachten, retourcodes en sessie-ID's in enterprise search

U kunt opdrachten gebruiken om een diagnose te maken van problemen, de status vast te stellen van de verschillende delen van het systeem, sessies te starten en te stoppen of het systeem te starten en te stoppen.

In een installatie met meerdere servers kunt u de opdrachten uitvoeren vanaf elke server in het systeem. U wordt echter aangeraden de opdrachten uit te voeren vanaf de indexserver. De indexserver, of controllerserver, kan informatie verkrijgen van alle andere servers in het systeem.

De meeste opdrachten hebben de volgende indelingen:

```
esadmin opdracht_naam parameters
esadmin sessie_ID actie -optie
```

Voor meer informatie over opdrachten geeft u `esadmin help` op. Voor meer informatie over een bepaalde opdracht geeft u `esadmin actie help` op.

### esadmin-opdrachten in enterprise search

Geef op een regel de volgende opdrachten op.

Tabel 9. *esadmin-opdrachten in enterprise search*

Optie	Beschrijving
<code>esadmin system startall</code>	Hiermee worden de enterprise search-componenten op alle enterprise search-servers gestart, inclusief de webserver, het programma ESSearchServer en het Informatiecentrum op de zoekservers, de crawlersessies op de crawlerserver en de indexsessies op de indexserver. Met deze opdracht wordt de CCL (Common Communication Layer) alleen op de lokale server gestart. Om de CCL opnieuw te starten, moet u de CCL handmatig stoppen en opnieuw starten op elke enterprise search-server op afstand.  <b>Voorbeeldopdracht:</b> <code>esadmin system startall</code>
<code>esadmin system stopall</code>	Hiermee worden de enterprise search-componenten op alle enterprise search-servers gestopt, inclusief het Informatiecentrum, het programma ESSearchServer en de webserver op de zoekservers, de crawlersessies op de crawlerserver en de indexsessies op de indexserver. Met deze opdracht wordt alleen de CCL op de lokale server gestopt. Om de CCL opnieuw te starten, moet u de CCL handmatig stoppen en opnieuw starten op elke enterprise search-server op afstand.  <b>Voorbeeldopdracht:</b> <code>esadmin system stopall</code>
<code>esadmin system checkall</code>	Hiermee wordt de status van alle enterprise search-componenten op alle enterprise search-servers gecontroleerd.  <b>Voorbeeldopdracht:</b> <code>esadmin system checkall</code>

Tabel 9. esadmin-opdrachten in enterprise search (vervolg)

Optie	Beschrijving
esadmin <i>crawler_sessie_id</i> start	<p>Hiermee wordt een crawlersessie gestart. Met deze opdracht wordt geen crawlactiviteit gestart.</p> <p><b>Voorbeeldopdracht:</b> esadmin col1.WEB1.esadmin start</p> <p><b>Voorbeeldberichten en -retourcodes:</b> FFQC5310I WEBCrawler1 (sid: col1.WEB1.esadmin) is niet actief. FFQC5314I Resultaat: 0</p>
esadmin <i>crawler_sessie_id</i> startCrawl	<p>Hiermee wordt het crawlen gestart.</p> <p><b>Voorbeeldopdracht:</b> esadmin col3.DB21.esadmin startCrawl</p> <p><b>Voorbeeldberichten en -retourcodes:</b> FFQC5303I DB2Crawler1 (sid: col3.DB21.esadmin) is al actief. PID: 23650 FFQC5314I Resultaat: 0</p>
esadmin <i>crawler_sessie_id</i> pause	<p>Hiermee wordt het crawlen onderbroken.</p> <p><b>Voorbeeldopdracht:</b> esadmin col3.DB21.esadmin pause</p> <p><b>Voorbeeldberichten en -retourcodes:</b> FFQC5303I DB2Crawler1 (sid: col3.DB21.esadmin) is al actief. PID: 23650 FFQC5314I Resultaat: 0</p>
esadmin <i>crawler_sessie_id</i> resume	<p>Hiermee wordt het crawlen hervat.</p> <p><b>Voorbeeldopdracht:</b> esadmin col3.DB21.esadmin resume</p> <p><b>Voorbeeldberichten en -retourcodes:</b> FFQC5303I DB2Crawler1 (sid: col3.DB21.esadmin) is al actief. PID: 23650 FFQC5314I Resultaat: 0</p>
esadmin <i>crawler_sessie_id</i> stopCrawl	<p>Hiermee wordt het crawlen gestopt.</p> <p><b>Voorbeeldopdracht:</b> esadmin col3.DB21.esadmin stopCrawl</p> <p><b>Voorbeeldberichten en -retourcodes:</b> FFQC5303I DB2Crawler1 (sid: col3.DB21.esadmin) is al actief. PID: 23650 FFQC5314I Resultaat: 0</p>
esadmin <i>crawler_sessie_id</i> stop	<p>Hiermee wordt een crawlersessie gestopt.</p> <p><b>Voorbeeldopdracht:</b> esadmin col3.DB21.esadmin stop</p> <p><b>Voorbeeldberichten en -retourcodes:</b> FFQC5303I DB2Crawler1 (sid: col3.DB21.esadmin) is al actief. PID: 23650 FFQC5314I Resultaat: 0</p>

Tabel 9. esadmin-opdrachten in enterprise search (vervolg)

Optie	Beschrijving
esadmin <i>crawler_sessie_id</i> getCrawlerStatus	<p>Hiermee wordt de status van een crawler opgehaald. Welke gegevens worden teruggezonden, is afhankelijk van het feit of de crawler een webcrawler of een crawler voor alle overige gegevensbronnen is.</p> <p><b>Voorbeeld voor een webcrawler:</b>  esadmin col1.WEB1.esadmin getCrawlerStatus</p> <p><b>Mogelijke retourcodes en berichten voor een webcrawler:</b>  FFQC5303I WebCrawler1 (sid: col1.WEB1.esadmin)  is al actief. PID: 23650</p> <p><b>Voorbeeld voor een niet-webcrawler:</b>  esadmin col3.DB21.esadmin getCrawlerStatus</p> <p><b>Mogelijke retourcodes en berichten voor een niet-webcrawler:</b>  FFQC5303I db2crawler (sid: db2col.DB2_96945)  is al actief. PID: 5936</p> <p>Voor meer informatie over statusberichten die worden weergegeven, raadpleegt u "Gedetailleerde informatie voor statusopdrachten" op pagina 342.</p>
esadmin <i>dscrawler_sessie_id</i> getCrawlSpaceStatus esadmin <i>web_crawler_sessie_id</i> getCrawlStatus -selections <i>waarde</i>	<p>Hiermee wordt de algemene crawlruimtestatus opgehaald voor alle niet-webcrawlers.</p> <p><b>Voorbeeldopdracht:</b>  esadmin col3.DB21.esadmin getCrawlSpaceStatus</p> <p><b>Voorbeeldberichten en -retourcodes:</b>  FFQC5303I DB2Crawler1 (sid: col3.DB21.esadmin)  is al actief. PID: 23650</p> <p>Hiermee wordt de algemene crawlruimtestatus voor de webcrawler opgehaald.</p> <p><b>Voorbeeldopdracht:</b>  esadmin col1.WEB1.esadmin getCrawlStatus</p> <p>Voor meer informatie over statusberichten die worden weergegeven, raadpleegt u "Gedetailleerde informatie voor statusopdrachten" op pagina 342.</p>

Tabel 9. esadmin-opdrachten in enterprise search (vervolg)

Optie	Beschrijving
<pre>esadmin dscrawler_sessie_id getCrawlSpaceStatusDetail -ts doel_server_id esadmin webcrawler_sessie_id getCrawlDetailsPerSite -url tekenreeks -selections num -threshold num</pre>	<p>Hiermee wordt de gedetailleerde crawlruimtestatus opgehaald voor alle niet-webcrawlers. Als u de optie voor de doelservers niet opgeeft, worden gegevens voor alle doelservers weergegeven. Als de DB2-crawler bijvoorbeeld de databases FOUNTAIN en SAMPLE crawl en u niet de optie voor de doelservers opgeeft, wordt de status van alle tabellen in de databases FOUNTAIN en SAMPLE weergegeven.</p> <p><b>Voorbeeldopdracht:</b></p> <pre>esadmin col3.DB21.esadmin getCrawlSpaceStatusDetail -ts FOUNTAIN</pre> <p><b>Voorbeeldberichten en -retourcodes:</b></p> <pre>FFQC5303I DB2Crawler1 (sid: col3.DB21.esadmin) is al actief. PID: 23650</pre>
<pre>esadmin monitor getCollectionParserMonitorStatus -cid collectie_ID</pre>	<p>Hiermee wordt de parserstatus opgehaald.</p> <p><b>Voorbeeldopdracht:</b></p> <pre>esadmin col1.WEB1.esadmin getCrawlDetailsPerSite</pre> <p>Voor meer informatie over statusberichten die worden weergegeven, raadpleegt u "Gedetailleerde informatie voor statusopdrachten" op pagina 342.</p> <hr/> <p>Hiermee wordt de parserstatus opgehaald.</p> <p><b>Voorbeeldopdracht:</b></p> <pre>esadmin monitor getCollectionParserMonitorStatus -cid coll</pre> <p><b>Voorbeeldberichten en -retourcodes:</b></p> <pre>FFQC5303I Monitor (node1) (sid: monitor) is al actief. PID: 12543</pre>
<pre>esadmin startMain -cid collectie_id</pre>	<p>Hiermee wordt het opbouwen van de hoofdindex gestart.</p> <p><b>Voorbeeldopdracht:</b></p> <pre>esadmin monitor getCollectionParserMonitorStatus -cid coll</pre> <p><b>Voorbeeldberichten en -retourcodes:</b></p> <pre>FFQC5303I Controller (node1) (sid: controller) is al actief. PID: 25917 FFQC5314I Resultaat: 1117671147056</pre>
<pre>esadmin startDelta -cid collectie_id</pre>	<p>Hiermee wordt het opbouwen van een delta-index gestart.</p> <p><b>Voorbeeldopdracht:</b></p> <pre>esadmin startMain startMain -cid coll</pre> <p><b>Voorbeeldberichten en -retourcodes:</b></p> <pre>FFQC5303I Controller (node1) (sid: controller) is al actief. PID: 4548 FFQC5314I Resultaat: 1117670603408</pre>

Tabel 9. esadmin-opdrachten in enterprise search (vervolg)

Optie	Beschrijving
esadmin monitor getCollectionIndexMonitorStatus -cid <i>collectie_id</i> -buildType [main   delta] -numrecords <i>laatsteNrecords</i>	Hiermee wordt de status van het opbouwproces voor een hoofd- of delta-index opgehaald. De optie numrecords toont de laatste <i>N</i> statusrecord voor het opbouwproces van de index. Als numrecords wordt weggelaten, wordt de status voor de laatste twintig opbouwprocessen voor de index weergegeven.  <b>Voorbeeldopdracht:</b> esadmin monitor getCollectionIndexMonitorStatus -cid coll -buildType main -numrecords 4  <b>Voorbeeldberichten en -retourcodes:</b> FFQC5303I Monitor (node1) (sid: monitor) is al actief. PID: 12649  Voor meer informatie over statusberichten die worden weergegeven, raadpleegt u "Gedetailleerde informatie voor statusopdrachten" op pagina 342.
esadmin startSearch -cid <i>collectie_id</i>	Hiermee worden de zoekserverprocessen gestart.  <b>Voorbeeldopdracht:</b> esadmin startSearch -cid coll  <b>Voorbeeldberichten en -retourcodes:</b> FFQC5303I Controller (node1) (sid: controller) is al actief. PID: 25917 FFQC5314I Resultaat: 0
esadmin stopSearch -cid <i>collectie_id</i>	Hiermee worden de zoekserverprocessen gestopt.  <b>Voorbeeldopdracht:</b> esadmin stopSearch -cid coll  <b>Voorbeeldberichten en -retourcodes:</b> FFQC5303I Controller (node1) (sid: controller) is al actief. PID: 15292 FFQC5314I Resultaat: 0

Tabel 9. esadmin-opdrachten in enterprise search (vervolg)

Optie	Beschrijving
esadmin monitor getCollectionSearchMonitorStatus -cid <i>collectie_id</i> esadmin <i>zoekbeheerder_sessie_id</i> getStatus -cid <i>collectie_id</i>	Hiermee wordt de status van de zoekserver opgehaald.  <b>Voorbeeldopdracht:</b> esadmin monitor getCollectionSearchMonitorStatus -cid coll  <b>Voorbeeldberichten en -retourcodes:</b> FFQC5303I Monitor (node1) (sid: monitor) is al actief. PID: 12649  Hiermee worden gedetailleerde statusgegevens van een zoekindex voor een collectie op een bepaalde zoekserver weergegeven. Per zoekserver is één zoekbeheersersessie aanwezig. Elke zoekbeheersersessie is verantwoordelijk voor het bewaken en werken met de zoekindexen op een bepaalde zoekserver.  <b>Voorbeeldopdracht:</b> esadmin searchmanager.node1 getStatus -cid coll  <b>Voorbeeldberichten en -retourcodes:</b> FFQC5303I Search Manager (node1) (sid: searchmanager.node1) is al actief. PID: 15711 FFQC5314I Resultaat: PID=18390 CacheHits=3 QueryRate=1 Port=44008 SessionId=coll.runtime.node1 CacheHitRate=0.333 ResponseTime=70 Status=1 SessionName=coll.runtime.node1.1  Voor meer informatie over statusberichten die worden weergegeven, raadpleegt u "Gedetailleerde informatie voor statusopdrachten".

## Gedetailleerde informatie voor statusopdrachten

Met sommige opdrachten kan uitgebreide informatie als resultaat worden gegeven. In dit gedeelte wordt beschreven welke gegevens kunnen worden teruggezonden voor de crawlerstatus en de crawlruimtestatus. De tabel in het gedeelte "esadmin-opdrachten in enterprise search" op pagina 337 bevat gegevens die op basis van elke esadmin-opdracht kunnen worden teruggezonden. In dit gedeelte worden de gegevens beschreven die op basis van de volgende opdrachten worden teruggezonden:

- Webcrawlerstatus
- Status van niet-webcrawlers
- Status van crawlruimte voor de webcrawler
- Status van crawlruimte voor niet-webcrawlers
- Gedetailleerde status van crawlruimte voor de webcrawler
- Gedetailleerde status van crawlruimte voor niet-webcrawlers
- Parserstatus
- Opbouwstatus van index

- Status van zoekserver
- Gedetailleerde status van zoekserver

**Status van webcrawler:** als u de opdracht uitvoert om de status van de webcrawler op te halen, worden de gegevens in een XML-documentindeling teruggezonden. De volgende gegevens kunnen met de opdracht voor de status van de webcrawler worden teruggezonden:

```
FFQC5314I Resultaat: <?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/"
NumActURLs="355"
NumProcURLs="350"
TimeRem="5" Duration="1195"/>
. . .
</ActiveBucketList>
<CrawlRate Value="75"/>
<RecentlyCrawledURLList Count="40">
<RecentlyCrawledURL URL="http://w3.ibm.com/foo.html"/>
<RecentlyCrawledURL URL="http://w3.ibm.com/foo.html"/>
<NumURLsThisSession Value="160000"/>
</CrawlerStatus>
```

In de volgende tabel worden de verschillende XML-elementen beschreven met de mogelijke kenmerken die door de statusopdracht voor de webcrawler worden teruggezonden:

Tabel 10. Statusinformatie van de webcrawler

Element	Attributen	Beschrijving
CrawlerStatus	<ul style="list-style-type: none"> <li>• CrawlerThreadStateDist</li> <li>• ActiveBucketList</li> <li>• CrawlRate</li> <li>• RecentlyCrawledURLList</li> <li>• NumURLsThisSession</li> </ul>	Crawlerstatus.
CrawlerRunLevel-waarde	<ul style="list-style-type: none"> <li>• Tekenreeks (Engels) "Not started": The crawler session exists, but it has not yet received the start message to process documents.</li> <li>• "Started": The crawler is starting.</li> <li>• "Running": The crawler finished initialization and startup and is actively crawling.</li> <li>• "Paused": The crawler was told to suspend active crawling, but not to exit.</li> <li>• "Stopping": The crawler received the stop signal and is going to stop.</li> <li>• "Error": The crawler is in an unrecoverable state, and it must be stopped and restarted to resume crawling.</li> </ul>	Informatie over de acties waarmee de crawler bezig is.

Tabel 10. Statusinformatie van de webcrawler (vervolg)

Element	Attributen	Beschrijving
CrawlerThreadState-status	Tekenreeks (Engels)	Crawlerthreadactiviteit. In dit veld wordt aangegeven welke acties door de thread of threads worden uitgevoerd.
ActiveBucket	<ul style="list-style-type: none"> <li>URL: tekenreeks (URL-specifiek) Het protocol, de host en de poort waarvoor de URL's worden gecrawld.</li> <li>NumActURLs: geheel getal (positief) Het aantal URL's in de bucket wanneer deze beschikbaar is gemaakt om de crawlen (geactiveerd).</li> <li>NumProcURLs: geheel getal (niet negatief) Het aantal URL's uit de bucket dat tot nu toe is verwerkt (gecrawld of afgewezen).</li> <li>TimeRem: geheel getal Het aantal resterende seconden voordat een timeout voor de bucket optreedt.</li> <li>Duration: geheel getal (niet negatief) Het aantal seconden sinds de bucket is geactiveerd.</li> </ul>	De huidige activiteit voor de opgegeven website.
CrawlRate	<p>Waarde: geheel getal (niet negatief)</p> <p>Aantal pagina's per seconde dat wordt gecrawld (alle buckets samen).</p>	De meting voor de crawlerdoorvoer.
RecentlyCrawledURL	<p>URL: tekenreeks (URL-specifiek)</p> <p>Tekenreeks waarmee het protocol, de host, de port en het bestand worden aangegeven die zijn gecrawld.</p>	Een pagina die recent is gecrawld.
NumURLsThisSession	Waarde: geheel getal (niet negatief)	Het aantal URL's dat is gemaakt sinds deze instance van de crawler (proces) is begonnen met crawlen.

**Status van niet-webcrawler:** als u de opdracht uitvoert om de status voor een niet-webcrawler op te halen, worden de gegevens in een XML-documentindeling teruggezonden. De volgende gegevens kunnen met de opdracht `getCrawlerStatus` voor niet-webcrawlers worden teruggezonden:

```
FFQC5314I Resultaat: <?xml version='1.0' encoding='UTF-8'?>
<GeneralStatus>
<Status>0</Status>
<StatusMessage>Idle</StatusMessage>
<NumberOfServers>1</NumberOfServers>
<NumberOfCompletedServers>1</NumberOfCompletedServers>
<NumberOfTargets>3</NumberOfTargets>
<NumberOfCompletedTargets>3</NumberOfCompletedTargets>
<NumberOfCrawledRecords>115</NumberOfCrawledRecords>
<RunningThreads>0</RunningThreads>
</GeneralStatus>
```

In de volgende tabel worden de XML-elementen en kenmerken voor alle enterprise search-crawlers beschreven, met uitzondering van de webcrawler. Deze gegevens worden teruggezonden met behulp van de opdracht voor de crawlerstatus.



Tabel 11. Crawlerstatusgegevens voor de NNTP-, DB2-, JDBC-database- en Notes-crawlers

Element- en kenmerknaam	NNTP-crawler	DB2- en JDBC-databasecrawlers	Notes-crawler
Status	Status (0, 1, 2, -1)	Status (0, 1, 2, -1)	Status (0, 1, 2, -1)
StatusMessage	Status: 0 - Niet actief, 1 - Actief, 2 - Onderbroken, -1 - Fout	Status: 0 - Niet actief, 1 - Actief, 2 - Onderbroken, -1 - Fout	Status: 0 - Niet actief, 1 - Actief, 2 - Onderbroken, -1 - Fout
NumberOfServers	Het aantal NNTP-servers in de crawlruimte.	Het aantal databases in de crawlruimte.	Het aantal databases in de crawlruimte.
NumberOfCompletedServers	Het aantal gecrawlde NNTP-servers.	Het aantal gecrawlde databases.	Het aantal gecrawlde databases.
NumberOfTargets	Het aantal nieuwsgroepen in de crawlruimte.	Het aantal databases in de crawlruimte.	Het aantal views en mappen in de crawlruimte.
NumberOfCompletedTargets	Het aantal gecrawlde nieuwsgroepen.	Het aantal gecrawlde tabellen.	Het aantal gecrawlde views en mappen.
NumberOfCompletedRecords	Het aantal gecrawlde artikelen.	Het aantal gecrawlde records.	Het aantal gecrawlde documenten.
RunningThreads	Het aantal crawlerthreads.	Het aantal crawlerthreads.	Het aantal crawlerthreads.

Tabel 12. Crawlerstatusgegevens voor de Exchange Server-, DB2 Content Manager- en Content Edition-crawlers

Element- en kenmerknaam	Exchange Server-crawler	DB2 Content Manager-crawler	Content Edition-crawler
Status	Status (0, 1, 2, -1)	Status (0, 1, 2, -1)	Status (0, 1, 2, -1)
StatusMessage	Status: 0 - Niet actief, 1 - Actief, 2 - Onderbroken, -1 - Fout	Status: 0 - Niet actief, 1 - Actief, 2 - Onderbroken, -1 - Fout	Status: 0 - Niet actief, 1 - Actief, 2 - Onderbroken, -1 - Fout
NumberOfServers	Het aantal Exchange Server-servers in de crawlruimte.	Het aantal Content Manager-servers in de crawlruimte.	Het aantal repository's in de crawlruimte.
NumberOfCompletedServers	Het aantal gecrawlde Exchange Server-servers.	Het aantal gecrawlde Content Manager-servers.	Het aantal gecrawlde repository's.
NumberOfTargets	Het aantal submappen in de crawlruimte.	Het aantal itemtypen in de crawlruimte.	Het aantal klassen in de crawlruimte.
NumberOfCompletedTargets	Het aantal gecrawlde submappen.	Het aantal gecrawlde itemtypen.	Het aantal gecrawlde itemklassen.
NumberOfCompletedRecords	Het aantal gecrawlde documenten.	Het aantal gecrawlde documenten.	Het aantal gecrawlde documenten.
RunningThreads	Het aantal crawlerthreads.	Het aantal crawlerthreads.	Het aantal crawlerthreads.

Tabel 13. Crawlerstatusgegevens voor de QuickPlace-, Domino Document Manager-, UNIX-bestandssysteem- en Windows-bestandssysteemcrawlers

Element- en kenmerknaam	QuickPlace-crawler	Domino Document Manager-crawler	UNIX- en Windows-bestandssysteemcrawlers
Status	Status (0, 1, 2, -1)	Status (0, 1, 2, -1)	Status (0, 1, 2, -1)
StatusMessage	Status: 0 - Niet actief, 1 - Actief, 2 - Onderbroken, -1 - Fout	Status: 0 - Niet actief, 1 - Actief, 2 - Onderbroken, -1 - Fout	Status: 0 - Niet actief, 1 - Actief, 2 - Onderbroken, -1 - Fout
NumberOfServers	Het aantal ruimten in de crawlruimte.	Het aantal bibliotheken in de crawlruimte.	Vaste waarde van 1.

Tabel 13. Crawlerstatusgegevens voor de QuickPlace-, Domino Document Manager-, UNIX-bestandssysteem- en Windows-bestandssysteemcrawlers (vervolg)

Element- en kenmerknaam	QuickPlace-crawler	Domino Document Manager-crawler	UNIX- en Windows-bestandssysteemcrawlers
NumberOfCompletedServers	Het aantal gecrawlde ruimten.	Het aantal gecrawlde bibliotheken.	0 of 1 als alle subdirectory's zijn gecrawld.
NumberOfTargets	Het aantal ruimtedatabases en roomdatabases in de crawlruimte.	Het aantal kabinetten in de crawlruimte.	Het aantal subdirectory's in de crawlruimte.
NumberOfCompletedTargets	Het aantal gecrawlde ruimtedatabases en roomdatabases.	Het aantal gecrawlde kabinetten.	Het aantal gecrawlde subdirectory's.
NumberOfCompletedRecords	Het aantal gecrawlde documenten.	Het aantal gecrawlde documenten.	Het aantal gecrawlde bestanden.
RunningThreads	Het aantal crawlerthreads.	Het aantal crawlerthreads.	Het aantal crawlerthreads.

Tabel 14. Crawlerstatusgegevens voor de WebSphere Portal- en Web Content Management-crawlers

Element- en kenmerknaam	WebSphere Portal-crawler	Web Content Management-crawler
Status	Status (0, 1, 2, -1)	Status (0, 1, 2, -1)
StatusMessage	Status: 0 - Niet actief, 1 - Actief, 2 - Onderbroken, -1 - Fout	Status: 0 - Niet actief, 1 - Actief, 2 - Onderbroken, -1 - Fout
NumberOfServers	Het aantal servers in de crawlruimte.	Het aantal sites in de crawlruimte.
NumberOfCompletedServers	Het aantal gecrawlde servers.	Het aantal gecrawlde sites.
NumberOfTargets	Het aantal servers in de crawlruimte.	Het aantal sites in de crawlruimte.
NumberOfCompletedTargets	Het aantal gecrawlde servers.	Het aantal gecrawlde sites.
NumberOfCompletedRecords	Het aantal gecrawlde documenten.	Het aantal gecrawlde documenten.
RunningThreads	Het aantal crawlerthreads.	Het aantal crawlerthreads.

**Status van de crawlruimte voor de webcrawler:** als u de opdracht uitvoert om de status van de crawlruimte voor een webcrawler op te halen, worden de gegevens in een XML-documentindeling teruggezonden. De volgende gegevens kunnen met de opdracht voor de status van de crawlruimte voor de webcrawler worden teruggezonden:

Tabel 15. Selectiemaskerwaarden voor de opdracht voor de status van de crawlruimte voor de webcrawler

Maskerbit	Selectie
1	Aantal pagina's in een opslagplek voor onbewerkte gegevens.
2	Aantal ontdekte sites.
4	Aantal sites met DNS.
8	Aantal sites zonder DNS.
16	Aantal ontdekte URL's.
32	Aantal unieke opgeslagen pagina's.
64	Aantal gecrawlde URL's.
128	Aantal niet-gecrawlde URL's.

Tabel 15. Selectiemaskerwaarden voor de opdracht voor de status van de crawlruimte voor de webcrawler (vervolg)

Maskerbit	Selectie
256	Aantal te late URL's.
512	Distributie van HTTP-statuscodes.

Alle waarden geven cumulatieve totalen aan voor alle sessies die gebruikmaken van de huidige interne database:

```
<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"/>
    <HTTPCode Code="301" Count="1000"/>
    <HTTPCode Code="404" Count="10000"/>
    <HTTPCode Code="780" Count="20000"/>
  </HTTPCode Code="780" Count="20000">
</CrawlStatus>
```

De gegevens die worden teruggezonden, bevatten enkele of alle (mogelijk geen) van de volgende elementen:

Tabel 16. Statusgegevens over crawlruimte voor de webcrawler

Element	Kenmerk	Beschrijving
CrawlerStatus	<ul style="list-style-type: none"> <li>NumPagesInRDS</li> <li>NumSitesDiscovered</li> <li>NumSitesWithDNS</li> <li>NumSitesWithoutDNS</li> <li>NumURLsDiscovered</li> <li>NumUniquePagesSaved</li> <li>NumURLsCrawled</li> <li>NumURLsUncrawled</li> <li>NumURLsOverdue</li> <li>HTTPCodeDist</li> </ul>	Gegevens die snel kunnen worden opgehaald over de cumulatieve status van de crawl (alle sessies).
NumPagesInRDS	<p>Waarde: niet-negatief geheel getal</p> <p>Het aantal pagina's dat momenteel in het RDS-staginggebied (Raw Data Store) zijn opgeslagen (alleen van deze crawler).</p>	Hoe vol de RDS (Raw Data Store) wordt (alleen van bijdragen van deze crawler).
NumSitesDiscovered	<p>Waarde: niet-negatief geheel getal</p> <p>Het aantal hosts dat is ontdekt met crawlen (of is geleverd via basiswaarden).</p>	Een meting van de crawlerdekking van het domein dat moet worden gecrawld (hosttelling).
NumSitesWithDNS	<p>Waarde: niet-negatief geheel getal</p> <p>Het aantal hosts waaraan IP-adressen zijn gekoppeld (die door de crawler op de achtergrond worden omgezet).</p>	Een meting van de mate van effectiviteit waarop de crawler IP-adressen kan ophalen voor hosts die op basis van de DNS-namen in URL's zijn ontdekt.

Tabel 16. Statusgegevens over crawlruimte voor de webcrawler (vervolg)

Element	Kenmerk	Beschrijving
NumSitesWithoutDNS	<p>Waarde: niet-negatief geheel getal</p> <p>Het aantal hosts waaraan geen IP-adressen zijn gekoppeld (die door de crawler op de achtergrond worden omgezet).</p>	Een meting van de mate van effectiviteit waarop de crawler IP-adressen kan ophalen voor hosts die op basis van de DNS-namen in URL's zijn ontdekt.
NumURLsDiscovered	<p>Waarde: niet-negatief geheel getal</p> <p>Het aantal unieke URL's dat door de crawler is bezocht.</p>	Een meting van de crawlerdekking van het domein dat moet worden gecrawld (URL-telling).
NumUniquePagesSaved	<p>Waarde: niet-negatief geheel getal</p> <p>Het aantal unieke pagina's dat naar de RDS is geschreven voor verdere verwerking door andere enterprise search-componenten.</p>	De bijdrage van de crawler aan de grootte van de index.
NumURLsCrawled	<p>Waarde: niet-negatief geheel getal</p> <p>Het aantal unieke URL's dat door de crawler is gecrawld.</p>	Een meting van de mogelijkheid van de A crawler om (achter elkaar) gegevens te verwerken. Dit aantal verschilt van het aantal pagina's dat naar de RDS wordt geschreven, omdat niet alle gecrawlde pagina's ook daadwerkelijk naar de RDS worden geschreven.
NumURLsOverdue	<p>Waarde: niet-negatief geheel getal</p> <p>Het aantal unieke URL's dat in aanmerking komt om opnieuw te worden gecrawld.</p>	Een meting van de mogelijkheid van de crawler om door de webruimte te bladeren.

**Status van de crawlruimte voor niet-webcrawlers:** als u de opdracht uitvoert om de status van de crawlruimte voor een niet-webcrawler op te halen, worden de gegevens in een XML-documentindeling teruggezonden. De volgende gegevens kunnen met de opdracht `getCrawlSpaceStatus` voor niet-webcrawlers worden teruggezonden:

```
FFQC5314I Resultaat: <?xml version='1.0' encoding='UTF-8'?>
<ServerStatus>
  <Server Name ="FOUNTAIN">
    <Status>5</Status>
    <StatusMessage>Scheduled</StatusMessage>
    <NumberOfTargets>1</NumberOfTargets>
    <NumberOfCompletedTargets>1</NumberOfCompletedTargets>
    <NumberOfErrors>0</NumberOfErrors>
    <StartTime>1118354510512</StartTime>
    <EndTime>1118354514386</EndTime>
    <ScheduleConfigured>2</ScheduleConfigured>
    <ScheduleTime>1118393377000</ScheduleTime>
    <TotalTime>3874</TotalTime>
  </Server>
</ServerStatus>
```

In de volgende tabel worden de XML-elementen en kenmerken voor alle enterprise search-crawlers beschreven, met uitzondering van de webcrawler. Deze gegevens worden teruggezonden met behulp van de opdracht voor de status van de crawlruimte. Voor Notes-crawlers, als het aggregatieniveau 0 is, is de waarde voor `Server@Name` servernaam + databasenaam. Als het aggregatieniveau 1 is, is de waarde voor `Server@Name` servernaam + directorynaam.

Tabel 17. Statusgegevens over de crawlruimte voor de NNTP-, DB2-, JDBC-database- en Notes-crawlers

Element- en kenmerknaam	NNTP-crawler	DB2- en JDBC-databasecrawlers	Notes-crawler
Server@Name	Naam nieuwsserver	Naam database	Naam database of directory
Server/Status	Status: (0, 1, 2, 3, 4, 5, -1) <ul style="list-style-type: none"> <li>• 0: Niet gecrawld</li> <li>• 1: bezig met crawlen</li> <li>• 2: Voltooid (niet gepland)</li> <li>• 3: Wachten</li> <li>• 4: Onderbroken</li> <li>• 5: Gepland</li> <li>• -1: Fout</li> </ul>	Status (0, 1, 2, 3, 4, 5, -1) <ul style="list-style-type: none"> <li>• 0: Niet gecrawld</li> <li>• 1: bezig met crawlen</li> <li>• 2: Voltooid (niet gepland)</li> <li>• 3: Wachten</li> <li>• 4: Onderbroken</li> <li>• 5: Gepland</li> <li>• -1: Fout</li> </ul>	Status (0, 1, 2, 3, 4, 5, -1) <ul style="list-style-type: none"> <li>• 0: Niet gecrawld</li> <li>• 1: bezig met crawlen</li> <li>• 2: Voltooid (niet gepland)</li> <li>• 3: Wachten</li> <li>• 4: Onderbroken</li> <li>• 5: Gepland</li> <li>• -1: Fout</li> </ul>
Server/StatusMessage	<ul style="list-style-type: none"> <li>• 0: Niet gecrawld</li> <li>• 1: bezig met crawlen</li> <li>• 2: Voltooid (niet gepland)</li> <li>• 3: Wachten</li> <li>• 4: Onderbroken</li> <li>• 5: Gepland</li> <li>• -1: Fout</li> </ul>	<ul style="list-style-type: none"> <li>• 0: Niet gecrawld</li> <li>• 1: bezig met crawlen</li> <li>• 2: Voltooid (niet gepland)</li> <li>• 3: Wachten</li> <li>• 4: Onderbroken</li> <li>• 5: Gepland</li> <li>• -1: Fout</li> </ul>	<ul style="list-style-type: none"> <li>• 0: Niet gecrawld</li> <li>• 1: bezig met crawlen</li> <li>• 2: Voltooid (niet gepland)</li> <li>• 3: Wachten</li> <li>• 4: Onderbroken</li> <li>• 5: Gepland</li> <li>• -1: Fout</li> </ul>
Server/NumberOfTargets	Het aantal nieuwsgroepen in de crawlruimte.	Het aantal databases in de crawlruimte.	Het aantal views en mappen of directory's in de crawlruimte.
Server/NumberOfCompletedTargets	Het aantal gecrawlde nieuwsgroepen.	Het aantal gecrawlde tabellen.	Het aantal gecrawlde views en mappen of directory's.
Server/NumberOfErrors	Het aantal fouten.	Het aantal fouten.	Het aantal fouten.
Server/StartTime	De starttijd (indien van toepassing).	De starttijd (indien van toepassing).	De starttijd (indien van toepassing).
Server/EndTime	De eindtijd (indien van toepassing).	De eindtijd (indien van toepassing).	De eindtijd (indien van toepassing).
Server/ScheduleConfigured	0, 1, 2 <ul style="list-style-type: none"> <li>• 0: de crawler is niet geconfigureerd voor planning volgens de crawlerconfiguratiebestanden.</li> <li>• 1: de crawler is geconfigureerd voor planning, maar de planning is uitgeschakeld voor de sessie.</li> <li>• 2: de crawler is geconfigureerd voor planning en de planning is ingeschakeld voor de sessie.</li> </ul>	0, 1, 2 <ul style="list-style-type: none"> <li>• 0: de crawler is niet geconfigureerd voor planning volgens de crawlerconfiguratiebestanden.</li> <li>• 1: de crawler is geconfigureerd voor planning, maar de planning is uitgeschakeld voor de sessie.</li> <li>• 2: de crawler is geconfigureerd voor planning en de planning is ingeschakeld voor de sessie.</li> </ul>	0, 1, 2 <ul style="list-style-type: none"> <li>• 0: de crawler is niet geconfigureerd voor planning volgens de crawlerconfiguratiebestanden.</li> <li>• 1: de crawler is geconfigureerd voor planning, maar de planning is uitgeschakeld voor de sessie.</li> <li>• 2: de crawler is geconfigureerd voor planning en de planning is ingeschakeld voor de sessie.</li> </ul>
Server/ScheduleTime	Geplande tijd (indien van toepassing).	Geplande tijd (indien van toepassing).	Geplande tijd (indien van toepassing).

Tabel 17. Statusgegevens over de crawlruimte voor de NNTP-, DB2-, JDBC-database- en Notes-crawlers (vervolg)

Element- en kenmerknaam	NNTP-crawler	DB2- en JDBC-databasecrawlers	Notes-crawler
Server/TotalTime	De totale tijd (indien van toepassing).	De totale tijd (indien van toepassing).	De totale tijd (indien van toepassing).
Server/AggregationLevel	0: de crawler crawlt documenten in de normale werkstand.	0: de crawler crawlt documenten in de normale werkstand.	0, 1: <ul style="list-style-type: none"> <li>0: de Notes-crawler crawlt documenten in de normale werkstand. (Met de overige crawlers wordt altijd de waarde 0 teruggezonden, met uitzondering van de Notes-crawler.)</li> <li>1: de Notes-crawler crawlt documenten met de directorywerkstand.</li> </ul>

Tabel 18. Statusgegevens van de crawlruimte voor de Exchange Server-, DB2 Content Manager- en Content Edition-crawlers

Element- en kenmerknaam	Exchange Server-crawler	DB2 Content Manager-crawler	Content Edition-crawler
Server@Name	Naam van de Exchange Server-server.	DB2 Content Manager-servers.	Repositorynaam.
Server/Status	Status (0, 1, 2, 3, 4, 5, -1) <ul style="list-style-type: none"> <li>0: Niet gecrawld</li> <li>1: bezig met crawlen</li> <li>2: Voltooid (niet gepland)</li> <li>3: Wachten</li> <li>4: Onderbroken</li> <li>5: Gepland</li> <li>-1: Fout</li> </ul>	Status (0, 1, 2, 3, 4, 5, -1) <ul style="list-style-type: none"> <li>0: Niet gecrawld</li> <li>1: bezig met crawlen</li> <li>2: Voltooid (niet gepland)</li> <li>3: Wachten</li> <li>4: Onderbroken</li> <li>5: Gepland</li> <li>-1: Fout</li> </ul>	Status (0, 1, 2, 3, 4, 5, -1) <ul style="list-style-type: none"> <li>0: Niet gecrawld</li> <li>1: bezig met crawlen</li> <li>2: Voltooid (niet gepland)</li> <li>3: Wachten</li> <li>4: Onderbroken</li> <li>5: Gepland</li> <li>-1: Fout</li> </ul>
Server/StatusMessage	<ul style="list-style-type: none"> <li>0: Niet gecrawld</li> <li>1: bezig met crawlen</li> <li>2: Voltooid (niet gepland)</li> <li>3: Wachten</li> <li>4: Onderbroken</li> <li>5: Gepland</li> <li>-1: Fout</li> </ul>	<ul style="list-style-type: none"> <li>0: Niet gecrawld</li> <li>1: bezig met crawlen</li> <li>2: Voltooid (niet gepland)</li> <li>3: Wachten</li> <li>4: Onderbroken</li> <li>5: Gepland</li> <li>-1: Fout</li> </ul>	<ul style="list-style-type: none"> <li>0: Niet gecrawld</li> <li>1: bezig met crawlen</li> <li>2: Voltooid (niet gepland)</li> <li>3: Wachten</li> <li>4: Onderbroken</li> <li>5: Gepland</li> <li>-1: Fout</li> </ul>
Server/NumberOfTargets	Het aantal submappen in de crawlruimte.	Het aantal itemtypen in de crawlruimte.	Het aantal itemklassen in de crawlruimte.
Server/NumberOfCompletedTargets	Het aantal gecrawlde submappen.	Het aantal gecrawlde itemtypen.	Het aantal gecrawlde itemklassen.
Server/NumberOfErrors	Het aantal fouten.	Het aantal fouten.	Het aantal fouten.
Server/StartTime	De starttijd (indien van toepassing).	De starttijd (indien van toepassing).	De starttijd (indien van toepassing).
Server/EndTime	De eindtijd (indien van toepassing).	De eindtijd (indien van toepassing).	De eindtijd (indien van toepassing).

Tabel 18. Statusgegevens van de crawlruimte voor de Exchange Server-, DB2 Content Manager- en Content Edition-crawlers (vervolg)

Element- en kenmerknaam	Exchange Server-crawler	DB2 Content Manager-crawler	Content Edition-crawler
Server/ScheduleConfigured	0, 1, 2 <ul style="list-style-type: none"> <li>0: de crawler is niet geconfigureerd voor planning volgens de crawlerconfiguratiebestanden.</li> <li>1: de crawler is geconfigureerd voor planning, maar de planning is uitgeschakeld voor de sessie.</li> <li>2: de crawler is geconfigureerd voor planning en de planning is ingeschakeld voor de sessie.</li> </ul>	0, 1, 2 <ul style="list-style-type: none"> <li>0: de crawler is niet geconfigureerd voor planning volgens de crawlerconfiguratiebestanden.</li> <li>1: de crawler is geconfigureerd voor planning, maar de planning is uitgeschakeld voor de sessie.</li> <li>2: de crawler is geconfigureerd voor planning en de planning is ingeschakeld voor de sessie.</li> </ul>	0, 1, 2 <ul style="list-style-type: none"> <li>0: de crawler is niet geconfigureerd voor planning volgens de crawlerconfiguratiebestanden.</li> <li>1: de crawler is geconfigureerd voor planning, maar de planning is uitgeschakeld voor de sessie.</li> <li>2: de crawler is geconfigureerd voor planning en de planning is ingeschakeld voor de sessie.</li> </ul>
Server/ScheduleTime	Geplande tijd (indien van toepassing).	Geplande tijd (indien van toepassing).	Geplande tijd (indien van toepassing).
Server/TotalTime	De totale tijd (indien van toepassing).	De totale tijd (indien van toepassing).	De totale tijd (indien van toepassing).
Server/AggregationLevel	0: de crawler crawlt documenten in de normale werkstand.	0: de crawler crawlt documenten in de normale werkstand.	0: de crawler crawlt documenten in de normale werkstand.

Tabel 19. Statusgegevens van de crawlruimte voor de QuickPlace-, Domino Document Manager-, UNIX-bestandssysteem- en Windows-bestandssysteemcrawlers

Element- en kenmerknaam	QuickPlace-crawler	Domino Document Manager-crawler	UNIX- en Windows-bestandssysteemcrawlers
Server@Name	Ruimtedirectory	Bibliotheekdatabase	Een vaste waarde van localhost.
Server/Status	Status (0, 1, 2, 3, 4, 5, -1) <ul style="list-style-type: none"> <li>0: Niet gecrawld</li> <li>1: bezig met crawlen</li> <li>2: Voltooid (niet gepland)</li> <li>3: Wachten</li> <li>4: Onderbroken</li> <li>5: Gepland</li> <li>-1: Fout</li> </ul>	Status (0, 1, 2, 3, 4, 5, -1) <ul style="list-style-type: none"> <li>0: Niet gecrawld</li> <li>1: bezig met crawlen</li> <li>2: Voltooid (niet gepland)</li> <li>3: Wachten</li> <li>4: Onderbroken</li> <li>5: Gepland</li> <li>-1: Fout</li> </ul>	Status (0, 1, 2, 3, 4, 5, -1) <ul style="list-style-type: none"> <li>0: Niet gecrawld</li> <li>1: bezig met crawlen</li> <li>2: Voltooid (niet gepland)</li> <li>3: Wachten</li> <li>4: Onderbroken</li> <li>5: Gepland</li> <li>-1: Fout</li> </ul>
Server/StatusMessage	<ul style="list-style-type: none"> <li>0: Niet gecrawld</li> <li>1: bezig met crawlen</li> <li>2: Voltooid (niet gepland)</li> <li>3: Wachten</li> <li>4: Onderbroken</li> <li>5: Gepland</li> <li>-1: Fout</li> </ul>	<ul style="list-style-type: none"> <li>0: Niet gecrawld</li> <li>1: bezig met crawlen</li> <li>2: Voltooid (niet gepland)</li> <li>3: Wachten</li> <li>4: Onderbroken</li> <li>5: Gepland</li> <li>-1: Fout</li> </ul>	<ul style="list-style-type: none"> <li>0: Niet gecrawld</li> <li>1: bezig met crawlen</li> <li>2: Voltooid (niet gepland)</li> <li>3: Wachten</li> <li>4: Onderbroken</li> <li>5: Gepland</li> <li>-1: Fout</li> </ul>

Tabel 19. Statusgegevens van de crawlruimte voor de QuickPlace-, Domino Document Manager-, UNIX-bestandsstelsel- en Windows-bestandssysteemcrawlers (vervolg)

Element- en kenmerknaam	QuickPlace-crawler	Domino Document Manager-crawler	UNIX- en Windows-bestandssysteemcrawlers
Server/NumberOfTargets	Het aantal ruimtedatabases en roomdatabases in de crawlruimte.	Het aantal kabinetten in de crawlruimte.	Het aantal subdirectory's in de crawlruimte.
Server/NumberOfCompletedTargets	Het aantal gecrawelde ruimtedatabases en roomdatabases.	Het aantal gecrawelde kabinetten.	Het aantal subdirectory's in de crawlruimte.
Server/NumberOfErrors	Het aantal fouten.	Het aantal fouten.	Het aantal fouten.
Server/StartTime	De starttijd (indien van toepassing).	De starttijd (indien van toepassing).	De starttijd (indien van toepassing).
Server/EndTime	De eindtijd (indien van toepassing).	De eindtijd (indien van toepassing).	De eindtijd (indien van toepassing).
Server/ScheduleConfigured	0, 1, 2 <ul style="list-style-type: none"> <li>0: de crawler is niet geconfigureerd voor planning volgens de crawlerconfiguratiebestanden.</li> <li>1: de crawler is geconfigureerd voor planning, maar de planning is uitgeschakeld voor de sessie.</li> <li>2: de crawler is geconfigureerd voor planning en de planning is ingeschakeld voor de sessie.</li> </ul>	0, 1, 2 <ul style="list-style-type: none"> <li>0: de crawler is niet geconfigureerd voor planning volgens de crawlerconfiguratiebestanden.</li> <li>1: de crawler is geconfigureerd voor planning, maar de planning is uitgeschakeld voor de sessie.</li> <li>2: de crawler is geconfigureerd voor planning en de planning is ingeschakeld voor de sessie.</li> </ul>	0, 1, 2 <ul style="list-style-type: none"> <li>0: de crawler is niet geconfigureerd voor planning volgens de crawlerconfiguratiebestanden.</li> <li>1: de crawler is geconfigureerd voor planning, maar de planning is uitgeschakeld voor de sessie.</li> <li>2: de crawler is geconfigureerd voor planning en de planning is ingeschakeld voor de sessie.</li> </ul>
Server/ScheduleTime	Geplande tijd (indien van toepassing).	Geplande tijd (indien van toepassing).	Geplande tijd (indien van toepassing).
Server/TotalTime	De totale tijd (indien van toepassing).	De totale tijd (indien van toepassing).	De totale tijd (indien van toepassing).
Server/AggregationLevel	0: de crawler crawl documenten in de normale werkstand.	0: de crawler crawl documenten in de normale werkstand.	0: de crawler crawl documenten in de normale werkstand.

Tabel 20. Statusgegevens van de crawlruimte voor de WebSphere Portal- en Web Content Management-crawlers

Element- en kenmerknaam	WebSphere Portal-crawler	Web Content Management-crawler
Server@Name	WebSphere Portal-server	Begin-URL voor zoekopdrachten van Web Content Management
Server/Status	Status (0, 1, 2, 3, 4, 5, -1) <ul style="list-style-type: none"> <li>0: Niet gecrawld</li> <li>1: bezig met crawlen</li> <li>2: voltooid (niet gepland)</li> <li>3: Wachten</li> <li>4: Onderbroken</li> <li>5: Gepland</li> <li>-1: Fout</li> </ul>	Status (0, 1, 2, 3, 4, 5, -1) <ul style="list-style-type: none"> <li>0: Niet gecrawld</li> <li>1: bezig met crawlen</li> <li>2: voltooid (niet gepland)</li> <li>3: Wachten</li> <li>4: Onderbroken</li> <li>5: Gepland</li> <li>-1: Fout</li> </ul>



Tabel 20. Statusgegevens van de crawlruimte voor de WebSphere Portal- en Web Content Management-crawlers (vervolg)

Element- en kenmerknaam	WebSphere Portal-crawler	Web Content Management-crawler
Server/StatusMessage	<ul style="list-style-type: none"> <li>• 0: Niet gecrawld</li> <li>• 1: Bezig met crawlen</li> <li>• 2: Voltooid (niet gepland)</li> <li>• 3: Wachten</li> <li>• 4: Onderbroken</li> <li>• 5: Gepland</li> <li>• -1: Fout</li> </ul>	<ul style="list-style-type: none"> <li>• 0: Niet gecrawld</li> <li>• 1: Bezig met crawlen</li> <li>• 2: Voltooid (niet gepland)</li> <li>• 3: Wachten</li> <li>• 4: Onderbroken</li> <li>• 5: Gepland</li> <li>• -1: Fout</li> </ul>
Server/NumberOfTargets	Het aantal servers in de crawlruimte.	Het aantal sites in de crawlruimte.
Server/NumberOfCompletedTargets	Het aantal gecrawlde servers.	Het aantal gecrawlde sites.
Server/NumberOfErrors	Het aantal fouten.	Het aantal fouten.
Server/StartTime	De starttijd (indien van toepassing).	De starttijd (indien van toepassing).
Server/EndTime	De eindtijd (indien van toepassing).	De eindtijd (indien van toepassing).
Server/ScheduleConfigured	0, 1, 2 <ul style="list-style-type: none"> <li>• 0: de crawler is niet geconfigureerd voor planning volgens de crawlerconfiguratiebestanden.</li> <li>• 1: de crawler is geconfigureerd voor planning, maar de planning is uitgeschakeld voor de sessie.</li> <li>• 2: de crawler is geconfigureerd voor planning en de planning is ingeschakeld voor de sessie.</li> </ul>	0, 1, 2 <ul style="list-style-type: none"> <li>• 0: de crawler is niet geconfigureerd voor planning volgens de crawlerconfiguratiebestanden.</li> <li>• 1: de crawler is geconfigureerd voor planning, maar de planning is uitgeschakeld voor de sessie.</li> <li>• 2: de crawler is geconfigureerd voor planning en de planning is ingeschakeld voor de sessie.</li> </ul>
Server/ScheduleTime	Geplande tijd (indien van toepassing).	Geplande tijd (indien van toepassing).
Server/TotalTime	De totale tijd (indien van toepassing).	De totale tijd (indien van toepassing).
Server/AggregationLevel	0: de crawler crawlt documenten in de normale werkstand.	0: de crawler crawlt documenten in de normale werkstand.

**Gedetailleerde status van de crawlruimte voor de webcrawler:** als u de opdracht uitvoert om de gedetailleerde status van de crawlruimte voor een webcrawler op te halen, worden de gegevens in een XML-documentindeling teruggezonden. De volgende gegevens kunnen met de opdracht voor de gedetailleerde status van de crawlruimte worden teruggezonden:

Tabel 21. Selectiemaskerwaarden voor de opdracht voor de status van de crawlruimte voor de webcrawler

Maskerbit	Selectie
1	Aantal pagina's in een opslagplek voor onbewerkte gegevens.
2	Aantal ontdekte sites.
4	Aantal sites met DNS.
8	Aantal sites zonder DNS.
16	Aantal ontdekte URL's.
32	Aantal unieke opgeslagen pagina's.

Tabel 21. Selectiemaskerwaarden voor de opdracht voor de status van de crawlruimte voor de webcrawler (vervolg)

Maskerbit	Selectie
64	Aantal gecrawlde URL's.
128	Aantal niet-gecrawlde URL's.
256	Aantal te late URL's.
512	Distributie van HTTP-statuscodes.

Voorbeeld van teruggezonden gegevens:

```
<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>
  <HTTPCodeDist Count="4" Total="1031000"/>
    <HTTPCode Code="200" Count ="1000000"/>
    <HTTPCode Code="301" Count ="1000"/>
    <HTTPCode Code="404" Count ="10000"/>
    <HTTPCode Code="780" Count="20000"/>
  </HTTPCodeDist>
</CrawlDetailsPerSite>
```

In de volgende tabel worden de velden beschreven die voor de gedetailleerde status van de crawlruimte voor de webcrawler worden teruggezonden:

Tabel 22. Gedetailleerde statusinformatie van de crawlruimte voor de webcrawler

Element	Attributen	Beschrijving
CrawlDetailsPerSite	<ul style="list-style-type: none"> <li>LastActivationTime:</li> <li>LastActivationDuration:</li> <li>IPAddressList:</li> <li>RobotsContent:</li> <li>HTTPCodeDist:</li> </ul>	Gegevens die snel kunnen worden opgehaald over de gedetailleerde status van een site.
Site	URL	URL van de siterootpagina.
NumURLsDiscovered	Waarde	Het aantal URL's dat via de site is ontdekt.
NumURLsOverdue	Waarde	Het aantal URL's dat in aanmerking komt om opnieuw te worden gecrawld via de site.
NumURLsCrawled	Waarde	Het aantal URL's dat is gecrawld voor de site.
NumURLsUncrawled	Waarde	Het aantal URL's dat nog niet is gecrawld voor de site.

Tabel 22. Gedetailleerde statusinformatie van de crawlruimte voor de webcrawler (vervolg)

Element	Attributen	Beschrijving
NumURLsOverdueBy	Drempel, waarde: geheel getal (positief of negatief)  De waarde geeft het aantal URL's aan dat in aanmerking komt om opnieuw te worden gecrawld. De drempel geeft aan hoeveel tijd de URL's moeten hebben gewacht om opnieuw te worden gecrawld. De drempel wordt gemeten als het aantal seconden vanaf de actuele tijd. Als de drempel negatief is, betekent dat dat er al URL's opnieuw gecrawld hadden moeten worden. Is de drempel positief, dan betekent dit dat er nog URL's opnieuw moeten worden gecrawld.	Het aantal URL's dat in aanmerking is gekomen om ten minste een aantal seconden geleden opnieuw te worden gecrawld, of die in aanmerking komen om in de komende zoveel seconden opnieuw te worden gecrawld.
NumURLsActivated	Waarde	Het aantal URL's dat in het geheugen is opgeslagen tijdens de laatste scanactie voor de site en dat beschikbaar is gemaakt voor crawlerthreads.
LastActivationTime	Waarde	Het aantal seconden sinds de epoche waarna de URL's van deze site voor het laatst in het geheugen zijn opgeslagen.
LastActivationDuration	Waarde	Het aantal seconden waarna de URL's van deze site voor het laatst in het geheugen zijn opgeslagen en beschikbaar zijn gemaakt voor crawlerthreads.
IPAddressList	IPAddress	Alle bekende IP-adressen voor de serverhost van deze site.
IPAddress	Waarde	IPv4-adressen (puntnotatie) voor de serverhost van deze site.
RobotsContent	Tekst	Tekst uit het robots-bestand (als er tekst aanwezig is).
HTTPCodeDist	HTTPCode	Distributie van HTTP-codes op basis van geprobeerde downloads van deze site.
HTTPCode	Code: geheel getal  Een HTTP-statuscode of een andere interne code.	Het aantal keer dat een bepaalde HTTP-statuscode optreedt tijdens het crawlen van deze site.

**Gedetailleerde status van de crawlruimte voor niet-webcrawlers:** als u de opdracht uitvoert om de gedetailleerde status van de crawlruimte voor niet-webcrawlers op te halen, worden de gegevens in een XML-documentindeling teruggezonden. De volgende gegevens kunnen met de opdracht `getCrawlSpaceStatusDetail` voor niet-webcrawlers worden teruggezonden:

```
FFQC5314I Resultaat: <?xml version='1.0' encoding='UTF-8'?>
<TargetStatus>
  <Target Name ="escmgr.crawlerinstances">
    <Status>2</Status>
```

```

<StatusMessage>Completed</StatusMessage>
<NumberOfRecords></NumberOfRecords>
<NumberOfCrawledRecords>117</NumberOfCrawledRecords>
<NumberOfInsertedRecords>21</NumberOfInsertedRecords>
<NumberOfUpdatedRecords>45</NumberOfUpdatedRecords>
<StartTime>1118354510727</StartTime>
<EndTime>1118354514386</EndTime>
<AggregationLevel>0<AggregationLevel>
<Target>
</TargetStatus>

```

Tabel 23. Gedetailleerde statusgegevens over de crawlruimte voor de NNTP-, DB2-, JDBC-database- en Notes-crawlers

Element- en kenmerknaam	NNTP-crawler	DB2- en JDBC-databasecrawlers	Notes-crawler
Target@Name	Nieuwsgroepnaam	Tabelnaam	View- of mapnaam
Target@CrawlType	Niet van toepassing.	0,1 (DB2); 0 (JDBC-database) <ul style="list-style-type: none"> <li>0: Actieve crawl (normaal)</li> <li>1: Passieve crawl (DB2 Event Publishing)</li> </ul>	0
Target/Status	Status: (0, 1, 2, 3, 4, -1) <ul style="list-style-type: none"> <li>0: Niet gecrawld</li> <li>1: bezig met crawlen</li> <li>2: Voltooid (niet gepland)</li> <li>3: Wachten</li> <li>4: Onderbroken</li> <li>-1: Fout</li> </ul>	Status: (0, 1, 2, 3, 4, -1) <ul style="list-style-type: none"> <li>0: Niet gecrawld</li> <li>1: bezig met crawlen</li> <li>2: Voltooid (niet gepland)</li> <li>3: Wachten</li> <li>4: Onderbroken</li> <li>-1: Fout</li> </ul>	Status: (0, 1, 2, 3, 4, -1) <ul style="list-style-type: none"> <li>0: Niet gecrawld</li> <li>1: bezig met crawlen</li> <li>2: Voltooid (niet gepland)</li> <li>3: Wachten</li> <li>4: Onderbroken</li> <li>-1: Fout</li> </ul>
Target/StatusMessage	<ul style="list-style-type: none"> <li>0: Niet gecrawld</li> <li>1: bezig met crawlen</li> <li>2: Voltooid (niet gepland)</li> <li>3: Wachten</li> <li>4: Onderbroken</li> <li>-1: Fout</li> </ul>	<ul style="list-style-type: none"> <li>0: Niet gecrawld</li> <li>1: bezig met crawlen</li> <li>2: Voltooid (niet gepland)</li> <li>3: Wachten</li> <li>4: Onderbroken</li> <li>-1: Fout</li> </ul>	<ul style="list-style-type: none"> <li>0: Niet gecrawld</li> <li>1: bezig met crawlen</li> <li>2: Voltooid (niet gepland)</li> <li>3: Wachten</li> <li>4: Onderbroken</li> <li>-1: Fout</li> </ul>
Target/NumberOf Records	Het nummer van het laatste artikel op de server.	Het aantal gecrawlde records.	Het aantal gecrawlde documenten.
Target/NumberOf CompletedRecords	Het aantal gecrawlde artikelen.	Het aantal gecrawlde records.	Het aantal gecrawlde documenten.
Target/NumberOf InsertedRecords	Het aantal nieuw geplaatste artikelen.	Het aantal ingevoegde records.	Het aantal ingevoegde records.
Target/NumberOf UpdatedRecords	Niet van toepassing.	Het aantal bijgewerkte records.	Het aantal bijgewerkte records.
Target/NumberOf DeletedRecords	Niet van toepassing.	Het aantal gewiste records.	Het aantal gewiste records.
Target/StartTime	De datum en tijd waarop de crawler voor het laatst is gestart.	De datum en tijd waarop de crawler voor het laatst is gestart.	De datum en tijd waarop de crawler voor het laatst is gestart.
Target/EndTime	De datum en tijd waarop het crawlen is voltooid.	De datum en tijd waarop het crawlen is voltooid.	De datum en tijd waarop het crawlen is voltooid.

Tabel 23. Gedetailleerde statusgegevens over de crawlruimte voor de NNTP-, DB2-, JDBC-database- en Notes-crawlers (vervolg)

Element- en kenmerknaam	NNTP-crawler	DB2- en JDBC-databasecrawlers	Notes-crawler
Target/TotalTime	De hoeveelheid tijd die de crawler aan het crawlen heeft besteed.	De hoeveelheid tijd die de crawler aan het crawlen heeft besteed.	De hoeveelheid tijd die de crawler aan het crawlen heeft besteed.
Target/AggregationLevel	0: de crawler crawlt documenten in de normale werkstand.	0: de crawler crawlt documenten in de normale werkstand.	0, 1: <ul style="list-style-type: none"> <li>0: de crawler crawlt documenten in de normale werkstand.</li> <li>De crawler crawlt documenten met de directorywerkstand.</li> </ul>
Target/LastUpdatedTime	Niet van toepassing.	Laatst bijgewerkt: <ul style="list-style-type: none"> <li>0: Actieve crawl (normaal)</li> <li>1: Passieve crawl (DB2 Event Publishing)</li> </ul>	Niet van toepassing.
Target/LastResetTime	Niet van toepassing.	Tijdstip waarop de cijfers voor het laatst opnieuw zijn ingesteld: <ul style="list-style-type: none"> <li>0: Actieve crawl (normaal)</li> <li>1: Passieve crawl (DB2 Event Publishing)</li> </ul>	Niet van toepassing.

Tabel 24. Gedetailleerde statusgegevens van de crawlruimte voor de Exchange Server-, DB2 Content Manager- en Content Edition-crawlers

Element- en kenmerknaam	Exchange Server-crawler	DB2 Content Manager-crawler	Content Edition-crawler
Target@Name	Naam submap	Naam itemtype	Naam itemklasse
Target@CrawlType	0	0	0
Target/Status	Status (0, 1, 2, 3, 4, -1) <ul style="list-style-type: none"> <li>0: Niet gecrawld</li> <li>1: bezig met crawlen</li> <li>2: Voltooid (niet gepland)</li> <li>3: Wachten</li> <li>4: Onderbroken</li> <li>-1: Fout</li> </ul>	Status (0, 1, 2, 3, 4, -1) <ul style="list-style-type: none"> <li>0: Niet gecrawld</li> <li>1: bezig met crawlen</li> <li>2: Voltooid (niet gepland)</li> <li>3: Wachten</li> <li>4: Onderbroken</li> <li>-1: Fout</li> </ul>	Status (0, 1, 2, 3, 4, -1) <ul style="list-style-type: none"> <li>0: Niet gecrawld</li> <li>1: bezig met crawlen</li> <li>2: Voltooid (niet gepland)</li> <li>3: Wachten</li> <li>4: Onderbroken</li> <li>-1: Fout</li> </ul>
Target/StatusMessage	<ul style="list-style-type: none"> <li>0: Niet gecrawld</li> <li>1: bezig met crawlen</li> <li>2: Voltooid (niet gepland)</li> <li>3: Wachten</li> <li>4: Onderbroken</li> <li>-1: Fout</li> </ul>	<ul style="list-style-type: none"> <li>0: Niet gecrawld</li> <li>1: bezig met crawlen</li> <li>2: Voltooid (niet gepland)</li> <li>3: Wachten</li> <li>4: Onderbroken</li> <li>-1: Fout</li> </ul>	<ul style="list-style-type: none"> <li>0: Niet gecrawld</li> <li>1: bezig met crawlen</li> <li>2: Voltooid (niet gepland)</li> <li>3: Wachten</li> <li>4: Onderbroken</li> <li>-1: Fout</li> </ul>
Target/NumberOf Records	Niet van toepassing.	Niet van toepassing.	Niet van toepassing.

Tabel 24. Gedetailleerde statusgegevens van de crawlruimte voor de Exchange Server-, DB2 Content Manager- en Content Edition-crawlers (vervolg)

Element- en kenmerknaam	Exchange Server-crawler	DB2 Content Manager-crawler	Content Edition-crawler
Target/NumberOfCompletedRecords	Het aantal gecrawlde documenten.	Het aantal gecrawlde documenten.	Het aantal gecrawlde documenten.
Target/NumberOfInsertedRecords	Het aantal ingevoegde records.	Het aantal ingevoegde records.	Het aantal ingevoegde records.
Target/NumberOfUpdatedRecords	Niet van toepassing.	Het aantal bijgewerkte records.	Het aantal bijgewerkte records.
Target/NumberOfDeletedRecords	Niet van toepassing.	Het aantal gewiste records.	Het aantal gewiste records.
Target/StartTime	De datum en tijd waarop de crawler voor het laatst is gestart.	De datum en tijd waarop de crawler voor het laatst is gestart.	De datum en tijd waarop de crawler voor het laatst is gestart.
Target/EndTime	De datum en tijd waarop het crawlen is voltooid.	De datum en tijd waarop het crawlen is voltooid.	De datum en tijd waarop het crawlen is voltooid.
Target/TotalTime	De hoeveelheid tijd die de crawler aan het crawlen heeft besteed.	De hoeveelheid tijd die de crawler aan het crawlen heeft besteed.	De hoeveelheid tijd die de crawler aan het crawlen heeft besteed.
Target/AggregationLevel	0: de crawler crawlt documenten in de normale werkstand.	0: de crawler crawlt documenten in de normale werkstand.	0: de crawler crawlt documenten in de normale werkstand.
Target/LastUpdatedTime	Niet van toepassing.	Niet van toepassing.	Niet van toepassing.
Target/LastResetTime	Niet van toepassing.	Niet van toepassing.	Niet van toepassing.

Tabel 25. Gedetailleerde statusgegevens van de crawlruimte voor de QuickPlace-, Domino Document Manager-, UNIX-bestandssysteem- en Windows-bestandssysteemcrawlers

Element- en kenmerknaam	QuickPlace-crawler	Domino Document Manager-crawler	UNIX- en Windows-bestandssysteemcrawlers
Target@Name	Naam van ruimtedatabase of roomdatabase	Naam van kabinetdatabase	Naam van subdirectory
Target@CrawlType	0	0	0
Target/Status	Status (0, 1, 2, 3, 4, -1) <ul style="list-style-type: none"> <li>• 0: Niet gecrawld</li> <li>• 1: bezig met crawlen</li> <li>• 2: Voltooid (niet gepland)</li> <li>• 3: Wachten</li> <li>• 4: Onderbroken</li> <li>• -1: Fout</li> </ul>	Status (0, 1, 2, 3, 4, -1) <ul style="list-style-type: none"> <li>• 0: Niet gecrawld</li> <li>• 1: bezig met crawlen</li> <li>• 2: Voltooid (niet gepland)</li> <li>• 3: Wachten</li> <li>• 4: Onderbroken</li> <li>• -1: Fout</li> </ul>	Status (0, 1, 2, 3, 4, -1) <ul style="list-style-type: none"> <li>• 0: Niet gecrawld</li> <li>• 1: bezig met crawlen</li> <li>• 2: Voltooid (niet gepland)</li> <li>• 3: Wachten</li> <li>• 4: Onderbroken</li> <li>• -1: Fout</li> </ul>
Target/StatusMessage	<ul style="list-style-type: none"> <li>• 0: Niet gecrawld</li> <li>• 1: bezig met crawlen</li> <li>• 2: Voltooid (niet gepland)</li> <li>• 3: Wachten</li> <li>• 4: Onderbroken</li> <li>• -1: Fout</li> </ul>	<ul style="list-style-type: none"> <li>• 0: Niet gecrawld</li> <li>• 1: bezig met crawlen</li> <li>• 2: Voltooid (niet gepland)</li> <li>• 3: Wachten</li> <li>• 4: Onderbroken</li> <li>• -1: Fout</li> </ul>	<ul style="list-style-type: none"> <li>• 0: Niet gecrawld</li> <li>• 1: bezig met crawlen</li> <li>• 2: Voltooid (niet gepland)</li> <li>• 3: Wachten</li> <li>• 4: Onderbroken</li> <li>• -1: Fout</li> </ul>
Target/NumberOfRecords	Niet van toepassing.	Niet van toepassing.	Niet van toepassing.

Tabel 25. Gedetailleerde statusgegevens van de crawlruimte voor de QuickPlace-, Domino Document Manager-, UNIX-bestandssysteem- en Windows-bestandssysteemcrawlers (vervolg)

Element- en kenmerknaam	QuickPlace-crawler	Domino Document Manager-crawler	UNIX- en Windows-bestandssysteemcrawlers
Target/NumberOf CompletedRecords	Het aantal gecrawlde documenten.	Het aantal gecrawlde documenten.	Het aantal gecrawlde bestanden.
Target/NumberOf InsertedRecords	Het aantal ingevoegde records.	Het aantal ingevoegde records.	Het aantal ingevoegde records.
Target/NumberOf UpdatedRecords	Het aantal bijgewerkte records.	Het aantal bijgewerkte records.	Het aantal bijgewerkte records.
Target/NumberOf DeletedRecords	Het aantal gewiste records.	Het aantal gewiste records.	Het aantal gewiste records.
Target/StartTime	De datum en tijd waarop de crawler voor het laatst is gestart.	De datum en tijd waarop de crawler voor het laatst is gestart.	De datum en tijd waarop de crawler voor het laatst is gestart.
Target/EndTime	De datum en tijd waarop het crawlen is voltooid.	De datum en tijd waarop het crawlen is voltooid.	De datum en tijd waarop het crawlen is voltooid.
Target/TotalTime	De hoeveelheid tijd die de crawler aan het crawlen heeft besteed.	De hoeveelheid tijd die de crawler aan het crawlen heeft besteed.	De hoeveelheid tijd die de crawler aan het crawlen heeft besteed.
Target/AggregationLevel	0: de crawler crawl't documenten in de normale werkstand.	0: de crawler crawl't documenten in de normale werkstand.	0: de crawler crawl't documenten in de normale werkstand.
Target/LastUpdatedTime	Niet van toepassing.	Niet van toepassing.	Niet van toepassing.
Target/LastResetTime	Niet van toepassing.	Niet van toepassing.	Niet van toepassing.

Tabel 26. Gedetailleerde statusgegevens van de crawlruimte voor de WebSphere Portal- en Web Content Management-crawlers

Element- en kenmerknaam	WebSphere Portal-crawler	Web Content Management-crawler
Target@Name	Naam van WebSphere Portal-server	De begin-URL voor zoekopdrachten die de site vertegenwoordigt
Target@CrawlType	0	0
Target/Status	Status: (0, 1, 2, 3, 4, -1) <ul style="list-style-type: none"> <li>• 0: Niet gecrawld</li> <li>• 1: bezig met crawlen</li> <li>• 2: Voltooid (niet gepland)</li> <li>• 3: Wachten</li> <li>• 4: Onderbroken</li> <li>• -1: Fout</li> </ul>	Status: (0, 1, 2, 3, 4, -1) <ul style="list-style-type: none"> <li>• 0: Niet gecrawld</li> <li>• 1: bezig met crawlen</li> <li>• 2: Voltooid (niet gepland)</li> <li>• 3: Wachten</li> <li>• 4: Onderbroken</li> <li>• -1: Fout</li> </ul>
Target/StatusMessage	<ul style="list-style-type: none"> <li>• 0: Niet gecrawld</li> <li>• 1: bezig met crawlen</li> <li>• 2: Voltooid (niet gepland)</li> <li>• 3: Wachten</li> <li>• 4: Onderbroken</li> <li>• -1: Fout</li> </ul>	<ul style="list-style-type: none"> <li>• 0: Niet gecrawld</li> <li>• 1: bezig met crawlen</li> <li>• 2: Voltooid (niet gepland)</li> <li>• 3: Wachten</li> <li>• 4: Onderbroken</li> <li>• -1: Fout</li> </ul>
Target/NumberOf Records	Niet van toepassing.	Niet van toepassing.
Target/NumberOf CompletedRecords	Het totaal aantal gecrawlde records.	Het totaal aantal gecrawlde records.
Target/NumberOf InsertedRecords	Het aantal ingevoegde records.	Het aantal ingevoegde records.

Tabel 26. Gedetailleerde statusgegevens van de crawlruimte voor de WebSphere Portal- en Web Content Management-crawlers (vervolg)

Element- en kenmerknaam	WebSphere Portal-crawler	Web Content Management-crawler
Target/NumberOf UpdatedRecords	Het aantal bijgewerkte records.	Het aantal bijgewerkte records.
Target/NumberOf DeletedRecords	Het aantal gewiste records.	Het aantal gewiste records.
Target/StartTime	De datum en tijd waarop de crawler voor het laatst is gestart.	De datum en tijd waarop de crawler voor het laatst is gestart.
Target/EndTime	De datum en tijd waarop het crawlen is voltooid.	De datum en tijd waarop het crawlen is voltooid.
Target/TotalTime	De hoeveelheid tijd die de crawler aan het crawlen heeft besteed.	De hoeveelheid tijd die de crawler aan het crawlen heeft besteed.
Target/AggregationLevel	0: de crawler crawlt documenten in de normale werkstand.	0: de crawler crawlt documenten in de normale werkstand.
Target/LastUpdatedTime	Niet van toepassing.	Niet van toepassing.
Target/LastResetTime	Niet van toepassing.	Niet van toepassing.

**Parserstatus:** als u de opdracht uitvoert om de parserstatus op te halen, worden de gegevens in een XML-documentindeling teruggezonden. De volgende gegevens kunnen met de opdracht voor de parserstatus worden teruggezonden:

```
FFQC5314I Resultaat:
<Monitor Type="Parser">
<ParserStatus>
  <Status>1</Status>
  <State>Parsing</State>
  <NumberOfDocsToBeIndexed>231974</NumberOfDocsToBeIndexed>
  <ParseRate>0</ParseRate>
  <ParseRateMBPerHour>0</ParseRateMBPerHour>
  <NumberOfCpmThreads>3</NumberOfCpmThreads>
  <ParserServiceSession>parserservice.1</ParserServiceSession>
</ParserStatus>
<CrawlerStatus>
  <Name>WEBCrawler1</Name>
  <Crawlerid>coll.WEB1.esadmin</Crawlerid>
  <Type>WEB</Type>
  <ParserStatus>1</ParserStatus>
  <NumberOfDocsAlreadyParsed>29</NumberOfDocsAlreadyParsed>
</CrawlerStatus>
<CrawlerStatus>
  <Name>Data Listener (server1)</Name>
  <Crawlerid>datalistener</Crawlerid>
  <Type>datalistener</Type>
  <ParserStatus>0</ParserStatus>
  <NumberOfDocsAlreadyParsed>0</NumberOfDocsAlreadyParsed>
</CrawlerStatus>
</Monitor>
```

In de volgende tabel worden de XML-elementen beschreven voor de gegevens die met de opdracht voor de parserstatus worden teruggezonden:

Tabel 27. Elementen voor de opdracht voor de parserstatus

Element	Beschrijving
Status	<ul style="list-style-type: none"> <li>• 0: de parsersessie voor deze collectie is gestopt.</li> <li>• 1: de parsersessie voor deze collectie is actief.</li> </ul>



Tabel 27. Elementen voor de opdracht voor de parserstatus (vervolg)

Element	Beschrijving
Status	<p>Mogelijke statuswaarden zijn: Bezig met initialiseren, Niet actief, Opnieuw starten, Bezig met analyseren, Gestopt, Onderbroken, Wordt hervat en GeenParserServiceBeschikbaar.</p> <p>De status Bezig met initialiseren houdt in dat de parser wordt gestart en bezig is met het initialiseren van de parserstatus.</p> <p>De status Niet actief houdt in dat de parser gedurende <math>N</math> minuten niet actief is en wacht op meer documenten van de crawlers in de collectie. De standaardwaarde is 300 seconden.</p> <p>De status Opnieuw starten houdt in dat de parser wacht totdat de JVM voor analyse/tokenisatie opnieuw is gestart. De JVM voor analyse/tokenisatie wordt in een aparte sessie uitgevoerd en op de JVM worden de documenten uiteindelijk verwerkt.</p> <p>Met de status Bezig met analyseren wordt aangegeven dat de documenten momenteel door de parser worden verwerkt.</p> <p>De status Onderbroken houdt in dat de parser is onderbroken door het opbouwproces van de index voor deze collectie.</p> <p>De status status Wordt hervat houdt in dat de status van de parser is gewijzigd van Onderbroken in Bezig met analyseren door de opbouwsessie van de index voor deze collectie.</p> <p>De status GeenParserServiceBeschikbaar houdt in dat er geen JVM's voor analyse/tokenisatie beschikbaar zijn voor de verwerking van de documenten voor deze collectie. Dit betekent dat alle JVM's voor analyse/tokenisatie door andere collecties worden gebruikt.</p>
NumberOfDocsToBeIndexed	Het aantal documenten in het archief voor deze collectie. Dit aantal bevat ook de documenten die zijn gemarkeerd om te worden gewist tijdens het volgende opbouwproces van de index.
ParseRate	De analysesnelheid in documenten per seconde.
ParseRateMBPerHour	De analysesnelheid MB per uur.
NumberOfCpmThreads	Het aantal CPM-threads dat door de JVM voor analyse/tokenisatie wordt gebruikt voor de verwerking van documenten voor deze collectie.
ParserServiceSession	De naam van de JVM voor analyse/tokenisatie waarmee de documenten voor deze collectie worden verwerkt. Dit veld is alleen beschikbaar als de status van de parser Bezig met analyseren is.
Naam	De naam van de crawler.
Crawlerid	Het ID dat voor deze crawler door het systeem is gemaakt.
Type	Het type (Web, NNTP, DB2, enzovoort).
ParserStatus	<ul style="list-style-type: none"> <li>• 0: documenten van deze crawler worden niet geanalyseerd (de parsersessie is gestopt).</li> <li>• 1: documenten van deze crawler worden geanalyseerd (de parsersessie is actief).</li> </ul>
NubmerOfDocsAlreadyParsed	Het aantal documenten van deze crawler dat is geanalyseerd.

**Status van opbouwproces index:** als u de opdracht uitvoert om de status van het opbouwproces van de index op te halen, worden de gegevens in een XML-documentindeling teruggezonden. De volgende gegevens kunnen met de opdracht voor de status van het opbouwproces voor de index worden teruggezonden:

```

<?xml version="1.0"?>
<Monitor Type="MainIndexHistory" Count="1">
  <IndexStatus Id="1">
    <StartTime>1131987633901</StartTime>
    <Progress>0</Progress>
    <CurrentPhase>0</CurrentPhase>
    <TotalPhase>3</TotalPhase>
    <IndexCopyTime>49822</IndexCopyTime>
    <CurrentServer>0</CurrentServer>
    <TotalServer>0</TotalServer>
    <IndexBuildTime>46158</IndexBuildTime>
    <Status>0</Status>
    <JobID>1131987633899</JobID>
    <MessagesAvailable>false</MessagesAvailable>
    <StopTime>1131987734199</StopTime>
    <TotalTime>100298</TotalTime>
    <NumberOfDocuments>43</NumberOfDocuments>
  </IndexStatus>
  <CurrentIndexWildcardSupport/>
  <NextIndexWildcardSupport Type="None" Limit="0"/>
  <ScheduleStatus>
    <Status>1</Status>
    <ScheduledTime Enabled="false"></ScheduledTime>
  </ScheduleStatus>
</Monitor>

```

In de volgende tabel worden de verschillende XML-elementen beschreven voor de gegevens die met de opdracht voor de status van het opbouwproces voor de index worden teruggezonden:

Tabel 28. Elementen voor de opdracht voor de status van het opbouwproces voor de index

Element	Beschrijving
IndexStatusId	Het indexstatus-ID.
StartTime	De tijd in seconden sinds 1970 toen het opbouwen van deze index is gestart. Om de huidige tijd te berekenen die met deze tijd wordt aangegeven, gebruikt u de formule January 1, 1970 %2B StartTime. Meer informatie over epoche-tijden vindt u op <a href="http://en.wikipedia.org/wiki/Unix_epoch">http://en.wikipedia.org/wiki/Unix_epoch</a> .
Voortgang	Het voltooiingspercentage voor het opbouwproces van deze index.
CurrentPhase	<ul style="list-style-type: none"> <li>• 1: fase van opnieuw schrijven naar archief</li> <li>• 2: fase van algemene analyse</li> <li>• 3: fase van opbouwen index</li> </ul>
TotalPhase	Het aantal fases voor het opbouwproces voor deze index. Deze waarde is momenteel 3.
IndexCopyProgress	Het voltooiingspercentage voor het kopieerproces van deze index. Met dit proces wordt de opgebouwde index van de server waarop het opbouwproces is uitgevoerd, naar de zoekservers gekopieerd.
CurrentServer	De zoekserver waarnaar de indexkopie wordt gekopieerd.
TotalServer	Het aantal zoekservers waarnaar de index wordt gekopieerd.
IndexCopyTime	De totale tijd voor het kopiëren van de index naar alle zoekservers.
IndexBuildTime	De totale tijd voor alle fases van het opbouwen van de index.
Status	<ul style="list-style-type: none"> <li>• 0: index opbouwen en kopiëren</li> <li>• -1: opdracht voor opbouwen van index is mislukt</li> <li>• 1: opbouwen of kopiëren van de index (of beide) is in uitvoering.</li> </ul>

Tabel 28. Elementen voor de opdracht voor de status van het opbouwproces voor de index (vervolg)

Element	Beschrijving
JobID	Een uniek ID dat us gekoppeld aan elke opdracht voor het opbouwen van een index.
MessagesAvailable	Een booleaanse waarde waarmee wordt aangegeven of er foutberichten beschikbaar zijn (als er fouten zijn opgetreden).
StopTime	De eindtijd voor het opbouwen van de index (alle fases) en voor het kopiëren van de index.
TotalTime	De periode tussen de starttijd en de stoptijd.
NumberOfDocuments	Het aantal documenten in de index.
CurrentIndexWildcardSupport	De jokertekeninstelling die voor het volgende opbouwproces voor de index moet worden gebruikt. Mogelijke waarden zijn None, QueryExpansion of IndexExpansion.
ScheduleStatus	<ul style="list-style-type: none"> <li>• 0 als voor deze collectie en dit indextype geen rooster is ingeschakeld.</li> <li>• 1 als voor deze collectie en dit indextype een rooster is ingeschakeld.</li> </ul>
ScheduledTimeEnabled	De tijd in seconden sinds 1970 waarop het volgende opbouwproces voor index voor deze collectie en dit indextype wordt uitgevoerd. Om de huidige tijd te berekenen die met deze tijd wordt aangegeven, gebruikt u de formule January 1, 1970 %2B ScheduledTimeEnabled. Meer informatie over epoche-tijden vindt u op <a href="http://en.wikipedia.org/wiki/Unix_epoch">http://en.wikipedia.org/wiki/Unix_epoch</a> .

**Status van zoekserver:** als u de opdracht uitvoert om de status van de zoekserver op te halen, worden de gegevens in een XML-documentindeling teruggezonden. De volgende gegevens kunnen met de opdracht voor de status van de zoekserver worden teruggezonden:

```
FFQC5314I Resultaat: <?xml version="1.0"?>
<Monitor Type="Search" Count="1">
<SearchStatus Name="Search Manager (node1)" SearchID=
"searchmanager.node1" HostName="myComputer.svl.ibm.com">
<Status>1</Status>
</SearchStatus>
</Monitor>
```

In de volgende tabel worden de XML-elementen beschreven voor de gegevens die met de opdracht voor de status van de zoekserver worden teruggezonden:

Tabel 29. Elementen voor de opdracht voor de status van de zoekserver

Element	Beschrijving
SearchStatusName	De naam en het ID van de zoekbeheersessie waarmee de zoekindex voor deze collectie wordt bewaakt en beheerd.
HostName	De hostnaam van de server waarop de zoekindex wordt uitgevoerd.
Status	<ul style="list-style-type: none"> <li>• 0 als de zoekindex voor deze collectie niet actief is.</li> <li>• 1 als de zoekindex voor deze collectie actief is.</li> </ul>

**Gedetailleerde status van zoekserver status:** met de opdracht voor de status van de zoekserver kunnen de volgende gegevens worden teruggezonden:

```
FFQC5303I Search Manager (node1) (sid: searchmanager.node1)
is al actief. PID: 15711
FFQC5314I Resultaat: PID=18390
CacheHits=3
QueryRate=1
Port=44008
```

```

SessionId=coll.runtime.node1
CacheHitRate=0.333
ResponseTime=70
Status=1
SessionName=coll.runtime.node1.1

```

In de volgende tabel worden de items beschreven in de gegevens die met de opdracht voor de gedetailleerde status van de zoekserver worden teruggezonden:

Tabel 30. Items voor de opdracht voor de gedetailleerde status van de zoekserver

Item	Beschrijving
CacheHits	Het aantal resultaten dat uit de zoekcache is opgehaald.
QueryRate	Het aantal query's dat in het laatste tijdsinterval is ontvangen. Standaard is het tijdsinterval vijf minuten.
Port	Het poortnummer dat door de zoekserver wordt gebruikt om te luisteren naar query's of om query's te ontvangen.
SessionId	Het sessie-ID voor de zoekindex voor deze collectie.
CacheHitRate	Het aantal resultaten dat uit de zoekcache is opgehaald, uitgedrukt als percentage van alle zoekresultaten.
ResponseTime	De gemiddelde responstijd in milliseconden voor het opgegeven tijdsinterval. (De standaardwaarde is vijf minuten.)
Status	<ul style="list-style-type: none"> <li>• 0 als de zoekindex voor deze collectie niet actief is.</li> <li>• 1 als de zoekindex voor deze collectie actief is.</li> </ul>
SessionName	De sessienaam voor de zoekindex voor deze collectie.

## Retourcodes voor esadmin-opdrachten

De volgende codes kunnen worden teruggezonden voor esadmin-opdrachten:

Tabel 31. Retourcodes voor esadmin-opdrachten

Code	Naam	Beschrijving
0	CODE_ERROR_NONE	De opdracht is correct voltooid.
102	CODE_ERROR_INSTANTIATION_EXCEPTION	Er is een fout opgetreden bij het inschakelen van een opdrachthandler.
103	CODE_ERROR_ACCESS_EXCEPTION	Er is een toegangsfout opgetreden bij het inschakelen van een opdrachthandler.
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	

## Sessie-ID's ophalen

Met de opdracht esadmin check kunt u een overzicht afbeelden met enterprise search-componenten en de bijbehorende sessie-ID's. In de volgende tabel vindt u een overzicht van algemene sessies, de bijbehorende ID's, de servers waarop de sessies worden uitgevoerd en de statuswaarden van de sessies.

Tabel 32. Voorbeelden van sessienamen, servers van herkomst, sessie-ID's en sessiestatuswaarden

Sessie	Server waarop de sessie wordt uitgevoerd	Sessie-ID	Sessiestatus
configmanager	indexserver	10433	Gestart
controller	indexserver	10464	Gestart
customcommunication	indexserver	Niet van toepassing	Niet van toepassing
datalistener	indexserver	10582	Gestart
discovery	indexserver	10649	Gestart
bewaken	indexserver	10682	Gestart
parserservice	indexserver	10718	Gestart
resource.node1	indexserver	10759	Gestart
samplecpp	indexserver	10827	Gestart
sampletest	indexserver	10857	Gestart
scheduler	indexserver	10889	Gestart
searchmanager.node1	indexserver	10927	Gestart
utilities.node1	indexserver	10384	Gestart

#### Verwante onderwerpen

“Backup en herstel van een enterprise search-systeem” op pagina 333  
Met scripts voor backup en herstel kunt u een backup van het enterprise search-systeem maken en het systeem later weer herstellen.

“Activiteit van enterprise search beheren” op pagina 299

Tijdens het bewaken van systeem- en collectieactiviteiten kunt u de status van verschillende processen bekijken, mogelijke problemen controleren of de configuratie-instellingen aanpassen om de prestaties te verbeteren.

#### Verwante taken

“Crawlers bewaken” op pagina 302

U kunt algemene gegevens bekijken over de status van elke crawler in een collectie of opties selecteren om gedetailleerde informatie te bekijken over de crawleractiviteiten.

“Een enterprise search-systeem starten” op pagina 291

Om in te stellen dat gebruikers een collectie kunnen doorzoeken, moet u de systeemprocessen starten en vervolgens de servers starten waarmee de collectie wordt gecrawld, geanalyseerd en geïndexeerd (de zoekservers worden automatisch gestart).

“sEen enterprise search-systeem stoppen” op pagina 294

Als u wijzigingen hebt aangebracht in een enterprise search-server of als u problemen moet oplossen, moet u de server mogelijk stoppen en opnieuw starten.

“Zoekservers in de stand-alone werkstand beheren” op pagina 296

Om te zorgen dat de zoekservers altijd beschikbaar zijn, kunt u de zoekservers voor afzonderlijke collecties starten zelfs als de indexserver niet actief is.



---

## Hoofdlettergevoeligheid in enterprise search

Hoofdletters en kleine letters worden in de verschillende onderdelen van enterprise search (zoals de syntaxis voor zoekopdrachten, snellinks, veldnamen, enzovoort) op verschillende manieren behandeld.

### Syntaxis voor zoekopdrachten

In zoekopdrachten wordt geen onderscheid gemaakt tussen hoofdletters en kleine letters, met uitzondering van de volgende gevallen:

#### XML-elementnamen en kenmerknamen

Hoofdlettergevoelig. Termen en kenmerkwaarden zijn hoofdlettergevoelig, zelfs in XML-query's. Bijvoorbeeld, in dit document:

```
<book>
  <Author>
    <Name>Ferdinand</Name>
    <Contact Type="eMail">ferdi@nand.org</Contact Type>
    <Contact Type="Phone">+1 408 876 4242</Contact Type>
  </Author>
</book>
```

Het document wordt niet als resultaat gegeven met de volgende query's:

- @xmlns::'author[Name ftcontains ("Ferdinand")]'
- @xmlns::'//contact[@type="eMail"]'
- @xmlf2::'<author><name>Ferdinand</name></author>'
- @xmlf2::'<CONTACT TYPE="email">ferdi</contact>'

Het document wordt wel als resultaat gegeven met de volgende query's:

- @xmlns::'Author[Name ftcontains ("ferdinand")]'
- @xmlns::'//Contact[@Type="email"]'
- @xmlf2::'<Author><Name>ferdinand</Author><Name>'
- @xmlf2::'<Contact Type="email">ferdi</Contact>'

#### Toeganglijsten (ACL's)

Hoofdlettergevoelig.

#### URL's in termen van het type docid: en samegroupas:

Hoofdlettergevoelig. URL-delen in termen van het type site: of url: zijn echter niet hoofdlettergevoelig. Bijvoorbeeld, in een document met URL <http://www.here.com/HR/>:

- Het document wordt niet als resultaat gegeven met de query docid:http://www.here.com/hr.
- Het document wordt als resultaat gegeven met de query's url:hr en url:HERE.
- Het document wordt als resultaat gegeven met de query site:HERE.com.

#### Categorie-ID's en taxonomie-ID's

Niet hoofdlettergevoelig. In de query taxonomie\_id::categorie\_id wordt geen onderscheid gemaakt tussen hoofdletters en kleine letters in taxonomie\_id en categorie\_id. De query RuleBased::c42 komt overeen met

de categorie c42 in de regelgebaseerde taxonomie en komt ook overeen met rulebased::C42. U kunt niet zoeken op categorienaam, maar u kunt wel zoeken op categorie-ID.

### **Scopes**

Niet hoofdlettergevoelig. Met zowel Scope:RESEARCH als scope::research worden documenten uit de scope Research als resultaat gegeven.

### **Jokertekens**

Niet hoofdlettergevoelig. De termen Fer\*n\*d en fer\*n\*d zijn gelijk.

### **Veldnamen**

Niet hoofdlettergevoelig. De query;s Title:Expenses, TITLE:expenses en title:expenses zijn gelijk. In veldnamen wordt geen onderscheid gemaakt tussen hoofdletters en kleine letters, zelfs niet als deze afkomstig zijn uit een XML-toewijzingsbestand. In externe bronnen kunnen veldnamen echter hoofdlettergevoelig zijn, volgens de bijbehorende querysemantiek.

### **Snellinks**

Niet hoofdlettergevoelig.

### **Regelgebaseerde categoriseringsmodule**

URL-regels zijn hoofdlettergevoelig, maar documentinhoudregels zijn niet hoofdlettergevoelig.

### **Samengevouwen URI's en wegingsdefinities op basis van URI-patronen**

URI's zijn hoofdlettergevoelig, maar samengevouwen URI-groepsnamen zijn niet hoofdlettergevoelig. U kunt niet twee samengevouwen URI-groepsnamen definiëren waarvan alleen het hoofdlettergebruik verschilt.

### **Veldnamen**

Niet hoofdlettergevoelig. Deze regel is van toepassing op veldnamen in query's, XML-toewijzingen, definities van wegingsklassen en alle andere beheerinterfaces waarvoor veldnamen worden opgegeven. Als u in de interface een veldnaam definieert met hoofdletters of met hoofdletters en kleine letters, worden de letters door het systeem mogelijk omgezet in kleine letters zodat de veldnamen in kleine letters worden weergegeven wanneer u de configuratie de volgende keer bekijkt. Daarnaast kan een veldnaam op basis van het onderscheid tussen hoofdletters en kleine letters door een externe bron worden geïnterpreteerd.

### **Woordenboeken**

Niet hoofdlettergevoelig. Deze regel is van toepassing op synoniemenwoordenboeken, stopwoordenboeken, woordenboeken met spelling-suggesties en gewogen woordenboeken.

### **Collectienamen en -ID's**

Hoofdlettergevoelig. Als u een collectienaam of een collectie-ID in beheer- of zoek-API opgeeft, moet de naam of het ID exact overeenkomen met het hoofdlettergebruik van de collectie met die naam of dat ID. Ook al zijn collectie-ID's hoofdlettergevoelig, u kunt niet twee collectie-ID's opgeven waarvan alleen het hoofdlettergebruik verschilt. Dezelfde beperking geldt voor de namen en ID's van crawlers en gegevensbronnen.

### **Zoekprogramma's**

Hoofdlettergevoelig. U kunt niet twee programmanamen of -ID's opgeven waarvan alleen het hoofdlettergebruik verschilt.



---

## Documentatie bij enterprise search

U kunt de documentatie bij OmniFind Enterprise Edition als PDF- of HTML-document lezen.

Het installatieprogramma van OmniFind Enterprise Edition installeert het Informatiecentrum van IBM Content Discovery automatisch. Dit omvat tevens HTML-versies van de documentatie van OmniFind Enterprise Edition, Versie 8.4 en WebSphere Information Integrator Content Edition Versie 8.3. Als u met meerdere servers werkt, wordt het Informatiecentrum op alle zoekservers geïnstalleerd. Als u het Informatiecentrum niet hebt geïnstalleerd en op Help klikt, wordt het Informatiecentrum geopend op de website van IBM.

Als u de geïnstalleerde versie van de PDF-documenten wilt bekijken, gaat u naar `ES_INSTALL_ROOT/docs/locale/pdf`. Als u de documenten bijvoorbeeld in het Engels wilt bekijken, gaat u naar `ES_INSTALL_ROOT/docs/en_US/pdf`.

Om de PDF-versie van de documentatie in alle beschikbare talen te zien, gaat u naar de website met de documentatie van OmniFind Enterprise Edition Versie 8.4.

Op de website OmniFind Enterprise Edition Support hebt u ook toegang tot downloads, fixpacks, technotes en het Informatiecentrum.

In de volgende tabel ziet u de beschikbare documentatie, bestandsnamen en locaties.

Tabel 33. Documentatie voor enterprise search

Title	Bestandsnaam	Locatie
Informatiecentrum		<a href="http://publib.boulder.ibm.com/infocenter/discover/v8r4/">http://publib.boulder.ibm.com/infocenter/discover/v8r4/</a>
<i>Installation Guide for Enterprise Search</i>	iiysi.pdf	<code>ES_INSTALL_ROOT/docs/locale/pdf/</code>
<i>Quick Start Guide</i> (Dit document is tevens beschikbaar als gedrukt boekje in het Engels, Frans en Japans.)	QuickStartGuide_ <i>twee letters van taal</i> .pdf	<code>ES_INSTALL_ROOT/docs/locale/pdf/</code>
<i>Installation Requirements for Enterprise Search</i>	iiysr.txt of iiysr.htm	<code>ES_INSTALL_ROOT/docs/locale/</code> (u kunt ook naar dit bestand gaan via het startvenster (launchpad) voor de installatie)
<i>Enterprise Search beheren</i>	iiysa.pdf	<code>ES_INSTALL_ROOT/docs/locale/pdf/</code>
<i>Programming Guide and API Reference for Enterprise Search</i>	iiysp.pdf	<code>ES_INSTALL_ROOT/docs/en_US/pdf/</code>
<i>Troubleshooting Guide and Messages Reference</i>	iiysm.pdf	<code>ES_INSTALL_ROOT/docs/locale/pdf/</code>
<i>Integratie van tekstanalyse</i>	iiyst.pdf	<code>ES_INSTALL_ROOT/docs/locale/pdf/</code>
<i>Plugin voor Google Desktop Search</i>	iiysg.pdf	<code>ES_INSTALL_ROOT/docs/locale/pdf/</code>

*Tabel 33. Documentatie voor enterprise search (vervolg)*

<b>Title</b>	<b>Bestandsnaam</b>	<b>Locatie</b>
<i>Release Notes</i>	iiysn.pdf	Beschikbaar op de website met documentatie voor OmniFind Enterprise Edition, Versie 8.4 (u kunt ook naar dit bestand gaan via het startvenster (launchpad) voor de installatie)

---

## Toegankelijkheidsfuncties in WebSphere Information Integrator OmniFind Edition

De gebruikersinterfaces en documentatie voor IBM WebSphere Information Integrator OmniFind Edition zijn toegankelijk.

### Installatieprogramma

U kunt gebruikmaken van sneltoetsen om door het installatieprogramma van WebSphere Information Integrator OmniFind Edition te navigeren. In de volgende tabel worden enkele sneltoetsen beschreven.

*Tabel 34. Sneltoetsen voor het installatieprogramma*

Actie	Sneltoets
Een keuzerondje markeren	Cursortoets
Een keuzerondje selecteren	Tabtoets
Een opdrachtknop markeren	Tabtoets
Een opdrachtknop selecteren	Enter-toets
Naar het volgende of vorige venster gaan of annuleren	Een opdrachtknop markeren met de tabtoets en vervolgens op Enter drukken
Het actieve venster deactiveren	Ctrl + Alt + Esc

### Beheerconsole en Informatiecentrum van enterprise search

De beheerconsole en het Informatiecentrum bevatten een op browsers gebaseerde interface die u kunt bekijken in Microsoft Internet Explorer of Mozilla FireFox. Raadpleeg de online Help bij Internet Explorer of FireFox voor een overzicht van sneltoetsen en andere toegankelijkheidsopties voor uw browser.

### documentatie in PDF-indeling

U kunt alle documentatie bij enterprise search bekijken in PDF-indeling. De PDF-documenten zijn toegankelijk via Adobe Acrobat versie 6.0. De PDF-documenten zijn gestructureerd en kunnen door de meeste schermlezers worden gelezen.



---

## Woordenlijst met termen voor enterprise search

In deze woordenlijst vindt u de termen die in de interfaces en documentatie van enterprise search worden gebruikt.

### **Portal Document Manager (PDM)**

Biedt gebruikers de mogelijkheid om met één centrale documentenrepository te werken, ten behoeve van samenwerking binnen een team. Beheerders hebben de mogelijkheid hun documenten effectief te beheren en kunnen invloed uitoefenen op de manier waarop gebruikers interactief kunnen werken met informatie.

### **aangepast tekstanalyseprogramma**

Een tekstanalyseprogramma dat is gemaakt met behulp van de Unstructured Information Management Architecture (UIMA) software development kit (SDK) en dat kan worden toegevoegd aan de set van standaard tekstanalyseprogramma's van enterprise search (ook bekend als de basisannotators van enterprise search). Zie ook tekstanalyseprogramma.

### **Afsluitend teken**

Een teken dat de laatste positie in een woord aangeeft.

### **Analyseprogramma**

Zie Tekstanalyseprogramma.

### **Analyseresultaten**

De gegevens die door de annotators worden geproduceerd. Analyse-resultaten worden geschreven in een gegevensstructuur die Common Analysis Structure wordt genoemd. De resultaten van analyse die is uitgevoerd door tekstanalyseprogramma's op maat (annotators) kunnen beschikbaar worden gesteld voor zoeken door ze op te nemen in de enterprise search-index.

### **Annotator**

Een softwarecomponent waarmee specifieke taalkundige analysetaken worden uitgevoerd en annotaties worden geproduceerd en vastgelegd. Een annotator is de analyselogica-component in een analyseprogramma.

### **Begripsextractie**

Een tekstanalysefunctie waarmee significante vocabulaire-items (zoals personen, plaatsen of producten) in tekstdocumenten worden geïdentificeerd, waarna een lijst met deze items als resultaat wordt gegeven. Zie ook Thema-extractie.

### **beheerdersrol**

Een classificatie van een gebruiker waarmee wordt bepaald welke functies de gebruiker kan uitvoeren in de beheerconsole van enterprise search. Daarnaast wordt met de rol bepaald welke collecties de gebruiker kan beheren.

### **Beveiligingstoken**

Informatie over de identiteit en beveiliging die wordt gebruikt om toegang te verlenen tot de documenten in een collectie. Verschillende typen gegevensbronnen bieden ondersteuning voor verschillende typen beveiligingstokens. Voorbeelden zijn gebruikersrollen, gebruiker-ID's en andere informatie die kan worden gebruikt om de toegang tot bepaalde inhoud te bepalen.

**bewaken**

Een enterprise search-gebruiker die de machtiging heeft voor het bewaken van processen op collectieniveau.

**Bibliotheek**

Een systeemobject dat dienst doet als directory voor andere objecten. Zie ook Domino Document Manager-bibliotheek.

**Booleaanse zoekopdracht**

Een zoekopdracht waarin een of meer zoektermen worden gecombineerd door gebruik te maken van operatoren zoals EN, NIET en OF.

**bouwen van delta-index**

Proces waarbij nieuwe gegevens aan een bestaande index in een enterprise search-systeem worden toegevoegd. Tegenstelling met bouwen van hoofd-index.

**bouwen van hoofdindex**

Proces waarbij de hele index in een enterprise search-systeem wordt samengesteld. Tegenstelling met bouwen van delta-dindex.

**CAS-consumer**

Een CAS-consumer (CAS = Common Analysis Structure) voert de laatste verwerkingsstappen uit op de analyseresultaten die zijn opgeslagen in de Common Analysis Structure. Een consumer kan bijvoorbeeld de inhoud van de Common Analysis Structure in een zoekmachine indexeren of een relationele database vullen met specifieke analyseresultaten.

**Categorie**

Een groep documenten met vergelijkbare eigenschappen.

**Categoriestructuur**

Een hiërarchie met categorieën die in de beheerconsole van enterprise search wordt weergegeven.

**Certificaat**

Een digitaal document waarmee een openbare sleutel wordt verbonden aan de identiteit van de eigenaar van het certificaat, zodat de eigenaar kan worden geverifieerd. Een certificaat wordt uitgegeven door een certificaatgever.

**Certificaatgever**

Een organisatie die certificaten uitgeeft en de entiteiten (individuen of organisaties) verifieert die betrokken zijn bij elektronische transacties. Certificaatgevers garanderen dat de twee partijen die gegevens uitwisselen, daadwerkelijk zijn wie ze claimen te zijn.

**Cliticum**

Een woord dat syntactisch gezien zelfstandig functioneert, maar fonetisch gezien verbonden is met een ander woord. Een cliticum kan verbonden zijn met het woord of zelfstandig worden geschreven van het woord waaraan het is verbonden. Veelvoorkomende voorbeelden van een cliticum zijn het laatste deel van een samentrekking in het Engels (*wouldn't* of *you're*).

**Collectie**

Een set gegevensbronnen en opties voor het crawlen, analyseren, indexeren en doorzoeken van deze gegevensbronnen.

**Common Analysis Structure (CAS)**

Een structuur waarin de content en metagegevens van een document worden opgeslagen, alsmede alle analyseresultaten die worden geproduceerd

door een tekstanalyseprogramma. Alle gegevensuitwisseling tijdens de analyse van documenten vindt plaats met behulp van de Common Analysis Structure.

#### **common communication layer (CCL)**

De communicatie-infrastructuur waarin de verschillende componenten (controller, parser, crawler, indexserver) van WebSphere Information Integrator OmniFind Edition worden samengevoegd.

#### **Crawler**

Een softwareprogramma waarmee documenten uit gegevensbronnen worden opgehaald en gegevens worden verzameld die kunnen worden gebruikt voor het maken van zoekindexen.

#### **Crawlruimte**

Een set met bronnen waarvan de opgegeven patronen overeenkomen (zoals URL's, databasenames, bestandssysteempaden, domeinnamen en IP-adressen) en waaruit de crawler gegevens leest om items op te halen voor indexeringsdoeleinden.

#### **datastore**

Een gegevensstructuur waarin de documenten in geanalyseerde vorm worden bewaard. De parser schrijft gegevens naar de datastore. De datastore wordt gebruikt voor het bouwen van de index en voor het genereren van samenvattingen. Niet te verwarren met de raw data store.

#### **Diakritische tekens**

Een teken dat aan een letter wordt toegevoegd om de uitspraak van een woord te wijzigen of om onderscheid te maken tussen vergelijkbare woorden, zoals een accentteken of de umlaut.

#### **DN-naam**

De unieke naam waarmee een vermelding in een woordenboek wordt aangegeven. Een DN-naam bestaat uit kenmerk:waardeparen, gescheiden door komma's. Daarnaast is de DN-naam een unieke set met naam/waardeparen (zoals CN=naam van persoon en C=land of regio) waarmee een entiteit in een digitaal certificaat wordt aangegeven.

#### **Documentobjectmodel**

Een systeem waarin een gestructureerd document, zoals een XML-bestand, wordt weergegeven als een structuur met objecten die via een programma kan worden geopend en bijgewerkt.

#### **Domino Document Manager-bibliotheek**

Een Domino Document Manager-database die het ingangspunt is voor Domino Document Manager.

#### **Domino Document Manager-kabinet**

Een Domino Document Manager-database die wordt gebruikt om documenten te organiseren. De Domino-databases worden opgeslagen in kabinetten.

#### **Domino Internet Inter-ORB Protocol (DIIOP)**

Een servertaak die op de server wordt uitgevoerd en samenwerkt met de Domino-objectaanvraagbroker om communicatie toe te staan tussen Java-applets die zijn gemaakt met de Notes Java-classes en de Domino-server. Gebruikers van browsers en Domino-servers maken gebruik van DIIOP voor de communicatie en voor het uitwisselen van objectgegevens.

#### **dynamische ranking**

Een type ranking waarin de termen in de query worden geanalyseerd op

basis van de documenten die worden doorzocht om de rang van de resultaten te bepalen. Zie ook Op tekst gebaseerde scores. Vergelijk met Statistische ranking.

#### **Dynamisch samenvatten**

Een manier van samenvatten waarbij de zoektermen worden gemarkeerd en de zoekresultaten zinnen bevatten die het meest overeenkomen met de begrippen in het document die de gebruiker zoekt. Vergelijk met Statisch samenvatten.

#### **enterprise search basisannotators**

Een set standaard tekstanalyseprogramma's die in enterprise search wordt gebruikt voor standaardanalyse van documenten.

#### **enterprise search-beheerder**

Een beheerdersrol waarmee een gebruiker het volledige enterprise search-systeem kan beheren.

#### **Escapeteken**

Een teken waarmee een speciale betekenis voor een of meer volgende tekens wordt onderdrukt of geselecteerd.

#### **expressieannotator**

De expressieannotator detecteert entiteiten of eenheden informatie in een tekstdocument, bijvoorbeeld telefoonnummers, productnummers, namen van werknemers, of adressen, op basis van een expressie die het exacte patroon beschrijft dat in de tekst van documenten moet worden gezocht. Als een van de expressies een match oplevert in de tekst van een document, brengt de expressieannotator de dienovereenkomstige annotaties die de match (of een deel daarvan) beslaan. Deze geannoteerde expressies worden dan opgeslagen, hetzij in de enterprise search-index (met behulp van een indextoewijzingsbestand), hetzij in een JDBC-database (met behulp van een databasetoewijzingsbestand).

#### **Externe gegevensbron**

Een federatieve gegevensbron die niet wordt gecrawld, geanalyseerd of geïndexeerd door WebSphere Information Integrator OmniFind Edition. Zoekopdrachten voor externe gegevensbronnen worden overgedragen aan de query-API van deze gegevensbronnen.

#### **Featurepad**

Een pad dat wordt gebruikt om toegang te krijgen tot de waarde van een feature in een UIMA-featurestructuur (UIMA = Unstructured Information Management Architecture).

#### **Featurestructuur**

De onderliggende gegevensstructuur waarin de resultaten van een tekstanalyse worden aangegeven. Een featurestructuur is een structuur met kenmerken en de bijbehorende waarden. Elke featurestructuur heeft een bepaald type en elk type heeft een opgegeven set met geldige features of kenmerken, vergelijkbaar met een Java-klasse.

#### **Federatie**

Een proces waarin de naamgevingssystemen worden gecombineerd, zodat in het samengevoegde systeem samengestelde namen kunnen worden verwerkt die in de verschillende naamgevingssystemen worden gebruikt.

#### **Federatieve zoekopdracht**

Een zoekoptie waarmee kan worden gezocht in verschillende zoekservices en waarmee een geconsolideerde lijst met zoekresultaten als resultaat wordt gegeven.



**Gebruikersagent**

Een programma waarmee op internet wordt gezocht en informatie over de agent zelf wordt achtergelaten op de bezochte websites. In enterprise search is de webcrawler een gebruikersagent.

**Geen index (instructie)**

Een instructie op een webpagina waarmee robots (zoals de webcrawler) worden geïnstrueerd de inhoud van deze pagina's niet in de index op te nemen.

**Gegevensbron**

Een opbergplaats voor gegevens waaruit documenten kunnen worden opgehaald, zoals het web, relationele en niet-relationele databases en contentbeheersystemen.

**Gegevensbrontype**

Een groepering gegevensbronnen volgens het protocol dat wordt gebruikt om toegang tot de gegevens te krijgen.

**Gegevensextractie**

Een type begripsextractie waarmee significante vocabulaire-items automatisch worden herkend in tekstdocumenten, zoals namen, termen en uitdrukkingen.

**Gewogen woord**

Een woord dat van invloed kan zijn op de relatieve ranking van een document in de zoekresultaten. Tijdens de queryverwerking kan het belang van een document waarin een gewogen woord voorkomt, worden vergroot of verkleind afhankelijk van de score die vooraf is gedefinieerd voor het woord.

**Het vinden van de stam**

Zie Het vinden van de woordstam.

**Het vinden van de woordstam**

Een proces van taalkundige normalisatie waarin de verschillende vormen van een woord worden gereduceerd tot een algemene vorm. Woorden als *verbindingen*, *verbindend* en *verbonden* worden bijvoorbeeld gereduceerd tot *verbind*.

**Hybride zoekopdracht**

Een combinatie van een booleaanse zoekopdracht en een vrije zoekopdracht.

**Identiteitenbeheer**

De mogelijkheid om de actuele legitimatiegegevens van een gebruiker te controleren met behulp van native toegangsbesturing. Als een gegevensbron wordt beschermd middels een product dat enkelvoudige aanmelding (single sign-on, SSO) ondersteunt en de crawler geconfigureerd is voor het gebruiken van SSO-beveiliging, worden er SSO-mechanismen gebruikt om de identiteit van de gebruiker te controleren. Is dat niet het geval, dan worden de legitimatiegegevens in versleutelde vorm vastgelegd in een veilige kluis die kan worden bijgewerkt wanneer de native toegangsbesturing verandert.

**Index** Zie Volledige tekstindex.

**indexeerwachtrij**

Een lijst van opdrachten voor het bouwen van hoofd- en delta-indexen.

**In wachtrij plaatsen**

Items in een wachtrij plaatsen.

**IP-adres**

Het unieke 32-bits adres waarmee een host in het netwerk wordt aangegeven.

**Java Database Connectivity (JDBC)**

Een industriestandaard voor databaseonafhankelijke connectiviteit tussen het Java-platform en een groot aantal databases. De JDBC-interface biedt een API op oproepniveau voor op SQL gebaseerde databasetoegang.

**JavaScript**

Een webscripttaal die in browsers en webservern wordt gebruikt.

**JavaServer Pages (JSP)**

Een serverscriptingtechnologie waarmee Java-code dynamisch in webpagina's (HTML-bestanden) wordt ingesloten en wordt uitgevoerd wanneer de pagina wordt verzonden, waarna dynamische inhoud aan een client als resultaat kan worden gegeven.

**Java virtual machine (JVM)**

Een software-implementatie van een processor waarmee gecompileerde Java-code wordt uitgevoerd (applets en toepassingen).

**Jokerteken**

Een teken dat wordt gebruikt om optionele tekens aan te geven aan het begin, in het midden of aan het einde van een zoekterm.

**Katakana**

Een tekenset die bestaat uit symbolen die worden gebruikt in een van de twee veelvoorkomende Japanse fonetische alfabetten en primair wordt gebruikt om buitenlandse woorden fonetisch te schrijven.

**Koppelingsanalyse**

Een methode die is gebaseerd op de analyse van hyperlinks tussen documenten en wordt gebruikt om te bepalen welke pagina's in de collectie belangrijk zijn voor gebruikers.

**Legitimatiegegevens**

Gedetailleerde informatie die tijdens de verificatie wordt verkregen en waarmee de gebruiker, groeps koppelingen en andere beveiligingsgerelateerde identiteitskenmerken worden beschreven. Met legitimatiegegevens kunnen verschillende services worden uitgevoerd, zoals autorisaties, controles en overdrachtstaken.

**lemma**

De basisvorm van een woord. Lemma's zijn van groot belang in talen waarin veel woordvervoegingen voorkomen, zoals het Tsjechisch.

**Lemmatisering**

Proces waarbij het lemma voor een bepaald woord in een woordenboek wordt gezocht. Het verschil tussen lemmatisering en het vinden van de stam is dat het vinden van de stam op algoritmische basis plaatsvindt en in het algemeen geen gebruikmaakt van een woordenboek met de woorden van een taal.

**Lexicale verwantschap**

De relatie tussen zoekwoorden in een document die qua betekenis dicht bij elkaar staan. Lexicale verwantschap wordt gebruikt om de relevantie van een resultaat te berekenen.

**Ligatuur**

Twee of meer tekens die verbonden zijn zodat ze als één teken worden weergegeven (bijvoorbeeld: de a en e vormen de ligatuur æ).

**Lightweight Directory Access Protocol (LDAP)**

Een open protocol waarvoor TCP/IP wordt gebruikt om toegang te verlenen tot directory's die ondersteuning bieden voor een X.500-model en waarvoor de bronvereisten van de complexere X.500-DAP (Directory Access Protocol) niet nodig zijn.

**Lokale federator**

Een clientfederator voor een set doorzoekbare objecten.

**Lotus QuickPlace-place**

Een webvenue die wordt geleverd door Lotus QuickPlace waarmee deelnemers die op verschillende plekken werken, kunnen samenwerken aan projecten en online kunnen communiceren in een gestructureerd en beveiligd werkgebied.

**Lotus QuickPlace-room**

Een gepartitioneerd gebied van een Lotus QuickPlace-place dat beperkt toegankelijk is voor bevoegde leden die een gemeenschappelijke interesse delen en moeten samenwerken.

**Maskeringsteken**

Een teken dat wordt gebruikt om optionele tekens aan te geven aan het begin, in het midden en aan het einde van een zoekterm. Maskeringstekens worden meestal gebruikt om variaties te vinden voor een term in een index. Zie ook Jokerteken.

**MIME-type**

Een internetstandaard voor de identificatie van het type object dat via internet wordt overgedragen.

**Nabijheidszoekopdracht**

Een type zoekopdracht waarin wordt gezocht naar bepaalde woorden in dezelfde zin, dezelfde alinea of hetzelfde document.

**Natuurlijke-taalquery**

Een type zoekopdracht waarmee geschreven expressies (zoals "Wie is verantwoordelijk voor de financiële transacties?") worden geanalyseerd, in plaats van een eenvoudige collectie trefwoorden.

**n-gramsegmentering**

Een analysemethode waarin overlappende reeksen van een bepaald aantal tekens als één woord worden beschouwd in plaats van wtruimten te gebruiken om woorden te scheiden (zoals in op Unicode gebaseerde wtruimtesegmentering).

**Niet-lokale federator**

Een serverfederator voor een set doorzoekbare objecten.

**Niet volgen (instructie)**

Een instructie op een webpagina waarmee robots (zoals de webcrawler) worden geïnstrueerd om de koppelingen op deze pagina's niet te volgen.

**Notes Remote Procedure Call (NRPC)**

De communicatiemechanisme van Lotus Notes dat wordt gebruikt voor alle Notes-naar-Notes-communicatie.

**Ontdekker**

Een functie van een crawler waarmee wordt bepaald uit welke gegevensbronnen de crawler gegevens kan ophalen.

**operator**

Een enterprise search-gebruiker die is gemachtigd voor het bewaken, starten en beëindigen van processen op collectieniveau.

**Op regels gebaseerde categorie**

Categorieën die zijn gemaakt op basis van regels waarmee wordt aangegeven welke documenten aan welke categorieën zijn gekoppeld. U kunt bijvoorbeeld regels definiëren om documenten te koppelen die bepaalde woorden bevatten of die voldoen aan een URI-patroon, met specifieke categorieën.

**op tekst gebaseerde score**

Proces waarbij een geheel getal aan een document wordt toegewezen en waarmee de relevantie van het document wordt aangegeven ten opzichte van de termen in een query. Een hoger getal geeft een grotere overeenkomst met de query aan. Zie ook Dynamische ranking.

**Op Unicode gebaseerde witruimtesegmentering**

Een tokenisatiemethode waarvoor Unicode-tekeneigenschappen worden gebruikt om een onderscheid te maken tussen tokens en scheidingstekens.

**parametrische zoekopdracht**

Een type zoekopdracht waarmee wordt gezocht naar objecten die een numerieke waarde of kenmerk bevatten, zoals een datum of geheel getal, of andere numerieke gegevenssoorten binnen een opgegeven reeks.

**Parser** Een programma waarin documenten worden geïnterpreteerd die aan de enterprise search-opslagplek zijn toegevoegd. Met de parser worden gegevens uit de documenten geëxtraheerd, waarna de documenten worden voorbereid voor indexerings-, zoek- en ophaaltaken.

**parserdriver**

Een enterprise search-service die de parserservice voorziet van documenten. Voor elke collectie is er één parserdriver. De parserdriverservice van een collectie correspondeert met de parser van de collectie in de beheerconsole van enterprise search.

**parserservice**

De enterprise search-service die alle documentanalyse (parsing) en tekstanalyse van documentcollecties afhandelt. Er is te allen tijde minimaal één parserservice actief.

**PEAR-archief (Processing Engine ARchive)**

Een .pear-archiefbestand (zip) dat een UIMA-analyseprogramma en alle bronnen bevat die nodig zijn voor aangepaste analyses in enterprise search.

**Populariteitsranking**

Een type ranking dat wordt toegevoegd aan de bestaande ranking van een document, op basis van de populariteit van het document.

**Proxyserver**

Een server die dienst doet als tussenliggend apparaat voor HTTP-webaanvragen van een programma of een webserver. Een proxyserver werkt als vervanger voor de inhoudsservers in het bedrijf.

**Ranking**

Proces waarin aan elk document dat als resultaat van een query wordt gegeven, een geheel getal wordt toegewezen. De volgorde van de documenten in de zoekresultaten is gebaseerd op de relevantie van de query. Een groter getal geeft een nauwkeurigere overeenkomst aan. Zie ook Dynamische ranking en Statische ranking.

**raw data store**

Een gegevensstructuur waarin gecrawlde documenten worden opgeslagen voordat ze naar de parser worden gezonden. Crawlers schrijven gegevens naar de raw data store en de parser leest de gegevens uit de raw data store. Nadat documenten door de parser zijn geanalyseerd, worden ze uit de raw data store verwijderd. Niet te verwarren met datastore.

**Robots Exclusion Protocol**

Een protocol waarmee websitebeheerders kunnen aangeven welke gedeelten van de site niet toegankelijk zijn voor de robot.

**Room** Een programma waarmee gebruikers documenten kunnen maken voor andere gebruikers, waarmee gebruikers kunnen reageren op de opmerkingen van andere gebruikers en de projectstatus en deadlines kunnen bekijken. Gebruikers kunnen ook met andere gebruikers in dezelfde room chatten. Zie ook Lotus QuickPlace-room.

**Ruimte**

Een virtuele locatie in de portal waar individuen en groepen elkaar ontmoeten zodat ze kunnen samenwerken. In een portal heeft elke gebruiker een persoonlijke ruimte voor eigen werk en hebben de individuen en groepen toegang tot verschillende gemeenschappelijke ruimten (openbare of beperkt toegankelijke ruimten). Zie ook Lotus QuickPlace-place.

**Samenvatten**

Proces waarbij zinnen in zoekresultaten worden opgenomen, zodat de inhoud van een document kort wordt beschreven. Zie ook Dynamisch samenvatten en Statisch samenvatten.

**Scope** Een groep gerelateerde URI's die worden gebruikt om de reikwijdte van een zoekopdracht te definiëren.

**Secure Sockets Layer (SSL)**

Een beveiligingsprotocol waarmee privacy tijdens communicatie wordt geleverd.

**seedlistpagina**

In WebSphere Portal: een XML-pagina die links bevat naar pagina's die beschikbaar zijn in de portal. Crawlers gebruiken de seedlist om erachter te komen welke documenten kunnen worden gecrawld. De seedlist-pagina bevat ook metagegevens die zijn opgeslagen bij de gecrawlde documenten in de enterprise search-index.

**Segmentering**

De verdeling van tekst in afzonderlijke lexicale eenheden. Niet op woordenboeken gebaseerde verwerking omvat witruimte- en n-gram-segmentering, en op woordenboeken gebaseerde segmentering omvat woord-, zin- en alineasegmentering, en lemmatisering.

**semantische zoekopdracht**

Semantische zoekopdrachten vormen een uitbreiding van zoekopdrachten met sleutelwoorden, in die zin dat ze meer kennis over taalkunde en de zoekoplossing omvatten. De technologie die deze kennis omvat en toepast, staat bekend als tekstanalyse.

**Servlet**

Een Java-programma dat op een webserver wordt uitgevoerd en de functionaliteit van de server uitbreidt door dynamische inhoud te genereren als antwoord op opdrachten van webclients. Servlets worden veel gebruikt voor databaseverbindingen met het web.

**shingle**

Een reeks opeenvolgende tokens (woorden) die uit een zin worden gehaald. Bijvoorbeeld: in "Dit is een heel kort zinnetje", komen de volgende 3-woords shingles (of trigrammen) voor:

Dit is een  
is een heel  
een heel kort  
heel kort zinnetje

Shingles kunnen worden gebruikt voor statistische taalkunde. Als in twee verschillende teksten bijvoorbeeld zeer veel dezelfde shingles voorkomen, bestaat er waarschijnlijk een bepaald verband tussen die twee teksten.

**Sleutelruimtebestand**

Een databasebestand dat de openbare sleutels bevat die zijn opgeslagen als certificaten van de ondertekenaar en persoonlijke sleutels die zijn opgeslagen in persoonlijke certificaten.

**snellink**

Een koppeling tussen een URI en trefwoorden of termen.

**Soft error-pagina**

Een speciale pagina met een gedetailleerde beschrijving van het probleem die wordt weergegeven als een HTTP-server de pagina die door een client is aangevraagd, niet kan verzenden en waarmee de HTTP-server zodanig wordt geconfigureerd dat deze speciale pagina wordt verzonden (in plaats van een antwoord met alleen een code waarmee wordt aangegeven wat het probleem is).

**start-URL**

Het startpunt voor het crawlen.

**statische ranking**

Een type ranking waarmee de rang wordt verhoogd aan de hand van de factoren voor de documenten waarvoor de ranking wordt uitgevoerd, zoals de datum, het aantal koppelingen dat naar het document verwijzen, enzovoort. Vergelijk met Dynamische ranking.

**Statisch samenvatten**

Een manier van samenvatten waarbij de opgegeven, opgeslagen samenvatting van het document in de zoekresultaten wordt weergegeven. Vergelijk met Dynamisch samenvatten.

**Stopwoord**

Een woord dat veel wordt gebruikt, zoals *de*, *een* of *en*, en dat in het zoekprogramma wordt genegeerd.

**Stopwoorden verwijderen**

Proces waarbij stopwoorden uit de query worden verwijderd zodat veelgebruikte woorden worden genegeerd en relevantere resultaten worden weergegeven.

**Synoniemenwoordenboek**

Een woordenboek waarin gebruikers kunnen zoeken naar synoniemen van de query tijdens het doorzoeken van een collectie.

**Taalidentificatie**

Een functie in enterprise search waarmee de taal van een document wordt bepaald.

**Taalkundige zoekopdracht**

Een type zoekopdracht waarvoor documenten wordt doorzocht, opgehaald en geïndexeerd en de termen worden gereduceerd tot de basisvorm (zodat *mice* bijvoorbeeld wordt geïndexeerd als *muis*) of worden uitgebreid met de basisvorm (zoals bij samengestelde woorden).

**taxonomy**

Een classificatie van objecten in groepen op basis van de overeenkomsten. In enterprise search worden gegevens in categorieën en subcategorieën verdeeld met een taxonomie. Zie ook Categoriestructuur.

**Tekennormalisatie**

Een proces waarin de verschillende vormen van een teken, zoals hoofdlettergebruik en diakritische tekens, worden omgezet in de algemene vorm.

**Teken voor terugloop met regelopschuiving**

Een stuurcode waarmee de positie in de afdruk of op het scherm één regel omlaag wordt verplaatst. Voor bepaalde systemen zijn meerdere tekens vereist.

**tekstanalyse**

Proces waarbij semantische gegevens en andere gegevens uit de tekst worden geëxtraheerd, zodat de gegevens in een collectie beter kunnen worden gevonden.

**tekstanalyseprogramma**

Een softwarecomponent die verantwoordelijk is voor het zoeken en weergeven van de context en de semantische inhoud in de tekst.

**tekstsegmentering**

Zie segmentering.

**Thema-extractie**

Een type begripsextractie waarmee significante vocabulaire-items automatisch worden herkend in tekstdocumenten, zodat het thema of onderwerp van een document kan worden opgehaald. Zie ook Begripsextractie.

**toegangslijst (access control list, ACL)**

Een lijst die bestaat uit een of meer gebruikers-ID's of -groepen en de bijbehorende bevoegdheden. Toegangslijsten worden gebruikt om de toegang van gebruikers tot items en objecten te besturen.

**Toelichting**

Informatie over een tekstspanne. Zo kan met een toelichting worden aangegeven dat een tekstspanne voor een bedrijfsnaam staat. In Unstructured Information Management Architecture (UIMA) is een annotatie een speciaal soort featurestructuur.

**Token**

De tekstuele basiseenheden die in enterprise search worden geïndexeerd. Tokens kunnen de woorden in een taal zijn of andere teksteenheden die in aanmerking komen voor indexering.

**tokenisering**

Zie segmentering.

**tokenizer**

Een tekstsegmenteringsprogramma waarmee tekst wordt gescand en wordt bepaald of en wanneer een reeks tekens als token kan worden herkend.

**typesysteem**

Het typesysteem definieert de typen objecten (featurestructuren) die door

een tekstanalyseprogramma kunnen worden ontdekt in een document. Het typesysteem definieert alle mogelijke featurestructuren in termen van typen en features. U kunt in een typesysteem een willekeurig aantal verschillende typen definiëren. Een typesysteem is specifiek voor een bepaald domein en een bepaald programma.

**Uit wachtrij verwijderen**

Items uit een wachtrij verwijderen.

**Uniform Resource Identifier (URI)**

Een compacte tekenreeks waarmee een abstracte of fysieke bron wordt aangegeven.

**Uniform Resource Locator (URL)**

Een tekenreeks waarmee informatiebronnen op een computer of in een netwerk (zoals internet) worden aangegeven. Deze tekenreeks bestaat uit de afkorting van het protocol waarmee toegang wordt verkregen tot de informatiebron, alsmede de informatie die het protocol gebruikt om de informatiebron op te zoeken.

**Universal Resource Name (URN)**

Een internetprotocolelement dat bestaat uit een korte tekenreeks die voldoet aan een bepaalde syntaxis. De tekenreeks bevat een naam die kan worden gebruikt om naar een resource te verwijzen.

**Unstructured Information Management Architecture (UIMA)**

Een IBM-architectuur waarmee een framework wordt gedefinieerd voor de implementatie van systemen voor de analyse van niet-gestructureerde gegevens.

**Veld** Het kleinste identificeerbare deel van een record.

**veldzoekopdracht**

Een query die beperkt is tot een bepaald veld.

**Volledige tekstindex**

Een gegevensstructuur waarmee naar gegevensitems wordt verwezen zodat met de zoekopdracht snel de documenten worden gevonden waarin de zoektermen voorkomen.

**vrije zoekopdracht**

Een zoekopdracht waarin de zoekterm wordt uitgedrukt als vrije tekst.

**webcrawler**

Een robotsoftwareklasse waarmee het internet wordt doorzocht door een webdocument op te halen en de koppelingen in dat document te volgen.

**wegingsklasse**

Een specificatie die van invloed kan zijn op de relatieve ranking van een document in de zoekresultaten.

**XML Path-taal (XPath)**

Een taal waarmee de unieke adresgedeelten van een bron-XML-document worden aangegeven. XPath biedt ook basisvoorzieningen voor de manipulatie van strings, getallen en booleaanse operatoren.

**zoekcache**

Een buffer waarin de gegevens en resultaten van eerdere zoekopdrachten zijn opgeslagen.

**Zoeken bij benadering**

Een zoekopdracht waarmee woorden als resultaat worden gegeven waarvan de spelling vergelijkbaar is met de spelling van de zoekterm.



**Zoekindexbestanden**

De set met bestanden waarin een index wordt opgeslagen in de zoekmachine.

**Zoekmachine**

Een programma waarin een zoekopdracht wordt geaccepteerd, waarna een lijst met documenten voor de gebruiker als resultaat wordt gegeven.

**Zoekopdracht met gewogen termen**

Een query waarin aan bepaalde termen een groter belang wordt toegekend.

**zoekprogramma**

Een programma waarmee query's worden verwerkt, de index wordt doorzocht, de zoekresultaten worden teruggezonden en de brondocumenten worden opgehaald voor collecties in een enterprise search-systeem.

**Zoekresultaten**

Een lijst met de documenten die aan de zoekopdracht voldoen.



---

## Toegang krijgen tot informatie over Content Management en Discovery

Informatie over IBM Content Management- en Discovery-producten is beschikbaar via de telefoon en op internet.

De telefoonnummers in dit document zijn van toepassing op de Verenigde Staten:

- Als u andere producten wilt bestellen of algemene informatie wilt opvragen, belt u 1-800-IBM-CALL (1-800-426-2255)
- Als u publicaties wilt bestellen, belt u 1-800-879-2755

Op internet kunt u informatie over IBM Content Management- en Discovery-producten vinden op <http://www.ibm.com/software/sw-bycategory/subcategory/SWB40.html>. Deze site bevat links waarmee u:

- Meer te weten kunt komen over de producten
- De producten kunt kopen
- Kunt deelnemen aan betatests van de producten
- Productondersteuning kunt aanvragen

Ga als volgt te werk om productdocumentatie op te vragen:

1. Ga naar de internetpagina <http://www.ibm.com/software/sw-bycategory/subcategory/SWB40.html>.
2. Selecteer een product waarover u meer informatie wenst, bijvoorbeeld WebSphere Information Integrator OmniFind Edition. Deze site bevat links naar:
  - Productdocumentatie, waaronder opmerkingen bij de release ("release notes") en online informatiecentra
  - Systeemvereisten
  - Productdownloads
  - Fixpacks
  - Nieuws over producten
  - Ondersteunend informatiemateriaal, zoals white papers en IBM Redbooks
  - Nieuwsgroepen en gebruikersgroepen
  - Instructies voor het bestellen van boeken
3. Klik op de link Support (links op de pagina).
4. Selecteer in de sectie Learn het type documentatie dat u wilt zien. Als voor het geselecteerde product een informatiecentrum beschikbaar is, kunt u de link naar het Informatiecentrum selecteren.

---

## Commentaar op de documentatie

U kunt uw commentaar op deze informatie of andere documentatie van IBM insturen.

Dankzij uw feedback kan IBM de documentatie steeds verder verbeteren. Mocht u commentaar hebben op deze informatie of op de andere documentatie van die IBM Software Development bij producten levert, laat ons dat dan weten. Dit kunt u op de volgende manieren doen:

1. U kunt uw commentaar insturen met behulp van het online formulier dat u kunt vinden op [www.ibm.com/software/awdtools/rcf/](http://www.ibm.com/software/awdtools/rcf/).
2. U kunt uw commentaar per e-mail verzenden naar [comments@us.ibm.com](mailto:comments@us.ibm.com). Vergeet niet de naam van het product te vermelden, alsmede het versienummer van het product en de naam en het onderdeelnummer van de informatie (indien van toepassing). Als u commentaar hebt op een specifiek stuk tekst, geef dan aan waar die tekst zich bevindt (bijvoorbeeld: de titel, het nummer van een tabel of het paginanummer).

---

## Contact opnemen met IBM

Als u contact wilt opnemen met de klantenservice van IBM in de Verenigde Staten of Canada, belt u 1-800-IBM-SERV (1-800-426-7378).

Als u meer informatie wilt over de beschikbare serviceopties, belt u een van de volgende nummers:

- In de Verenigde Staten: 1-888-426-4343
- In Canada: 1-800-465-9600

Als u een IBM-vestiging in uw land of regio zoekt, raadpleegt u de wereldwijde adressenlijst van IBM op [www.ibm.com/planetwide](http://www.ibm.com/planetwide).

---

## Kennisgevingen en handelsmerken

---

### Kennisgevingen

Deze informatie is ontwikkeld voor producten en diensten die in de Verenigde Staten worden aangeboden. Mogelijk levert IBM niet alle in dit document genoemde producten, diensten of functies in alle andere landen. Neem contact op met uw IBM-vertegenwoordiger voor informatie over de producten en diensten die bij u beschikbaar zijn. Verwijzing in deze publicatie naar producten of diensten van IBM houdt niet in dat uitsluitend IBM-producten of -diensten gebruikt kunnen worden. Functioneel gelijkwaardige producten of diensten kunnen in plaats daarvan worden gebruikt, mits dergelijke producten of diensten geen inbreuk maken op intellectuele eigendomsrechten of andere rechten van IBM. De gebruiker is verantwoordelijk voor de samenwerking van IBM-producten of -diensten met producten of diensten van anderen, tenzij uitdrukkelijk anders aangegeven door IBM.

Mogelijk heeft IBM octrooien of octrooi-aanvragen met betrekking tot bepaalde in deze publicatie genoemde producten. Aan het feit dat deze publicatie aan u ter beschikking is gesteld, kan geen recht op licentie of ander recht worden ontleend. Voor vragen over licenties kunt u zich wenden tot: IBM Director of Licensing IBM Corporation North Castle Drive Armonk, NY 10504-1785 U.S.A.

Voor licentievragen over DBCS-informatie (Double Byte Character Set) neemt u contact op met het IBM Intellectual Property Department in uw land of stelt u de vragen schriftelijk aan: IBM World Trade Asia Corporation Licensing 2-31 Roppongi 3-chome, Minato-ku Tokyo 106-0032, Japan

De volgende alinea is niet van toepassing op het Verenigd Koninkrijk, noch op enig ander land waar dergelijke bepalingen niet stroken met de lokale wetgeving: INTERNATIONAL BUSINESS MACHINES BIEDT DEZE PUBLICATIE AAN OP "AS IS"-BASIS, ZONDER ENIGE GARANTIE, UITDRUKKELIJK NOCH STILZWIJGEND, MET INBEGRIIP VAN, MAAR NIET BEPERKT TOT, IMPLICIETE GARANTIES VAN VERHANDELBAARHEID, GESCHIKTHEID VOOR EEN BEPAALD DOEL OF DE GARANTIE DAT DEZE PUBLICATIE GEEN INBREUK MAAKT OP DE RECHTEN VAN DERDEN. In bepaalde rechtsgebieden is het uitsluiten of beperken van uitdrukkelijke of stilzwijgende garanties niet toegestaan; zodat het bovenstaande mogelijk niet op u van toepassing is.

Deze informatie kan technische onjuistheden en/of drukfouten bevatten. IBM kan zonder voorafgaand bericht wijzigingen en/of verbeteringen aanbrengen in de producten en/of programma's die in deze publicatie worden beschreven.

Verwijzingen in deze publicatie naar niet-IBM websites mogen niet worden opgevat als een aanbeveling van die websites. Het materiaal op dergelijke websites maakt geen deel uit van het materiaal voor dit IBM-product en het gebruik van dergelijke websites is geheel voor eigen risico.

IBM mag informatie die door u wordt verstrekt gebruiken en distribueren op elke manier die haar goeddunkt zonder daarbij verplichtingen jegens u aan te gaan.

Licentiehouders die informatie over dit programma willen ontvangen over: (i) het uitwisselen van informatie tussen in eigen beheer gemaakte programma's en

andere programma's (waaronder dit programma) en (ii) het gemeenschappelijk gebruik van de uitgewisselde informatie, dienen contact op te nemen met:

IBM Corporation J46A/G4  
555 Bailey Avenue  
San Jose, CA 95141-1003 U.S.A.

Dergelijke informatie kan beschikbaar zijn onder bepaalde voorwaarden en bepalingen waaronder, in bepaalde gevallen, betaling van een vergoeding.

Het gelicentieerde programma dat in dit document wordt beschreven en al het bij dit programma behorende materiaal, wordt door IBM geleverd onder de voorwaarden omschreven in de IBM Klantenovereenkomst, de IBM Internationale programmalicentie-overeenkomst of enige andere gelijkwaardige overeenkomst.

Prestatiegegevens die hierin worden vermeld, zijn verzameld in een gecontroleerde omgeving. De resultaten die in een andere verwerkingsomgeving worden behaald, kunnen hiervan derhalve aanzienlijk afwijken. Bepaalde metingen zijn verricht op systemen die nog in ontwikkeling, waren en er kan geen garantie worden gegeven dat deze metingen op algemeen verkrijgbare systemen gelijk zullen zijn. Bovendien zijn sommige metingen niet meer dan schattingen die door extrapolatie tot stand zijn gekomen. De feitelijke resultaten kunnen hiervan afwijken. Gebruikers van dit document dienen de feitelijke gegevens in hun eigen specifieke omgeving vast te stellen.

Informatie over niet door IBM geleverde producten is verkregen van de leveranciers van de desbetreffende producten, uit de publicaties van deze leveranciers of uit andere publiek toegankelijke bronnen. IBM heeft die producten niet getest en kan niet bevestigen dat de gegevens op het gebied van prestaties, compatibiliteit of enig ander gebied, correct zijn. Vragen met betrekking tot de mogelijkheden van niet-IBM producten dienen te worden gericht aan de leveranciers van die producten.

Alle uitspraken met betrekking tot de strategie of plannen van IBM kunnen zonder voorafgaand bericht worden gewijzigd of ingetrokken. Dergelijke uitspraken geven uitsluitend doelstellingen aan.

Deze informatie bevat voorbeelden van gegevens en rapporten die in de dagelijkse bedrijfsvoering worden gebruikt. Voor een zo volledig mogelijke illustratie zijn in deze voorbeelden namen van personen, bedrijven, merken en producten opgenomen. Al deze namen zijn fictief en enige gelijkenis met de namen en/of adressen van werkelijke personen of bedrijven berust uitsluitend op toeval.

#### AUTEURSRECHTLICENTIE:

Deze informatie bevat voorbeeldprogramma's in brontaal ter illustratie van de programmeertechnieken op diverse besturingsplatforms. U mag deze voorbeeldprogramma's zonder betaling aan IBM in elke vorm kopiëren, aanpassen en distribueren, mits dit tot doel heeft het ontwikkelen, gebruiken, verkopen of distribueren van toepassingsprogramma's die voldoen aan de Application Programming Interface voor het besturingsplatform waarvoor de voorbeeldprogramma's zijn geschreven. Deze voorbeelden zijn niet onder alle omstandigheden uitvoering getest. IBM kan de betrouwbaarheid, onderhoudbaarheid en functionaliteit van deze programma's derhalve niet garanderen of impliceren. U mag deze voorbeeldprogramma's zonder betaling aan IBM in elke vorm kopiëren, aanpassen en distribueren, mits dit tot doel heeft het ontwikkelen, gebruiken, verkopen of distribue-

ren van toepassingsprogramma's die voldoen aan de Application Programming Interface voor het besturingsplatform waarvoor de voorbeeldprogramma's zijn geschreven.

Bij elk exemplaar of onderdeel van deze voorbeeldprogramma's, of afgeleide versies hiervan, moet een auteursrechtvermelding worden opgenomen, volgens het onderstaande voorbeeld:

Outside In (®) Viewer Technology, © 1992-2006 Stellent, Chicago, IL., Inc. Alle rechten voorbehouden.

IBM XSLT Processor Gelicentieerd materiaal - Eigendom van IBM ©Copyright IBM Corp., 1999-2006. Alle rechten voorbehouden.

---

## Handelsmerken

In dit onderwerp vindt u een overzicht van de IBM-handelsmerken en enkele niet-IBM-handelsmerken.

Zie <http://www.ibm.com/legal/copytrade.shtml> voor meer informatie over IBM-handelsmerken.

De volgende termen zijn handelsmerken of geregistreerde handelsmerken van andere bedrijven:

Java en alle op Java gebaseerde merken en logo's zijn handelsmerken of geregistreerde handelsmerken van Sun Microsystems, Inc. in de Verenigde Staten en/of andere landen.

Microsoft, Windows, Windows NT en het Windows-logo zijn handelsmerken van Microsoft Corporation in de Verenigde Staten en/of andere landen.

Intel, Intel Inside (logo's), MMX en Pentium zijn handelsmerken van Intel Corporation in de Verenigde Staten en/of andere landen.

UNIX is een handelsmerk van The Open Group in de Verenigde Staten en andere landen.

Linux is een handelsmerk van Linus Torvalds in de Verenigde Staten en/of andere landen.

Andere namen van bedrijven, producten en diensten kunnen handelsmerken zijn van derden.





# Trefwoordenregister

## A

aangepaste tekstanalyse  
 analyseresultaten toewijzen aan de index 140  
 analyseresultaten toewijzen aan een relationele database 141  
 Common Analysis Structure toewijzen aan de index 140  
 Common Analysis Structure toewijzen aan een relationele database 141  
 description 136  
 tekstanalyseprogramma's 137, 138  
 XML-elementen toewijzen 139  
aanmelden bij de beheerconsole 21  
Aanpassingstool Zoekprogramma  
 config.properties-bestand 220  
 starten 220  
actieve websites, bewaken 304, 306  
AdminLinkBarInclude.jsp-bestand 276  
adresregels voor webcrawlers 84  
AIX-besturingssystemen  
 configuratie van DB2 Content Manager-crawler 55  
 Content Edition-crawler-configuratie 43  
 DB2-crawlerconfiguratie 48  
 Domino Document Manager-crawler-configuratie 69  
 Event Publishing-configuratie 48  
 Notes-crawlerconfiguratie 69  
 QuickPlace-crawlerconfiguratie 69  
alertsignalen  
 beschikbare ruimte op servers 325  
 collectieniveau 299, 324  
 description 323  
 e-mail ontvangen voor 328  
 e-mailopties 324, 325  
 gecrawelde documenten 324  
 geïndexeerde documenten 324  
 indexbeperkingen 299  
 responstijd zoekopdrachten 324  
 SMTP-serverconfiguratie 327  
 systeemniveau 325  
alertsignalen voor beschikbare ruimte 325  
algemene beveiliging in WebSphere  
 Aanpassingstool Zoekprogramma 220  
 eigenschappen van zoekprogramma 219  
 uitschakelen 265  
algemene webcrawlruijme 96  
analyse van ankerktext  
 collectiebeveiliging 248  
 description 231  
 documenten indexeren 249  
 globale analyse 248  
annotators 136  
API's  
 Data Listener 11  
 description 11

API's (*vervolg*)  
 Search and Index 11, 203  
archiefbestanden  
 crawlen 110  
 ondersteunde indelingen 110  
 URI-indelingen 111  
automatische detectie  
 codetabellen 155  
 talen 154  
**B**  
backup maken van enterprise  
 search 333, 334  
backscripts  
 description 333  
 uitvoeren 334  
banner.jspf, bestand 284  
banner\_searchControl.jspf, bestand 284  
bedrijfstoepassingen  
 ESAdmin-programma 265  
 ESSearchApplication-programma 265  
 ESSearchServer-programma 265  
beheerconsole  
 aanmelden 21  
 description 10  
 interface 17  
 taakoverzicht 17  
beheerdersrollen  
 bewaken 234, 235  
 collectiebeheerder 234, 235  
 configureren 235  
 description 234  
 enterprise search-beheerder 234, 235  
 operator 234, 235  
beheerderswachtwoord  
 wijzigen op één server 22  
 wijzigen op meerdere servers 23  
bekijken  
 logbestanden 330  
 logbestanden voor gewiste documenten 320  
 URI-gegevens 301  
belang van documenten  
 gewogen woordenboeken 194  
 in gemigreerde collecties 287  
 inschakelen voor een collectie 31  
 standaardwaarden herstellen 190  
 statisch 190  
 URI-patronen 194, 195  
 wegingsklassen 196, 200  
beperken van webcrawlruijme 84  
bestandsextensies  
 ondersteund door collectieparsers 147, 149  
 ondersteund door Stellent-parsers 149  
uitsluiten van webcrawlruijmen 84  
beveiliging  
 algemeen, WebSphere Application Server 250

beveiliging (*vervolg*)  
 algemene beveiliging in WebSphere 265  
 analyse van ankerktext 248  
 beheerdersrollen 235  
 collectieniveau 236, 267  
 crawlerplug-ins 108  
 description 231  
 documentniveau 237, 238, 239, 241, 247, 267  
 duplicaatdocumenten, detectie 236  
 gebruikersprofielen 244  
 HTTPS-configuratie voor zoeken 224  
 Identiteitenbeheer 242, 247  
 inschakelen voor een collectie 31, 231  
 inschakelen voor enterprise search 250  
 LDAP-gebruikersregister 250  
 Lotus Domino-documenten 259, 260  
 samengevouwen zoekresultaten 268  
 setup met een enkele server 252  
 setup met meerdere servers 253  
 SSL-configuratie voor zoeken 224  
 SSO-ondersteuning (Single Sign-On) 246  
 toegangsbesturing 233  
 toegangsinstellingen op documentniveau negeren 267  
 uitschakelen voor bedrijfs-toepassing 265  
 verificatie 233, 265  
 Windows-domeinen 263  
 zoekprogramma-ID's 237  
Beveiliging (view), beschrijving 17  
beveiliging op collectieniveau  
 analyse van ankerktext 248  
 description 231, 236  
 duplicaatdocumenten, detectie 236  
 inschakelen 31  
 programma-ID's 237  
beveiliging op documentniveau  
 beveiligingstokens 239  
 crawlerconfiguratie 35  
 crawlerplug-ins 108  
 description 231, 237  
 gebruikersprofielen 244  
 geïndexeerde toegangsbesturing-functies 239  
 Identiteitenbeheer 242, 247  
 Lotus Domino-documenten 260  
 resultaten achteraf filteren 238  
 resultaten vooraf filteren 238  
 SSO-ondersteuning (Single Sign-On) 246  
 validatie in real-time 241  
 validatie van actuele legitimatiegegevens 241  
 voor Lotus Domino-documenten 259  
 voor Windows-bestandssystemen 263  
beveiligingsoptie Lokale gebruiker, QuickPlace-crawlers 261

- beveiligingstokens
  - beveiliging op documentniveau 239
  - crawlerconfiguratie 239
  - uitschakelen voor een collectie 267
- bewaken
  - Actieve sites van de webcrawler 306
  - collecties 300
  - crawlers 302
  - crawlsnelheid webcrawler 306
  - Data Listener 317
  - description 234
  - enterprise search 299
  - gewiste documenten 318, 320
  - historie van responstijd 315
  - logbestanden 330
  - parsers 312
  - populaire query's 315
  - recente query's 315
  - rolconfiguratie 235
  - threadgegevens webcrawler 305
  - URI-gegevens 301
  - Web-crawlers 304
  - zoekservers 315
- bewerken
  - collecties 32
  - crawlereigenschappen 38
  - crawlruimten 38
  - Data Listener-programma's 107
  - eigenschappen van zoekprogramma 206, 219
- bos.iocp.rte-module 73

## C

- categorieën
  - maken 128
- categorieregels
  - configureren 128
  - documentcontent 124, 128
  - URI-patronen 124, 128
- categories
  - categoriestructuren 126
  - description 124
  - migreren vanuit WebSphere Portal 287
  - nesten van subcategorieën 126
  - regelgebaseerd 124, 127
  - type categorisering 127
  - URI-indelingen 111
  - zoeken 124
- categoriestructuren
  - description 126
  - migreren vanuit WebSphere Portal 287
- CCLServer\_date.log-bestand 27
- Chinees
  - n-gramsegmentering 156
  - verwijderen van nieuwe-regeltekens 156
- codetabellen
  - automatische detectie 155
  - ondersteund 155
- collectie-ID, syntaxisregels 31
- collectiebeheerder
  - description 234
  - rolconfiguratie 235

- collecties
  - analyse 123
  - beveiliging 236
  - beveiliging van ankertekst 248
  - beveiliging van programma-ID 237
  - bewaken 300
  - bewerken 32
  - concept 29
  - description 3
  - duplicaatdocumenten, beveiliging 236
  - Federatie 29
  - koppelen aan zoekprogramma's 204
  - maken met de Collectiewizard 29
  - maken met view Collecties 31
  - manieren om te maken 29
  - MigrationWizard.log-bestand 287
  - migreren vanuit WebSphere Portal 287
  - schatting van grootte 31
  - schatting van resources 299
  - standaardmigratie-instellingen 289
  - systemstatus 300
  - toeganginstellingen op documentniveau negeren 267
  - wissen 34
  - zoeken 164
  - zoekservers 177
- Collectiewizard 29
- Common Analysis Structure
  - description 136
  - toewijzen aan de index 140
  - toewijzen aan relationele databases 141
  - XML-elementen toewijzen aan 139
- config.properties-bestand 252, 253
  - aanpassen 220
  - beschrijving van eigenschappen 206
  - bewerken 219
  - klonen 221
- Content Edition-crawlers
  - configuratie 40
  - directe werkstand 42
  - instellen in Solaris-besturingsomgeving 43
  - instellen onder Windows 44
  - instellen op AIX-besturingsstelsysteem 43
  - instellen op Linux-besturingsstelsysteem 43
  - serverwerkstand 43
  - URI-indelingen 111
- cookies.ini-bestand
  - configureren 95
  - description 94
  - indeling 95
- cookies voor webcrawling
  - configureren 95
  - description 94
  - indeling 95
- crawl.rules-bestand 96
- Crawler (pagina), beschrijving 17
- crawlereigenschappen
  - bewerken 38
  - description 4
- crawlerhistorierapporten
  - description 304

- crawlerhistorierapporten (*vervolg*)
  - maken 307
  - rapport met HTTP-statuscodes 307
  - siterapport 307
- crawlerplug-ins 108
- crawlers
  - archiefbestanden 110
  - basiswaarden voor 35, 37
  - beveiliging op documentniveau 237
  - beveiliging op documentniveau configureren 35
  - bewaken 302
  - combineren van crawlertypen 35
  - configuratieoverzicht 35
  - Content Edition 40, 42, 43
  - crawlereigenschappen bewerken 38
  - crawlruimten bewerken 38
  - Data Listener-programma's 107
  - DB2 45
  - DB2 Content Manager 54
  - description 4
  - Domino Document Manager 58
  - Exchange Server 61, 258
  - JDBC-database 62
  - maken 37
  - NNTP 64
  - ondersteund voor extern gebruik 11, 107
  - Opmerkingen 65, 67
  - plannen 35, 40
  - plug-ins 108
  - QuickPlace 75
  - standaardmigratie-instellingen 289
  - systemstatus 302
  - UNIX-bestandssysteem 79
  - URI-indelingen 111
  - Web 80
  - Web Content Management 100, 104
  - WebSphere Portal 102, 104
  - Windows-bestandssysteem 105
  - wissen 39
- crawlerservers
  - starten 291, 302
  - stoppen 294, 302
- crawlertypen
  - basiswaarden voor 35
  - combineren in een collectie 35
- Crawlruimte
  - alertsignalen over 324
  - bewerken 38
  - description 4
  - webcrawlerconfiguratie 84
  - crawlsnelheid, bewaken 306

## D

- Data Listener
  - bewaken 317
  - configureren 107
  - opnieuw starten 107, 317
- Data Listener API 11
- DB2 Content Manager-crawlers
  - configuratie 54
  - instellen in Solaris-besturingsomgeving 55
  - instellen onder Windows 57

- DB2 Content Manager-crawlers (*vervolg*)
  - instellen op AIX-besturings-systeem 55
  - instellen op Linux-besturings-systeem 55
  - URI-indelingen 111
- DB2-crawlers
  - configuratie 45
  - configuratie van WebSphere II Event Publisher Edition 49
  - Event Publishing 45
  - Event Publishing-configuratie 48
  - instellen onder AIX 48
  - instellen onder Linux 48
  - instellen onder Solaris 48
  - instellen onder Windows 48
  - URI-indelingen 111
  - WebSphere II Classic Federation 53
  - WebSphere MQ-configuratie 52
  - WebSphere MQ-installatie onder AIX 48
  - WebSphere MQ-installatie onder Linux 48
  - WebSphere MQ-installatie onder Solaris 48
  - WebSphere MQ-installatie onder Windows 48
- DB2 UDB-databases
  - toegang als externe bron 227
  - toegang met DB2-crawlers 45
  - toegang met JDBC-databasecrawlers 62
- Default.jsp, bestand 284
- Default.jsp-bestand 276
- delta-indexen
  - description 7, 159
  - gelijktijdig opbouwen 162
  - plannen 160
- DIIOp-protocol, crawlerconfiguratie 72
- directe werkstand, Content Edition-repository's 42
- Directory Assistance-configuratie 262
- discovery 4
- documentatie
  - HTML 369
  - PDF 369
  - toegankelijkheid 371
  - zoeken 369
- Documentatie in PDF-indeling voor enterprise search 369, 371
- documentcontent, beschrijving 164
- documentranking
  - standaardwaarden herstellen 190
- documentsamenvattingen
  - aanpassen 184
  - eigenschappen bewerken voor 185
- documenttracering
  - description 317
  - inschakelen 318
  - logbestandconfiguratie 318
  - logbestanden 320
  - rapporten 318
  - uitschakelen 318
- documenttypen
  - analyse 146
  - detecteren 145
  - voor parserservices 147, 149

- documenttypen (*vervolg*)
  - voor Stellent-parsers 149
- domeinregels voor webcrawlers 84
- Domino Document Manager-crawlers
  - configuratie 58
  - DIIOp-protocolconfiguratie 72
  - instellen in Solaris-besturings-omgeving 69
  - instellen onder Windows 71
  - instellen op AIX-besturings-systeem 69
  - instellen op Linux-besturings-systeem 69
  - IOCP-configuratie 73
  - NRPC-protocol 69, 71
  - URI-indelingen 111
- Domino-gebruikersconfiguratie, QuickPlace-crawlers 261
- Dublin Core-elementen 134
- duplicaatdocumenten, detectie
  - beveiliging inschakelen 236
  - description 231
  - globale analyse 236
  - wegingsklassen, configuratie 196
- Dynamisch samenvatten 184
- dynamische ranking 189

## E

- e-mailmeldingen
  - SMTP-serverconfiguratie 327
  - voor alertsignalen 328
  - voor berichten 328
- EAR-bestanden
  - ESAdmin-programma 265
  - ESSearchApplication-programma 265
  - ESSearchServer-programma 265
- enterprise search
  - API's 11
  - backups maken 334
  - backupscripts 333
  - beheerconsole 10
  - beheerdersrollen 234
  - beveiliging 231
  - beveiliging op collectieniveau 236
  - beveiliging op documentniveau 237
  - bewaken 299
  - componenten 3
  - crawlerservers 4, 35
  - herstelscripts 333
  - indexservers 7, 159
  - integratie met WebSphere Portal 271
  - logbestanden 323
  - opdrachten 337
  - overzicht 1
  - parsers 5, 123
  - poortnummerconfiguratie 27
  - retourcodes 337
  - sessie-ID's 337
  - starten van servers 291
  - starten van zoekservers 296
  - stoppen van servers 291, 294
  - stoppen van zoekservers 296
  - stroomschema 13
  - terugzetten vanaf een backup 335
  - URI-indelingen 111
  - zoekprogramma's 12

- enterprise search (*vervolg*)
  - zoekservers 9, 177
- enterprise search-beheerder
  - description 234
  - rolconfiguratie 235
  - wachtwoord wijzigen op één server 22
  - wachtwoord wijzigen op meerdere servers 23
- es.cfg-bestand 22, 23, 224, 252, 253
- ES\_INSTALL\_ROOT, beschrijving 22, 23
- ES\_NODE\_ROOT, beschrijving 22, 23
- es.search.provider.jar-bestand 279
- es.security.jar-bestand 273, 279
- es\_special\_field.default\_field, gereserveerd veld 200
- es\_special\_field.default\_metadata\_field, gereserveerd veld 200
- es\_special\_field.regular\_text, gereserveerd veld 200
- es.wp5.install.jar-bestand 273
- es.wp6.install.jar-bestand 279
- esadmin-opdracht 337
- ESAdmin-programma
  - aanmelden bij 21
  - beveiliging uitschakelen 265
- esadmin startSearch, opdracht 296
- esadmin stopSearch, opdracht 296
- esadmin system startall-opdracht 337
- esadmin system stopall-opdracht 337
- esapi.jar-bestand 273, 279
- esbackup.bat, script 334
- esbackup.sh, script 334
- eschangepw-script 22, 23
- eschangetrustpw, opdracht 224
- eschangewaspw, opdracht
  - configuratie met een enkele server 252
  - configuratie met meerdere servers 253
- escrcm.sh-script 55
- escrcm.vbs-script 57
- escrdb2.sh-script 48
- escrdb2.vbs-script 48
- escrnote.sh-script 69
- escrnote.vbs-script 71
- escrvbr.sh-script 43
- escrvbr.vbs-script 44
- ESPACServer.ear-bestand 273, 279
- esrestore.bat-script 335
- esrestore.sh-script 335
- ESSearchAdapter.ear-bestand 273
- ESSearchApplication-programma
  - beveiliging uitschakelen 265
  - config.properties-bestand 206, 219
  - starten 223
- ESSearchPortlet.war-bestand 273, 279
- ESSearchRegistrationPortlet.war-bestand 273
- ESSearchServer-programma
  - beveiliging uitschakelen 265
  - opnieuw starten 219, 220
- Event Publishing
  - DB2-crawlerconfiguratie 49, 52
  - description 45
  - instellen in Solaris-besturings-omgeving 48

- Event Publishing (*vervolg*)
  - instellen onder Windows 48
  - instellen op AIX-besturings-systeem 48
  - instellen op Linux-besturings-systeem 48
- Exchange Server-crawlers
  - beveiligde documenten 258
  - configuratie 61
  - URI-indelingen 111
- externe bronnen
  - beveiliging van programma-ID 237
  - configuratie 227
  - description 227
  - koppelen aan zoekprogramma's 230
- externe crawlers
  - configureren 107
  - Data Listener API 11
  - Data Listener-programma's 107
- externe JDBC-bronnen
  - bewerken 227
  - configuratie 227
  - JDBC-drivers 227
  - wissen 227
- externe LDAP-bronnen
  - bewerken 227
  - configureren 227
  - wissen 227

## F

- federatieve collecties 29
- firewalls, Exchange Server-documenten
  - crawlen 258
- followindex.rules-bestand
  - configureren 98
  - description 98
- Form-Based Authentication 91, 93
- foutberichten
  - e-mail ontvangen voor 326, 328
  - logbestanden bekijken 330
  - logbestanden voor gewiste documenten bekijken 320
  - SMTP-serverconfiguratie 327

## G

- gebruikersagents 81
- gebruikersbeveiligingscontext, tekenreeks 242
- gebruikersprofielen
  - configuratie 247
  - description 244
- gecrawld documenten, datums
  - configureren voor webcrawlers 99
- gegevensbronnen, typen
  - CA-Datacom-databases 45
  - Content Edition-repository's 40, 42, 43
  - DB2 Content Manager, itemtypen 54
  - DB2 UDB-databases 45, 62, 227
  - DB2 UDB for iSeries-databases 45
  - DB2 UDB for z/OS 45
  - Domino Document Manager-databases 58

- gegevensbronnen, typen (*vervolg*)
  - Exchange Server, openbare mappen 61
  - IMS-databases 45
  - Informix-databases 45
  - JDBC-databases 62, 227
  - NNTP-nieuwsgroepen 64
  - Notes-databases 65, 67
  - ondersteund door enterprise search 2
  - ondersteund voor extern gebruik 2, 11
  - Oracle-databases 45, 62, 227
  - QuickPlace-databases 75
  - relationele databases 45
  - Software AG Adabas-databases 45
  - SQL Server-databases 45, 62
  - Sybase-databases 45
  - UNIX-bestandssystemen 79
  - VSAM-databases 45
  - Web Content Management-sites 100
  - websites 80
  - WebSphere Portal-sites 102
  - Windows-bestandssystemen 105
- gegevensstroom, enterprise search-systeem 13
- gelijktijdig opbouwen van index 162
- gewiste documenten
  - description 317
  - logbestandconfiguratie 318
  - logbestanden voor 320
  - rapporten over 318
- gewogen woordenboeken
  - description 192
  - koppelen aan een collectie 194
  - toevoegen aan het systeem 193
- global.rules-bestand 96
- globale analyse
  - analyse van ankerstekst 231, 248
  - description 7
  - duplicaatdocumenten, detectie 231, 236

## H

- herstel van enterprise search 333
- herstelscripts
  - description 333
  - uitvoeren 335
- historie van responstijd, bewaken 315
- hoofdindexen
  - description 7, 159
  - gelijktijdig opbouwen 162
  - plannen 160
- HTML-documentatie voor enterprise search 369
- HTML-documenten, zoeken 133
- HTML-zoekvelden
  - description 133
  - Dublin Core-elementen 134
  - elementen toewijzen aan 133, 134
  - maken 134
- HTTP Basic Authentication 91
- HTTP-proxyservers 93
- HTTP-statuscodes
  - ontvangen door webcrawlers 307
  - webcrawlerrapport 307
- HTTPS, zoekserverconfiguratie 224

## I

- I/O Completion Port-module, crawler-configuratie 73
- Identiteitenbeheer
  - configuratie 247
  - description 242
  - gebruikersbeveiligingscontext 242
  - gebruikersprofielen 244
  - groepsextractie 242
  - SSO-ondersteuning (Single Sign-On) 246
  - uitschakelen 242
  - XML-querytekenreeks 242
- ideografische talen 153
- Index (pagina), beschrijving 17
- index opbouwen
  - description 159
  - gelijktijdig 162
  - parallel 162
  - plannen 161
  - starten 313
  - stoppen 313, 314
  - systeemstatus 314
- indexeerwachtrij 314
- indexen
  - alertsignalen over 324
  - ankertekst 249
  - bewaken 313, 314
  - description 7, 159
  - gelijktijdig opbouwen 162
  - invloed van jokertekens 170
  - jokertekens 163, 168, 171
  - parallel opbouwen 162
  - plannen 160
  - rooster inschakelen 161, 313
  - rooster uitschakelen 161, 313
  - rooster wijzigen 161
  - samengevouwen URI's 163, 174, 175
  - scopes 163, 172
  - URI-indelingen 111
  - URI's verwijderen 163, 175
  - wissen uit wachtrij 314
- indexen opbouwen 159
- indexroosters inschakelen 161
- indexroosters uitschakelen 161
- indexservers
  - starten 291
  - stoppen 294
- indexuitbreiding
  - description 168
  - invloed op indexgrootte 170
  - invloed op tijd voor opbouwen index 170
- installatiescripts
  - WebSphere Portal 272
- integratie met WebSphere Portal
  - description 271
  - es.wp5.install.jar-bestand 273
  - es.wp6.install.jar-bestand 279
  - installatiescripts 272
  - wp5\_install-script 273
  - wp6\_install-script 279
- interval voor opnieuw crawlen voor webcrawlers 89
- IOCP, crawlerconfiguratie 73
- IP-adresregels voor webcrawlers 84

## J

- Japans
  - n-gramsegmentering 156
  - verwijderen van nieuwe-regeltekens 156
- Java-connector voor DB2 Content Manager 55, 57
- JavaScript-ondersteuning in webcrawlers 84
- JDBC-databasecrawlers
  - configuratie 62
  - ondersteunde drivers 62
  - URI-indelingen 111
- JDBC-drivers
  - voor externe JDBC-bronnen 227
  - voor JDBC-databasecrawlers 62
- jokertekens
  - in query's 168
  - indexuitbreiding 168, 170, 171
  - query-uitbreiding 168, 171

## K

- klonen
  - crawlers 35
  - zoekprogramma's 221
- Koreaans
  - analyse van samengestelde termen 143
  - n-gramsegmentering 156

## L

- LDAP-gebruikersregister 250
- Linux-besturingssysteem
  - configuratie van DB2 Content Manager-crawler 55
  - Content Edition-crawler-configuratie 43
  - DB2-crawlerconfiguratie 48
  - Domino Document Manager-crawler-configuratie 69
  - Event Publishing-configuratie 48
  - Notes-crawlerconfiguratie 69
  - QuickPlace-crawlerconfiguratie 69
  - Solaris-besturingsomgeving
    - Event Publishing-configuratie 48
- locale
  - analyse 153
  - zoeken 153
- logbestanden
  - bekijken 330
  - bewaken 320, 330
  - description 323
  - e-mailopties 328
  - filteren 330
  - gewiste documenten bekijken 320
  - maximumgrootte 326
  - migratiewizard 290
  - overschrijven 326
  - severityniveaus 326
  - SMTP-serverconfiguratie 327
  - standaardlocatie 323
  - voor documenttracering 318
- Loggen (pagina), beschrijving 17
- Lotus Domino-domeinen 259, 260

Lotus Domino Trusted Servers 260

## M

- machtiging, beschrijving 233
- maken
  - collecties 29, 31
  - crawlers 37
  - HTML-zoekvelden 134
  - regelgebaseerde categorieën 128
  - scopes 173
  - snellinks 186
  - web crawlerrapporten 307
  - XML-zoekvelden 130
- maximuminterval voor opnieuw crawlen 89
- met wachtwoord beveiligde websites 91
  - Form-Based Authentication 93
  - HTTP Basic Authentication 91
- migratiewizard
  - collecties 287
  - description 287
  - logbestand 290
  - regelgebaseerde taxonomieën 287
  - standaardcollectie-instellingen 289
  - standaardcrawlerinstellingen 289
  - starten 287
- migreren
  - collecties 287
  - regelgebaseerde taxonomie 287
- MIME-typen, opnemen in webcrawluimten 84
- minimuminterval voor opnieuw crawlen 89
- Monitor (view), beschrijving 17
- multiple-bytecodering 155

## N

- n-gramsegmentering 156
- native XML-zoekopdrachten 144
- nieuwe-regeltekens verwijderen 156
- NNTP-crawlers, configureren 64
- no-follow (instructies)
  - configureren 98
  - description 98
- no-index (instructies)
  - configureren 98
  - description 98
- nodes.ini, bestand 224
- Notes-crawlers
  - configuratie 65
  - configuratie van beveiliging op documentniveau 259
  - DIIOp-protocolconfiguratie 72
  - instellen in Solaris-besturingsomgeving 69
  - instellen onder Windows 71
  - instellen op AIX-besturings-systeem 69
  - instellen op Linux-besturings-systeem 69
  - IOCP-configuratie 73
  - Lotus Domino Trusted Server 260
  - NRPC-protocol 69, 71
  - tips voor gebruik 67

- Notes-crawlers (*vervolg*)
  - URI-indelingen 111
  - validatie van actuele legitimatiegegevens 260
  - veldtoewijzingsregels 67
- NRPC-protocol, crawlerconfiguratie 69, 71

## O

- op tekst gebaseerde score 189
- opdrachten, enterprise search 337
- operator
  - description 234
  - rolconfiguratie 235
- Oracle-databases
  - toegang als externe bron 227
  - toegang met DB2-crawlers 45
  - toegang met JDBC-databasecrawlers 62

## P

- parallel opbouwen van index 162
- parametrische zoekopdracht, beschrijving 164
- Parser (pagina), beschrijving 17
- parsers
  - analyse van samengestelde termen 143
  - bewaken 312
  - codetabeldetectie 155
  - description 5, 123
  - documentindeling, detectie 145
  - documenttypen analyseren 146
  - documenttypen voor parser-services 147, 149
  - documenttypen voor Stellent-parsers 149
  - gegevensanalyse 5
  - n-gramsegmentering 156
  - native XML-zoekopdrachten 144
  - nieuwe-regeltekens verwijderen 156
  - ondersteunde talen 153
  - parsertype, selectie 145
  - starten 312
  - stoppen 312
  - systemstatus 312
  - taaldetectie 154
  - taalkundige verwerking 153
  - threads 142
  - witruimten verwijderen 156
- parserservers
  - threadconfiguratie 142
- parserTypes.cfg-bestand 145, 147
- plannen
  - crawlers 35, 40
  - index opbouwen 160, 161
- plug-ins, voor crawlers 108
- poortnummer, enterprise search 27
- populaire query's, bewaken 315
- portlets
  - description 271
  - enterprise search 271
  - integratie met WebSphere Portal 5.1 273

- portlets (*vervolg*)
  - integratie met WebSphere Portal 6 279
  - verwijderen uit WebSphere Portal 5.1 278
  - verwijderen uit WebSphere Portal versie 6 285
- prefixregels voor webcrawlers 84
- programma-ID's 237
- proxyservers 93

## Q

- query's met een hoge terughalwaarde
  - description 196
  - standaardwegingsfactoren 200
- query's met een lage terughalwaarde
  - description 196
  - standaardwegingsfactoren 200
- query-uitbreiding
  - description 168
  - invloed op indexgrootte 170
  - invloed op tijd voor opbouwen index 170
- queryvalidatie 241
- QuickPlace-crawlers
  - beveiligingsoptie Lokale gebruiker 261
  - configuratie 75
  - DIIOp-protocolconfiguratie 72
  - Directory Assistance-configuratie 262
  - Domino-gebruikersconfiguratie 261
  - instellen in Solaris-besturingsomgeving 69
  - instellen onder Windows 71
  - instellen op AIX-besturingsstelsysteem 69
  - instellen op Linux-besturingsstelsysteem 69
  - IOCP-configuratie 73
  - NRPC-protocol 69, 71
  - URI-indelingen 111

## R

- ranking van zoekresultaten
  - description 189
  - dynamisch 189
  - gewogen woordenboeken 194
  - op tekst gebaseerde score 189
  - runtime.properties-bestand 190
  - standaardwaarden herstellen 190
  - statisch 190
  - URI-patronen 194, 195
  - wegingsklassen 196, 198, 199, 200
- recente query's, bewaken 315
- recentelijk gecrawld URL's, bewaken 304
- regelgebaseerde categorieën
  - description 124
  - maken 128
  - type categorisering selecteren 127
- regelgebaseerde taxonomie, migreren vanuit WebSphere Portal 287
- removeCjNewlineChars-optie 156
- removeCjNewlineCharsMode-optie 156

- responstijd zoekopdrachten
  - alertsignalen over 324
  - bewaken 315
- retourcodes, enterprise search 337
- Robots Exclusion Protocol
  - identificatie van gebruikersagent 81
  - webcrawler, naleving 81
- robots.txt-bestanden
  - identificatie van gebruikersagent 81
  - webcrawler, naleving 81
- runtime-generic.properties-bestand 185

## S

- samengestelde termen, analyseren 143
- samengevoegen URI's
  - beveiligingsbeperkingen 268
  - configureren 175
  - description 174
- samengevoegen zoekresultaten
  - beveiligingsbeperkingen 268
  - configureren 175
  - description 174
- samenvattingen
  - aanpassen 184, 185
  - dynamisch 184
- schatting maken van systeemresources 299
- scopes
  - description 172
  - maken 173
  - URI-indelingen 111
  - URI-patronen 172, 173
  - zoeken 172
- scripts
  - esbackup.bat 334
  - esbackup.sh 334
  - escrm.sh 55
  - escrm.vbs 57
  - escrdb2.sh 48
  - escrdb2.vbs 48
  - escrnote.sh 69
  - escrnote.vbs 71
  - escrvbr.sh 43
  - escrvbr.vbs 44
  - esrestore.bat 335
  - esrestore.sh 335
  - startcl 335
- Search and Index API 11, 203
- SearchBarInclude.jsp-bestand 276
- semantische zoekopdracht 136, 139, 144
- serverwerkstand, Content Edition-repository's 43
- sessie-ID's, enterprise search 337
- SI-API (Search and Index API) 11, 203
- siapi.jar-bestand 273
- single-bytecodering 155
- siteregeversrapporten
  - description 304
  - maken 307
- sleutelruimtebestanden 224
- SMTP-serverconfiguratie 327
- snellinks
  - description 186
  - maken 186
  - URI-indelingen 111
  - zoeken 186

- soft error-pagina's, webcrawlers 90
- Solaris-besturingsomgeving
  - configuratie van DB2 Content Manager-crawler 55
  - Content Edition-crawler-configuratie 43
  - Domino Document Manager-crawler-configuratie 69
  - Notes-crawlerconfiguratie 69
  - QuickPlace-crawlerconfiguratie 69
- Solaris-besturingsstelsysteem
  - DB2-crawlerconfiguratie 48
- sorteerbare velden, beschrijving 164
- SQL Server-databases
  - toegang met DB2-crawlers 45
  - toegang met JDBC-databasecrawlers 62
- SSL, zoekserverconfiguratie 224
- SSO-beveiliging (Single Sign-On) configuratie 247
- Identiteitenbeheer 246
- standaardzoekprogramma 219
- start-URL's voor webcrawlers 84, 89
- startcl-script 335
- starten
  - Aanpassingstool Zoekprogramma 220
  - crawlerservers 302
  - Data Listener 317
  - enterprise search-servers 291
  - index opbouwen 313
  - migratiewizard 287
  - parsers 312
  - zoekprogramma's 223
  - zoekservers 296, 315
- statische ranking
  - description 190
  - in gemigreerde collecties 287
  - inschakelen voor een collectie 31
- Stellent-parser
  - description 145
  - documenttypen analyseren 146
  - documenttypes koppelen 149
- stellent.properties-bestand 149
- stellenttypes.cfg-bestand 149
- stellentTypes.cfg-bestand 145
- stoppen
  - crawlerservers 302
  - enterprise search-servers 291, 294
  - index opbouwen 313, 314
  - parsers 312
  - zoekservers 296, 315
- stopwoordenboeken
  - description 181
  - koppelen aan een collectie 183
  - toevoegen aan het stelsysteem 183
- synoniemenwoordenboeken
  - description 179
  - koppelen aan een collectie 181
  - toevoegen aan het stelsysteem 180
- Systeem (view), beschrijving 17
- stelsysteembackup 333, 334
- stelsysteemherstel 333, 335
- systeemresources
  - schatting 299
- stelsysteemstatus
  - collecties 300

stelsysteemstatus (*vervolg*)  
crawlers 302  
index opbouwen 314  
parsers 312  
Web-crawlers 304  
zoekerservers 315

## T

taakoverzicht, beheerconsole 17  
taalkundige ondersteuning  
aangepaste tekstanalyse 136  
codetabeldetectie 155  
gewogen woordenboeken 192  
locale 153  
n-gramsegmentering 156  
native XML-zoekopdrachten 144  
semantische zoekopdracht 136, 144  
stopwoordenboeken 181  
synoniemenwoordenboeken 179  
taalcodes 153  
taaldetectie 154  
witruimten verwijderen 156  
talen  
automatische detectie 154  
codes met twee tekens 153  
ondersteund 153, 154  
zoeken 153  
talen met eenvoudige tekst 153  
talen met ingewikkelde tekst 153  
tar-bestanden  
crawlen 110  
URI-indelingen 111  
taxonomieën, migreren vanuit WebSphere Portal 287  
tekstanalyse  
Common Analysis Structure 140, 141  
tekstanalyseprogramma's 137, 138  
XML-elementen toewijzen 139  
tekstanalyseprogramma's  
analyseresultaten toewijzen aan de index 140  
analyseresultaten toewijzen aan relationele databases 141  
Common Analysis Structure toewijzen aan relationele databases 141  
description 136  
koppelen aan collecties 138  
toevoegen aan het systeem 137  
XML-elementen toewijzen 139  
tekstverwerking  
annotators 136  
Common Analysis Structure 136  
tekstanalyseprogramma's 136  
threadgegevens, bewaken 304  
threads  
Parser 142  
webcrawler 305  
toegangsbesturing  
beveiliging op documentniveau 239  
description 233  
Identiteitenbeheer 242, 244  
SSO-beveiliging (Single Sign-On) 246  
uitschakelen voor een collectie 267  
validatie van actuele gebruiker 241  
vereisten voor Lotus Domino 259, 260

toegangsbesturing (*vervolg*)  
vereisten voor Windows-bestandsystemen 263  
toegankelijkheid 371  
toewijzen  
analyseresultaten toewijzen aan relationele databases 141  
Common Analysis Structure aan relationele databases 141  
Common Analysis Structures aan de index 140  
HTML-zoekvelden 134  
velden aan wegingsklassen 198  
XML-elementen aan de Common Analysis Structure 139  
XML-zoekvelden 130  
trefwoorden in snellinks 186  
Trusted Server-configuratie 260  
type categorisering  
regelgebaseerd 124  
selecteren 31, 127

## U

UIMA  
analyseresultaten toewijzen aan de index 140  
analyseresultaten toewijzen aan relationele databases 141  
Common Analysis Structure 140, 141  
Common Analysis Structure toewijzen aan de index 140  
Common Analysis Structure toewijzen aan relationele databases 141  
description 136  
koppelen aan collecties 138  
tekstanalyseprogramma's aan het systeem toevoegen 137  
XML-elementen toewijzen 139  
Unicode-codering 155  
UNIX-bestandssysteemcrawlers  
configuratie 79  
URI-indelingen 111  
URI-gegevens  
bewaken 301  
gewiste documenten 318  
URI's  
categorieregels 124, 128  
gegevens bekijken over 301  
indelingen in enterprise search 111  
samengevoegen in zoekresultaten 174, 175  
scopes 172, 173  
snellinks 186  
statische scores beïnvloeden 194, 195  
verwijderen uit de index 175  
URI's uit de index verwijderen 175  
URL-paddiepte 84  
URL's zo snel mogelijk bezoeken 89  
URL's zo snel mogelijk opnieuw bezoeken 89  
USC-tekenreeks 242

## V

validatie van actuele legitimatiegegevens 241, 259, 260, 263  
vbr\_access\_services.jar-bestand 43, 44  
velden toewijzen aan wegingsklassen 198  
veldzoekopdracht, beschrijving 164  
verificatie  
description 233  
uitschakelen voor bedrijfs-toepassingen 265  
view Collecties  
collecties maken 31  
description 17  
voorbeeldzoekprogramma  
beveiliging uitschakelen 265  
config.properties-bestand 206, 219  
description 12, 204  
HTTPS afdwingen 224  
klonen 221  
SSL afdwingen 224  
standaardimplementatie 219  
zoekfuncties 203, 204  
vrije zoekopdracht, beschrijving 164

## W

wachtwoord, enterprise search-beheerder 22, 23  
Web Content Management-crawlers  
configuratie 100  
URI-indelingen 111  
URL's van sites kopiëren 104  
Web-crawlers  
actieve sites 304, 306  
algemene crawlruimte 96  
beperken van crawlruimte 84  
bewaken 304  
configuratie 80  
cookie-indeling 95  
cookieconfiguratie 95  
cookies 94  
crawlerhistorie 304  
crawlregels 84  
crawl snelheid 306  
followindex.rules-bestand 98  
gebruikersagents 81  
interval voor opnieuw crawlen 89  
JavaScript-ondersteuning 84  
met wachtwoord beveiligde websites 91, 93  
no-follow (instructies) 98  
no-index (instructies) 98  
proxyservers 93  
rapporten maken over 307  
recentelijk gecrawelde URL's 304  
robots.txt-bestanden 81  
sitegegevens 304  
soft error-pagina's 90  
start-URL's 84, 89  
stelsysteemstatus 304  
threadgegevens 304, 305  
URL's zo snel mogelijk bezoeken 89  
URL-status 304  
webcrawlers  
gecrawelde documenten, datums 99

- WebSphere Application Server-gebruiker wachtwoord voor configuratie met een enkele servers 252
  - wachtwoord voor configuratie met meerdere servers 253
  - WebSphere II Classic Federation 53
  - WebSphere II Event Publisher Edition, DB2-crawlerconfiguratie 49
  - WebSphere II OmniFind Edition
    - API's 11
    - beheerconsole 10
    - componenten 3
    - crawlerservers 4
    - indexservers 7
    - integratie met WebSphere Portal 271
    - opdrachten 337
    - overzicht 1
    - parsers 5
    - poortnummerconfiguratie 27
    - retourcodes 337
    - sessie-ID's 337
    - stroomschema 13
    - wachtwoord wijzigen op één server 22
    - wachtwoord wijzigen op meerdere servers 23
    - zoekprogramma's 12
    - zoekservers 9
  - WebSphere MQ, crawlerserver-configuratie 48
  - WebSphere MQ, DB2-crawler-configuratie 52
  - WebSphere Portal
    - installatiescripts voor enterprise search 272
    - integratie met enterprise search 271
    - migratie van categoriestructuur 287
    - migratie van taxonomieën 287
    - migraties van collecties 287
    - standaardmigratie-instellingen 289
    - versie 5.1, configuratie van zoekbalk 276
    - versie 5.1, enterprise search verwijderen 278
    - versie 5.1, integratiescripts 273
    - versie 6, configuratie van zoekbalk 284
    - versie 6, configuratie van Zoekcentrum 282
    - versie 6, enterprise search verwijderen 285
    - versie 6, integratiescripts 279
    - zoekbalk, beschrijving 271
    - Zoekcentrum, beschrijving 271
  - WebSphere Portal-crawlers
    - configuratie 102
    - URI-indelingen 111
  - wegingsfactoren
    - voor gewogen woordenboeken 192
    - voor URI-patronen 194, 195
    - voor wegingsklassen 198, 200
  - wegingsklassen, configuratie 196, 199
  - wegingsklassen
    - configuratie 198, 199
    - description 196
    - duplicaatdocumenten, detectie 196
  - wegingsklassen (*vervolg*)
    - query's met een hoge terughaalwaarde 196, 200
    - query's met een lage terughaalwaarde 196, 200
    - standaardwaarden 200
    - velden toewijzen aan 198
  - Windows-bestandssysteemcrawler
    - configuratie 105
    - configuratie van beveiliging op documentniveau 263
    - URI-indelingen 111
  - Windows-besturingssysteem
    - configuratie van DB2 Content Manager-crawler 57
    - Content Edition-crawler-configuratie 44
    - crawlerconfiguratie 71
    - DB2-crawlerconfiguratie 48
    - Event Publishing-configuratie 48
  - Windows-domeinen 263
  - wissen
    - collecties 34
    - crawlers 39
    - indexen uit wachtrij 314
  - witruimten verwijderen 156
  - wp5\_install-script 273
  - wp5\_uninstall-script 278
  - wp6\_install-script 279
  - wp6\_uninstall-script 285
  - WpsMigratorLog-bestand 290
- ## X
- XML-documenten
    - native XML-zoekopdrachten 144
    - zoeken 130
  - XML-elementen
    - toewijzen aan de Common Analysis Structure 139
    - toewijzen aan zoekvelden 130
    - zoeken 130, 139
  - XML-fragmenten, native XML-zoekopdrachten 144
  - XML-queriesyntaxis, native 144
  - XML-zoekvelden
    - description 130, 139
    - elementen toewijzen aan 130, 139
    - maken 130
  - XPath, native XML-zoekopdrachten 144
- ## Z
- zip-bestanden
    - crawlen 110
    - URI-indelingen 111
  - zoekbalk, WebSphere Portal
    - versie 5.1, doorsturen naar enterprise search 276
    - versie 6, doorsturen naar enterprise search 284
  - zoekcache
    - configureren 178
    - description 178
  - Zoekcentrum, WebSphere Portal
    - description 271
  - Zoekcentrum, WebSphere Portal (*vervolg*)
    - versie 6, integratie met enterprise search 282
  - zoeken
    - categories 124
    - collecties 164
    - HTML-documenten 133, 134
    - snellinks 186
    - XML-documenten 130, 139
  - Zoeken (pagina), beschrijving 17
  - zoekopties
    - documentcontent 164
    - parametrische zoekopdracht 164
    - sorteerbare velden 164
    - veldzoekopdracht 164
    - volledige overeenkomst 164
    - voor zoekresultaten 164
    - vrije zoekopdracht 164
  - zoekportlet implementeren
    - es.wp5.install.jar-bestand 273
    - es.wp6.install.jar-bestand 279
    - wp5\_install-script 273
    - wp5\_uninstall-script 278
    - wp6\_install-script 279
    - wp6\_uninstall-script 285
  - zoekprogramma's
    - aangepast 203
    - aanpassen 220
    - beveiliging op collectieniveau 237
    - description 12
    - koppelen aan collecties 204
    - koppelen aan externe bronnen 230
    - programma-ID's 237
    - starten 223
    - toegang 223
    - voorbeeld 203, 204
  - zoekprogramma's aanpassen 219, 220
  - zoekresultaatvelden, beschrijving 164
  - Zoekresultaten
    - achteraf filteren 238
    - description 189
    - Dynamisch samenvatten 184
    - dynamische ranking 189
    - groeperen 174, 175
    - jokertekens 168
    - jokertekenuitbreiding 171
    - op tekst gebaseerde score 189
    - Ranking 195
    - samenvattingen 184, 185
    - samenvattingen aanpassen 184, 185
    - samenvouwen 174, 175, 268
    - statische ranking 190
    - URI-patroonconfiguratie 194
    - vooraf filteren 238
    - wegingsklassen, configuratie 196, 198, 199, 200
  - zoekservers
    - bewaken 315
    - description 9, 177
    - gewogen woordenboeken 192
    - historie van responstijd 315
    - HTTPS-configuratie 224
    - koppelen aan gewogen woordenboeken 193, 194
    - populaire query's 315
    - recente query's 315
    - SSL-configuratie 224



zoekservers (*vervolg*)  
starten 291, 296, 315  
stoppen 294, 296, 315  
stopwoordenboeken 181  
stopwoordenboeken koppelen 183  
synoniemenwoordenboeken 179, 180  
synoniemenwoordenboeken koppelen 181  
systeemstatus 315  
zoekcache 178  
zoekvelden met volledige overeenkomst,  
beschrijving 164







Gedrukt in Nederland



**Java**<sup>™</sup>  
**COMPATIBLE**

SC18-9283-03



Spine information:

OmniFind Enterprise Edition

**Versie 8.4**

**Enterprise Search beheren**

