



Verwaltung der Unternehmenssuche



Verwaltung der Unternehmenssuche

Hinweis

Vor Verwendung dieser Informationen und des darin beschriebenen Produkts sollten die Informationen in „Bemerkungen und Marken“ auf Seite 455 gelesen werden.

Diese Veröffentlichung ist eine Übersetzung des Handbuchs
IBM OmniFind Enterprise Edition Version 8.5, Administering Enterprise Search,
IBM Form SC18-9283-04,
herausgegeben von International Business Machines Corporation, USA

© Copyright International Business Machines Corporation 2008
© Copyright IBM Deutschland GmbH 2008

Informationen, die nur für bestimmte Länder Gültigkeit haben und für Deutschland, Österreich und die Schweiz nicht zutreffen, wurden in dieser Veröffentlichung im Originaltext übernommen.

Möglicherweise sind nicht alle in dieser Übersetzung aufgeführten Produkte in Deutschland angekündigt und verfügbar; vor Entscheidungen empfiehlt sich der Kontakt mit der zuständigen IBM Geschäftsstelle.

Änderung des Textes bleibt vorbehalten.

Herausgegeben von:
SW TSC Germany
Kst. 2877
Februar 2008

Inhaltsverzeichnis

ibm.com und zugehörige Ressourcen	vii	Löschen eines Crawlers	45
Senden von Kommentaren	vii	Crawlerzeitpläne	46
Kontaktaufnahme mit IBM	viii	Content Edition-Crawler	46
		Zugriff auf Content Edition-Repositorys im Direktmodus	48
		Zugriff auf WebSphere II Content Edition-Repositorys im Servermodus	49
Was ist Unternehmenssuche?	1	DB2-Crawler	52
Von der Unternehmenssuche unterstützte Datenquellentypen	2	Konfigurieren des Crawler-Servers unter UNIX für DB2-Crawler	54
Komponentenübersicht für die Unternehmenssuche	3	Konfigurieren des Crawler-Servers unter Windows für DB2-Crawler	55
Crawler für die Unternehmenssuche	4	Konfigurieren von WebSphere Information Integrator Event Publisher Edition für DB2-Crawler	56
Parser für die Unternehmenssuche	5	Konfigurieren von WebSphere MQ für DB2-Crawler	59
Indizes für die Unternehmenssuche	7	Durchsuchen von DB2-Datenbanken auf einem Server für klassische Datenquellen	61
Suchserver für die Unternehmenssuche	8	DB2 Content Manager-Crawler	62
Administrationskonsole für die Unternehmenssuche	9	Konfigurieren des Crawler-Servers unter UNIX für DB2 Content Manager-Crawler	63
Überwachen eines Systems für die Unternehmenssuche	10	Konfigurieren des Crawler-Servers unter Windows für DB2 Content Manager-Crawler	65
Protokolldateien für die Unternehmenssuche	10	Domino Document Manager-Crawler	66
Anpassen der Unternehmenssuche	11	Exchange Server-Crawler	69
Mustersuchanwendung für die Unternehmenssuche	12	JDBC-Datenbankcrawler	70
Der Datenfluss für die Unternehmenssuche	13	Beziehungszuordnungen für JDBC-Datenbanken	72
		Durchsuchen mehrerer strukturierter JDBC-Datenbanktabellen	74
Verwaltung von Systemen für die Unternehmenssuche	15	NNTP-Crawler	78
Anmelden an der Administrationskonsole	19	Notes-Crawler	79
Ändern des Kennworts des Administrators für die Unternehmenssuche in einer Einzelserverkonfiguration	20	Tipps für das Durchsuchen von Lotus Domino-Datenbanken	82
Ändern des Kennworts des Administrators für die Unternehmenssuche in einer Konfiguration mit mehreren Servern	21	Konfigurieren des Crawler-Servers unter UNIX für das Durchsuchen von Lotus Domino-Quellen	83
TCP-Portnummern für die Unternehmenssuche	24	Konfigurieren des Crawler-Servers unter Windows für das Durchsuchen von Lotus Domino-Quellen	85
Ändern der Portnummer für das System für die Unternehmenssuche	25	Konfigurieren von Servern, die das DIIOP-Protokoll verwenden	86
Ändern von Hostnamen oder IP-Adressen von Servern für die Unternehmenssuche	26	Konfigurieren des I/O Completion Ports unter AIX für das Durchsuchen von Lotus Domino-Quellen	88
Konfigurieren der Unterstützung für doppelte IP-Adressen	28	QuickPlace-Crawler	89
Aktivieren der Unterstützung für das IPv6-Protokoll	29	Crawler für Einstiegspunklisten	93
		Crawler für UNIX-Dateisysteme	95
		Web-Crawler	96
		Benutzeragentenkonfiguration	97
		Vorgehensweise des Web-Crawlers bei der Verwendung des Protokolls zum Sperren von Webseitbereichen für Robots	98
		Unterstützung für JavaScript	100
		Regeln zum Begrenzen des Web-Crawler-Bereichs	101
		Testen von URL-Verbindungen mit dem Web-Crawler	105
Objektgruppen für die Unternehmenssuche	33		
Erstellen einer Objektgruppe mit dem Objektgruppenassistenten	33		
Erstellen einer Objektgruppe mit der Sicht 'Objektgruppen'	35		
Bearbeiten einer Objektgruppe	37		
Löschen einer Objektgruppe	38		
Ermitteln der Objektgruppen-ID	39		
Crawlerverwaltung	41		
Erstellen eines Crawlers	43		
Bearbeiten von Crawlermerkmalen	44		
Bearbeiten eines Crawlerbereichs	44		

Intervalleinstellungen für erneute Suche im Web-Crawler	106
Optionen für das Aufrufen von URL-Adressen mit dem Web-Crawler	106
Vorgehensweise des Web-Crawlers bei Seiten mit detaillierten Fehlerhinweisen	107
Unterstützung für die Crawlersuche auf sicheren Websites.	109
Über Proxy-Server bereitgestellte Websites.	111
Cookieverwaltung.	112
Konfiguration eines globalen Web-Crawler-Bereichs	114
Die Anweisungen 'no-follow' und 'no-index' Außerkräftsetzen der Anweisungen 'no-follow' und 'no-index' auf Webseiten	116
Konfigurieren des vom Web-Crawler für durchsuchte Dokumente verwendeten Datums	117
Web Content Management-Crawler	118
WebSphere Portal-Crawler	120
Kopieren der zu durchsuchenden URL-Adresse von WebSphere Portal	122
Crawler für das Windows-Dateisystem	123
Konfigurieren der Unterstützung für Anwendungen der Datenempfangsfunktion	125
Benutzerdefinierte Crawler-Plug-ins	127
Unterstützung für die Crawlersuche in Archivdateien	129
URI-Formate in einem Index für die Unternehmenssuche	130

Parserverwaltung. 141

Funktionsweise von Kategorien	142
Regelbasierte Kategorien	143
Kategoriebäume	145
Auswählen des Kategorisierungstyps	145
Konfigurieren von Kategorien	146
Funktionsweise von XML-Suchfeldern	147
XML-Suchfelder	148
Zuordnen von XML-Elementen zu Suchfeldern	149
Funktionsweise von HTML-Suchfeldern	151
HTML-Suchfelder	151
Zuordnen von HTML-Metadatenelementen zu Suchfeldern	152
Benutzerdefinierte Textverarbeitung	154
Hinzufügen von Textanalysesteuerkomponenten zum System	156
Zuordnen einer Textanalysesteuerkomponente zu einer Objektgruppe	157
Zuordnen von XML-Elementen zur allgemeinen Analysestruktur	158
Zuordnen der allgemeinen Analysestruktur zum Index	159
Zuordnen der allgemeinen Analysestruktur zu einer relationalen Datenbank	160
Konfigurieren von Threads für den Parser-Service	162
Aktivieren der erweiterten Analyse für zusammengesetzte Begriffe	162
Aktivieren der Unterstützung für eine native XML-Suche	163

Dokumentformaterkennung 165

Standardmäßig unterstützte Dokumenttypen	166
Dokumenttypen für Objektgruppenparser und Stellent-Parser	167
Zuordnen von Dokumenttypen zu einem Objektgruppenparser	168
Standardserviceregeln für Objektgruppenparser	169
Syntaxanalyse für unbekannte Dokumenttypen	170
Ändern der Ersetzungsregeln für einige HTML-Tags	171
Standard-HTML-Ersetzungsregeln	172
Zuordnen von Dokumententypen zu einem Stellent-Parser	173
Standardsyntaxanalyserregeln für Stellent-Parser	176

Sprachen- und Codepage-Unterstützung 181

Automatische Spracherkennung	182
Automatische Codepage-Erkennung.	183
Linguistische Analyse von Dokumenten in chinesischer, japanischer und koreanischer Sprache	184
N-Gram-Segmentierung	184
Entfernen von Leerraum aus Text	185

Indexverwaltung 187

Terminieren von Indexerstellungen	188
Ändern des Indexzeitplans	189
Aktivieren und Inaktivieren der Indexzeitpläne	189
Konfigurieren von gleichzeitig ablaufenden Indexerstellungen	190
Indexerstellung nur nach der Erkennung von Änderungen.	192
Stoppen von Indexerstellungen	193
Optionen zum Beeinflussen der durchsuchbaren Sicht des Index.	194
Indexierte Optionen für das Durchsuchen von Dokumenten	194
Erkennung von doppelten Dokumenten	199
Platzhalterzeichen in Abfragen	199
Bereiche	203
Konfigurieren von Bereichen	204
Ausgeblendete URIs	205
Ausblenden von URIs in den Suchergebnissen	206
Entfernen von URIs aus dem Index	207

Verwaltung des Suchservers. 209

Suchcaches	210
Konfigurieren eines Suchcaches	210
Benutzerdefinierte Synonymverzeichnisse	210
Hinzufügen von Synonymverzeichnissen zum System	212
Zuordnen eines Synonymverzeichnisses zu einer Objektgruppe	213
Benutzerdefinierte Verzeichnisse von Stoppwörtern	213
Hinzufügen von Verzeichnissen von Stoppwörtern zum System	215
Zuordnen eines Verzeichnisses von Stoppwörtern zu einer Objektgruppe	215
Erneutes Implementieren von benutzerdefinierten Wörterverzeichnissen.	216
Dynamische Zusammenfassung	217

Anpassen von Dokumentzusammenfassungen in der Administrationskonsole	218
Anpassen von Dokumentzusammenfassungen durch Bearbeitung von Merkmalen	219
Funktionsweise von Quick Links	220
Quick Links	220
Konfigurieren von Quick Links	220

Dokumentrangfolge. 223

Textbasierte Bewertung	223
Statische Rangfolge	224
Wiederherstellung der Standardwerte für die statische Dokumentrangfolge	225
Benutzerdefinierte Verzeichnisse von Boostwörtern	227
Hinzufügen von Verzeichnissen von Boostwörtern zum System	229
Zuordnen eines Verzeichnisses von Boostwörtern zu einer Objektgruppe	229
Auf URI-Mustern basierende Dokumentrangfolge	230
Beeinflussen der Bewertungen von Dokumenten, die mit URI-Mustern übereinstimmen	231
Auf Boostklassen basierende Dokumentrangfolge	232
Zuordnen von Feldern zu Boostklassen	233
Konfigurieren von Boostfaktoren für Boostklassen	235
Standardwerte für Boostklassen	235

Suchanwendungen für die Unternehmenssuche. 239

Zuordnen von Suchanwendungen zu Objektgruppen	240
Funktionen der Mustersuchanwendung	240
Suchanwendungsmerkmale	242
Bearbeiten der Merkmale der Mustersuchanwendung	260
Anpassen von Suchanwendungen	261
Klonen der Mustersuchanwendung	263
Analysieren der häufigsten Ergebnisse	265
Zugreifen auf Suchanwendungen	268
Konfigurieren der Suchserver für die ausschließliche Annahme von sicheren Suchanforderungen (SSL)	269
Konfigurieren der Suchserver für die Annahme von Anforderungen über einen Proxy-Server	271

Unterstützung für externe Quellen 273

Hinzufügen von externen Quellen zum System	273
Zuordnen von Suchanwendungen zu externen Quellen	276

Sicherheit für die Unternehmenssuche 277

Installationssicherheit	279
Authentifizierung und Zugriffssteuerung	280
Verwaltungsrollen	280
Konfigurieren von Benutzern mit Verwaltungsaufgaben	281
Sicherheit auf Objektgruppenebene	282
Analyse von doppelten Dokumenten und Objektgruppensicherheit	282
Suchanwendungs-IDs	283

Sicherheit auf Dokumentebene	284
Vor- und Nachfiltern von Suchergebnissen	284
Prüfung durch gespeicherte Sicherheitstoken	286
Prüfung aktueller Berechtigungsnachweise während der Abfrageverarbeitung	287
Ankertextanalyse	295
Aktivieren der Sicherheit für die Unternehmenssuche	297
Konfigurieren der globalen Sicherheit und einer LDAP-Benutzerregistry in WebSphere Application Server	298
Aktivieren der Sicherheit für ein System für die Unternehmenssuche mit einem Server	300
Aktivieren der Sicherheit für ein System für die Unternehmenssuche mit mehreren Servern	301
Anforderungen der Crawlerkonfiguration zur Unterstützung der Sicherheit	302
Zugriffsprüfung zum Schutz von Exchange Server-Dokumenten	307
Umsetzung der Sicherheit auf Dokumentebene für Lotus Domino-Dokumente	307
Umsetzung der Sicherheit auf Dokumentebene für Dokumente des Windows-Dateisystems	311
Inaktivieren der Sicherheit für die Unternehmenssuche	314
Inaktivieren der Sicherheit für eine Unternehmensanwendung in WebSphere Application Server	314
Inaktivieren der Sicherheit auf Dokumentebene	316
Inaktivieren der Sicherheit für ausgeblendete Suchergebnisse	317

Starten und Stoppen eines Systems für die Unternehmenssuche 319

Starten eines Systems für die Unternehmenssuche	319
Stoppen eines Systems für die Unternehmenssuche	321
Steuern, welche Komponenten gestartet oder gestoppt werden sollen	323
Verwalten der Suchserver im Standalone-Modus	324

Überwachen der Aktivitäten der Unternehmenssuche. 327

Schätzen der Anzahl Dokumente in einer Objektgruppe	327
Überwachen einer Objektgruppe	328
Anzeigen von Details zu einem URI	328
Überwachen von Crawlern	330
Anzeigen von Details zur Web-Crawler-Aktivität	331
Details von Web-Crawler-Threads	332
Aktive Web-Crawler-Sites	333
Suchrate des Web-Crawlers	334
Erstellen von Web-Crawler-Berichten	334
An den Web-Crawler zurückgegebene HTTP-Statuscodes	336
Überwachen des Parsers	340
Überwachen der Indexaktivität für eine Objektgruppe	341
Überwachen der Indexierungwarteschlange für die Unternehmenssuche	342
Überwachen der Suchserver	343

Ändern der Berechnung von Abfragestatistiken	345
Überwachen der Datenempfangsfunktion	346
Dokumentüberwachung	346
Konfigurieren von Protokolldateien für die Dokumentüberwachung	347
Anzeigen von Berichten über gelöschte Dokumente	348
Anzeigen von Protokolldateien zu gelöschten Dokumenten	349

Protokolldateien und Alerts 351

Alerts	351
Konfigurieren von Alerts auf Objektgruppenebene	352
Konfigurieren von Alerts auf Systemebene	354
Konfigurieren von Protokolldateien	354
Konfigurieren von SMTP-Serverinformationen	356
Empfangen von E-Mail zu protokollierten Nachrichten	357
Ändern der Größe des Abfrageprotokolls	359
Anzeigen von Protokolldateien	360

Durchführen von Backup und Restore für ein System für die Unternehmenssuche. 363

Sichern des Systems für die Unternehmenssuche	364
Wiederherstellung des Systems für die Unternehmenssuche	365
Exportieren und Importieren von Objektgruppenkonfigurationen	366

Integration in Lotus Notes Version 8 371

Erstellen der Update-Site für das Plug-in für die Unternehmenssuche	371
Installieren des Plug-ins für die Unternehmenssuche in einem Lotus Notes-Client der Version 8	371

Integration in WebSphere Portal 373

Setup-Scripts zum Integrieren der Unternehmenssuche in WebSphere Portal	374
Konfigurieren der Unternehmenssuche in WebSphere Portal Version 5.1	375
Konfigurieren der Suchleiste von WebSphere Portal Version 5.1 Search für die Verwendung der Unternehmenssuche	378
Entfernen der Unternehmenssuche aus WebSphere Portal Version 5.1	380

Konfigurieren der Unternehmenssuche in WebSphere Portal Version 6	381
Konfigurieren der Suchzentrale von WebSphere Portal Version 6 für die Unternehmenssuche	384
Konfigurieren der Suchleiste von WebSphere Portal Version 6 Search für die Verwendung der Unternehmenssuche	385
Konfigurieren des Portlets für die Unternehmenssuche für Lotus Quickr	387
Entfernen der Unternehmenssuche aus WebSphere Portal Version 6	388
Integration der Unternehmenssuche in WebSphere Portal-Clustersysteme	389
Konfigurieren der Unternehmenssuche in einem WebSphere Portal-Clustersystem	390
Entfernen der Unternehmenssuche aus einem WebSphere Portal-Clustersystem	393

Migration von WebSphere Portal in die Unternehmenssuche. 397

Migration einer Objektgruppe aus WebSphere Portal	397
Einstellungen für migrierte Objektgruppen	399
Protokolldatei des Migrationsassistenten	400

Befehle, Rückkehrcodes und Sitzungs-IDs der Unternehmenssuche . . 401

Groß-/Kleinschreibung bei der Unternehmenssuche. 433

Dokumentation für die Unternehmenssuche. 437

Funktionen zur behindertengerechten Bedienung 439

Glossar der Begriffe für die Unternehmenssuche. 441

Bemerkungen und Marken. 455

Bemerkungen	455
Marken	457

Index 459

ibm.com und zugehörige Ressourcen

Produktunterstützung und -dokumentation ist unter [ibm.com](http://www.ibm.com) verfügbar.

Unterstützung

Produktunterstützung ist im Web verfügbar.

IBM OmniFind Enterprise Edition

<http://www.ibm.com/software/data/enterprise-search/omnifind-enterprise/support.html>

IBM OmniFind Discovery Edition

<http://www.ibm.com/software/data/enterprise-search/omnifind-discovery/support.html>

IBM OmniFind Yahoo! Edition

<http://www.ibm.com/software/data/enterprise-search/omnifind-yahoo/support.html>

Informationszentrale

Sie können die Produktdokumentation mit einem Web-Browser in einer Eclipse-basierten Informationszentrale anzeigen. Sie finden die Informationszentrale unter <http://publib.boulder.ibm.com/infocenter/discover/v8r5m0/>.

PDF-Veröffentlichungen

Sie können die PDF-Dateien online anzeigen, indem Sie Adobe Acrobat Reader für Ihr Betriebssystem verwenden. Wenn Sie Acrobat Reader nicht installiert haben, können Sie ihn von der Adobe-Website unter <http://www.adobe.com> herunterladen.

Rufen Sie die folgenden Websites mit PDF-Veröffentlichungen auf:

Produkt	Adresse der Website
OmniFind Enterprise Edition Version 8.5	http://www.ibm.com/support/docview.wss?rs=63&uid=swg27010938
OmniFind Discovery Edition Version 8.4	http://www.ibm.com/support/docview.wss?rs=3035&uid=swg27008552
OmniFind Yahoo! Edition Version 8.4	http://www.ibm.com/support/docview.wss?rs=3193&uid=swg27008932

Senden von Kommentaren

Ihre Rückmeldung ist eine wichtige Hilfe dabei, präzise und qualitativ hochwertige Informationen bereitzustellen.

Senden Sie Ihre Kommentare, indem Sie das online verfügbare Formular für Leserkommentare unter https://www14.software.ibm.com/webapp/iwm/web/signup.do?lang=en_US&source=swg-rcf verwenden.

Kontaktaufnahme mit IBM

Unter 0180 3 313233 erreichen Sie Hallo IBM, wo Sie Antworten zu allgemeinen Fragen erhalten.

Telefonische Unterstützung erhalten Sie über folgende Nummern:

- Unter 0180 3 313233 erreichen Sie Hallo IBM, wo Sie Antworten zu allgemeinen Fragen erhalten.
- Unter 0180 5 426014 erreichen Sie die DB2 Helpline, wo Sie Antworten zu DB2-spezifischen Problemen erhalten.

Weitere Informationen dazu, wie Sie mit IBM in Kontakt treten, finden Sie auf der IBM Website **Kontakt** unter <http://www.ibm.com/contact/de/>.

Was ist Unternehmenssuche?

Ein System für die Unternehmenssuche stellt erweiterte Funktionen für das Durchsuchen einer beliebigen Anzahl strukturierter und unstrukturierter Datenquellen in einer einzelnen Abfrage bereit. Schnelle Abfrageantwortzeiten und eine konsolidierte Ergebnisliste mit Rangordnung, die auf einer umfangreichen Textanalyse basiert, ermöglichen Ihnen nicht nur das Suchen nach interessanten Dokumenten sondern auch das Extrahieren von aussagekräftigem Dokumentinhalt.

Die Komponenten für die Unternehmenssuche, die mit IBM OmniFind Enterprise Edition installiert werden, erfassen Informationen aus Ihrem gesamten Unternehmen. Durch Eingabe einer Abfrage in einen Web-Browser können Sie lokale und ferne Datenbanken, Collaborationsysteme, Content-Management-Systeme, Dateisysteme und interne und externe Websites gleichzeitig durchsuchen.

Ein System für die Unternehmenssuche, das für die nahtlose Integration in Ihre vorhandenen Systeme entworfen wurde, ist für die Logistik zuständig, die erforderlich ist, um Daten von verschiedenen Quellen zu erfassen und die Daten für ein schnelles Abrufen zu indexieren. Die Unternehmenssuche kann durch die Anwendung linguistischer Analysen und anderer Analysetypen auf Daten äußerst relevante Suchergebnisse liefern. Sie müssen sich nicht mit unterschiedlichen Schnittstellen befassen, um verschiedene Repository-Typen zu durchsuchen.

Sie können Unterstützung für das Durchsuchen von Datenquellen hinzufügen, die Sie nicht in einen Index für die Unternehmenssuche einschließen möchten. Mit der föderierten Suchfunktion der Unternehmenssuche können Sie beim Durchsuchen von indexierten Datenquellen diese externen Quellen gleichzeitig mit durchsuchen.

Suchqualität

OmniFind Enterprise Edition unterstützt IBM Unstructured Information Management Architecture (UIMA), um sicherzustellen, dass Benutzer die gesuchten Informationen finden. UIMA ist ein offenes Framework, das eine allgemeine Standardschnittstelle für die Textanalyse definiert. Dank einer umfassenden semantischen Analyse ist die Unternehmenssuche in der Lage, Konzepte, verborgene Bedeutungen, Abhängigkeiten, Fakten und andere relevante Daten zu erkennen, die häufig in unstrukturierten Texten verborgen vorkommen. Mit den bei der Analyse extrahierten Informationen lässt sich die Qualität von Suchergebnissen oder die Qualität von anderen Anwendungen (z. B. Business-Intelligence und Data-Mining) verbessern.

Sicherheit

Die Sicherheit ist ein fest integriertes Element für die Unternehmenssuche. Nur Benutzer, die für die Verwaltung des Systems berechtigt sind, können Verwaltungsaufgaben ausführen. Sie können mit den in IBM WebSphere Application Server verfügbaren Sicherheitsmechanismen Verwaltungsrollen konfigurieren und steuern, welche Benutzer Zugriff auf verschiedene Verwaltungsfunktionen haben.

Sie können außerdem Optionen angeben, um Daten bei der Datenerfassung Sicherheitstoken zuzuordnen. Wenn Ihre Suchanwendungen die Sicherheit aktivieren, können Sie mit diesen Token, die mit Dokumenten im Index gespeichert werden,

die Zugriffssteuerung umsetzen und sicherstellen, dass nur Benutzer mit den entsprechenden Berechtigungsnachweisen die Daten abfragen und Suchergebnisse anzeigen können.

Bei bestimmten Datenquellentypen können Sie Optionen konfigurieren, um die Anmeldeberechtigungen eines Benutzers während der Abfrageverarbeitung mit aktuellen Zugriffssteuerungen prüfen. Mit dieser zusätzlichen Sicherheitsschicht wird sichergestellt, dass die Prüfung der Zugriffsrechte eines Benutzers in Echtzeit mit der nativen Datenquelle erfolgt. Diese Option kann Instanzen verhindern, in denen sich die Berechtigungsnachweise eines Benutzers ändern, nachdem ein Dokument und die zugehörigen Sicherheitstoken indexiert wurden.

Lernprogramm für das Produkt

Ein Onlinelernprogramm steht unter <http://www.ibm.com/developerworks/edu/dm-dw-dm-0503buehler-i.html> zur Verfügung. Im Lernprogramm werden Installations- und Konfigurationsschritte beschrieben. Außerdem wird gezeigt, wie Sie verschiedene Datenquellentypen durchsuchen können, und es wird beschrieben, wie Sie die Anwendungsprogrammierschnittstellen des Produkts verwenden können, um die Unternehmenssuche zu erweitern. Das Lernprogramm wurde für eine ältere Version von OmniFind Enterprise Edition erstellt. Viele Konzepte und Prozeduren sind jedoch immer noch gültig.

Zugehörige Konzepte

„Komponentenübersicht für die Unternehmenssuche“ auf Seite 3

„Der Datenfluss für die Unternehmenssuche“ auf Seite 13

Sicherheit für die Unternehmenssuche



Integration der benutzerdefiniert Textanalyse



Basiskonzepte bei der Verarbeitung der Textanalyse

Zugehörige Verweise

„Von der Unternehmenssuche unterstützte Datenquellentypen“

Von der Unternehmenssuche unterstützte Datenquellentypen

Für das Durchsuchen einer Vielzahl von Datenquellentypen ist eine vordefinierte Unterstützung verfügbar.

Nach der Installation von IBM OmniFind Enterprise Edition können Sie mit dem Erfassen von Daten aus den folgenden Datenquellentypen beginnen:

- IBM DB2 Content Manager-Elementtypen (Dokumente, Ressourcen und Elemente)
- IBM DB2-Datenbanken
- IBM Domino Document Manager-Datenbanken (früher Domino.Doc)
- IBM Lotus Notes-Datenbanken
- IBM Lotus QuickPlace-Datenbanken
- IBM Lotus Quickr-Inhaltsbibliotheken
- IBM WebSphere Information Integrator Content Edition-Repositorys, einschließlich Documentum, FileNet Panagon Content Services, FileNet P8 Content Manager, Hummingbird Document Management (DM), Microsoft SharePoint, Open-Text Livelink Enterprise Server und WebSphere Portal Document Manager (PDM)

- IBM WebSphere Information Integrator-Kurznamentabellen für viele Datenbanksystemtypen, einschließlich IBM DB2 für z/OS, IBM Informix, Microsoft SQL Server, Oracle und Sybase
- IBM WebSphere Portal-Sites
- IBM Workplace Web Content Management-Sites
- Allgemein zugängliche Ordner von Microsoft Exchange Server
- Microsoft SQL Server-Datenbanken
- Microsoft Windows-Dateisysteme
- NNTP-Newsgruppen (NNTP - Network News Transfer Protocol)
- Oracle-Datenbanken
- UNIX-Dateisysteme
- Websites im Internet oder in Ihrem Intranet

Sie können außerdem Unterstützung für das Durchsuchen der folgenden externen Quellentypen hinzufügen, ohne dem Index für die Unternehmenssuche Dokumente dieser Quellen hinzufügen zu müssen:

- Datenbanken, die das JDBC-Protokoll (Java Database Connectivity) unterstützen (nur DB2- und Oracle-Datenbanksysteme). Für jede Tabelle, die Sie für die Suche aktivieren, wird eine separate externe Quelle erstellt.
- LDAP-Server (Lightweight Directory Access Protocol). Für jeden LDAP-Server wird eine externe Quelle erstellt.

Die neuesten Informationen zu unterstützten Datenquellentypen und zu den unterstützten Produktversionen finden Sie auf der Seite mit Systemvoraussetzungen auf der Unterstützungswebsite für OmniFind Enterprise Edition.

Zugehörige Konzepte

Unterstützung für externe Quellen

„Der Datenfluss für die Unternehmenssuche“ auf Seite 13

Komponentenübersicht für die Unternehmenssuche

Die Komponenten für die Unternehmenssuche erfassen Daten aus Ihrem gesamten Unternehmen, analysieren, bewerten und kategorisieren die Informationen und erstellen einen Index, den Benutzer durchsuchen können.

Eine *Objektgruppe* für die Unternehmenssuche stellt die Gruppe von Quellen dar, die Benutzer in einer einzelnen Abfrage durchsuchen können. Wenn Sie eine Objektgruppe erstellen, geben Sie an, welche Quellen eingeschlossen werden sollen, und Sie konfigurieren die Optionen dafür, wie Benutzer die indexierten Daten durchsuchen können.

Sie können mehrere Objektgruppen erstellen, und jede Objektgruppe kann Daten von einer Vielzahl von Datenquellen enthalten. Sie können z. B. eine Objektgruppe erstellen, die Dokumente aus IBM DB2 Universal Database-, IBM Lotus Notes- und IBM DB2 Content Manager-Datenbanken enthält. Wenn Benutzer diese Objektgruppe durchsuchen, schließen die Suchergebnisse potenziell Dokumente aus jeder der Datenquellen ein.

Da ein Durchsuchen von föderierten Objektgruppen unterstützt wird, können Benutzer mehrere Objektgruppen in einer einzigen Abfrage durchsuchen. Die Suchergebnisse enthalten möglicherweise Dokumente aus allen Objektgruppen und externen Quellen in Ihrem System für die Unternehmenssuche.

Zum Erstellen und Verwalten einer Objektgruppe gehören die folgenden Aktivitäten:

Erfassen von Daten

Die *Crawler*-Komponenten erfassen Dokumente aus Datenquellen entweder kontinuierlich oder anhand eines von Ihnen angegebenen Zeitplans. Häufig ausgeführte Crawler-Suchen stellen sicher, dass Benutzer immer auf die neuesten Informationen zugreifen können.

Analysieren von Daten

Die *Syntaxanalyse*-Komponenten extrahieren Text aus den Dokumenten und führen linguistische Analysen und andere Analysetypen für jedes Dokument aus, das ein Crawler durchsucht. Die detaillierte Inhaltsanalyse verbessert die Qualität der Suchergebnisse.

Indexieren von Daten

Die *Index*-Komponenten werden auf regelmäßig terminierter Basis ausgeführt, um dem Index Informationen zu neuen und geänderten Dokumenten hinzuzufügen. Die Indexkomponenten führen auch globale Analysen für die Dokumente in einer Objektgruppe aus, um die Qualität der Suchergebnisse zu verbessern.

Suchen nach Daten

Die *Such*-Komponenten durchsuchen den Index und arbeiten zusammen mit Ihren Suchanwendungen, um Suchanforderungen zu verarbeiten und Suchergebnisse zurückzugeben.

Andere OmniFind Enterprise Edition-Komponenten ermöglichen es Ihnen, Sicherheitseinstellungen anzugeben, die Systemaktivität zu überwachen und aufgetretene Fehler zu beheben. Das Produkt stellt außerdem ein funktionierendes Beispiel einer Suchanwendung bereit, das Sie als Schablone verwenden können, um Ihre eigenen Suchanwendungen zu erstellen.

Zugehörige Konzepte

Crawlerverwaltung

Parserverwaltung

Indexverwaltung

Suchserververwaltung

„Der Datenfluss für die Unternehmenssuche“ auf Seite 13

Crawler für die Unternehmenssuche

Crawler für die Unternehmenssuche erfassen Dokumente aus Datenquellen, sodass die Dokumente analysiert, indiziert und durchsucht werden können.

Die Crawlerkomponente, die zusammen mit OmniFind Enterprise Edition bereitgestellt wird, verfügt über die folgenden Funktionen:

- Wenn Sie einen Crawler konfigurieren, suchen die *Discovery*-Prozesse Informationen zu den Quellen, die für das Durchsuchen verfügbar sind, wie z. B. die Namen aller Sichten und Ordner in einer Lotus Notes-Datenbank oder die Namen aller Dateisysteme auf einem UNIX-Server.
- Nachdem Sie die Quellen, die Sie durchsuchen wollen, ausgewählt und den Crawler gestartet haben, erfassen die Crawlerkomponenten Daten von den Quellen, sodass die Daten analysiert und indiziert werden können.

Eine einzelne Objektgruppe kann über mehrere Crawler verfügen, und jeder Crawler ist dazu entworfen, Daten von einem bestimmten Datenquellentyp zusammen-

zustellen. Sie können z. B. drei Crawler erstellen, um Daten von Dateisystemen, Notes-Datenbanken und relationalen Datenbanken in derselben Objektgruppe zu kombinieren. Oder Sie könnten verschiedene Crawler desselben Typs erstellen und für sie unterschiedliche Zeitpläne für die Crawleruche entsprechend der Änderungshäufigkeit der von dem jeweiligen Crawler erfassten Daten festlegen.

Nachdem Sie den Web-Crawler gestartet haben, wird er fortlaufend ausgeführt. Sie geben an, welche URL-Adressen (URL - Uniform Resource Locator) Sie durchsuchen wollen, und der Crawler kehrt regelmäßig zurück, um zu prüfen, ob neue oder geänderte Daten vorhanden sind. Sie können andere Crawlertypen manuell starten oder stoppen, oder Sie können Zeitpläne für die Crawleruche konfigurieren. Wenn Sie einen Crawler terminieren, geben Sie an, wann er das erste Mal ausgeführt wird und wie oft er die Datenquellen erneut aufrufen soll, um neue oder geänderte Dokumente zu durchsuchen.

Crawlermerkmale sind eine Gruppe von Regeln, die das Verhalten eines bestimmten Crawlers regeln, wenn er eine Crawleruche ausführt. Sie geben z. B. Regeln an, um zu steuern, wie der Crawler Systemressourcen verwendet. Die Gruppe der Quellen, die zum Durchsuchen auswählbar ist, stellt den *Crawlerbereich* eines Crawlers dar. Nach dem Erstellen eines Crawlers können Sie die Crawlermerkmale jederzeit bearbeiten, um die Datenerfassung durch den Crawler zu ändern. Sie können auch den Crawlerbereich bearbeiten, um den Crawlerzeitplan zu ändern, neue Quellen hinzuzufügen oder Quellen zu entfernen, die Sie nicht mehr durchsuchen wollen.

Zugehörige Konzepte

Crawlerverwaltung

Zugehörige Tasks

Überwachen von Crawlern

Parser für die Unternehmenssuche

Ein Parser für die Unternehmenssuche analysiert Dokumente, die durch einen Crawler erfasst wurden, und bereitet sie zur Indexierung vor.

Die Parserkomponente, die zusammen mit OmniFind Enterprise Edition ausgeliefert wird, analysiert Dokumentinhalte und Dokumentmetadaten. Sie speichert die Ergebnisse der Analyse in einem Datenspeicher, damit die Indexkomponente darauf zugreifen kann. Der Parser führt die folgenden Tasks aus:

- Extrahiert Text aus dem jeweiligen Dokumentformat. Der Parser extrahiert z. B. Text aus den Tags in XML- und HTML-Dokumenten. Wenn Sie Oracle (früher Stellent) Outside In Content Access für IBM OmniFind Enterprise Edition verwenden, extrahiert der Parser auch Text aus Binärformaten, wie z. B. Microsoft Word- und Adobe Acrobat PDF-Dokumente.
- Ermittelt die Zeichensatzcodierung jedes Dokuments. Vor der linguistischen Analyse verwendet der Parser diese Informationen, um den gesamten Text in Unicode umzuwandeln.
- Ermittelt die Quellsprache jedes Dokuments.
- Extrahiert Text und fügt Token hinzu, um die Abrufbarkeit von Daten zu verbessern. Der Parser führt während dieser Phase die folgenden Tasks aus:
 - Zeichennormalisierung, wie z. B. die Normalisierung der Großschreibung und der diakritischen Zeichen, wie z. B. der deutschen Umlaute.
 - Analyse der Struktur von Absätzen, Sätzen, Wörtern und Leerzeichen. Durch die linguistische Analyse zerlegt der Parser Wortkomposita und ordnet Token zu, die die Suchfunktion für Wörterbücher und Synonyme aktivieren.

- Wendet die Syntaxanalyseregeln an, die Sie für die Objektgruppe angegeben haben. Bei der Parserkonfiguration können Sie die folgenden Syntaxanalyseaktivitäten konfigurieren:

Feldzuordnungsregeln für XML- und HTML-Dokumente

Mit dieser Option können Benutzer strukturierten und unstrukturierten Inhalt in XML- und HTML-Dokumenten durchsuchen. Wenn Sie XML-Elemente oder HTML-Metadatenelemente Suchfeldern im Index für die Unternehmenssuche zuordnen, können Benutzer die Feldnamen in Abfragen angeben und bestimmte Teile von XML- und HTML-Dokumenten durchsuchen. Abfragen, die bestimmte Felder durchsuchen, können präzisere Suchergebnisse liefern als Freiformatabfragen, die den gesamten Dokumentinhalt durchsuchen.

Kategorien

Mit dieser Option können Benutzer anhand der Kategorien, zu denen die Dokumente gehören, Dokumente durchsuchen. Benutzer können Kategorien auch in den Suchergebnissen auswählen und nur nach Dokumenten suchen, die zur selben Kategorie gehören.

Wenn Sie eine Objektgruppe erstellen, wählen Sie aus, ob Sie eine Kategorisierung verwenden wollen. Bei *regelbasierten* Kategorien werden Dokumente den Kategorien entsprechend den definierten Regeln zugeordnet. Sie können regelbasierte Kategorien mit von Ihnen erstellten Objektgruppen für die Unternehmenssuche sowie mit Objektgruppen konfigurieren, die Sie von IBM WebSphere Portal migrieren.

Benutzerdefinierte Textanalyse

Anwendungsentwickler können benutzerdefinierte Analyseprogramme erstellen, um komplexe linguistische Analysen der Daten durchzuführen, die Sie durchsuchen müssen. Sie können diese Programme in das System für die Unternehmenssuche integrieren und sie für Kommentare zum Inhalt Ihrer Objektgruppen verwenden. Durch die Indexierung der Kommentare ist ein semantisches Durchsuchen der Objektgruppen möglich.

Benutzer können z. B. nach Abfragebegriffen suchen, die in Zusammenhang stehen oder in demselben Satz auftreten; oder sie können nach Abhängigkeiten zwischen Abfragebegriffen suchen. Beispiel: Ein Benutzer benötigt möglicherweise Dokumente, in denen es um den IBM Verkäufer mit dem Namen Schmidt geht und nicht um den IBM Entwickler mit dem Namen Schmidt.

Unterstützung für N-Gram-Segmentierung

Zur Verbesserung der Abrufbarkeit von Dokumenten in chinesischer, japanischer und koreanischer Sprache können Sie die N-Gram-Segmentierungsmethode für die lexikalische Analyse aktivieren. Bei dieser Form der Analyse werden keine Leerzeichen als Wortbegrenzer verwendet. Sie können die Segmentierungsmethode nach der Erstellung einer Objektgruppe nicht ändern.

Unterstützung für das Durchsuchen von XML-Dokumenten mit nativer XML-Suche Eine native XML-Suche kann genauere Suchergebnisse erzielen, indem XML-Markupdateien durchsucht werden. In einer Abfrage könnte beispielsweise angegeben werden, dass ein Wort in einem bestimmten XML-Element vorhanden sein muss.

Klassen zum Boosten der Bewertung des relativen Rangs von Feldern

Wenn Sie Felder Boostklassen zuordnen, können Sie beeinflussen, welchen Rang Dokumente in den Suchergebnissen erhalten. Sie können z. B. die Bewertung von Titelfeldern erhöhen, um sicherzustellen, dass

Dokumente, bei denen der Abfragebegriff im Titel vorkommt, in den Suchergebnissen einen höheren Rang erhalten.

Zugehörige Konzepte

Funktionsweise von Kategorien

XML-Suchfelder

HTML-Suchfelder

 Integration der benutzerdefiniert Textanalyse

 Textanalyse innerhalb der Unternehmenssuche

Zugehörige Tasks

Überwachen des Parsers

Indizes für die Unternehmenssuche

Die Indexierungskomponenten für die Unternehmenssuche werden auf der Basis von regelmäßigen Zeitplänen ausgeführt, um dem Index Informationen zu neuen und geänderten Dokumenten hinzuzufügen.

Damit sichergestellt ist, dass Benutzer stets Zugriff auf die neuesten Informationen in den Quellen haben, die sie durchsuchen, umfasst die Erstellung eines Index zwei Arbeitsabschnitte:

Erstellen des Hauptindex

Während einer Hauptindexerstellung, wird der gesamte Index erneut erstellt, sodass die Struktur optimal organisiert ist. Die Indexierungsprozesse lesen alle Daten, die von Crawlern erfasst und vom Parser analysiert wurden.

Erstellen von Deltaindizes

Während einer Deltaindexerstellung werden dem Index seit der letzten Hauptindexerstellung von Crawlern erfasste Informationen hinzugefügt.

Wenn Sie Indexoptionen für eine Objektgruppe konfigurieren, geben Sie Zeitpläne für das Erstellen von Haupt- und Deltaindizes an. Die Häufigkeit der Indexerstellung hängt von Ihren Systemressourcen und davon ab, ob die indexierten Quellen statischen oder dynamischen Inhalt aufweisen.

Terminieren Sie eine häufige Deltaindexerstellung, um so sicherzustellen, dass neue Informationen verfügbar sind. Terminieren Sie regelmäßig eine Hauptindexerstellung, um die neuen Informationen zu konsolidieren, neuen Inhalt zu analysieren und die Indexleistung zu optimieren.

Sie können die Indexierungsprozesse auch starten, ohne sie zu terminieren. Wenn Sie z. B. bestimmte Syntaxanalyseregeln ändern und diese Änderungen für Ihre Suchanwendungen verfügbar sein sollen, können Sie eine Hauptindexerstellung starten, nachdem die Daten erneut durchsucht und syntaktisch analysiert wurden, anstatt darauf zu warten, dass die Indexerstellung zum terminierten Zeitpunkt startet.

Zur Steuerung der Ressourcennutzung können Sie steuern, wie viele Objektgruppen gleichzeitig Indexierungsprozesse gemeinsam nutzen und Indexstellungsanforderungen übergeben können. Mit der gleichzeitigen Indexerstellung stellen Sie sicher, dass die Erstellung eines sehr großen Hauptindex nicht die Erstellung von Deltaindizes für andere Objektgruppen blockiert. Die Indexer-

stellung kann ein ressourcenintensiver Prozess sein. Daher müssen Sie bei Großsystemen die Systembelastung überwachen, um die Terminpläne für die Haupt- und Deltaindexerstellung anzupassen.

Bei der Erstellung eines Index führen die Indexierungsprozesse eine globale Dokumentanalyse durch. Während dieser Phase werden Algorithmen angewendet, um Dokumentkopien zu erkennen, um die Linkstruktur von Dokumenten zu analysieren und um Ankertext (d. h. Text, der die Zielseite in einer Hypertextverbindung beschreibt) in Webdokumenten in besonderer Weise zu verarbeiten.

Sie können Optionen für die folgenden Indexierungsaktivitäten angeben:

- Damit Benutzer Platzhalterzeichen angeben können, können Sie eine Unterstützung für die Erweiterung der Abfragebegriffe im Index erstellen oder angeben, dass die Abfragebegriffe während der Abfrageverarbeitung erweitert werden sollen. Bei Ihrer Entscheidung müssen Sie zwischen Ressourcennutzung und Abfrageantwortzeit abwägen.
- Sie können Bereiche konfigurieren. Mit einem *Bereich* können Sie einschränken, was Benutzer in der Objektgruppe anzeigen können. Sie könnten z. B. einen Bereich erstellen, der die URIs für Dokumente Ihrer technischen Unterstützungsabteilung enthält. Sie könnten ferner einen weiteren Bereich für die URIs von Dokumenten aus Ihrer Personalabteilung erstellen. Wenn die Suchanwendung Bereiche unterstützt, können Benutzer Dokumente nur aus diesen Untergruppen der Objektgruppe durchsuchen und abrufen.
- Sie können Optionen für das Ausblenden von Suchergebnisdokumenten angeben, die über dasselbe URI-Präfix verfügen. Sie können auch einen Gruppennamen angeben, sodass Dokumente mit verschiedenen URI-Präfixen in den Suchergebnissen gemeinsam ausgeblendet werden.
- Nachdem ein Index erstellt wurde, können Sie URIs entfernen, die von Benutzern nicht durchsucht werden sollen.

Zugehörige Konzepte

Indexverwaltung

Platzhalterzeichen in Abfragen

Bereiche

Ausgeblendete URIs

Auf URI-Mustern basierende Dokumentrangfolge

Zugehörige Tasks

Terminierung von Indexerstellungen

Konfigurieren von gleichzeitig ablaufenden Indexerstellungen

Entfernen von URIs aus dem Index

Überwachen der Indexaktivität für eine Objektgruppe

Überwachen der Indexierungswarteschlange für die Unternehmenssuche

Suchserver für die Unternehmenssuche

Die Suchserver für die Unternehmenssuche arbeiten mit Ihren Suchanwendungen zusammen, um Abfragen zu verarbeiten, den Index zu durchsuchen und Suchergebnisse zurückzugeben.

Die Suchserver für die Unternehmenssuche werden bei der Installation von OmniFind Enterprise Edition installiert. Wenn Sie die Suchserver für eine Objektgruppe konfigurieren, können Sie Optionen angeben, wie die Objektgruppe durchsucht werden soll:

- Sie können einen Suchcache konfigurieren, der häufig angeforderte Suchergebnisse enthält. Ein Suchcache kann die Such- und Abrufleistung verbessern.
- Sie können eine voreingestellte Sprache für das Durchsuchen von Dokumenten in der Objektgruppe angeben.
- Wenn Ihre Anwendungsentwickler benutzerdefinierte Wörterverzeichnisse erstellen, können Sie diese den Objektgruppen zuordnen:
 - Wenn Benutzer eine Objektgruppe abfragen, die ein *Synonymverzeichnis* verwendet, werden Dokumente, die Synonyme zu den Abfragebegriffen enthalten, in die Suchergebnisse eingeschlossen.
 - Wenn Benutzer eine Objektgruppe abfragen, die ein *Verzeichnis von Stoppwörtern* verwendet, werden die Stoppwörter aus der Abfrage gelöscht, bevor diese verarbeitet wird.
 - Wenn Benutzer eine Objektgruppe abfragen, die ein *Verzeichnis von Boostwörtern* verwendet, erhalten Dokumente, die diese Boostwörter enthalten, einen höheren oder niedrigeren Rang, je nachdem welcher Boostfaktor dem Wort im Wörterverzeichnis zugeordnet ist.
- Wenn Sie vorher festlegen, dass bestimmte Dokumente für bestimmte Abfragen relevant sind, können Sie Quick Links konfigurieren. Ein *Quick Link* ordnet einem bestimmten URI spezifische Schlüsselwörter und Phrasen zu. Wenn eine Abfrage beliebige in einer Quick Link-Definition angegebene Schlüsselwörter oder Phrasen enthält, wird der zugeordnete URI automatisch in den Suchergebnissen zurückgegeben.

Bei einer Konfiguration mit mehreren Servern steht ein Ausfallschutz nicht nur auf der Serverebene, sondern auch auf der Objektgruppenebene zur Verfügung. Wenn eine Objektgruppe auf einem Suchserver aus irgendeinem Grund nicht mehr zur Verfügung steht, werden die Abfragen für diese Objektgruppe automatisch an den anderen Suchserver weitergeleitet.

Zugehörige Konzepte

Suchanwendungen für die Unternehmenssuche

Suchcaches

Benutzerdefinierte Synonymverzeichnisse

Benutzerdefinierte Verzeichnisse von Stoppwörtern

Benutzerdefinierte Verzeichnisse von Boostwörtern

Quick Links

Zugehörige Tasks

Überwachen der Suchserver

Administrationskonsole für die Unternehmenssuche

Die Administrationskonsole für die Unternehmenssuche wird in einem Browser ausgeführt, das heißt, dass Benutzer mit Verwaltungsaufgaben jederzeit von einem beliebigen Standort aus auf sie zugreifen können. Sicherheitsmechanismen stellen sicher, dass nur die Benutzer auf sie zugreifen, die für den Zugriff auf Verwaltungsfunktionen berechtigt sind.

Die Administrationskonsole für die Unternehmenssuche wird auf den Suchservern installiert, wenn Sie OmniFind Enterprise Edition installieren.

Die Administrationskonsole enthält Assistenten, die Sie bei der Ausführung verschiedener primärer Verwaltungstasks unterstützen können. Der Objektgruppenassistent unterstützt Sie z. B. bei der Erstellung einer Objektgruppe und ermöglicht Ihnen, Ihre Arbeit im Entwurfsmodus zu speichern. Crawlerassistenten sind spezi-

fisch für einen Datenquellentyp und unterstützen Sie bei der Auswahl der Quellen, die Sie Ihren Benutzern für Suchvorgänge verfügbar machen wollen.

Sie können für weitere Verwaltungstasks einzelne Elemente auswählen, die Sie verwalten wollen. Wenn Sie z. B. eine Objektgruppe bearbeiten, können Sie die Seite **Index** auswählen, um den Indexzeitplan zu ändern, oder die Seite **Syntaxanalyse** auswählen, um eine Regel für die Syntaxanalyse von XML-Dokumenten zu modifizieren.

Zugehörige Konzepte

Systemverwaltung für die Unternehmenssuche

Verwaltungsrollen

Zugehörige Tasks

Anmelden an der Administrationskonsole

Überwachen eines Systems für die Unternehmenssuche

Sie können mit der Administrationskonsole für die Unternehmenssuche Systemaktivitäten überwachen und Unternehmensaktivitäten nach Bedarf anpassen.

Nachdem Sie OmniFind Enterprise Edition installiert und mindestens eine Objektgruppe erstellt haben, können Sie detaillierte Statistikdaten für jede Hauptaktivität (Crawlersuche, Syntaxanalyse, Indexierung und Suche) anzeigen. Die Informationen umfassen durchschnittliche Antwortzeiten und den Verarbeitungsfortschritt, z. B., wie viele Dokumente während einer bestimmten Crawlersuche oder einer Indexerstellungssitzung durchsucht oder indexiert wurden.

Sie können die meisten Aktivitäten stoppen und starten. Sie können z. B. eine Aktivität anhalten, ihre Konfiguration ändern oder einen Fehler beheben und die Verarbeitung erneut starten, wenn Sie mit der Aktivität fortfahren wollen.

Sie können auch Alerts konfigurieren, durch die Sie E-Mails zu bestimmten Aktivitäten empfangen können, sobald ein überwachtes Ereignis auftritt. Sie können z. B. einen Alert empfangen, wenn die Suchantwortzeit einen bestimmten Grenzwert überschreitet.

Wenn ein Dokument aus dem System für die Unternehmenssuche gelöscht wurde, können Sie das Dokument überwachen und ermitteln, wann, wo und warum es gelöscht wurde. Möglicherweise kann ein Parser ein Dokument nicht syntaktisch analysieren, oder ein Administrator entfernt ein Dokument aus dem Index.

Zugehörige Konzepte

Überwachen der Aktivitäten der Unternehmenssuche

Starten und Stoppen eines Systems für die Unternehmenssuche

Protokolldateien für die Unternehmenssuche

Protokolldateien werden für einzelne Objektgruppen und für Sitzungen auf Systemebene erstellt.

Wenn Sie Protokollierungsoptionen für eine Objektgruppe für die Unternehmenssuche oder für das System konfigurieren, geben Sie die Nachrichtentypen an, die protokolliert werden sollen, wie z. B. Fehlermeldungen und Warnungen. Sie geben ferner an, wie oft das System die alten Protokolldateien turnusmäßig wechseln soll, um Platz für die aktuellen Nachrichten zu schaffen. Sie können Optionen für den Empfang von E-Mails zu bestimmten Nachrichten (einschließlich Alerts) bzw. zu allen Fehlermeldungen auswählen, sobald diese auftreten.

Wenn Sie Protokolldateien anzeigen, wählen Sie die Protokolldatei aus, die angezeigt werden soll. Der Dateiname umfasst Informationen dazu, wann die Datei erstellt wurde und welche Komponente die Nachrichten abgesetzt hat. Sie können auch Anzeigefilter angeben. Sie können z. B. auswählen, dass nur Fehlernachrichten oder nur Nachrichten von einer bestimmten Sitzung der Unternehmenssuche angezeigt werden.

Zugehörige Konzepte

Protokolldateien und Alerts

Alerts



Nachrichten für die Unternehmenssuche

Zugehörige Tasks

Konfigurieren von Protokolldateien

Konfigurieren von SMTP-Serverinformationen

Empfangen von E-Mail zu protokollierten Nachrichten

Anzeigen von Protokolldateien

Anpassen der Unternehmenssuche

Mit den Anwendungsprogrammierschnittstellen (APIs) für die Unternehmenssuche können Sie benutzerdefinierte Suchanwendungen, benutzerdefinierte Anwendungen zur Aktualisierung des Inhalts von Objektgruppen, benutzerdefinierte Programme für die Textanalyse und benutzerdefinierte Verzeichnisse für Synonyme, Stoppwörter und Boostwörter erstellen.

Nach der Installation von OmniFind Enterprise Edition sind die folgenden API-Familien zur Erweiterung von Objektgruppen für die Unternehmenssuche verfügbar:

Search and Index API (SI-API)

Erstellen Sie mit dieser API benutzerdefinierte Suchanwendungen und eine benutzerdefinierte Verwaltungsschnittstelle.

Crawler-Plug-ins

Mit Plug-in-APIs fügen Sie Dokumenten Metadaten hinzu, wenn diese durchsucht werden, oder ordnen Sicherheitstoken zu, die die Geschäfts- und Sicherheitsregeln Ihres Unternehmens umsetzen.

Sie können die Abrufbarkeit von Informationen verbessern, indem Sie benutzerdefinierte Programme für die linguistische Analyse in Ihre Objektgruppen für die Unternehmenssuche integrieren. Nachdem Sie dem System benutzerdefinierte Textanalysesteuerkomponenten hinzugefügt haben, können Sie diese den Objektgruppen zuordnen. Wenn Benutzer eine Objektgruppe abfragen, nutzen sie die Wortassoziationen, die von Ihren benutzerdefinierten Programmen im Index erstellt werden. Benutzer können z. B. nach Konzepten sowie nach Abhängigkeiten zwischen Begriffen suchen und nicht nur nach den Begriffen selbst.

Sie können die Abrufbarkeit von Informationen auch verbessern, indem Sie benutzerdefinierte Wörterverzeichnisse integrieren, die z. B. Akronyme, Abkürzungen und Vokabular enthalten, die speziell für Ihre Branche gelten. Nachdem Sie dem System Wörterverzeichnisse hinzugefügt haben, können Sie diese den Objektgruppen zuordnen. Das Abfragen einer Objektgruppe bietet Benutzern folgende Vorteile:

- Wenn eine Abfrage Wörter enthält, die als Synonyme definiert sind, werden die Dokumente, die Synonyme zu den Abfragebegriffen enthalten, in die Suchergebnisse eingeschlossen.
- Wenn eine Abfrage Stoppwörter enthält, werden die Stoppwörter aus der Abfrage entfernt, um zu verhindern, dass nicht relevante Dokumente in den Suchergebnissen zurückgegeben werden.
- Wenn eine Abfrage Boostwörter enthält, erhalten Dokumente, in denen diese Boostwörter vorkommen, einen höheren oder niedrigeren Rang in den Suchergebnissen, je nachdem welcher Boostwert dem Wort im Wörterverzeichnis zugeordnet ist.

Zugehörige Konzepte

Suchanwendungen für die Unternehmenssuche

Benutzerdefinierte Synonymverzeichnisse

Benutzerdefinierte Verzeichnisse von Stoppwörtern

Benutzerdefinierte Verzeichnisse von Boostwörtern



SI-API-Übersicht (Search and Index API)



Crawler-Plug-ins

Mustersuchanwendung für die Unternehmenssuche

Sie können die Mustersuchanwendung für die Unternehmenssuche als Schablone für die Entwicklung benutzerdefinierter Suchanwendungen verwenden.

Bei der Installation von OmniFind Enterprise Edition wird eine Mustersuchanwendung installiert. Die Mustersuchanwendung veranschaulicht die meisten Such- und Abruffunktionen, die für die Unternehmenssuche verfügbar sind. Die Anwendung ist auch ein funktionierendes Beispiel, mit dem Sie alle aktiven Objektgruppen und externen Quellen in Ihrem System für die Unternehmenssuche durchsuchen können. Sie können die Musteranwendung verwenden, um neue Objektgruppen und externe Quellen zu testen, bevor Sie die Objektgruppen und externen Quellen Benutzern zur Verfügung stellen.

Die Mustersuchanwendung veranschaulicht, wie die föderierte Suche unterstützt wird, indem sie es Ihnen ermöglicht, mehrere Objektgruppen und externe Quellen gleichzeitig zu durchsuchen.

Für bestimmte Crawlertypen können Sie die Komponente für das Identitätsmanagement für die Unternehmenssuche verwenden, um aktuelle Berechtigungsnachweise zu prüfen, wenn Benutzer auf die Suchanwendung zugreifen. Wenn die zu durchsuchende Domäne durch SSO-Sicherheit (Single Sign-on) geschützt ist, können SSO-Mechanismen verwendet werden, um den Benutzer über die gesamte Suchsitzung zu validieren. Anderenfalls kann die Komponente für das Identitätsmanagement die Benutzerberechtigung verschlüsseln und in einem Profil speichern sowie die Berechtigungsnachweise verwenden, um verbotene Dokumente aus den Suchergebnissen auszuschließen.

Wenn Sie die Mustersuchanwendung anpassen wollen, können Sie die Anpassungsfunktion für Suchanwendungen verwenden. Dies ist eine grafische Benutzerschnittstelle, mit der Sie die Auswirkungen Ihrer Änderungen anzeigen können, während Sie sie vornehmen. Sie können die Suchanwendung auch anpassen, indem Sie die Konfigurationsdatei für die Anwendung bearbeiten.

Wenn Sie eine benutzerdefinierte Suchanwendung erstellen möchten, verwenden Sie Search and Index API für die Unternehmenssuche.

Zugehörige Konzepte

Suchanwendungen für die Unternehmenssuche

Funktionen der Mustersuchanwendung

 SI-API-Übersicht (Search and Index API)

Zugehörige Tasks

Zugriff auf Suchanwendungen

Bearbeiten der Merkmale der Mustersuchanwendung

Anpassen der Suchanwendungen

Der Datenfluss für die Unternehmenssuche

Die Komponenten für die Unternehmenssuche, die Sie zusammen mit OmniFind Enterprise Edition installieren, interagieren eng miteinander, um den Datenfluss durch das System sicherzustellen.

Crawler stellen Dokumente aus den Datenquellen Ihres gesamten Unternehmens zusammen. Die Parser extrahieren nützliche Informationen aus den durchsuchten Dokumenten und generieren Token, die z. B. Dokumente Kategorien zuordnen und die Relevanz von Dokumenten anhand der Bedingungen in einer Suchanforderung ermitteln können. Der Index speichert die Daten für ein effizientes Abrufen.

Mithilfe eines Web-Browsers und einer Suchanwendung durchsuchen Benutzer indizierte Objektgruppen und externe Quellen. Die Suchanwendung kann eine Liste der Ergebnisse anzeigen, die Benutzer in einem Browser anklicken können, oder die Anwendung kann komplexer sein und dynamisch generierten Inhalt zurückgeben, der auf Informationen aus verschiedenen Quellen basiert.

Eine Katalogsuchanwendung kann z. B. die Anzeige der Produkte anpassen, die eine Suchanforderung erfüllen. Eine einzelne Abfrage kann Dokumente aus unterschiedlichen Datenquellentypen durchsuchen, wie z. B. eine Kombination aus Dokumenten in IBM DB2 Content Manager- und Lotus Notes-Repositorys.

Administratoren legen fest, welche Daten erfasst werden und wie diese durchsucht, syntaktisch analysiert und indiziert werden. Durch Überwachen der Systemaktivität nehmen Administratoren zudem Anpassungen zur Optimierung des Datendurchsatzes vor.

Das folgende Diagramm zeigt den Informationsfluss durch ein System für die Unternehmenssuche.

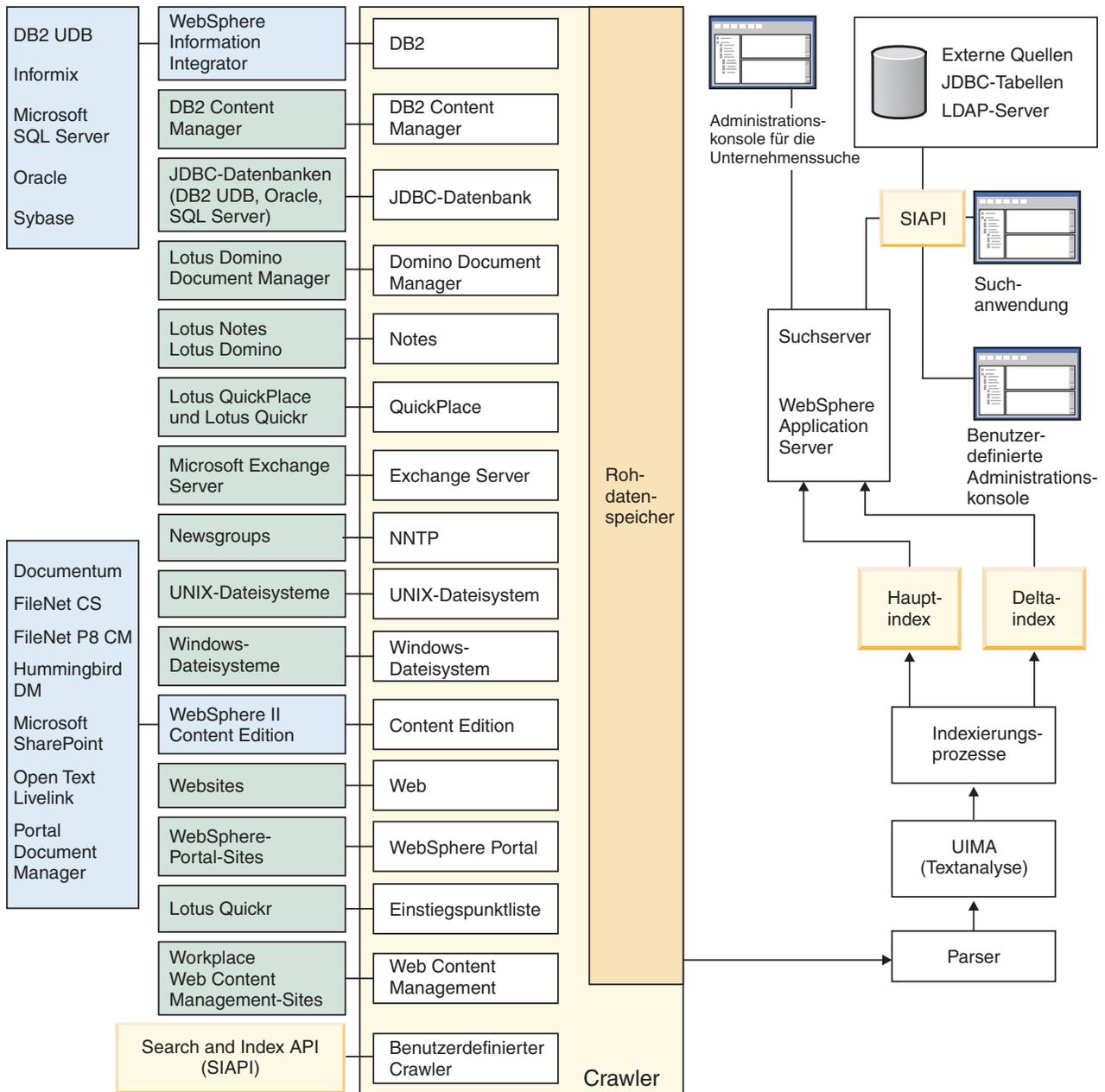


Abbildung 1. Datenfluss durch ein System für die Unternehmenssuche

Zugehörige Konzepte

„Was ist Unternehmenssuche?“ auf Seite 1

„Komponentenübersicht für die Unternehmenssuche“ auf Seite 3

Zugehörige Verweise

„Von der Unternehmenssuche unterstützte Datenquellentypen“ auf Seite 2

Verwaltung von Systemen für die Unternehmenssuche

Über die Administrationskonsole für die Unternehmenssuche können Sie Objektgruppen und externe Quellen erstellen und verwalten, Komponenten starten und stoppen, Systemaktivitäten und Protokolldateien überwachen, Benutzer mit Verwaltungsaufgaben konfigurieren, Objektgruppen und externen Quellen Suchanwendungen zuordnen und Informationen zur Umsetzung der Sicherheit angeben.

Tipp: Unter <http://www.ibm.com/developerworks/edu/dm-dw-dm-0503buehler-i.html> ist ein Onlinelernprogramm verfügbar. Im Lernprogramm werden die Installations- und Konfigurationsschritte beschrieben. Es wird außerdem gezeigt, wie verschiedene Typen von Datenquellen durchsucht werden, und es wird beschrieben, wie Sie die Anwendungsprogrammierschnittstellen des Programms verwenden können, um die Unternehmenssuche zu erweitern. Das Lernprogramm beschreibt eine ältere Version von OmniFind Enterprise Edition, aber viele Konzepte und Vorgehensweisen sind noch gültig.

Detaillierte Beispiele für die Konfiguration von Crawlern und für die Aktivierung der Sicherheit in kleinen, mittleren und großen Unternehmen finden Sie im IBM Redbook IBM OmniFind Enterprise Edition Configuration and Implementation Scenarios.

Sicht 'Objektgruppen'

Beachten Sie bei der Erstellung der ersten Objektgruppe und bei der Verwaltung des Systems die folgenden Richtlinien.

Anmelden

Melden Sie sich an der Administrationskonsole für die Unternehmenssuche an. Die Sicht **Objektgruppen** ist der Ausgangspunkt für das Erstellen und Verwalten von Objektgruppen.

Tipp: Um Unterstützung zur Administrationskonsole zu erhalten, können Sie jederzeit **Hilfe** in der Funktionsleiste oder **Hilfe für diese Seite** anklicken. Wenn ausführliche Erläuterungen und Fehlerberichtigungen zu den Nachrichten verfügbar sind, können Sie diese durch Anklicken des Symbols  **Weitere Informationen** am Ende der Nachricht anzeigen.

Erstellen einer Objektgruppe

Erstellen Sie eine neue, leere Objektgruppe anhand einer der folgenden Vorgehensweisen:

- Zum Erstellen einer Objektgruppe unter Verwendung des Objektgruppenassistenten klicken Sie **Objektgruppenassistent** an, und folgen Sie der Bedienerführung des Assistenten.
- Zum Erstellen einer Objektgruppe unter Verwendung der Sicht **Objektgruppen** klicken Sie **Objektgruppe erstellen** an, füllen Sie die Felder auf der Seite **Objektgruppe erstellen** aus, und klicken Sie dann **OK** an.

Konfigurieren der Objektgruppe

Sie müssen eine neue, leere Objektgruppe bearbeiten, um ihr Inhalt hinzuzufügen und um Optionen dazu anzugeben, wie Sie Daten durchsuchen und die Daten für Suchvorgänge aktivieren möchten.

Klicken Sie  **Bearbeiten** für Ihre neue Objektgruppe an, und wählen Sie dann eine Seite aus, um Optionen für die Objektgruppe anzugeben.

Achtung: Wenn Sie zur vorherigen Seite zurückkehren oder Informationen in der Administrationskonsole aktualisieren wollen, klicken Sie **Zurück** und **Aktualisieren** in der Administrationskonsole für die Unternehmenssuche an. Wenn Sie im Web-Browser den Knopf **Aktualisieren** oder **Zurück** anklicken, kann es zu inkonsistenten Ergebnissen und zu Datenverlust kommen.

- Auf der Seite **Allgemein** können Sie Optionen angeben, die für die gesamte Objektgruppe gelten:
 - Sie können allgemeine Optionen bearbeiten, um den Namen oder die Beschreibung der Objektgruppe zu ändern oder um die geschätzte Größe der Objektgruppe zu ändern.
 - Sie können Informationen zur Objektgruppe anzeigen, die Sie nicht ändern können, wie z. B. die Objektgruppen-ID oder die Methode für die statische Rangfolge für das Ordnen von Dokumenten in den Suchergebnissen.
 - Wenn für die Objektgruppe bei der Erstellung die Sicherheit aktiviert wurde, können Sie Sicherheitsmaßnahmen auf Dokumentebene aktivieren oder inaktivieren.
- Konfigurieren Sie auf der Seite **Crawler** mindestens einen Crawler. Eine einzelne Objektgruppe kann Daten von einer Vielzahl von Datenquellen enthalten. Sie müssen mindestens einen Crawler für jeden Datenquellentyp konfigurieren, den Sie mit einschließen wollen. Wenn Sie einen Crawler erstellen, werden Sie bei der Konfiguration des Crawlers von einem Assistenten unterstützt, der für den durchsuchten Datentyp spezifisch ist.
- Auf der Seite **Syntaxanalyse** können Sie Optionen dafür konfigurieren, wie die vom Crawler erfassten Daten syntaktisch analysiert werden sollen, sodass diese effektiv durchsucht werden können:
 - Sie können angeben, ob XML-Dokumente syntaktisch analysiert werden sollen, sodass sie mit einer nativen XML-Suche durchsucht werden können.
 - Sie können Dokumente Kategorien zuordnen, sodass Benutzer die Möglichkeit erhalten, eine Untergruppe der Objektgruppe zu durchsuchen oder Dokumente in den Suchergebnissen nach zugehörigen Kategorien zu durchsuchen.
 - Sie können Suchfeldern XML-Elemente und HTML-Metadaten-elemente im Index zuordnen, sodass Benutzer die Möglichkeit erhalten, die Feldnamen in Abfragen anzugeben und bestimmte Teile von Dokumenten zu durchsuchen.
 - Wenn Sie dem System für die Unternehmenssuche benutzerdefinierte Textanalysesteuerkomponenten hinzugefügt haben, können Sie eine Textanalysesteuerkomponente für die Verwendung mit der Objektgruppe auswählen und anschließend Textverarbeitungsoptionen angeben, um die Abrufbarkeit von Informationen zu verbessern und eine semantische Suche zu unterstützen.
 - Sie können Feldern Boostklassen zuordnen, um zu beeinflussen, welche Rangfolge Felder, die mit den Abfragebegriffen übereinstimmen, in den Suchergebnissen einnehmen.

- Konfigurieren Sie auf der Seite **Index** Zeitpläne für die Indexerstellung. Terminieren Sie häufige Indexerstellungen, sodass Ihre Benutzer immer auf die neuesten Informationen zugreifen können. Sie haben zudem die folgenden optionalen Möglichkeiten:
 - Sie können es Benutzern ermöglichen, Platzhalterzeichen in Abfragebegriffen anzugeben.
 - Sie können Bereiche konfigurieren, sodass Benutzer nur einen bestimmten Teil der Objektgruppe anstelle der gesamten Dokumente im Index durchsuchen können.
 - Sie können Suchergebnisse ausblenden, sodass Dokumente derselben Quelle in den Suchergebnissen ausgeblendet werden.
 - Durch die Zuordnung von Boostfaktoren können Sie beeinflussen, welche Rangfolge Dokumente, die mit einem URI-Muster übereinstimmen, in den Suchergebnissen einnehmen.
 - Sie können URIs aus dem Index entfernen. Zum Beispiel können Sie verhindern, dass Benutzer bestimmte Dokumente anzeigen, nachdem die Objektgruppe erstellt wurde.
- Auf der Seite **Suche** können Sie Optionen für das Durchsuchen von Dokumenten in der Objektgruppe angeben:
 - Sie können Cachespeicherplatz für Suchergebnisse reservieren und die voreingestellte Sprache für die Objektgruppe ändern.
 - Wenn Sie dem System für die Unternehmenssuche benutzerdefinierte Verzeichnisse für Synonyme, Stoppwörter und Boostwörter hinzugefügt haben, können Sie die Verzeichnisse auswählen, die verwendet werden sollen, wenn Benutzer die Objektgruppe durchsuchen.
 - Sie können eine Anzeigelänge für Dokumentzusammenfassungen in den Suchergebnissen angeben.
 - Wenn bestimmte URIs immer automatisch in den Suchergebnissen vorkommen sollen, wenn eine Abfrage bestimmte Schlüsselwörter oder Phrasen enthält, können Sie Quick Links konfigurieren.
- Auf der Seite **Protokoll** haben Sie die folgenden Möglichkeiten:
 - Sie können Optionen für die Nachrichtentypen angeben, die protokolliert werden sollen, und angeben, wie oft die Protokolldateien erneut gestartet werden sollen.
 - Sie können Optionen zum Empfangen von Alerts zur Objektgruppenaktivität angeben. Ein Alert kann Sie z. B. darüber informieren, wenn die durchschnittliche Antwortzeit bei Suchen einen angegebenen Grenzwert überschreitet.
 - Sie können Optionen für den Erhalt von E-Mails angeben, sobald bestimmte Nachrichten oder Nachrichtentypen protokolliert werden.
 - Sie können Optionen für das Protokollieren von Informationen angeben, anhand derer Sie feststellen können, wann, wo und warum ein Dokument aus dem System für die Unternehmenssuche gelöscht wurde.

Starten der Komponenten

Nachdem Sie die zu durchsuchenden Datenquellen und die Optionen für das Erfassen und Durchsuchen von Daten angegeben haben, können Sie die Prozesse für das Erstellen der Objektgruppe starten. Die Reihenfolge, in der Sie die Komponenten starten, ist entscheidend. Crawler müssen die Daten durchsuchen, bevor die Daten syntaktisch analysiert und indiziert werden können. Der Index muss erstellt

werden, bevor die Suchserver Suchanforderungen verarbeiten können.

Sicht 'Externe Quellen'

Wenn Sie Datenquellen durchsuchen wollen, ohne diese durch einen Crawler durchsuchen oder diese indexieren zu lassen, können Sie **Externe Quellen** in der Funktionsleiste anklicken, um Optionen zum Aktivieren der Datenquellen für Suchvorgänge anzugeben. Sie müssen Informationen angeben, die den Zugriff auf Ihre JDBC-Datenbanken (Java Database Connectivity) und Lightweight Directory Access Protocol-Server für den Zweck der Unternehmenssuche ermöglichen. Nachdem Sie den externen Quellen Suchanwendungen zugeordnet haben, können Benutzer diese Quellen gleichzeitig mit Objektgruppen durchsuchen, die Daten enthalten, die von einem Crawler durchsucht sowie syntaktisch analysiert und indexiert wurden.

Sicht 'System'

Wenn Sie die Rolle eines Administrators für die Unternehmenssuche haben, haben Sie durch Anklicken von **System** in der Funktionsleiste die nachfolgend aufgeführten Möglichkeiten. Objektgruppenadministratoren, Bediener und Überwachungsbeauftragte können auf diese Sicht nur zugreifen, wenn ihnen ein Administrator für die Unternehmenssuche die entsprechende Berechtigung erteilt.

- Hinzufügen von benutzerdefinierten Textanalysesteuerkomponenten zum System
- Hinzufügen von benutzerdefinierten Verzeichnissen für Synonyme, Stoppwörter und Boostwörter zum System
- Angeben der Anzahl Objektgruppen, die parallel Indizes erstellen können, und Angeben, ob Hauptindexierungen für eine einzelne Objektgruppe gleichzeitig mit Deltaindexierungen ausgeführt werden können
- Konfigurieren von Alerts für Ereignisse auf Systemebene
- Angeben von Optionen für das Protokollieren von Nachrichten, die bei Sitzungen auf Systemebene erstellt werden
- Angeben von Informationen zu Ihrem E-Mail-Server, sodass Sie E-Mails zu Aktivitäten der Unternehmenssuche empfangen können

Sicht 'Sicherheit'

Wenn Sie die Rolle eines Administrators für die Unternehmenssuche haben, können Sie in der Funktionsleiste **Sicherheit** anklicken, um die Sicherheitsoptionen anzugeben. Administratoren für die Objektgruppe, Bediener und Überwachungsbeauftragte können auf diese Sicht nicht zugreifen.

Wenn Sie die Sicherheit in WebSphere Application Server aktivieren, können Sie mit der Sicht **Sicherheit** Verwaltungsrollen konfigurieren. Durch das Konfigurieren von Verwaltungsrollen können Sie es mehr Benutzern ermöglichen, das System zu verwalten, und den Zugriff jedes Benutzers auf bestimmte Funktionen und Objektgruppen einschränken.

Darüber hinaus können Sie mit der Sicht **Sicherheit** Identitätsmanagementoptionen konfigurieren. Beispielsweise können Sie Optionen zum Speichern von Benutzerberechtigungen in Profilen angeben, mit denen die aktuellen Berechtigungsnachweise des Benutzers während der Abfrageverarbeitung geprüft werden können. Wenn die zu durchsuchende Quelle durch SSO-Sicherheit (Single Sign-on) geschützt ist, können Sie auch Optionen für die Verwendung von SSO-Authentifi-

zierungsmethoden angeben, mit denen die aktuellen Berechtigungsnachweise des Benutzers während der Abfrageverarbeitung geprüft werden können.

Bis zur Erstellung eigener Suchanwendungen können Sie die Mustersuchanwendung verwenden, um alle Objektgruppen und externen Quellen zu durchsuchen. Nachdem Sie eine benutzerdefinierte Suchanwendung erstellt haben, verwenden Sie die Sicht **Sicherheit**, um den Objektgruppen und externen Quellen Ihre Anwendung zuzuordnen, die sie durchsuchen kann.

Sicht 'Überwachen'

Sie können  **Überwachung** anklicken, um das System oder die Objektgruppenkomponenten jederzeit zu überwachen. Wenn Ihre Verwaltungsrolle dies zulässt, können Sie die Komponentenprozesse auch während der Überwachung starten und stoppen.

Zugehörige Konzepte

„Crawlerverwaltung“ auf Seite 41

„Überwachen der Aktivitäten der Unternehmenssuche“ auf Seite 327

Zugehörige Tasks

„Starten eines Systems für die Unternehmenssuche“ auf Seite 319

„Verwalten der Suchserver im Standalone-Modus“ auf Seite 324

„Stoppen eines Systems für die Unternehmenssuche“ auf Seite 321

„Erstellen einer Objektgruppe mit dem Objektgruppenassistenten“ auf Seite 33

„Erstellen einer Objektgruppe mit der Sicht 'Objektgruppen'“ auf Seite 35

Anmelden an der Administrationskonsole

Zur Verwaltung eines Systems für die Unternehmenssuche geben Sie eine URL in einen Web-Browser ein und melden sich anschließend an der Administrationskonsole an.

Vorbereitung

Sie müssen sich mit einer Benutzer-ID anmelden, die für den Zugriff auf die Administrationskonsole für die Unternehmenssuche berechtigt ist:

- Wenn Sie die globale Sicherheit in WebSphere Application Server nicht aktivieren, kann nur der Administrator für die Unternehmenssuche, der während der Installation von OmniFind Enterprise Edition angegeben wurde, auf die Administrationskonsole zugreifen.
- Wenn Sie die globale Sicherheit in WebSphere Application Server aktivieren, können Sie mit der Administrationskonsole für die Unternehmenssuche Verwaltungsrollen konfigurieren. Die Benutzer-IDs, die Sie konfigurieren, müssen in einer WebSphere Application Server-Benutzerregistry vorhanden sein. Wenn Sie Verwaltungsrollen konfigurieren, ermöglichen Sie mehr Benutzern, sich an der Administrationskonsole anzumelden, aber Sie können die Funktionen und Objektgruppen steuern, auf die jeder Benutzer mit Verwaltungsaufgaben zugreifen kann.

Vorgehensweise

Gehen Sie wie folgt vor, um sich an der Administrationskonsole für die Unternehmenssuche anzumelden:

1. Geben Sie die URL für die Administrationskonsole in Ihrem Web-Browser ein.
Beispiel:

`http://suchserver.com/ESAdmin/`

`suchserver.com` ist der Hostname des Suchservers für die Unternehmenssuche. Abhängig von Ihrer Web-Server-Konfiguration müssen Sie unter Umständen auch die Portnummer angeben. Beispiel:

`http://suchserver.com:9080/ESAdmin/`

2. Geben Sie auf der Begrüßungsseite Ihre Benutzer-ID und Ihr Kennwort ein, und klicken Sie **Anmelden** an.

Die Sicht **Objektgruppen** wird angezeigt. Sie ist Ihr Ausgangspunkt für das Verwalten des Systems und der Objektgruppen. Wenn Sie Verwaltungsrollen verwenden, hängen die Aktionen, die Sie ausführen können, und die Objektgruppen, die Sie anzeigen können, von Ihrer Verwaltungsrolle ab.

Wenn Ihre Sitzung für eine gewisse Zeit inaktiv ist, meldet das System Sie automatisch ab. Melden Sie sich erneut an, um mit dem Verwalten des Systems fortzufahren.

Nach der Verwaltung von Objektgruppen können Sie **Abmelden** anklicken, um sich von der Konsole abzumelden. Sie können sich anschließend mit einer anderen ID und einem anderen Kennwort anmelden oder den Web-Browser schließen, um die Administrationskonsole zu verlassen.

Zugehörige Konzepte

„Verwaltungsrollen“ auf Seite 280

Ändern des Kennworts des Administrators für die Unternehmenssuche in einer Einzelserverkonfiguration

Das Kennwort für den Administrator für die Unternehmenssuche wird in einem verschlüsselten Format gespeichert. Sie ändern das Kennwort mit dem Script **eschangepw**.

Einschränkungen

Kennwörter können die folgenden Sonderzeichen enthalten:

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

Wenn Sie auf AIX-, Linux- oder Solaris-Systemen ein Kennwort angeben, das Sonderzeichen enthält, müssen Sie das komplette Kennwort in einfache Anführungszeichen einschließen.

Beispiel: `'mypwd@$%'`

Wenn Sie auf einem Windows-System ein Kennwort angeben, das Sonderzeichen enthält, müssen Sie das komplette Kennwort in Anführungszeichen einschließen.

Beispiel: `"my?+!pwd"`

Informationen zu dieser Task

Das Kennwort für die erste ID des Administrators für die Unternehmenssuche wird bei der Installation von OmniFind Enterprise Edition angegeben.

Wenn Sie das Kennwort ändern möchten, müssen Sie das Script **eschangepw** ausführen, um die Änderung im gesamten System für die Unternehmenssuche zu verbreiten. Das Installationsprogramm erstellt zwei Umgebungsvariablen, die Sie mit dem Script **eschangepw** verwenden können:

ES_INSTALL_ROOT

Das Installationsverzeichnis für die Unternehmenssuche.

ES_NODE_ROOT

Das Datenverzeichnis für die Unternehmenssuche. Das Kennwort für die ID des Administrators für die Unternehmenssuche wird in diesem Verzeichnis in der Datei `es.cfg` gespeichert.

Vorgehensweise

Gehen Sie wie folgt vor, um das Kennwort des Administrators für die Unternehmenssuche in einer Einzelserverkonfiguration zu ändern:

1. Melden Sie sich als Administrator für die Unternehmenssuche an, und stoppen Sie das System für die Unternehmenssuche:

```
esadmin system stopall
```

Wichtig: Wenn das System gestoppt ist, können Benutzer keine Suchanforderungen übergeben.

2. Ändern Sie das Systemkennwort für die Benutzer-ID des Administrators für die Unternehmenssuche, indem Sie Betriebssystembefehle (unter AIX, Linux oder Solaris) oder die Funktion zur Kennwortänderung (unter Windows) verwenden.
3. Führen Sie das folgende Script aus. Dabei ist *neuer_wert* das Kennwort, das Sie in Schritt 2 angegeben haben:

AIX, Linux oder Solaris

```
eschangepw.sh neuer_wert
```

Windows

```
eschangepw neuer_wert
```

4. Starten Sie das System für die Unternehmenssuche erneut:

```
esadmin system startall
```

Zugehörige Verweise

„Befehle, Rückkehrcodes und Sitzungs-IDs der Unternehmenssuche“ auf Seite 401

Ändern des Kennworts des Administrators für die Unternehmenssuche in einer Konfiguration mit mehreren Servern

Das Kennwort für den Administrator für die Unternehmenssuche wird in einem verschlüsselten Format gespeichert. Wenn Sie das Kennwort ändern möchten, verwenden Sie das Script **eschangepw**, um das Kennwort auf allen Computern in Ihrem System für die Unternehmenssuche zu ändern.

Einschränkungen

Kennwörter können die folgenden Sonderzeichen enthalten:

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

Wenn Sie auf AIX-, Linux- oder Solaris-Systemen ein Kennwort angeben, das Sonderzeichen enthält, müssen Sie das komplette Kennwort in einfache Anführungszeichen einschließen.

Beispiel: 'mypwd@\$%'

Wenn Sie auf einem Windows-System ein Kennwort angeben, das Sonderzeichen enthält, müssen Sie das komplette Kennwort in Anführungszeichen einschließen.

Beispiel: "my?+!pwd"

Informationen zu dieser Task

Das Kennwort für die ID des Administrators für die Unternehmenssuche, die erstmalig bei der Installation von OmniFind Enterprise Edition angegeben wird, muss auf allen Servern für die Unternehmenssuche identisch sein.

Zum Ändern des Kennworts und zur Verbreitung der Änderung im gesamten System für die Unternehmenssuche müssen Sie das Script **eschangepw** auf jedem Computer ausführen, den Sie für die Unternehmenssuche verwenden. In der nachfolgend beschriebenen Vorgehensweise wird eine Reihenfolge für die Änderung der Kennwörter auf allen Servern vorgeschlagen. Sie brauchen diese Reihenfolge nicht einzuhalten, müssen aber die für jeden Servertyp erforderlichen Schritte ausführen.

Das Installationsprogramm erstellt zwei Umgebungsvariablen, die Sie mit dem Script **eschangepw** verwenden können:

ES_INSTALL_ROOT

Das Installationsverzeichnis für die Unternehmenssuche.

ES_NODE_ROOT

Das Datenverzeichnis für die Unternehmenssuche. Das Kennwort für die ID des Administrators für die Unternehmenssuche wird in diesem Verzeichnis in der Datei `es.cfg` gespeichert.

Vorgehensweise

Gehen Sie wie folgt vor, um das Kennwort des Administrators für die Unternehmenssuche in einer Konfiguration mit mehreren Servern zu ändern:

1. Melden Sie sich am Indexserver für die Unternehmenssuche als Administrator für die Unternehmenssuche an, und stoppen Sie das System für die Unternehmenssuche:

```
esadmin system stopall
```

Wichtig: Wenn das System gestoppt ist, können Benutzer keine Suchanforderungen übergeben.

- a. Ändern Sie das Systemkennwort für die Benutzer-ID des Administrators für die Unternehmenssuche, indem Sie Betriebssystembefehle (unter AIX, Linux oder Solaris) oder die Funktion zur Kennwortänderung (unter Windows) verwenden.
- b. Führen Sie das folgende Script aus. Dabei ist *neuer_wert* das Kennwort, das Sie in Schritt 1a angegeben haben:

AIX, Linux oder Solaris

```
eschangepw.sh neuer_wert
```

Windows

`eschangepw neuer_wert`

2. Führen Sie die folgenden Schritte auf den anderen Computern in Ihrem System für die Unternehmenssuche aus:
 - a. Melden Sie sich als Administrator für die Unternehmenssuche an.
 - b. Stoppen Sie die allgemeine Übertragungsschicht (Common Communications Layer, CCL) für die Unternehmenssuche:

AIX, Linux oder Solaris

`stopccl.sh`

Windows-Eingabeaufforderung

`stopccl`

Verwaltungstool für Windows-Services

- 1) Starten Sie die Windows-Services: **Start** → **Einstellungen** → **Systemsteuerung** → **Verwaltung** → **Dienste**.
 - 2) Klicken Sie **IBM OmniFind Enterprise Edition** mit der rechten Maustaste an, und klicken Sie **Beenden** an.
- c. Ändern Sie das Systemkennwort für die Benutzer-ID des Administrators für die Unternehmenssuche, indem Sie Betriebssystembefehle (unter AIX, Linux oder Solaris) oder die Funktion zur Kennwortänderung (unter Windows) verwenden. Dieses Kennwort muss mit dem Kennwort übereinstimmen, das Sie in Schritt 1a auf Seite 22 angegeben haben.
 - d. Führen Sie das folgende Script aus. Dabei ist *neuer_wert* das Kennwort, das Sie in Schritt 1a auf Seite 22 angegeben haben:

AIX, Linux oder Solaris

`eschangepw.sh neuer_wert`

Windows

`eschangepw neuer_wert`

- e. Starten Sie die allgemeine Übertragungsschicht erneut:

AIX, Linux oder Solaris

`startccl.sh -bg`

Windows-Eingabeaufforderung

`startccl`

Verwaltungstool für Windows-Services

Gehen Sie wie folgt vor, um die allgemeine Übertragungsschicht im Hintergrund zu starten:

- 1) Starten Sie die Windows-Services: **Start** → **Einstellungen** → **Systemsteuerung** → **Verwaltung** → **Dienste**.
 - 2) Klicken Sie **IBM OmniFind Enterprise Edition** mit der rechten Maustaste an, und klicken Sie **Eigenschaften** an.
 - 3) Klicken Sie die Indexzunge **Anmelden** an.
 - 4) Ändern Sie das Kennwort, indem Sie das neue Kennwort angeben. Klicken Sie anschließend **OK** an.
 - 5) Klicken Sie **IBM OmniFind Enterprise Edition** mit der rechten Maustaste an, und klicken Sie **Start** an.
3. Melden Sie sich am Indexserver für die Unternehmenssuche als Administrator für die Unternehmenssuche an, und starten Sie das System für die Unternehmenssuche erneut:

`esadmin system startall`

Zugehörige Verweise



Festlegen desselben verschlüsselten Administratorkennworts auf allen Servern

„Befehle, Rückkehrcodes und Sitzungs-IDs der Unternehmenssuche“ auf Seite 401

TCP-Portnummern für die Unternehmenssuche

Überprüfen Sie die standardmäßigen Portnummern, die in einem System für die Unternehmenssuche verwendet werden, damit Sie Portkonflikte vermeiden können, wenn Sie Ressourcen konfigurieren oder Portnummern anderen Anwendungen zuordnen.

Wenn Sie eine Firewall konfigurieren müssen Sie den Zugriff auf bestimmte Portnummern explizit aktivieren. Darüber hinaus müssen Sie sicherstellen, dass sich alle Server für die Unternehmenssuche innerhalb der Firewall befinden.

Tabelle 1. Portnummernkonfigurationen für die Unternehmenssuche

Portname	Portnummern	Wo konfiguriert
Allgemeine Übertragungsschicht (Common Communications Layer, CCL)	6002	ES_NODE_ROOT/nodeinfo/es.cfg und ES_NODE_ROOT/master_config/nodes.ini auf allen Servern für die Unternehmenssuche
HTTP auf den Suchservern	80	HTTP_SERVER_ROOT/conf/http.conf auf den Suchservern
Administrationskonsole von WebSphere Application Server Version 6	9060	Auf den Suchservern
Administrationskonsole von WebSphere Application Server Version 5.1	9090	Auf den Suchservern
DB2-Crawler	6000, 6001, 6002, 60003, 50000	Auf dem Crawler-Server
Informationszentrale	8888	Auf den Suchservern
Anonyme oder dynamische Ports für CCL, Dateiübertragungen (ESFTP) und Indexkopie	49152 bis 65535	Auf allen Servern für die Unternehmenssuche
Apache Derby Network Server	1527	Auf dem Crawler-Server
Benutzerdefinierte Kommunikation	8890	Auf dem Crawler-Server
Verbindungen ferner Clients zum DB2 Server (werden nur mit Releases vor OmniFind Enterprise Edition Version 8.4 verwendet)	50000	Auf dem Crawler-Server
WebSphere Information Integrator Content Edition FastObjects-Datenbank	6001 (alternativ 6002)	In der Administrationskonsole von WebSphere Information Integrator Content Edition

Tabelle 1. Portnummernkonfigurationen für die Unternehmenssuche (Forts.)

Portname	Portnummern	Wo konfiguriert
WebSphere Information Integrator Content Edition Proxy-Connector für Methodenaufrufe über Remotezugriff (Remote Method Invocation, RMI)	1251 (RMI-Port)	In der Administrationskonsole von WebSphere Information Integrator Content Edition

Ändern der Portnummer für das System für die Unternehmenssuche

Wenn die Portnummer, die das System für die Unternehmenssuche zur Kommunikation verwendet, mit einer Portnummer in Konflikt steht, die von einem anderen Produkt genutzt wird, muss die Portnummer für die Unternehmenssuche geändert werden.

Informationen zu dieser Task

Die Angabe einer Portnummer für das System für die Unternehmenssuche erfolgt bei der Installation von OmniFind Enterprise Edition. In einer Konfiguration mit mehreren Servern wird auf allen Servern dieselbe Portnummer angegeben.

Kann die Portnummer nicht verwendet werden (weil sie beispielsweise einem anderen Produkt auf demselben Server zugeordnet ist), wird wegen dieses Konflikts die folgende Fehlermeldung in die Datei `CCLServer_datum.log` geschrieben, wobei *datum* für das Datum steht, an dem die Protokolldatei erstellt wurde:

```
FFQ00273W Eine interne Warnung ist aufgetreten - Ausnahmeschichtnachricht: {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)
```

Vorgehensweise

Gehen Sie wie folgt vor, um die von der Unternehmenssuche verwendete Portnummer zu ändern:

1. Melden Sie sich am Computer, dessen Portnummer geändert werden muss, als Administrator für die Unternehmenssuche an, und stoppen Sie das System für die Unternehmenssuche:

```
esadmin system stopall
```

Wichtig: Wenn das System gestoppt ist, können Benutzer keine Suchanforderungen übergeben.

2. Bearbeiten Sie die Datei `ES_NODE_ROOT/nodeinfo/es.cfg`, suchen Sie das folgende Merkmal, geben Sie einen neuen Wert für die Portnummer ein, und speichern und schließen Sie dann die Datei:

```
CCLPort=neue_portnummer
```

3. Starten Sie die allgemeine Übertragungsschicht (Common Communications Layer, CCL) für die Unternehmenssuche erneut:

AIX, Linux oder Solaris

```
startccl.sh
```

Windows-Eingabeaufforderung

```
startccl
```

Verwaltungstool für Windows-Services

Gehen Sie wie folgt vor, um die allgemeine Übertragungsschicht im Hintergrund zu starten:

- a. Starten Sie die Windows-Services: **Start** → **Einstellungen** → **Systemsteuerung** → **Verwaltung** → **Dienste**.
 - b. Klicken Sie **IBM OmniFind Enterprise Edition** mit der rechten Maustaste an, und klicken Sie **Start** an.
4. Melden Sie sich am Indexserver als Administrator für die Unternehmenssuche an, und stoppen Sie die allgemeine Übertragungsschicht:

AIX, Linux oder Solaris

```
stopccl.sh
```

Windows-Eingabeaufforderung

```
stopccl
```

Verwaltungstool für Windows-Services

- a. Starten Sie die Windows-Services: **Start** → **Einstellungen** → **Systemsteuerung** → **Verwaltung** → **Dienste**.
 - b. Klicken Sie **IBM OmniFind Enterprise Edition** mit der rechten Maustaste an, und klicken Sie **Beenden** an.
5. Bearbeiten Sie die Datei `ES_NODE_ROOT/nodeinfo/es.cfg`:
- a. Suchen Sie das folgende Merkmal, wobei *computername* für den Namen des Computers steht, auf dem die Portnummer in Schritt 2 auf Seite 25 geändert wurde. Die Angabe *N* im Merkmal `nodeN` steht für einen Wert, der den Server für die Unternehmenssuche angibt.
`nodeN.destination=computername`
 - b. Suchen Sie das folgende Untermerkmal, geben Sie hier dieselbe Portnummer an, die Sie im Schritt 2 auf Seite 25 für den Server angegeben haben, und speichern und schließen Sie die Datei:
`nodeN.port=neue_portnummer`
6. Starten Sie das System für die Unternehmenssuche erneut:
- ```
esadmin system startall
```

### Zugehörige Verweise

„Befehle, Rückkehrcodes und Sitzungs-IDs der Unternehmenssuche“ auf Seite 401

„TCP-Portnummern für die Unternehmenssuche“ auf Seite 24

---

## Ändern von Hostnamen oder IP-Adressen von Servern für die Unternehmenssuche

Sie können die Hostnamen und IP-Adressen ändern, für deren Verwendung die Server für die Unternehmenssuche konfiguriert sind.

So kann es z. B. sein, dass Sie die IP-Adresse ändern wollen, wenn Sie mehrere Netzschnittstellenkarten auf jedem Server für die Unternehmenssuche haben und feststellen, dass der Indexserver für die Verwendung eines langsamen Netzes konfiguriert ist. Sie können die Einstellungen in den Konfigurationsdateien ändern, damit der Indexserver ein schnelleres Netz verwenden kann.

**Tipp:** Wenn Sie die Konfigurationsdateien nicht bearbeiten wollen, können Sie OmniFind Enterprise Edition erneut installieren und die neuen Hostnamen oder IP-Adressen bei der Ausführung des Installationsprogramms angeben.

### Vorgehensweise:

Gehen Sie wie folgt vor, um Hostnamen oder IP-Adressen zu ändern:

1. Melden Sie sich als Administrator für die Unternehmenssuche an. Wenn Sie eine Konfiguration mit mehreren Servern haben, melden Sie sich auf einem beliebigen Server für die Unternehmenssuche an.
2. Stoppen Sie die Sitzungen für die Unternehmenssuche:  
`esadmin system stopall`
3. Stoppen Sie die allgemeine Übertragungsschicht (Common Communications Layer, CCL). Wenn Sie eine Konfiguration mit mehreren Servern haben, verwenden Sie eine der folgenden Methoden, um die allgemeine Übertragungsschicht auf jedem Server für die Unternehmenssuche zu stoppen:

#### AIX, Linux oder Solaris

```
stopccl.sh
```

#### Windows-Eingabeaufforderung

```
stopccl
```

#### Verwaltungstool für Windows-Services

- a. Starten Sie die Windows-Services: **Start** → **Einstellungen** → **Systemsteuerung** → **Verwaltung** → **Dienste**.
  - b. Klicken Sie **IBM OmniFind Enterprise Edition** mit der rechten Maustaste an, und klicken Sie **Beenden** an.
4. Führen Sie die folgenden Schritte auf jedem Server für die Unternehmenssuche aus:
    - a. Bearbeiten Sie die Datei `ES_INSTALL_ROOT/configurations/ccl.properties`, und geben Sie den Hostnamen oder die IP-Adresse, den bzw. die Sie für diesen Server verwenden wollen, im Parameter `es_server_hostName` an.
    - b. Bearbeiten Sie die Datei `ES_NODE_ROOT/nodeinfo/es.cfg`, und geben Sie den Hostnamen oder die IP-Adresse, den bzw. die Sie für diesen Server verwenden wollen, im Parameter `LocalHostName` an.
    - c. Bearbeiten Sie die Datei `ES_NODE_ROOT/master_config/nodes.ini`, und ersetzen Sie alle Vorkommen des Parameters `nodeN.destination` durch den Hostnamen oder die IP-Adresse, den bzw. die Sie für diesen Server verwenden wollen.
    - d. Bearbeiten Sie die Datei `ES_NODE_ROOT/config/nodes.ini`, und ersetzen Sie alle Vorkommen des Parameters `nodeN.destination` durch den Hostnamen oder die IP-Adresse, den bzw. die Sie für diesen Server verwenden wollen.
  5. Starten Sie die allgemeine Übertragungsschicht erneut. Wenn Sie eine Konfiguration mit mehreren Servern haben, verwenden Sie eine der folgenden Methoden, um die allgemeine Übertragungsschicht auf jedem Server für die Unternehmenssuche erneut zu starten:

#### AIX, Linux oder Solaris

```
startccl.sh -bg
```

#### Windows-Eingabeaufforderung

```
startccl
```

### Verwaltungstool für Windows-Services

Gehen Sie wie folgt vor, um die allgemeine Übertragungsschicht im Hintergrund zu starten:

- a. Starten Sie die Windows-Services: **Start** → **Einstellungen** → **Systemsteuerung** → **Verwaltung** → **Dienste**.
  - b. Klicken Sie **IBM OmniFind Enterprise Edition** mit der rechten Maustaste an, und klicken Sie **Start** an.
6. Starten Sie die Sitzungen für die Unternehmenssuche erneut:  
esadmin system startall

---

## Konfigurieren der Unterstützung für doppelte IP-Adressen

Wenn die Server, auf denen Sie die Unternehmenssuche installieren, für die Unterstützung doppelter IP-Adressen konfiguriert sind, müssen Sie die Server für die Unternehmenssuche manuell für die Ausführung in dieser Umgebung konfigurieren.

So kann z. B. ein Desktop-Administrator einen Microsoft Loopback-Adapter installieren, um ein virtuelles Netzwerk zu erstellen, das die Voraussetzungen für den Netzbetrieb bestimmter Produkte unterstützt, wie z. B. Microsoft SQL Server.

### Vorgehensweise:

Gehen Sie wie folgt vor, um ein System für die Unternehmenssuche zu konfigurieren, damit es auf Servern ausgeführt werden kann, die doppelte IP-Adressen unterstützen:

1. Melden Sie sich als Administrator für die Unternehmenssuche an. Wenn Sie eine Konfiguration mit mehreren Servern haben, melden Sie sich auf einem beliebigen Server für die Unternehmenssuche an.
2. Stoppen Sie die Sitzungen für die Unternehmenssuche:  
esadmin system stopall
3. Stoppen Sie die allgemeine Übertragungsschicht (Common Communications Layer, CCL). Wenn Sie eine Konfiguration mit mehreren Servern haben, verwenden Sie eine der folgenden Methoden, um die allgemeine Übertragungsschicht auf jedem Server für die Unternehmenssuche zu stoppen:

#### AIX, Linux oder Solaris

```
stopccl.sh
```

#### Windows-Eingabeaufforderung

```
stopccl
```

### Verwaltungstool für Windows-Services

- a. Starten Sie die Windows-Services: **Start** → **Einstellungen** → **Systemsteuerung** → **Verwaltung** → **Dienste**.
  - b. Klicken Sie **IBM OmniFind Enterprise Edition** mit der rechten Maustaste an, und klicken Sie **Beenden** an.
4. Bearbeiten Sie auf jedem Server für die Unternehmenssuche, der doppelte IP-Adressen unterstützt, die Datei `ES_NODE_ROOT/nodeinfo/es.cfg`, und fügen Sie einen Parameter namens **LocalIPAddress** hinzu. Ordnen Sie als Wert eine IP-Adresse zu, die von Domain Name System aufgelöst werden kann.

5. Starten Sie die allgemeine Übertragungsschicht erneut. Wenn Sie eine Konfiguration mit mehreren Servern haben, verwenden Sie eine der folgenden Methoden, um die allgemeine Übertragungsschicht auf jedem Server für die Unternehmenssuche erneut zu starten:

**AIX, Linux oder Solaris**

```
startccl.sh -bg
```

**Windows-Eingabeaufforderung**

```
startccl
```

**Verwaltungstool für Windows-Services**

Gehen Sie wie folgt vor, um die allgemeine Übertragungsschicht im Hintergrund zu starten:

- a. Starten Sie die Windows-Services: **Start** → **Einstellungen** → **Systemsteuerung** → **Verwaltung** → **Dienste**.
  - b. Klicken Sie **IBM OmniFind Enterprise Edition** mit der rechten Maustaste an, und klicken Sie **Start** an.
6. Starten Sie die Sitzungen für die Unternehmenssuche erneut:  

```
esadmin system startall
```

---

## Aktivieren der Unterstützung für das IPv6-Protokoll

Sie müssen eine Umgebungsvariable definieren, die das System für die Unternehmenssuche anweist, nur IPv6-Socketadressen zu verwenden, um die Unterstützung für URL-Adressen zu aktivieren, die dem IP-Protokoll Version 6 (IPv6) unterliegen.

### Vorbereitung

1. Prüfen Sie, ob die Werte für das Merkmal **LocalHostName** und das Merkmal **LocalIPAddress** (falls angegeben) in %ES\_CFG% ein Hostname oder eine gültige IPv6-Adresse für das lokale System sind. Sie können diese Werte erforderlichenfalls manuell ändern und anschließend %ES\_CFG% speichern.
2. Prüfen Sie, ob der Wert für das Merkmal **DerbyServerHostName** in %ES\_CFG% ein Hostname ist. Dieser Wert darf keine IPv4-Adresse oder IPv6-Adresse sein. Wenn Sie dieses Merkmal ändern, speichern Sie die Datei %ES\_CFG%.
3. Prüfen Sie, ob die Werte für die Merkmale **destination** und **serverhost** in der Datei ES\_NODE\_ROOT/master\_config/nodes.ini auf dem Indexserver eine gültige IPv6-Adresse oder ein gültiger Hostname sind. Diese Werte dürfen keine IPv4-Adresse sein. Wenn Sie diese Datei ändern, speichern Sie die Änderungen.

### Einschränkungen

Die Unterstützung des IPv6-Protokolls ist nur in Systemen für die Unternehmenssuche verfügbar, die Sie auf Windows 2003-Servern installieren. Nachdem Sie die Unterstützung von IPv6-Adressen aktiviert haben, verwendet die Unternehmenssuche keine IPv4-Adressen mehr für die Socketkommunikation.

### Unterstützte Browser

Die Browser Internet Explorer und Mozilla Firefox handhaben IPv6-Adressen unterschiedlich.

#### Mozilla Firefox

Sie können die IPv6-Adresse oder den Hostnamen in der URL angeben, um die Administrationskonsole für die Unternehmenssuche oder die Suchanwendung auszuführen. Beispiel:

`http://[2001::db8]/ESAdmin`  
`http://suchserver.com/ESSearchApplication/`

### Internet Explorer

Sie dürfen die IPv6-Adresse nicht in der URL angeben, um die Administrationskonsole für die Unternehmenssuche oder die Suchanwendung auszuführen. Sie müssen das folgende Format verwenden und sicherstellen, dass der Hostname der IPv6-Adresse in Domain Name System oder in der Datei `c:\windows\system32\etc\hosts` zugeordnet ist:

`http://suchserver.com/ESAdmin/`  
`http://suchserver.com/ESSearchApplication/`

Wenn Sie eine IPv6-Adresse in der URL angeben, wird die Nachricht Fehler: ungültige Syntax angezeigt. Weitere Informationen zu dieser Einschränkung finden Sie unter <http://support.microsoft.com/kb/325414>.

### Unterstützung für IPv4-Datenquellen

Datenquellen, die auf einem IPv4-Server ausgeführt werden, werden von den Crawlern der Unternehmenssuche unterstützt. Wenn Sie den Web-Crawler konfigurieren und die Start-URL-Adressen im IPv6-Adressformat angeben, stellen Sie sicher, dass die URLs in eckigen Klammern eingeschlossen sind. Beispiel:

`http://[2001:db8:0:1:0:0:0:1]`  
`http://[2001:db8:0:1::1]`

### Vorgehensweise

Gehen Sie wie folgt vor, um die Unterstützung für das IPv6-Protokoll zu aktivieren:

1. Melden Sie sich am Indexserver für die Unternehmenssuche als Administrator für die Unternehmenssuche an, und stoppen Sie das System für die Unternehmenssuche:

```
esadmin system stopall
```

**Wichtig:** Wenn das System gestoppt ist, können Benutzer keine Suchanforderungen übergeben.

2. Stoppen Sie den Server der allgemeinen Übertragungsschicht (Common Communications Layer, CCL) auf allen Servern für die Unternehmenssuche:
  - a. Starten Sie die Windows-Services: **Start** → **Einstellungen** → **Systemsteuerung** → **Verwaltung** → **Dienste**.
  - b. Klicken Sie **IBM OmniFind Enterprise Edition** mit der rechten Maustaste an, und klicken Sie **Beenden** an.
3. Definieren Sie **ES\_IPV6=1** in den Systemumgebungsvariablen. Bei einer Installation mit mehreren Servern für die Unternehmenssuche müssen Sie diesen Schritt auf den Suchservern ausführen.
4. Aktualisieren Sie die benutzerdefinierten Merkmale von Java Virtual Machine in WebSphere Application Server. Bei einer Installation mit mehreren Servern für die Unternehmenssuche müssen Sie diese Schritte auf den Suchservern ausführen.
  - a. Falls er aktiv ist, stoppen Sie den Anwendungsserver **ESSearchServer** in WebSphere Application Server.

- b. Öffnen Sie die Administrationskonsole von WebSphere Application Server, und navigieren Sie zur Anzeige der benutzerdefinierten Merkmale von Java Virtual Machine. Wählen Sie **Server** → **Anwendungsserver** → **ESSearchServer** → **Java- und Prozessverwaltung** → **Prozessdefinition** → **Java Virtual Machine** → **Benutzerdefinierte Merkmale** aus.
  - c. Konfigurieren Sie die folgenden Merkmale:
 

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

Wenn das benutzerdefinierte Merkmal noch nicht aufgelistet ist, erstellen Sie ein neues Merkmal, und geben Sie den Merkmalnamen in das Feld **Name** und einen gültigen Wert in das Feld **Wert** ein.
  - d. Wenn Sie WebSphere Application Server Version 6.1 ausführen, fügen Sie der Datei `ES_NODE_ROOT/nodeinfo/es.cfg` das Merkmal **LocalIPAddress** hinzu, damit diese die IPv6-Adresse für diesen Suchserver enthält.
  - e. Starten Sie die Anwendung `ESSearchServer` erneut.
5. Führen Sie die folgenden Schritte aus, wenn Sie WebSphere Application Server Version 6.1 ausführen. Bei einer Installation mit mehreren Servern für die Unternehmenssuche müssen Sie diese Schritte auf den Suchservern ausführen.
    - a. Bearbeiten Sie die Datei `httpd.conf` für den IBM HTTP-Server.
    - b. Entfernen Sie das Kommentarzeichen `#` vom Anfang der folgenden Zeile, sodass der IBM HTTP-Server an Port 80 für IPv6-Verbindungen empfangsbereit sein kann:
 

```
Listen [::]:80
```
    - c. Starten Sie den IBM HTTP-Server erneut.
  6. Starten Sie die allgemeine Übertragungsschicht auf allen Servern für die Unternehmenssuche erneut:
    - a. Starten Sie die Windows-Services: **Start** → **Einstellungen** → **Systemsteuerung** → **Verwaltung** → **Dienste**.
    - b. Klicken Sie **IBM OmniFind Enterprise Edition** mit der rechten Maustaste an, und klicken Sie **Start** an.
  7. Starten Sie das System für die Unternehmenssuche erneut:
 

```
esadmin system startall
```
  8. Prüfen Sie die Datei `ES_NODE_ROOT/node/logs/cc10.log`, und prüfen Sie, ob der Server der allgemeinen Übertragungsschicht ohne Fehler gestartet wurde. Wurde IPv6 ordnungsgemäß aktiviert, werden ähnliche Nachrichten wie die folgende protokolliert:
 

```
INFO: CCL server ready for business. Now, waiting for external requests.
CCL host name is fe80::250:56ff:feb4:27d1
CCL host dot.ip address is fe80:0:0:0:250:56ff:feb4:27d1
CCL server port is 6002
Total JVM Runtime memory is 33022Kb
Current free memory is 21346Kb
java.net.preferIPv6Addresses=true
java.net.preferIPv4Stack=false
INFO: Session "TraceDaemonSession" was attached PID=4992
```
  9. Führen Sie die folgenden Schritte aus, um zu prüfen, ob die Such- und Verwaltungsanwendungen ordnungsgemäß ausgeführt werden:
    - a. Starten Sie die Suchserver für eine beliebige Objektgruppe, öffnen Sie anschließend die Mustersuchanwendung, und übergeben Sie eine Abfrage. Dieser Schritt prüft, ob die Abfrageübergabeverarbeitung funktioniert, selbst wenn keine zu durchsuchenden Dokumente im Index vorhanden sind.

- b. Öffnen Sie die Administrationskonsole, und prüfen Sie, ob Sie in der Lage sind, sich ohne Probleme anzumelden.

#### **Inaktivieren der Unterstützung für das IPv6-Protokoll**

1. Lesen Sie den Abschnitt **Vorbereitung**, und ersetzen Sie für jede Instanz einer IPv6-Adresse den Wert mit einem Hostnamen oder einer IPv4-Adresse.
2. Befolgen Sie die Anweisungen zum Aktivieren der Unterstützung für das IPv6-Protokoll mit den folgenden Ausnahmen:
  - Entfernen Sie in Schritt 3 auf Seite 30 die Umgebungsvariable **ES\_IPV6**.
  - Entfernen Sie in Schritt 4 auf Seite 30 die benutzerdefinierten Merkmale für Java Virtual Machine, die für die Anwendung ESSearchServer in WebSphere Application Server festgelegt wurden.
3. Entfernen Sie das Merkmal **LocalIPAddress** aus der Datei ES\_NODE\_ROOT/nodeinfo/es.cfg, wenn sie eine IPv6-Adresse enthält.
4. Befolgen Sie die Anweisungen in Schritt 9 auf Seite 31, um die Such- und Verwaltungsanwendungen zu prüfen.

---

## Objektgruppen für die Unternehmenssuche

Eine Objektgruppe für die Unternehmenssuche enthält den kompletten Satz der Quellen, die Benutzer in einer einzelnen Abfrage durchsuchen können. Durch eine Föderation können Benutzer mehrere Objektgruppen in einer einzigen Abfrage durchsuchen.

Wenn Sie eine Objektgruppe erstellen, geben Sie Optionen an, die auf die gesamte Objektgruppe angewendet werden. Die Objektgruppe bleibt so lange leer, bis Sie ihr Inhalte hinzufügen.

Sie können auf zwei verschiedene Arten einem System für die Unternehmenssuche Objektgruppen hinzufügen:

- Wenn Sie noch nicht mit der Administrationskonsole für die Unternehmenssuche vertraut sind oder wenn Sie noch lernen, wie die Objektgruppenkomponenten zusammen funktionieren, könnten Sie den Objektgruppenassistenten verwenden, um eine Objektgruppe zu erstellen. Der Objektgruppenassistent hilft Ihnen bei der Ausführung der Tasks und ermöglicht es Ihnen, Ihre Arbeit bei der Erstellung als Entwurfsobjektgruppe zu speichern.
- Wenn Sie mit der Verwendung der Administrationskonsole vertrauter sind, bevorzugen Sie unter Umständen die Erstellung von Objektgruppen, indem Sie bestimmte Seiten, die Sie verwalten möchten, in der Sicht **Objektgruppen** auswählen.

Nach dem Erstellen einer Objektgruppe verwenden Sie die Steuerelemente in der Sicht **Objektgruppen**, um die Objektgruppe, das System für die Unternehmenssuche und die Sicherheitsoptionen zu bearbeiten und zu überwachen.

### Föderation von Objektgruppen

Wenn die Föderation in der Suchanwendung unterstützt wird, können Benutzer mehrere Objektgruppen gleichzeitig durchsuchen. Durch die Föderation können Sie auch die maximale Größe einer Objektgruppe (20 Millionen Dokumente pro Objektgruppe) überschreiten. Benutzer können z. B. zwei Objektgruppen durchsuchen, die jeweils 20 Millionen Dokumente enthalten.

Die Suchqualität hängt von den Ergebnissen ab, die von den einzelnen Objektgruppen generiert werden, die anschließend zur endgültigen Ergebnismenge zusammengefügt werden. Die Ergebnisse sind mit den Ergebnissen identisch, die Sie erhalten, wenn Sie zwei separate Suchen übergeben und anschließend die Ergebnisse zusammenfügen und eine Rangfolge erstellen.

#### Zugehörige Tasks

„Überwachen einer Objektgruppe“ auf Seite 328

---

## Erstellen einer Objektgruppe mit dem Objektgruppenassistenten

Wenn Sie ein neuer Benutzer der Unternehmenssuche sind, können Sie den Objektgruppenassistenten verwenden, um eine Objektgruppe zu erstellen. Der Assistent enthält Details zu jedem Schritt des Prozesses und bietet Ihnen die Möglichkeit, Ihre Einstellungen dabei zu speichern.

### Vorbereitung

Sie müssen ein Administrator für die Unternehmenssuche sein, um eine Objektgruppe zu erstellen.

Sie müssen ein Administrator für die Unternehmenssuche oder ein Objektgruppenadministrator der Objektgruppe sein, um einer Objektgruppe Inhalte hinzuzufügen oder um Optionen anzugeben, wie der Inhalt einer Objektgruppe syntaktisch analysiert, indexiert oder durchsucht werden soll.

### Einschränkungen

Sie können mithilfe des Objektgruppenassistenten die folgenden Crawlertypen erstellen:

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

### Informationen zu dieser Task

Während Sie eine Objektgruppe erstellen, können Sie diese im Entwurfzustand speichern. Solange sie sich noch im Entwurfzustand befindet, kann jeder Administrator, der berechtigt ist, die Objektgruppe zu verwalten, Änderungen daran vornehmen. So können Sie zum Beispiel einen Objektgruppenadministrator, der mit Lotus Notes-Quellen Erfahrung hat, bitten, einen Notes-Crawler zu konfigurieren. Danach kann ein Objektgruppenadministrator, der mit UNIX-Systemen Erfahrung hat, den Objektgruppenentwurf bearbeiten und einen UNIX-Dateisystem-Crawler konfigurieren.

### Vorgehensweise

Gehen Sie wie folgt vor, um mithilfe des Objektgruppenassistenten eine Objektgruppe zu erstellen:

1. Klicken Sie **Objektgruppen** an, um die Sicht **Objektgruppen** zu öffnen.
2. Klicken Sie **Objektgruppenassistent** an.
3. Befolgen Sie die Anweisungen des Assistenten, um eine leere Objektgruppe zu erstellen und ihr Inhalte hinzuzufügen.

Sie müssen allgemeine Angaben für die Objektgruppe konfigurieren und mindestens einen Crawler erstellen. Sie können die Standardwerte der übrigen Konfigurationsoptionen übernehmen oder Optionen für die neue Objektgruppe angeben.

4. Wenn Sie eine Objektgruppe speichern wollen, bevor Sie fertig gestellt ist, klicken Sie **Als Entwurf speichern** an.

Ihre Objektgruppe wird mit anderen Objektgruppenentwürfen in der Sicht **Objektgruppen** aufgelistet. Wenn Sie die Sicherheit für die Objektgruppe aktiviert haben, wird das Symbol  **Objektgruppensicherheit ist aktiviert** neben dem Namen der Objektgruppe angezeigt.

5. Wenn Sie zu einer Objektgruppe zurückkehren wollen, deren Erstellung noch nicht abgeschlossen ist, klicken Sie **Zurück zum Assistenten** in der Sicht **Objektgruppen** an.
6. Klicken Sie **Fertig stellen** an, um die Objektgruppe zu erstellen.  
Ihre neue Objektgruppe wird mit anderen Objektgruppen in der Sicht **Objektgruppen** angezeigt.

Nachdem Sie eine Objektgruppe erstellt haben, müssen Sie den Prozess für die Crawlersuche, die Syntaxanalyse, das Indexieren und das Durchsuchen der Objektgruppe starten. Bis Sie der Objektgruppe eigene Suchanwendungen zum Durchsuchen zuordnen, können Sie die Mustersuchanwendung (Standard) verwenden, um die neue Objektgruppe zu durchsuchen.

---

## Erstellen einer Objektgruppe mit der Sicht 'Objektgruppen'

Verwenden Sie die Sicht **Objektgruppen**, um eine leere Objektgruppe zu erstellen. Anschließend können Sie die Objektgruppe bearbeiten, um Optionen zum Hinzufügen von Daten anzugeben und die Objektgruppe für die Suche freizugeben.

### Vorbereitung

Sie müssen ein Administrator für die Unternehmenssuche sein, um eine Objektgruppe zu erstellen.

Sie müssen ein Administrator für die Unternehmenssuche oder ein Objektgruppenadministrator der Objektgruppe sein, um einer Objektgruppe Inhalte hinzuzufügen oder um Optionen anzugeben, wie der Inhalt einer Objektgruppe syntaktisch analysiert, indexiert oder durchsucht werden soll.

### Informationen zu dieser Task

Wenn Sie wissen wollen, welche Werte Sie für eine neue Objektgruppe angeben können, klicken Sie **Hilfe** an, während Sie die Objektgruppe erstellen.

### Vorgehensweise

Gehen Sie wie folgt vor, um eine Objektgruppe in der Sicht **Objektgruppen** zu erstellen:

1. Klicken Sie **Objektgruppe erstellen** in der Sicht **Objektgruppen** an.
2. Geben Sie auf der Seite **Objektgruppe erstellen** Informationen ein, oder nehmen Sie in den folgenden Feldern eine Auswahl vor:
  - **Name der Objektgruppe.** Geben Sie einen beschreibenden Namen für den Inhalt oder den Zweck der Objektgruppe an.
  - **Objektgruppensicherheit.** Geben Sie an, ob Sie die Sicherheit für die Objektgruppe aktivieren wollen. Nachdem Sie die Objektgruppe erstellt haben, können Sie diese Einstellung nicht mehr ändern. Wenn die Objektgruppensicherheit aktiviert ist, können Sie später Optionen für die Umsetzung der Zugriffssteuerung auf Dokumentebene angeben.
  - **Dokumentrang (statisches Rangfolgemodell).** Geben Sie eine Strategie für die Zuordnung eines statischen Rangfolgefaktors an, der verwendet wird, um die Dokumente in den Suchergebnissen einzustufen. Nachdem Sie die Objektgruppe erstellt haben, können Sie diesen Wert nicht mehr ändern.
  - **Kategorisierungstyp.** Geben Sie an, ob Sie die Möglichkeit haben wollen, Dokumente nach den Kategorien zu suchen, zu denen sie gehören.

- **Zu verwendende Sprache.** Geben Sie die voreingestellte Sprache für das Suchen von Dokumenten in der Objektgruppe an.
3. Akzeptieren Sie die Standardwerte der folgenden Felder, oder geben Sie Optionen an, die Sie für die Objektgruppe verwenden wollen:
- **Beschreibung.** Standardmäßig wird keine Beschreibung erstellt.
  - **Geschätzte Anzahl Dokumente.** Der Standardwert für die geschätzte Größe der Objektgruppe ist 1 000 000 Dokumente. Das System verwendet diesen Wert, um die für die Objektgruppe erforderlichen Speicher- und Plattenressourcen einzuschätzen, nicht um die Größe der Objektgruppe einzuschränken.
  - **Speicherposition für Objektgruppendaten.** Die Standardspeicherposition von Dateien, die zur Objektgruppe gehören, ist auf dem Indexserver. Nachdem Sie die Objektgruppe erstellt haben, können Sie diesen Wert nicht mehr ändern.
  - **Objektgruppen-ID.** Die Standard-ID der Objektgruppe basiert auf dem Objektgruppennamen. Nachdem Sie die Objektgruppe erstellt haben, können Sie diesen Wert nicht mehr ändern. Wenn Sie eine benutzerdefinierte Objektgruppen-ID angeben, wird die Objektgruppe von Ihren Suchanwendungen mit dieser ID aufgerufen und nicht mit der möglicherweise kryptischen vom System erstellten ID.
  - **N-Gram-Segmentierung.** Die Standardsegmentierungsmethode basiert auf Unicode mit Leerzeichensegmentierung. Wählen Sie die Option zur Verwendung der N-Gram-Segmentierung nur aus, wenn Ihre Objektgruppe chinesische, japanische oder koreanische Dokumente enthält und wenn Sie wollen, dass der Parser die N-Gram-Segmentierung für die Wortbegrenzung verwendet. Nachdem Sie die Objektgruppe erstellt haben, können Sie diesen Wert nicht mehr ändern.

Weitere Informationen zur Konfiguration der Unterstützung der vollständigen N-Gram-Syntaxanalyse und der Aufbereitung in Objektgruppen für die Unternehmenssuche finden Sie unter <http://www.ibm.com/support/docview.wss?rs=63&uid=swg27011088>.

4. Klicken Sie **OK** an.

In der Sicht **Objektgruppen** wird Ihre neue Objektgruppe mit den anderen Objektgruppen in Ihrem System für die Unternehmenssuche aufgelistet. Wenn Sie die Sicherheit für die Objektgruppe aktiviert haben, wird das Symbol  **Objektgruppensicherheit ist aktiviert** neben dem Namen der Objektgruppe angezeigt.

Die Objektgruppe ist leer, bis Sie ihr Inhalt hinzufügen. Wenn Sie einer neuen Objektgruppe Inhalte hinzufügen wollen, wählen Sie die Objektgruppe in der Sicht **Objektgruppen** aus, bearbeiten Sie sie, erstellen Sie mindestens einen Crawler, und geben Sie Optionen an, wie Daten syntaktisch analysiert, indexiert und gesucht werden sollen.

Danach müssen Sie den Prozess für die Crawlersuche, die Syntaxanalyse, die Indexierung und das Durchsuchen der Objektgruppe starten. Sie können zum Durchsuchen der neuen Objektgruppe die Mustersuchanwendung verwenden, bis Sie Ihre benutzerdefinierten Suchanwendungen verwenden können.

---

## Bearbeiten einer Objektgruppe

Bearbeiten Sie Objektgruppen, um Informationen zu den Dokumenten anzugeben, die Sie in eine Objektgruppe aufnehmen wollen.

### Vorbereitung

Sie müssen ein Administrator für die Unternehmenssuche oder ein Objektgruppenadministrator der Objektgruppe sein, um eine Objektgruppe zu bearbeiten.

### Informationen zu dieser Task

Beim Bearbeiten einer Objektgruppe geben Sie Optionen für das Durchsuchen von Datenquellen, die Syntaxanalyse von Dokumenten, das Erstellen des Index, das Durchsuchen des indexierten Inhalts und das Protokollieren von Fehlernachrichten an. Wenn Sie eine Objektgruppe erstellen, müssen Sie diese bearbeiten, um ihr Inhalt hinzuzufügen. Später können Sie die Objektgruppe bearbeiten, um den Inhalt zu aktualisieren oder um die Art der Crawlersuche, Syntaxanalyse, Indexierung, Suche oder Protokollierung der Informationen zu ändern.

### Vorgehensweise

Gehen Sie wie folgt vor, um eine Objektgruppe zu bearbeiten:

1. Klicken Sie in der Funktionsleiste **Objektgruppen** an, um die Sicht **Objektgruppen** anzuzeigen.
2. Suchen Sie in der Objektgruppenliste die Objektgruppe, die Sie bearbeiten wollen, und klicken Sie  **Bearbeiten** an.
3. Nehmen Sie auf den folgenden Seiten die gewünschten Änderungen vor:

#### Crawler

Geben Sie die zu durchsuchenden Datenquellen an, und geben Sie Optionen an, wie der Inhalt durchsucht werden soll. Jede Objektgruppe muss mindestens einen Crawler enthalten, und eine einzelne Objektgruppe kann Daten aus unterschiedlichen Datenquellentypen enthalten. Sie müssen mindestens einen Crawler für jeden Datenquellentyp konfigurieren, der in der Objektgruppe enthalten ist.

#### Syntaxanalyse

Geben Sie Optionen an, wie durchsuchte Dokumente syntaktisch analysiert und ausgewertet werden sollen. Sie können Kategorien konfigurieren, die es den Benutzern ermöglichen, Untergruppen einer Objektgruppe zu durchsuchen. Sie können auch Regeln konfigurieren, damit die Benutzer bestimmte Abschnitte von XML- und HTML-Dokumenten durchsuchen können. Wenn Sie dem System für die Unternehmenssuche benutzerdefinierte Textanalysesteuerkomponenten hinzufügen, können Sie eine Textanalysesteuerkomponente für die Analyse und das Kommentieren von Inhalt in dieser Objektgruppe auswählen. Sie können Feldern auch Boostklassen zuordnen, um zu beeinflussen, welche Rangfolge Dokumente in den Suchergebnissen einnehmen.

#### Index

Geben Sie Zeitpläne zum Erstellen des Gesamtindex und zum Aktualisieren des Index mit neuen und geänderten Inhalten an. Sie können auch Optionen für die Verwendung von Platzhalterzeichen in Abfragen konfigurieren, die Anzeige der Objektgruppe auf einen bestimmten URI-Bereich einschränken, Suchergebnisse für eine komplette Website ausblenden und URIs aus dem Index entfernen.

**Suche** Geben Sie Optionen für die Suchvorgänge an, indem Sie beispielsweise einen Suchcache konfigurieren und eine Sprache für die Suche auswählen. Sie können auch Quick Links konfigurieren. Dabei handelt es sich um eine Funktion, die sicherstellt, dass zuvor festgelegte URIs immer zurückgegeben werden, wenn ein Benutzer bestimmte Wörter oder Phrasen in eine Abfrage einschließt. Wenn Sie dem System für die Unternehmenssuche benutzerdefinierte Wörterverzeichnisse hinzugefügt haben, können Sie die Wörterverzeichnisse auswählen, die zum Durchsuchen dieser Objektgruppe verwendet werden sollen.

#### **Protokoll**

Geben Sie die Nachrichtentypen an, die Sie protokollieren wollen, sowie Optionen zum Erstellen und Rollieren von Protokolldateien. Sie können Alerts konfigurieren, damit Sie informiert werden, wenn bestimmte Ereignisse auftreten, und Optionen angeben, damit Sie eine E-Mail empfangen, wenn bestimmte Nachrichten oder Nachrichtentypen protokolliert werden. Außerdem können Sie Optionen für das Protokollieren von Informationen zu Dokumenten angeben, die aus dem System für die Unternehmenssuche gelöscht werden.

#### **Allgemein**

Geben Sie allgemeine Informationen zur Objektgruppe an, und zeigen Sie Einstellungen an, die Sie nicht ändern können. Wenn beim Erstellen der Objektgruppe die Sicherheit aktiviert war, können Sie Sicherheitsoptionen auf Dokumentenebene konfigurieren.

---

## **Löschen einer Objektgruppe**

Wenn Sie eine Objektgruppe löschen, werden alle Informationen zu dieser Objektgruppe vollständig aus Ihrem System für die Unternehmenssuche entfernt.

#### **Vorbereitung**

Sie müssen ein Administrator für die Unternehmenssuche sein, um eine Objektgruppe zu löschen.

Sie müssen alle der Objektgruppe zugeordneten Prozesse stoppen, bevor Sie die Objektgruppe löschen können.

#### **Informationen zu dieser Task**

Das Löschen einer Objektgruppe kann ein zeitintensiver Prozess sein. Nach der Bestätigung, dass Sie die Objektgruppe löschen wollen, löscht das System alle zur Objektgruppe gehörigen Daten.

**Tipp:** Es wird möglicherweise eine Nachricht angezeigt, dass die angeforderte Operation das Zeitlimit überschritten hat, selbst wenn der Prozess noch im Hintergrund aktiv ist. Klicken Sie **Aktualisieren** in der Administrationskonsole an (klicken Sie nicht **Aktualisieren** im Web-Browser an), um zu ermitteln, ob die Task beendet wurde. Der Löschprozess ist abgeschlossen, wenn der Objektgruppenname nicht mehr in der Liste der Objektgruppen angezeigt wird.

#### **Vorgehensweise**

Gehen Sie wie folgt vor, um eine Objektgruppe zu löschen:

1. Klicken Sie **Objektgruppen** an, um die Sicht **Objektgruppen** zu öffnen.

- Suchen Sie in der Objektgruppenliste die Objektgruppe, die Sie löschen wollen, und klicken Sie  **Löschen** an.

---

## Ermitteln der Objektgruppen-ID

Für viele Verwaltungstasks brauchen Sie die Objektgruppen-ID.

### Vorbereitung

Sie müssen ein Administrator für die Unternehmenssuche sein, um die Objektgruppen-IDs anzuzeigen.

### Informationen zu dieser Task

Wenn Sie eine Objektgruppe erstellen, geben Sie einen Wert für die ID an, oder Sie lassen die ID vom System zuweisen. Wenn Sie nach dem Erstellen einer Objektgruppe die Objektgruppen-ID ermitteln wollen, können Sie die Administrationskonsole für die Unternehmenssuche verwenden oder eine Konfigurationsdatei anzeigen.

### Vorgehensweise

- Gehen Sie wie folgt vor, um die Objektgruppen-ID mithilfe der Administrationskonsole zu ermitteln:
  - Klicken Sie **Objektgruppen** an, um die Sicht **Objektgruppen** zu öffnen.
  - Klicken Sie **Objektgruppeneinstellungen anzeigen** auf der Seite **Allgemein** an.

Die Seite **Objektgruppeneinstellungen** enthält die Objektgruppen-ID, den vollständig qualifizierten Pfad zum Speicherort der Objektgruppe und das statische Rangfolgemodell, das für die Dokumente dieser Objektgruppe verwendet wird.

- Gehen Sie wie folgt vor, um die Objektgruppen-ID in einer Konfigurationsdatei anzuzeigen:

- Öffnen Sie die Datei `ES_NODE_ROOT/master_config/collections.ini`. Sortieren Sie die Datei, um die Anzeige übersichtlicher zu machen. In der folgenden Beispielausgabe ist `coll` die Objektgruppen-ID:

```
% sort $ES_NODE_ROOT/master_config/collections.ini | more
collection1.configfile=coll_config.ini
collection1.datadir=/home/eseach/node/data/coll
collection1.description=
collection1.displayname=Collection1
collection1.flags=0
collection1.id=coll
collection1.sectiontype=collection
collection1.type=1
...
```



---

## Crawlerverwaltung

Sie konfigurieren Crawler für die unterschiedlichen Datentypen, die Sie in eine Objektgruppe einschließen wollen. Eine einzelne Objektgruppe kann eine beliebige Anzahl Crawler enthalten.

**Tipp:** Unter <http://www.ibm.com/developerworks/edu/dm-dw-dm-0503buehler-i.html> ist ein Onlinelernprogramm verfügbar. Im Lernprogramm werden die Installations- und Konfigurationsschritte beschrieben. Es wird außerdem gezeigt, wie verschiedene Typen von Datenquellen durchsucht werden, und es wird beschrieben, wie Sie die Anwendungsprogrammierschnittstellen des Programms verwenden können, um die Unternehmenssuche zu erweitern. Das Lernprogramm beschreibt eine ältere Version von OmniFind Enterprise Edition, aber viele Konzepte und Vorgehensweisen sind noch gültig.

Detaillierte Beispiele für die Konfiguration von Crawlern und für die Aktivierung der Sicherheit in kleinen, mittleren und großen Unternehmen finden Sie im IBM Redbook IBM OmniFind Enterprise Edition Configuration and Implementation Scenarios.

### Konfigurieren von Crawlern

Sie verwenden die Administrationskonsole für die Unternehmenssuche, um Crawler zu erstellen, zu bearbeiten und zu löschen. In der Regel wird der Crawler von einem Experten für den zu durchsuchenden Datentyp konfiguriert. Wird z. B. ein Crawler für das Durchsuchen von Lotus Notes-Datenquellen konfiguriert, sollte der Objektgruppenadministrator entweder ein Notes-Administrator sein oder eng mit jemandem zusammenarbeiten, der mit den zu durchsuchenden Datenbanken vertraut ist.

Wenn Sie einen Crawler erstellen, werden Sie bei der Angabe von Merkmalen, die die Verwendung von Systemressourcen durch den Crawler steuern, von einem Assistenten für den durchsuchten Datentyp unterstützt. Der Assistent unterstützt Sie ebenfalls bei der Auswahl der Quellen, die Sie durchsuchen wollen.

Sie können jederzeit Änderungen an vorhandenen Crawlern vornehmen. Sie können Crawlermerkmale oder Teile des Crawlerbereichs bei Bedarf bearbeiten. Crawlerassistenten unterstützen Sie auch bei diesen Änderungen.

### Füllen eines neuen Crawlers mit Basiswerten

Sie können einen Crawler durch Verwenden der Systemstandardwerte oder durch Kopieren der Werte erstellen, die für einen anderen Crawler desselben Typs angegeben wurden. Wenn Sie einen vorhandenen Crawler als Basis für einen neuen Crawler verwenden, können Sie schnell mehrere Crawler mit ähnlichen Merkmalen erstellen und diese dann z. B. für das Durchsuchen verschiedener Quellen oder für die Ausführung in verschiedenen Zeitplänen für die Crawlersuche konfigurieren.

Indem Sie einen Crawler kopieren, können Sie die Auslastung der Crawlersuche auf mehrere Crawler verteilen, die dieselben Crawlersuchregeln verwenden. Sie könnten z. B. einen Notes-Crawler kopieren, weil Sie dieselben Merkmale und Crawlersuchregeln für Felder auf einem anderen Lotus Notes-Server verwenden wollen. Der einzige Unterschied sind möglicherweise die Datenbanken, die jeder

Crawler durchsucht, und die Einstellungen für die Sicherheit auf Dokumentenebene.

## **Kombinieren von Crawlertypen in einer Objektgruppe**

Crawler für die Unternehmenssuche sind so konzipiert, dass sie Informationen aus Datenquellen mit bestimmten Typen erfassen. Wenn Sie Crawler für eine Objektgruppe konfigurieren, müssen Sie entscheiden, wie diese unterschiedlichen Datenquellentypen kombiniert werden, sodass Benutzer problemlos Ihre Unternehmensdaten durchsuchen können. Wenn Sie z. B. möchten, dass Benutzer in der Lage sind, Microsoft Windows-Dateisysteme und Microsoft Exchange Server mit allgemein zugänglichen Ordnern in einer einzigen Abfrage zu durchsuchen, erstellen Sie eine Objektgruppe, die Crawler für das Windows-Dateisystem und Exchange Server-Crawler einschließt.

Wenn Sie mehrere Crawlertypen in einer einzigen Objektgruppe kombinieren, stellen Sie sicher, dass alle Crawler dieselbe statische Rangordnungsmethode verwenden können. (Sie geben die statische Rangordnungsmethode an, wenn Sie die Objektgruppe erstellen.) Wenn Sie z. B. Webquellen (die Dokumentverknüpfungen als Rangordnungsfaktor verwenden) und NNTP-Quellen (die normalerweise das Dokumentdatum als Rangordnungsfaktor verwenden) kombinieren, könnte sich die Qualität der Suchergebnisse verschlechtern.

## **Konfigurieren der Sicherheit auf Dokumentenebene**

Wenn Sie bei der Erstellung einer Objektgruppe die Sicherheit für diese Objektgruppe aktivieren, können Sie Optionen für die Sicherheit auf Dokumentenebene konfigurieren. Jeder Crawler kann den von ihm durchsuchten Dokumenten Sicherheitstoken zuordnen. Wenn Sie sich bei der Konfiguration des Crawlers für die Sicherheit auf Dokumentenebene entschieden haben, ordnet der Crawler die Sicherheitstoken zu, die Sie für jedes Dokument angegeben haben, und diese Token werden mit den Dokumenten dem Index hinzugefügt.

Wenn Sie in Ihren benutzerdefinierten Suchanwendungen die Sicherheit aktivieren, können Ihre Anwendungen die Sicherheitstoken verwenden, die die Crawler den Dokumenten zugeordnet haben, um Benutzer zu authentifizieren. Mit dieser Funktion können Sie den Zugriff auf einige Dokumente in einer Objektgruppe einschränken und zulassen, dass andere Dokumente von allen Benutzern durchsucht werden. Sie könnten z. B. in einer Objektgruppe allen Benutzern den Zugriff auf sämtliche Dokumente in Ihrem Microsoft Exchange Server mit allgemein zugänglichen Ordnern erteilen, aber nur Benutzern mit bestimmten Benutzer-IDs den Zugriff auf Dokumente in Ihren Lotus Notes-Datenbanken gestatten.

Sie können benutzerdefinierte Geschäftsregeln anwenden, um den Wert der Sicherheitstoken zu ermitteln, indem Sie die Regeln in einer Java-Klasse codieren. Bei der Konfiguration der Crawlermerkmale geben Sie den Namen des Plug-ins an, das der Crawler beim Durchsuchen von Dokumenten verwenden soll. Die Sicherheitstoken, die Ihr Plug-in hinzufügt, werden im Index gespeichert und können für die Steuerung des Zugriffs auf Dokumente verwendet werden.

Bei der Konfiguration bestimmter Crawlertypen können Sie zusätzliche Sicherheitsmaßnahmen angeben. Sie können beispielsweise angeben, dass Benutzer während der Abfrageverarbeitung geprüft werden sollen. Wenn Sie diese Option aktivieren, werden die Berechtigungsnachweise des Benutzers mit den aktuellen Zugriffssteuerungslisten verglichen, die von den zu durchsuchenden Datenquellen verwaltet werden. Diese Prüfung der aktuellen Berechtigungsnachweise kann anstelle

oder zusätzlich zu der Prüfung erfolgen, die auf Sicherheitstoken im Index für die Unternehmenssuche basiert.

**Zugehörige Konzepte**

„Sicherheit auf Dokumentebene“ auf Seite 284

**Zugehörige Tasks**

„Überwachen von Crawlern“ auf Seite 330

**Zugehörige Verweise**

„Anforderungen der Crawlerkonfiguration zur Unterstützung der Sicherheit“ auf Seite 302

---

## Erstellen eines Crawlers

Beim Erstellen eines Crawlers geben Sie an, welchen Crawlertyp Sie erstellen wollen. Mithilfe eines Assistenten können Sie Informationen zu den Daten angeben, die Sie der Objektgruppe hinzufügen wollen.

### Vorbereitung

Sie müssen ein Administrator für die Unternehmenssuche oder ein Objektgruppenadministrator der Objektgruppe sein, zu der der Crawler gehört, um einen Crawler zu erstellen.

### Informationen zu dieser Task

Sie müssen mindestens einen Crawler für eine Objektgruppe erstellen. Der von Ihnen erstellte Crawlertyp hängt davon ab, welche Datentypen die Objektgruppe enthalten soll. Ein Assistent für den erstellten Crawlertyp unterstützt Sie beim Angeben von Optionen für den Crawler. So hilft Ihnen der Assistent beispielsweise bei der Angabe der Optionen, wie der Crawler Systemressourcen verwenden soll. Außerdem können Sie mithilfe des Assistenten die Datenquellen auswählen, die Sie der Objektgruppe hinzufügen wollen.

### Vorgehensweise

Gehen Sie wie folgt vor, um einen Crawler zu erstellen:

1. Bearbeiten Sie eine Objektgruppe, wählen Sie die Seite **Crawler** aus, und klicken Sie **Crawler erstellen** an.
2. Wählen Sie den Crawlertyp und Basiswerte für den Crawler aus:

- a. Wählen Sie den Crawlertyp aus, der den zu durchsuchenden Datentyp unterstützt, wie zum Beispiel Websites, Lotus Notes-Datenbanken oder UNIX-Dateisysteme.

Nach der Auswahl eines Crawlertyps werden Optionen angezeigt, wie dieser zu erstellen ist.

- b. Wählen Sie die Basiswerte für den Crawler aus:

**Systemstandardwerte für den neuen Crawler verwenden**

Füllt die Anfangswerte für den Crawler mit Standardwerten für die Installation.

Wenn Sie diese Option auswählen, klicken Sie **Weiter** an, um mit der Konfiguration Ihres neuen Crawlers zu beginnen.

**Werte eines vorhandenen Crawlers für den neuen Crawler klonen**

Füllt die Anfangswerte für den Crawler mit Werten, die für einen anderen Crawler dieses Typs konfiguriert wurden.

Wenn Sie diese Option auswählen, wird eine Liste mit Crawlern angezeigt, die mit diesem Crawlertyp übereinstimmen. Wählen Sie den Crawler aus, den Sie für den neuen Crawler verwenden wollen, und klicken Sie anschließend **Weiter** an, um mit der Konfiguration Ihres neuen Crawlers zu beginnen.

Ein Assistent wird geöffnet, mit dessen Hilfe Sie den Typ des zu erstellenden Crawlers auswählen können. Befolgen Sie die Anweisungen des Assistenten, um den Crawler zu erstellen. Sie können auf jeder Seite des Assistenten **Hilfe** anklicken, um mehr über die Optionen zu erfahren, die Sie für den betreffenden Crawlertyp angeben können.

Ihr neuer Crawler wird auf der Seite **Crawler** mit anderen Crawlern aufgelistet, die zur Objektgruppe gehören. Wenn Sie Änderungen am Crawler vornehmen wollen, können Sie jederzeit die Optionen zum Bearbeiten der Crawlermerkmale und des Crawlerbereichs anklicken.

---

## Bearbeiten von Crawlermerkmalen

Sie können die Informationen zu einem Crawler und die Art ändern, wie er Daten durchsucht. Sie können z. B. ändern, wie der Crawler Systemressourcen verwendet.

### Vorbereitung

Sie müssen ein Administrator für die Unternehmenssuche oder ein Objektgruppenadministrator der Objektgruppe sein, zu der der Crawler gehört, um Crawlermerkmale zu bearbeiten.

### Informationen zu dieser Task

Wenn Sie wissen wollen, welche Änderungen Sie vornehmen können, klicken Sie **Hilfe** an, während Sie die Crawlermerkmale bearbeiten. Die Merkmale, die Sie bearbeiten können, hängen vom Crawlertyp ab.

### Vorgehensweise

Gehen Sie wie folgt vor, um die Merkmale eines Crawlers zu bearbeiten:

1. Bearbeiten Sie eine Objektgruppe, wählen Sie die Seite **Crawler** aus, suchen Sie den Crawler, den Sie bearbeiten wollen, und klicken Sie  **Crawlermerkmale** an.
2. Ändern Sie die Crawlermerkmale, und klicken Sie anschließend **OK** an.
3. Stoppen Sie den Crawler, und starten Sie ihn erneut, damit die Änderungen wirksam werden. (Wenn Sie nur die Beschreibung des Crawlers ändern, ist es nicht erforderlich, den Crawler erneut zu starten.)

---

## Bearbeiten eines Crawlerbereichs

Sie können die Informationen zu den Datenquellen ändern, die von einem Crawler durchsucht werden. So können Sie z. B. Datenquellen hinzufügen oder entfernen, den Zeitplan für die Crawleruche ändern oder die Regeln für die Crawleruche in Dokumenten für eine bestimmte Datenquelle ändern.

### Vorbereitung

Sie müssen ein Administrator für die Unternehmenssuche oder ein Objektgruppenadministrator der Objektgruppe sein, zu der der Crawler gehört, um einen Crawlerbereich zu bearbeiten.

### Informationen zu dieser Task

Wenn Sie wissen wollen, welche Änderungen Sie für den von Ihnen verwalteten Crawlertyp vornehmen können, klicken Sie **Hilfe** an, während Sie den Crawlerbereich bearbeiten.

### Vorgehensweise

Gehen Sie wie folgt vor, um einen Crawlerbereich zu bearbeiten:

1. Bearbeiten Sie eine Objektgruppe, wählen Sie die Seite **Crawler** aus, suchen Sie den Crawler, den Sie bearbeiten wollen, und klicken Sie  **Crawlerbereich** an.
2. Ändern Sie den Crawlerbereich, indem Sie die zu ändernden Optionen auswählen.

Die Optionen, die zur Auswahl verfügbar sind, hängen vom Crawlertyp ab. Bei einigen Optionen, wie dem Hinzufügen von Datenquellen zu einer Objektgruppe, wird ein Assistent geöffnet, der Ihnen hilft, den Crawlerbereich zu ändern.

3. Stoppen Sie den Crawler, und starten Sie ihn erneut, damit die Änderungen wirksam werden.

---

## Löschen eines Crawlers

Wenn Sie einen Crawler löschen, werden alle Informationen zu diesem Crawler aus Ihrem System für die Unternehmenssuche entfernt. Informationen, die von diesem Crawler zuvor durchsucht wurden, bleiben im Index, bis der nächste Hauptindex erstellt wird.

### Vorbereitung

Sie müssen ein Administrator für die Unternehmenssuche oder ein Objektgruppenadministrator der Objektgruppe sein, zu der der Crawler gehört, um einen Crawler zu löschen.

### Informationen zu dieser Task

Das Löschen eines Crawlers kann ein zeitintensiver Prozess sein. Nach der Bestätigung, dass Sie den Crawler löschen wollen, löscht das System alle zu diesem Crawler gehörigen Daten.

**Tipp:** Da der Abschluss dieser Task eine Weile dauert, wird möglicherweise eine Nachricht angezeigt, dass für die angeforderte Operation das Zeitlimit überschritten wurde, obwohl der Prozess im Hintergrund noch aktiv ist. Klicken Sie deshalb ab und zu in der Administrationskonsole **Aktualisieren** an, um zu prüfen, ob die Task abgeschlossen wurde. (Klicken Sie hierzu nicht im Web-Browser **Aktualisieren** an.) Der Löschvorgang ist beendet, wenn der Crawlername nicht mehr in der Crawlerliste angezeigt wird.

### Vorgehensweise

Gehen Sie wie folgt vor, um einen Crawler zu löschen:

1. Bearbeiten Sie eine Objektgruppe, und wählen Sie die Seite **Crawler** aus.

- Suchen Sie den Crawler, den Sie löschen wollen, und klicken Sie  **Löschen** an.

---

## Crawlerzeitpläne

Crawler, die Sie für Webquellen erstellen, werden kontinuierlich ausgeführt. Nachdem Sie einen Web-Crawler gestartet haben, müssen Sie ihn in der Regel nicht stoppen, sofern Sie die Konfiguration des Crawlers nicht ändern. Geben Sie für alle anderen Crawlertypen einen Zeitplan für die Crawlersuche an, wenn Sie den Crawler konfigurieren.

Für einige Datenquellentypen steuert ein einziger Zeitplan, wann der Crawler alle Datenquellen im Crawlerbereich aufruft. Für andere Datenquellentypen können Sie andere Zeitpläne für bestimmte Datenquellen angeben. Sie können z. B. unterschiedliche Zeitpläne für die Crawlersuche in jeder Lotus Notes-Datenbank angeben, die der Crawler durchsucht.

Beim Konfigurieren des Zeitplans geben Sie an, welcher Typ Crawlersuche ausgeführt werden soll. Sie können eine vollständige Crawlersuche aller Dokumente im Crawlerbereich, eine Crawlersuche, die alle Aktualisierungen (neue, geänderte und gelöschte Dokumente) im Crawlerbereich beinhaltet, oder eine Crawlersuche terminieren, bei der nur neue und geänderte Dokumente berücksichtigt werden. Eine vollständige Crawlersuche dauert am längsten. Werden bei einer Crawlersuche gelöschte Dokumente entfernt, dauert diese länger als eine Crawlersuche, bei der gelöschte Dokumente ignoriert werden.

Beim Bearbeiten des Crawlerbereichs eines Crawlers können Sie einen zweiten Zeitplan für die Crawlersuche angeben. Sie konfigurieren möglicherweise einen Zeitplan, um jeden Samstag Abend alle Dokumente im Crawlerbereich zu durchsuchen, und einen zweiten Zeitplan, der häufiger ausgeführt wird, bei dem neue und geänderte Dokumente durchsucht werden.

Durch das Erstellen mehrerer Crawlerzeitpläne können Sie besser steuern, wann der Crawler die Zielquellen aufruft. Sie können den Crawler für solche Zeiten terminieren, zu denen Benutzer wahrscheinlich nicht mehr arbeiten, um z. B. Datenbanken in unterschiedlichen Zeitzonen zu durchsuchen.

---

## Content Edition-Crawler

Wenn Sie IBM WebSphere Information Integrator Content Edition-Repositorys in eine Objektgruppe für die Unternehmenssuche einschließen möchten, müssen Sie einen Content Edition-Crawler konfigurieren.

Mit dem Content Edition-Crawler können Sie die folgenden Repositorytypen durchsuchen:

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

Wenn Sie den Crawler konfigurieren, geben Sie Optionen dafür an, wie der Crawler alle Repositorys im Crawlerbereich durchsuchen soll. Sie wählen außerdem die Elementklassen aus, die in jedem Repository durchsucht werden sollen.

### **Tipp:**

Detaillierte Beispiele für die Konfiguration von Connectorn und eines sicheren Content Edition-Crawlers finden Sie im Szenario für ein mittleres Unternehmen im IBM Redbook *IBM OmniFind Enterprise Edition Configuration and Implementation Scenarios*.

## **Crawler-Server-Konfiguration**

Wie Sie das Durchsuchen von Repositorys vorbereiten, hängt davon ab, ob Sie den Direktmodus oder den Servermodus verwenden wollen, um eine Verbindung zu den zu durchsuchenden Daten herzustellen. Wenn Sie den Direktmodus verwenden, müssen Sie einen Connector in WebSphere Information Integrator Content Edition konfigurieren. Wenn Sie den Servermodus verwenden, müssen Sie ein Script auf dem Crawler-Server ausführen. Dieses Script, das im Lieferumfang von OmniFind Enterprise Edition enthalten ist, ermöglicht es dem Content Edition-Crawler, mit WebSphere Information Integrator Content Edition-Servern zu kommunizieren.

Wenn Sie den Servermodus verwenden, führen Sie die für Ihre Umgebung geeignete Task aus, bevor Sie einen Content Edition-Crawler erstellen:

- „Konfigurieren des Crawler-Servers unter UNIX für WebSphere II Content Edition“ auf Seite 50.
- „Konfigurieren des Crawler-Servers unter Windows für WebSphere II Content Edition“ auf Seite 51.

Ausführliche Anweisungen, wie Sie Ihr System für die Unternehmenssuche für das Durchsuchen von WebSphere Information Integrator Content Edition-Repositorys konfigurieren können, finden Sie im IBM developerWorks-Artikel *Search WebSphere Portal Document Manager using WebSphere Information Integrator OmniFind Edition* unter der URL-Adresse <http://www-128.ibm.com/developerworks/db2/library/techarticle/dm-0606lee/>.

## **Konfigurationsüberblick**

Wenn Sie einen Content Edition-Crawler erstellen oder ändern möchten, müssen Sie sich an der Administrationskonsole für die Unternehmenssuche anmelden. Sie müssen die Rolle eines Administrators für die Unternehmenssuche haben oder ein Objektgruppenadministrator für die Objektgruppe sein, zu der der Crawler gehört.

Wenn Sie den Crawler erstellen, unterstützt Sie ein Assistent bei den folgenden Tasks:

- Angeben von Merkmalen, die steuern, wie der Crawler ausgeführt wird und Systemressourcen verwendet. Die Crawlermerkmale steuern, wie der Crawler alle Repositorys im Crawlerbereich durchsucht.
- Angeben, ob der Crawler im Direktmodus oder Servermodus auf Repositorys zugreifen soll. Für den Servermodus müssen Sie zudem Informationen angeben, mit denen der Crawler auf den Webanwendungsserver zugreifen kann.
- Auswählen der zu durchsuchenden Repositorys.

- Angeben der Benutzer-IDs und Kennwörter, die es dem Crawler ermöglichen, auf den Inhalt in den ausgewählten Repositories zuzugreifen.
- Definieren eines Zeitplans für das Durchsuchen der Repositories.
- Auswählen der Elementklassen, die in jedem Repository durchsucht werden sollen.
- Angeben von Optionen für das Aktivieren der Elementklassenmerkmale für Suchvorgänge. Sie können z. B. bestimmte Dokumenttypen aus dem Crawlerbereich ausschließen oder angeben, dass Sie eine bestimmte Version eines Repositories durchsuchen möchten.
- Konfigurieren von Optionen für die Sicherheit auf Dokumentebene. Wenn die Sicherheit bei der Erstellung der Objektgruppe aktiviert war, kann der Crawler Sicherheitsdaten Dokumenten im Index zuordnen. Mithilfe dieser Daten können Suchanwendungen die Zugriffssteuerung auf Basis der Zugriffssteuerungslisten oder Sicherheitstoken erzwingen.

Für Documentum-, FileNet Panagon Content Services-, Hummingbird DM-, Portal Document Manager- und SharePoint-Elementklassen können Sie außerdem eine Option auswählen, um Berechtigungsnachweise eines Benutzers bei der Übergabe einer Abfrage zu prüfen. In diesem Fall vergleicht das System die Benutzerberechtigungs-nachweise nicht mit indexierten Sicherheitsdaten, sondern mit den aktuellen Zugriffssteuerungslisten, die von den ursprünglichen Datenquellen verwaltet werden. Dieser Typ der Prüfung für aktuelle Berechtigungs-nachweise steht bei anderen Repositorytypen nicht zur Verfügung.

Klicken Sie **Hilfe** an, während Sie den Crawler erstellen, um Informationen zu den Feldern im Assistenten und zur Vorgehensweise abzurufen, wie Sie dem Crawler die für das Durchsuchen der Daten erforderlichen Informationen zur Verfügung stellen.

## Zugriff auf Content Edition-Repositories im Direktmodus

Sie können den Content Edition-Crawler so konfigurieren, dass er im Direktmodus auf WebSphere Information Integrator Content Edition-Repositories zugreift.

### Informationen zu dieser Task

Im Direktmodus verwendet der Crawler einen WebSphere Information Integrator Content Edition-Connector, der bei der Installation von OmniFind Enterprise Edition auf dem Crawler-Server installiert wird. Der Crawler verwendet Inhaltsintegrations-APIs, um eine direkte Verbindung zu den zu durchsuchenden Repositories herzustellen. Wenn der Inhaltsintegrationsserver im Direktmodus ausgeführt wird, ist nicht seine gesamte Funktionalität verfügbar. Informationen zum Ausführen des Inhaltsintegrationsservers im Direktmodus und zu den Unterschieden der Funktionalität im Vergleich mit einem Inhaltsintegrationsserver im Servermodus finden Sie in der WebSphere Information Integrator Content Edition-Dokumentation.

Diese Vorgehensweise fasst die Schritte zusammen, die erforderlich sind, um den Zugriff im Direktmodus zu konfigurieren. Ausführliche Anweisungen finden Sie im IBM developerWorks-Artikel *Search WebSphere Portal Document Manager using WebSphere Information Integrator OmniFind Edition* unter der URL-Adresse <http://www-128.ibm.com/developerworks/db2/library/techarticle/dm-0606lee/>.

### Vorgehensweise

Gehen Sie wie folgt vor, um das System so zu konfigurieren, dass der Crawler im Direktmodus auf Repositories zugreifen kann:

1. Überprüfen Sie, ob durch die Umgebungsvariablen VBR\_HOME und JAVA\_HOME in der Datei *ice\_installationsroot/bin/config.sh* (unter UNIX) bzw. in der Datei *ice\_installationsroot\bin\config.bat* (unter Microsoft Windows) das richtige Verzeichnis angegeben wird.
2. Um die WebSphere Information Integrator Content Edition-Administrationskonsole so zu konfigurieren, dass sie im Direktmodus ausgeführt wird, fügen Sie der Datei *ice\_installationsroot/bin/Admin.sh* (unter UNIX) bzw. der Datei *ice\_installationsroot\bin\Admin.bat* (unter Windows) das Java-Systemmerkmal **-Dvbr.as.operationMode=direct** hinzu. Beispiel:

**Datei Admin.sh**

```
java -classpath \
"$VBR_CLASSPATH" \
-Dvbr.home="$VBR_HOME" \
-Dvbr.as.operationMode=direct \
-Dlog4j.category.com.venetica.vbr.tools.admin=WARN \
com.venetica.vbr.tools.admin.AdminFrame $1 $2 $3 $4
```

**Datei Admin.bat**

```
java -classpath "%VBR_CLASSPATH%" ^
-Dvbr.home="%VBR_HOME%" ^
-Dvbr.as.operationMode=direct ^
-Dlog4j.category.com.venetica.vbr.tools.admin=WARN ^
com.venetica.vbr.tools.admin.AdminFrame %*
```

3. Starten Sie die WebSphere Information Integrator Content Edition-Administrationskonsole im Direktmodus, und konfigurieren Sie den Connector für den OmniFind Enterprise Edition-Crawler-Server.
4. Wählen Sie die Option für den Direktmodus aus, wenn Sie zur Konfiguration des Content Edition-Crawlers die Administrationskonsole für die Unternehmenssuche verwenden.

## Zugriff auf WebSphere II Content Edition-Repositorys im Servermodus

Sie können den Content Edition-Crawler so konfigurieren, dass er im Servermodus auf Repositorys zugreift.

Im Servermodus wird der WebSphere Information Integrator Content Edition-Connector, mit dem der Crawler auf Daten zugreift, als Unternehmensanwendung unter WebSphere Application Server installiert, und der Crawler greift über den Server auf Repositorys zu. Mithilfe dieser Methode können Sie die Vorteile von J2EE-Anwendungsserver-Umgebungen nutzen.

Bevor Sie den Crawler so konfigurieren, dass er im Servermodus auf WebSphere Information Integrator Content Edition-Repositorys zugreift, müssen Sie auf dem Crawler-Server ein Script ausführen. Dieses Script, das mit OmniFind Enterprise Edition bereitgestellt wird, ermöglicht es dem Content Edition-Crawler, auf Repositorys auf dem Server zuzugreifen.

Führen Sie die für Ihre Umgebung geeignete Task aus:

- „Konfigurieren des Crawler-Servers unter UNIX für WebSphere II Content Edition“ auf Seite 50.
- „Konfigurieren des Crawler-Servers unter Windows für WebSphere II Content Edition“ auf Seite 51.

## Konfigurieren des Crawler-Servers unter UNIX für WebSphere II Content Edition

Wenn Sie OmniFind Enterprise Edition auf einem Computer installieren, auf dem IBM AIX, Linux oder die Solaris-Betriebsumgebung ausgeführt wird, und den Content Edition-Crawler so konfigurieren, dass er im Servermodus auf Repositorys zugreift, müssen Sie ein Script ausführen, um den Crawler-Server zu konfigurieren. Mithilfe des Scripts kann der Content Edition-Crawler auf WebSphere Information Integrator Content Edition-Repositorys zugreifen.

### Informationen zu dieser Task

Der Content Edition-Crawler verwendet Java-Bibliotheken von WebSphere Information Integrator Content Edition als Java-Client. Im Servermodus erfordern diese Java-Bibliotheken EJB-bezogene Java-Bibliotheken von WebSphere Application Server. Nach der Installation von WebSphere Application Server führen Sie auf dem Crawler-Server ein Setup-Script aus, das von OmniFind Enterprise Edition bereitgestellt wird, um sicherzustellen, dass der Content Edition-Crawler mit Java-Bibliotheken zusammenarbeiten kann.

WebSphere Information Integrator Content Edition wird auf dem Crawler-Server installiert, wenn OmniFind Enterprise Edition installiert wird. Damit Sie den Content Edition-Crawler im Servermodus verwenden können, müssen Sie die Datei `vbr_access_services.jar` vom WebSphere Information Integrator Content Edition-Server auf den Crawler-Server kopieren.

### Vorgehensweise

Gehen Sie wie folgt vor, um den Crawler-Server zu konfigurieren, damit er WebSphere Information Integrator Content Edition-Repositorys durchsuchen kann:

1. Wenn OmniFind Enterprise Edition in einer Konfiguration mit mehreren Servern installiert ist, installieren und binden Sie die WebSphere Application Server-Java-Bibliotheken.
2. Führen Sie das Setup-Script für den Content Edition-Crawler auf dem Crawler-Server aus:
  - a. Melden Sie sich als Administrator für die Unternehmenssuche an.
  - b. Starten Sie das folgende Script, das im Verzeichnis `$ES_INSTALL_ROOT/bin` installiert wurde, und folgen Sie der Bedienungsführung:  
`escrvbr.sh`
3. Stoppen Sie das System für die Unternehmenssuche auf dem Crawler-Server, und starten Sie es erneut:

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

4. Kopieren Sie die Datei `vbr_access_services.jar` vom WebSphere Information Integrator Content Edition-Server auf den Crawler-Server.

#### Kopieren von:

Die Datei `vbr_access_services.jar` befindet sich an der folgenden Standardposition:

*was-installationsroot*/installedApps/*servername*/*anwendungsname*

*was-installationsroot* ist das WebSphere Application Server-Installationsverzeichnis, *servername* ist der Name, den Sie für den Server angegeben haben, und *anwendungsname* ist der Name, den Sie für

die WebSphere Information Integrator Content Edition-Anwendung in WebSphere Application Server angegeben haben.

**Kopieren nach:**

Das Zielverzeichnis auf dem Crawler-Server ist *ice\_installationsroot/lib*. Dabei steht *ice\_installationsroot* für das WebSphere Information Integrator Content Edition-Installationsverzeichnis auf dem Crawler-Server.

## **Konfigurieren des Crawler-Servers unter Windows für WebSphere II Content Edition**

Wenn Sie OmniFind Enterprise Edition auf einem Microsoft Windows-Computer installieren und den Content Edition-Crawler so konfigurieren, dass er im Servermodus auf Repositories zugreift, müssen Sie ein Script ausführen, um den Crawler-Server zu konfigurieren. Mithilfe des Scripts kann der Content Edition-Crawler auf WebSphere Information Integrator Content Edition-Repositories zugreifen.

### **Informationen zu dieser Task**

Der Content Edition-Crawler verwendet Java-Bibliotheken von WebSphere Information Integrator Content Edition als Java-Client. Im Servermodus erfordern diese Java-Bibliotheken EJB-bezogene Java-Bibliotheken von WebSphere Application Server. Nach der Installation von WebSphere Application Server führen Sie auf dem Crawler-Server ein Setup-Script aus, das von OmniFind Enterprise Edition bereitgestellt wird, um sicherzustellen, dass der Content Edition-Crawler mit Java-Bibliotheken zusammenarbeiten kann.

WebSphere Information Integrator Content Edition wird auf dem Crawler-Server installiert, wenn OmniFind Enterprise Edition installiert wird. Damit Sie den Content Edition-Crawler im Servermodus verwenden können, müssen Sie die Datei *vbr\_access\_services.jar* vom WebSphere Information Integrator Content Edition-Server auf den Crawler-Server kopieren.

### **Vorgehensweise**

Gehen Sie wie folgt vor, um den Crawler-Server zu konfigurieren, damit er WebSphere Information Integrator Content Edition-Repositories durchsuchen kann:

1. Wenn OmniFind Enterprise Edition in einer Konfiguration mit mehreren Servern installiert ist, installieren und binden Sie die WebSphere Application Server-Java-Bibliotheken.
2. Führen Sie das Setup-Script für den Content Edition-Crawler auf dem Crawler-Server aus:
  - a. Melden Sie sich mit der ID des Administrators für die Unternehmenssuche an (diese Benutzer-ID wurde während der Installation von OmniFind Enterprise Edition angegeben).
  - b. Starten Sie das folgende Script, das im Verzeichnis `%ES_INSTALL_ROOT%\bin` installiert wurde, und folgen Sie der Bedienungsführung:  
`escrvbr.vbs`
3. Stoppen Sie das System für die Unternehmenssuche auf dem Crawler-Server, und starten Sie es erneut:

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

4. Kopieren Sie die Datei *vbr\_access\_services.jar* vom WebSphere Information Integrator Content Edition-Server auf den Crawler-Server.

**Kopieren von:**

Die Datei `vbr_access_services.jar` befindet sich an der folgenden Standardposition:

`was-installationsroot\installedApps\servername\anwendungsname`

`was-installationsroot` ist das WebSphere Application Server-Installationsverzeichnis, `servername` ist der Name, den Sie für den Server angegeben haben, und `anwendungsname` ist der Name, den Sie für die WebSphere Information Integrator Content Edition-Anwendung in WebSphere Application Server angegeben haben.

**Kopieren nach:**

Das Zielverzeichnis auf dem Crawler-Server ist `ice_installationsroot\lib`. Dabei steht `ice_installationsroot` für das WebSphere Information Integrator Content Edition-Installationsverzeichnis auf dem Crawler-Server.

---

## DB2-Crawler

Sie verwenden den DB2-Crawler, um IBM DB2-Datenbanken in einer Objektgruppe einzuschließen.

Wenn Sie IBM WebSphere Information Integrator verwenden, um Kurznamen-Tabellen für die folgenden Datenbanksystemtypen zu föderieren und zu erstellen, können Sie die Tabellen mithilfe des DB2-Crawlers anhand der Kurznamen durchsuchen:

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

Sie müssen für jeden Datenbankserver, den Sie durchsuchen wollen, einen separaten Crawler konfigurieren. Wenn Sie den Crawler konfigurieren, geben Sie Optionen dafür an, wie der Crawler alle Datenbanken auf demselben Server durchsuchen soll. Sie wählen außerdem die jeweiligen Tabellen aus, die in jeder Datenbank durchsucht werden sollen.

Die Tabellen, die Sie für zum Durchsuchen auswählen, müssen Datenbanktabellen, Kurznamentabellen oder Sichten sein. Der DB2-Crawler unterstützt keine verknüpften Tabellen.

**Tipp:**

Detaillierte Beispiele für die Konfiguration eines sicheren DB2-Crawlers finden Sie im Szenario für ein großes Unternehmen im IBM Redbook *IBM OmniFind Enterprise Edition Configuration and Implementation Scenarios*.

## Crawler-Server-Konfiguration

Bevor Sie Datenbanktabellen durchsuchen können, müssen Sie DB2-Administration Client auf dem Crawler-Server installieren. Anschließend müssen Sie ein Script auf dem Crawler-Server ausführen. Dieses Script, das mit OmniFind Enterprise Edition bereitgestellt wird, ermöglicht dem DB2-Crawler die Kommunikation mit Datenbankservern.

Führen Sie die Ihrer Umgebung entsprechende Task aus, bevor Sie mit der Administrationskonsole für die Unternehmenssuche einen DB2-Crawler konfigurieren:

- „Konfigurieren des Crawler-Servers unter UNIX für DB2-Crawler“ auf Seite 54.
- „Konfigurieren des Crawler-Servers unter Windows für DB2-Crawler“ auf Seite 55.

## Ereignisveröffentlichung

Wenn Sie WebSphere Information Integrator Event Publisher Edition verwenden und die zu durchsuchenden Datenbanken Veröffentlichungswarteschlangenmasken zuordnen, kann der DB2-Crawler Aktualisierungen der Datenbanktabellen mithilfe der Masken durchsuchen.

Eine Veröffentlichungswarteschlangenmaske gibt eine WebSphere MQ-Warteschlange an, die XML-Nachrichten empfängt, wenn Aktualisierungen zu einer Datenbanktabelle veröffentlicht werden. Der Crawler ist bei dieser Warteschlange für Informationen zu diesen veröffentlichten Ereignissen empfangsbereit und aktualisiert den Crawlerbereich, wenn Tabellen aktualisiert werden (beim erstmaligen Durchsuchen einer Tabelle durchsucht der Crawler alle Dokumente).

Die Ereignisveröffentlichung sorgt dafür, dass neue und geänderte Dokumente schneller für Suchvorgänge verfügbar werden, als Dokumente, die der Crawler nach dem Crawlerzeitplan durchsucht.

Wenn für einige oder für alle Tabellen die Nutzung der Ereignisveröffentlichung konfiguriert ist, haben Sie bei der Konfiguration des Crawlers die Möglichkeit, diesem durch die Angabe von Informationen Zugriff auf WebSphere MQ und die Veröffentlichungswarteschlangenmasken zu erteilen.

Darüber hinaus müssen Sie sicherstellen, dass auf dem zu durchsuchenden Server WebSphere MQ und WebSphere Information Integrator Event Publisher Edition und auf dem Crawler-Server die WebSphere MQ-Clientmodule konfiguriert sind. Führen Sie die folgenden Tasks aus, um die Ereignisveröffentlichung bei einem DB2-Crawler zu verwenden:

- „Konfigurieren von WebSphere MQ für DB2-Crawler“ auf Seite 59.
- „Konfigurieren von WebSphere Information Integrator Event Publisher Edition für DB2-Crawler“ auf Seite 56.

## Konfigurationsüberblick

Wenn Sie einen DB2-Crawler erstellen oder ändern möchten, müssen Sie sich an der Administrationskonsole für die Unternehmenssuche anmelden. Sie müssen die Rolle eines Administrators für die Unternehmenssuche haben oder ein Objektgruppenadministrator für die Objektgruppe sein, zu der der Crawler gehört.

Wenn Sie den Crawler erstellen, unterstützt Sie ein Assistent bei den folgenden Tasks:

- Angeben von Merkmalen, die steuern, wie der Crawler ausgeführt wird und Systemressourcen verwendet. Die Crawlermerkmale steuern, wie der Crawler sämtliche Datenbanken auf einem bestimmten Datenbankserver durchsucht.
- Angeben der Informationen zu den Datenbanktypen, die durchsucht werden sollen.

Wenn Sie ferne Datenbanken durchsuchen möchten, die nicht auf dem lokalen Datenbankserver katalogisiert sind, müssen Sie den DB2-Verwaltungsserver auf dem fernen Server starten, bevor Sie diese Datenbanken mit dem DB2-Crawler durchsuchen können. Sie müssen außerdem den Hostnamen und den Port des fernen Datenbankservers angeben, wenn Sie den Crawler konfigurieren.

- Angeben der zu durchsuchenden Datenbanken.
- Angeben der Benutzer-IDs und Kennwörter, die dem Crawler den Zugriff auf Datenbanken ermöglichen, die Zugriffssteuerungen verwenden.
- Definieren eines Zeitplans für das Durchsuchen der Datenbanken.
- Auswählen der Tabellen, die in jeder Datenbank durchsucht werden sollen.

**Achtung:** Zum Optimieren der Leistung der Discoveryprozesse (und zum Verhindern des Überschreitens von Zeitlimits für den Crawlerkonfigurationsprozess) wählen Sie das Durchsuchen aller Tabellen nur aus, wenn die Datenbank wenige Tabellen enthält bzw. die Tabellen wenige Spalten enthalten. Wenn Sie jetzt nur einige Tabellen zum Durchsuchen auswählen, können Sie den Crawlerbereich später bearbeiten und der Objektgruppe weitere Tabellen hinzufügen.

- Auswählen der Tabellen, die durchsucht werden sollen, wenn in einer Ereignisveröffentlichungswarteschlange Aktualisierungen zu diesen Tabellen veröffentlicht werden, und Angeben von Informationen, die dem Crawler Zugriff auf die Ereignisveröffentlichungswarteschlange gestatten.
- Angeben von Optionen für das Aktivieren der Spalten für Suchvorgänge in bestimmten Tabellen. Sie können z. B. bestimmte Spalten für die Verwendung in parametrischen Suchen aktivieren oder angeben, welche Spalten in den Suchergebnissen zurückgegeben werden können.
- Konfigurieren von Optionen für die Sicherheit auf Dokumentenebene. Wenn die Sicherheit bei der Erstellung der Objektgruppe aktiviert war, kann der Crawler Sicherheitsdaten Dokumenten im Index zuordnen. Mithilfe dieser Daten können Suchanwendungen die Zugriffssteuerung auf Basis der gespeicherten Zugriffssteuerungslisten oder Sicherheitstoken erzwingen.

Klicken Sie **Hilfe** an, während Sie den Crawler erstellen, um Informationen zu den Feldern im Assistenten und zur Vorgehensweise abzurufen, wie Sie dem Crawler die für das Durchsuchen der Daten erforderlichen Informationen zur Verfügung stellen.

## Konfigurieren des Crawler-Servers unter UNIX für DB2-Crawler

Wenn Sie OmniFind Enterprise Edition auf einem Computer installieren, auf dem IBM AIX, Linux oder die Solaris-Betriebsumgebung ausgeführt wird, müssen Sie ein Script ausführen, um den Crawler-Server zu konfigurieren. Mithilfe des Scripts kann der DB2-Crawler mit Datenbankservern kommunizieren. Wenn Sie Ereignisveröffentlichung verwenden, kann der Crawler mithilfe des Scripts auch auf WebSphere MQ-Warteschlangenmanager und -Warteschlangen zugreifen.

### Informationen zu dieser Task

Führen Sie auf dem Crawler-Server ein von OmniFind Enterprise Edition bereitgestelltes Setup-Script aus (`escrdb2.sh`), um sicherzustellen, dass der DB2-Crawler Datenbanktabellen durchsuchen kann.

Bevor Sie das Script ausführen, müssen Sie sicherstellen, dass DB2 Administration Client auf dem Crawler-Server installiert ist.

Wenn Sie die Ereignisveröffentlichung verwenden, müssen Sie die WebSphere MQ 5.3-Module für Java Messaging auf dem Crawler-Server installieren, damit der DB2-Crawler auf WebSphere MQ-Warteschlangenmanager und -Warteschlangen zugreifen kann. Nach der Installation der WebSphere MQ-Module müssen Sie das Setup-Script `escrdb2.sh` ausführen.

### Vorgehensweise

Gehen Sie wie folgt vor, um den Crawler-Server für die Unterstützung der Crawleruche durch DB2-Crawler zu konfigurieren:

1. Optional: Wenn Sie planen, die Ereignisveröffentlichung zu verwenden, installieren Sie die WebSphere MQ 5.3-Module für Java Messaging auf dem Crawler-Server:
  - a. Melden Sie sich als Root an, und geben Sie den folgenden Befehl ein:  
`export LD_ASSUME_KERNEL=2.4.19`
  - b. Legen Sie die WebSphere MQ-CD ein.
  - c. Wechseln Sie in das Verzeichnis, in dem sich die MQ-Module für Java Messaging befinden.
  - d. Geben Sie den folgenden Befehl ein, um die Module zu installieren:  
`rpm -i MQSeriesJava-5.3.0-1.i386.rpm`
2. Führen Sie das Setup-Script für den DB2-Crawler auf dem Crawler-Server aus:
  - a. Melden Sie sich als Administrator für die Unternehmenssuche an (diese Benutzer-ID wurde während der Installation von OmniFind Enterprise Edition angegeben).
  - b. Starten Sie das folgende Script, das im Verzeichnis `$ES_INSTALL_ROOT/bin` installiert wurde, und folgen Sie der Bedienungsführung:  
`escrdb2.sh`
3. Stoppen Sie das System für die Unternehmenssuche auf dem Crawler-Server, und starten Sie es erneut:

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

## Konfigurieren des Crawler-Servers unter Windows für DB2-Crawler

Wenn Sie OmniFind Enterprise Edition auf einem Microsoft Windows-Computer installieren, müssen Sie ein Script ausführen, um den Crawler-Server zu konfigurieren. Mithilfe des Scripts kann der DB2-Crawler mit Datenbankservern kommunizieren. Wenn Sie Ereignisveröffentlichung verwenden, kann der Crawler mithilfe des Scripts auch auf WebSphere MQ-Warteschlangenmanager und -Warteschlangen zugreifen.

### Informationen zu dieser Task

Führen Sie auf dem Crawler-Server ein von OmniFind Enterprise Edition bereitgestelltes Setup-Script aus (escrdb2.vbs), um sicherzustellen, dass der DB2-Crawler Datenbanktabellen durchsuchen kann.

Bevor Sie das Script ausführen, müssen Sie sicherstellen, dass DB2 Administration Client auf dem Crawler-Server installiert ist.

Wenn Sie die Ereignisveröffentlichung verwenden, müssen Sie die WebSphere MQ 5.3-Module für Java Messaging auf dem Crawler-Server installieren, damit der DB2-Crawler auf WebSphere MQ-Warteschlangenmanager und -Warteschlangen zugreifen kann. Nach der Installation der WebSphere MQ-Module müssen Sie das Setup-Script escrdb2.vbs ausführen.

### Vorgehensweise

Gehen Sie wie folgt vor, um den Crawler-Server für die Unterstützung der Crawler-uche durch DB2-Crawler zu konfigurieren:

1. Optional: Wenn Sie planen, die Ereignisveröffentlichung zu verwenden, installieren Sie die WebSphere MQ 5.3-Module für Java Messaging auf dem Crawler-Server:
  - a. Legen Sie die WebSphere MQ-CD ein.
  - b. Starten Sie das WebSphere MQ-Installationsprogramm.
  - c. Wählen Sie im Fenster **Choose Product Features** die Installationsoption **Java Messaging** aus.
2. Führen Sie das Setup-Script für den DB2-Crawler auf dem Crawler-Server aus:
  - a. Melden Sie sich mit der ID des Administrators für die Unternehmenssuche an (diese Benutzer-ID wurde während der Installation von OmniFind Enterprise Edition angegeben).
  - b. Starten Sie das folgende Script, das im Verzeichnis %ES\_INSTALL\_ROOT%\bin installiert wurde, und folgen Sie der Bedienungsführung:  
escrdb2.vbs
3. Stoppen Sie das System für die Unternehmenssuche auf dem Crawler-Server, und starten Sie es erneut:

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

## Konfigurieren von WebSphere Information Integrator Event Publisher Edition für DB2-Crawler

Bevor Sie einen DB2-Crawler für die Verwendung der Ereignisveröffentlichung konfigurieren, stellen Sie sicher, dass IBM WebSphere Information Integrator Event Publisher Edition auf dem zu durchsuchenden Server konfiguriert ist.

### Informationen zu dieser Task

Beachten Sie die folgenden Richtlinien, wenn Sie WebSphere Information Integrator Event Publisher Edition für die Verwendung mit dem DB2-Crawler konfigurieren:

- In den Quellentabellen müssen geänderte und nicht geänderte Spalten für die Veröffentlichung ausgewählt werden.
- In den Quellentabellen müssen gelöschte Zeilen für die Veröffentlichung ausgewählt werden.

- Eine Warteschlange für die Ereignisveröffentlichung darf nicht von mehreren Datenbanken gemeinsam genutzt werden.
- Eine einzelne Datenbank kann über mehrere Warteschlangenmasken und Warteschlangen verfügen.
- In einer Tabelle sollte einer Veröffentlichungswarteschlangenmaske jeweils eine XML-Veröffentlichung zugeordnet sein. (In einer Tabelle sollten einer einzelnen Veröffentlichungswarteschlangenmaske nicht mehrere XML-Veröffentlichungen zugeordnet sein. Eine Tabelle kann über mehrere XML-Veröffentlichungen verfügen, wenn jede der XML-Veröffentlichungen einer anderen Veröffentlichungswarteschlange zugeordnet ist.)

### Vorgehensweise

Führen Sie die folgenden Schritte aus, um einen Datenbankserver so zu konfigurieren, dass der DB2-Crawler Zugriff auf Tabellenaktualisierungen hat, die in einer Ereignisveröffentlichungswarteschlange veröffentlicht werden. (Unterstützung zu diesen Schritten finden Sie in der WebSphere Information Integrator Publisher Edition-Dokumentation.)

1. Installieren Sie WebSphere Information Integrator Event Publisher Edition auf dem zu durchsuchenden Datenbankserver.
2. Starten Sie das Launchpad für die Replikationszentrale:

**AIX, Linux oder Solaris**  
db2rc

#### Windows-Eingabeaufforderung

Klicken Sie **Start** → **IBM DB2-Replikationszentrale** an.

3. Erstellen Sie Q Capture-Steuertabellen:
  - a. Wählen Sie **Event-Publishing** als Launchpadsicht aus, wählen Sie **Q Capture-Steuertabellen erstellen** aus, und klicken Sie dann **Weiter** an.
  - b. Wählen Sie im Feld **Q Capture-Server** aus der Liste der verfügbaren Datenbankserver den Server aus, der als Q Capture-Server verwendet werden soll, und klicken Sie **OK** an.
  - c. Geben Sie eine Benutzer-ID sowie ein Kennwort an, die bzw. das für den Zugriff auf den ausgewählten Q Capture-Server berechtigt ist. Ändern Sie das Q Capture-Schema, oder akzeptieren Sie den Standardschemanamen, und klicken Sie **Weiter** an.
  - d. Geben Sie den Namen des Warteschlangenmanagers, der Verwaltungswarteschlange und der Neustartwarteschlange an, den Sie bei der Konfiguration von WebSphere MQ auf diesem Datenbankserver angegeben haben, und klicken Sie **Weiter** an.
  - e. Klicken Sie **Fertig stellen** an. Wenn eine Seite mit Nachrichten und SQL-Scripts angezeigt wird, klicken Sie **Schließen** an.
  - f. Wählen Sie **Jetzt ausführen** als Verarbeitungsoption aus, und klicken Sie **OK** an. Wenn eine Nachricht angezeigt wird, dass die SQL-Scripts abgeschlossen sind, klicken Sie **Schließen** an.
4. Erstellen Sie eine XML-Veröffentlichung:
  - a. Wählen Sie **Event-Publishing** im Launchpad für die Replikationszentrale als Launchpadsicht aus, wählen Sie **XML-Veröffentlichung erstellen** aus, und klicken Sie dann **Weiter** an.
  - b. Klicken Sie **Weiter** auf der Seite **Starten** an.
  - c. Bestätigen Sie auf der Seite **Server und Warteschlangenmaske**, dass die Angaben für den Q Capture-Server und das Q Capture-Schema korrekt

- sind, und klicken Sie dann die Option neben dem Feld **Veröffentlichungswarteschlangenmaske** und anschließend **Neu** an, um eine Veröffentlichungswarteschlangenmaske zu erstellen.
- d. Geben Sie auf der Seite **Allgemein** einen Namen für die Warteschlangenmaske ein.
  - e. Geben Sie auf der Seite **Merkmale** den Namen der Sendewarteschlange an (beispielsweise den Namen der Datenwarteschlange, den Sie bei der Konfiguration von WebSphere MQ auf diesem Server angegeben haben), wählen Sie **Zeilenoperation** oder **Transaktion** als Nachrichteninhaltsstyp aus, inaktivieren Sie die Markierungsfelder für das Senden von Überwachungssignalnachrichten sowie das Hinzufügen von JMS-Nachrichtenheadern, und klicken Sie **OK** an.
  - f. Wenn eine Seite mit Nachrichten und SQL-Scripts angezeigt wird, klicken Sie **Schließen** an.
  - g. Wählen Sie **Jetzt ausführen** als Verarbeitungsoption aus, und klicken Sie **OK** an. Wenn eine Nachricht angezeigt wird, dass die SQL-Scripts abgeschlossen sind, klicken Sie **Schließen** an.
  - h. Wählen Sie auf der Seite **Veröffentlichungswarteschlangenmaske auswählen** die von Ihnen erstellte Warteschlangenmaske aus, und klicken Sie **OK** an.
  - i. Bestätigen Sie auf der Seite **Server- und Warteschlangenmaske**, dass der Name der Warteschlangenmaske korrekt ist, und klicken Sie **Weiter** an.
  - j. Klicken Sie **Hinzufügen** auf der Seite **Quellentabelle** an, klicken Sie **Alle abrufen** an, wählen Sie eine für die Ereignisveröffentlichung zu aktivierende Tabelle aus, und klicken Sie dann **OK** und anschließend **Weiter** an.
  - k. Wählen Sie auf der Seite **Spalten und Zeilen** die Spalten (bzw. alle Spalten) aus, die vom DB2-Crawler durchsucht werden sollen, und wählen Sie die Schlüsselspalten aus. Wählen Sie auf der Seite, auf der Sie die Zeilen (bzw. alle Zeilen) auswählen, die durchsucht werden sollen, die Option für das Veröffentlichens von Löschungen von Quellentabellen aus. Wenn Sie mit der Konfiguration dieser Optionen fertig sind, klicken Sie **Weiter** an.
  - l. Wählen Sie auf der Seite **Nachrichteninhalt** die Option für das Einschließen von geänderten und nicht geänderten Spalten für die Spaltendaten aus, und wählen Sie die Option **Nur neue Datenwerte** aus. Stellen Sie sicher, dass das Markierungsfeld für das automatische Starten von XML-Veröffentlichungen ausgewählt wurde, und klicken Sie **Weiter** an.
  - m. Klicken Sie **Weiter** auf der Seite **XML-Veröffentlichungen überprüfen und vervollständigen** an.
  - n. Klicken Sie **Fertig stellen** auf der Seite **Zusammenfassung** an. Wenn eine Seite mit Nachrichten und SQL-Scripts angezeigt wird, klicken Sie **Schließen** an.
  - o. Wählen Sie **Jetzt ausführen** als Verarbeitungsoption aus, und klicken Sie **OK** an. Wenn eine Nachricht angezeigt wird, dass die SQL-Scripts abgeschlossen sind, klicken Sie **Schließen** an.
5. Starten Sie den Q Capture-Server:
    - a. Schließen Sie das Launchpad für die Replikationszentrale, und starten Sie die Replikationszentrale.
    - b. Klicken Sie **Q Replication** → **Definitionen** → **Q Capture-Server** in der Objektbaumstruktur an.
    - c. Klicken Sie das Symbol für den von Ihnen konfigurierten Q Capture-Server mit der rechten Maustaste an, und wählen Sie **Datenbank für Q Replikation aktivieren** aus.

- d. Klicken Sie **OK** in der angezeigten Warnung an.
- e. Wenn eine Seite mit DB2-Nachrichten angezeigt wird, klicken Sie **Schließen** an.
- f. Klicken Sie in der Objektbaumstruktur das Symbol für den Q Capture-Server mit der rechten Maustaste an, und wählen Sie **Q Capture-Programm starten** aus.
- g. Wählen Sie **Jetzt ausführen** für die Verarbeitungsoption aus, geben Sie den Systemnamen, die Benutzer-ID und das Kennwort für den DB2-Benutzer, den Pfad für das Verzeichnis, in dem Protokolle gespeichert werden, sowie den DB2-Instanznamen an, und klicken Sie dann **OK** an.
- h. Wenn eine Nachricht angezeigt wird, dass die Anforderung übergeben wurde, klicken Sie **Schließen** an.
- i. Klicken Sie in der Objektbaumstruktur das Symbol für den Q Capture-Server mit der rechten Maustaste an, und wählen Sie **Status überprüfen** aus.  
Der Status des Q Capture-Servers wird angezeigt. Wenn Fehler aufgetreten sind, wird durch eine Statusnachricht angegeben, dass der Server möglicherweise nicht aktiv ist. Geben Sie den folgenden Befehl in einer Befehlszeile ein, um die Protokolle zu prüfen und die Ursache für eventuelle Fehler zu ermitteln:

```
asnqcap Capture_Server=name_des_capture-servers LOGSTDOUT=y
```

## Konfigurieren von WebSphere MQ für DB2-Crawler

Bevor Sie einen DB2-Crawler für die Verwendung der Ereignisveröffentlichung konfigurieren, stellen Sie sicher, dass IBM WebSphere MQ auf dem Server konfiguriert ist, auf dem der Crawler empfangsbereit ist.

### Vorbereitung

Stellen Sie sicher, dass DB2, WebSphere Information Integrator Event Publisher Edition und WebSphere MQ auf dem Zieldatenbankserver installiert sind.

### Einschränkungen

Wenn der Zieldatenbankserver auf einem Linux-Computer installiert ist, ist es erforderlich, dass alle Benutzer von DB2, von WebSphere MQ und von OmniFind Enterprise Edition die folgende Umgebungsvariable definieren:

```
export LD_ASSUME_KERNEL=2.4.19
```

Durch diese Umgebungsvariable ist es möglich, Implementierungen von Linux-Thread-Threadings von jeder Shell aus zu exportieren, auf der Installationen vorgenommen, WebSphere MQ-Steuerbefehle ausgegeben oder WebSphere MQ-Anwendungen ausgeführt werden. Für WebSphere MQ ist es erforderlich, dass diese Umgebungsvariable exportiert wird.

DB2-Crawler, die die Ereignisveröffentlichung verwenden, stellen über eine Clientverbindung eine Verbindung zu WebSphere MQ-Warteschlangen her. Zum Ermöglichen von Clientverbindungen melden Sie sich als WebSphere MQ-Administrator an, und führen Sie den folgenden Befehl aus, um die ID des codierten Zeichensatzes des Warteschlangenmanagers auf 819 zu setzen:

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

### Informationen zu dieser Task

Der DB2-Crawler unterstützt den Clientverbindungsmodus zum WebSphere MQ-Server. Der Crawler ist für XML-Nachrichten empfangsbereit, die in einer Ereignisveröffentlichungswarteschlange veröffentlicht werden. Für XML-Nachrichten, die über mehrere Warteschlangen übertragen werden, ist der Crawler nicht empfangsbereit.

Nach erfolgter Konfiguration von WebSphere MQ verwendet der DB2-Crawler den Namen des Warteschlangenmanagers, den Namen der Warteschlange, den Hostnamen des Servers, die Portnummer des Servers und den Namen des Serverkanals, um XML-Nachrichten aus einer Veröffentlichungswarteschlange abzurufen. Der Crawler führt eine Syntaxanalyse der Nachrichten durch und aktualisiert den Crawlerbereich mit aktuellen Informationen zu aktualisierten Tabellen.

### Vorgehensweise

Führen Sie die folgenden Schritte aus, um einen Datenbankserver so zu konfigurieren, dass der DB2-Crawler für eine Ereignisveröffentlichungswarteschlange empfangsbereit ist. (Unterstützung zu diesen Schritten finden Sie in der WebSphere MQ-Dokumentation.)

1. Melden Sie sich als WebSphere MQ-Administrator an, und geben Sie die folgenden Befehle ein, um einen Warteschlangenmanager sowie Warteschlangen zu erstellen.
  - a. Geben Sie den folgenden Befehl in einer Befehlszeile ein:  
`crtmqm QM1`
  - b. Wenn die Nachricht Setup completed angezeigt wird, geben Sie den folgenden Befehl ein:  
`strmqm QM1`
  - c. Wenn die Nachricht 'QM1' started angezeigt wird, geben Sie den folgenden Befehl ein:  
`runmqsc QM1`
  - d. Wenn die Nachricht Starting MQSC for queue manager QM1 angezeigt wird, geben Sie den folgenden Befehl ein, um eine Verwaltungswarteschlange zu erstellen:  
`DEFINE QLOCAL('ASN.QM1.ADMINQ')`
  - e. Wenn die Nachricht WebSphere MQ queue created angezeigt wird, geben Sie den folgenden Befehl ein, um eine Neustartwarteschlange zu erstellen:  
`DEFINE QLOCAL('ASN.QM1.RESTARTQ')`
  - f. Wenn die Nachricht WebSphere MQ queue created erneut angezeigt wird, geben Sie den folgenden Befehl ein, um eine Datenwarteschlange zu erstellen:  
`DEFINE QLOCAL('ASN.QM1.DATAQ')`
  - g. Wenn die Nachricht WebSphere MQ queue created erneut angezeigt wird, geben Sie zum Beenden den folgenden Befehl ein:  
`end`
2. Geben Sie den folgenden Befehl ein, um MQ Listener auf dem Datenbankserver zu starten (MQ Listener muss aktiv sein, wenn Sie einen DB2-Crawler erstellen, der die Ereignisveröffentlichung nutzt). In diesem Beispiel ist 1414 die Portnummer des Servers und es wird der Standardkanal SYSTEM.DEF.SVRCONN verwendet:  
`runmqlsr -m QM1 -t TCP -p 1414 &`

3. Geben Sie die folgenden Befehle ein, um einen DB2-Benutzer zu berechtigen, für die Ereignisveröffentlichung über die Message Queuing Interface (MQI) auf den Warteschlangenmanager und die Warteschlangen zuzugreifen (in diesem Beispiel lautet die Benutzer-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. Geben Sie die folgenden Befehle für die Benutzer-ID ein, die verwendet wird, um den DB2-Crawler mit Ereignisveröffentlichung zu erstellen und auszuführen. Mit diesen Befehlen wird die Benutzer-ID berechtigt, für die Ereignisveröffentlichung über die Message Queuing Interface (MQI) auf den Warteschlangenmanager und die Warteschlangen zuzugreifen. In diesem Beispiel wird die Benutzer-ID esuser verwendet:

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

## Durchsuchen von DB2-Datenbanken auf einem Server für klassische Datenquellen

Der DB2-Crawler kann eine DB2-Datenbank auf dem Server für klassische Datenquellen über WebSphere Information Integrator Classic Federation durchsuchen.

### Informationen zu dieser Task

Damit eine DB2-Datenbank auf dem Server für klassische Datenquellen durchsucht werden kann, muss die Datenbank unter Verwendung des ODBC-Wrappers von WebSphere Information Integrator Classic Federation mit einer DB2-Datenbank auf dem Server für nicht klassische Datenquellen föderiert werden.

### Vorgehensweise

Gehen Sie wie folgt vor, um die Datenbank zu föderieren:

1. Installieren Sie WebSphere Information Integrator Classic Federation auf dem Server für klassische Datenquellen.
2. Installieren Sie das Clientmodul von WebSphere Information Integrator Classic Federation auf dem Server für nicht klassische Datenquellen, der die Datenbank enthält, die mit der Datenbank auf dem Server für klassische Datenquellen föderiert wird.
3. Konfigurieren Sie den ODBC-Treiber von WebSphere Information Integrator Classic Federation, um eine Verbindung zum Server für klassische Datenquellen herzustellen.
4. Stellen Sie die Verbindung zur föderierenden Datenbank her, und erstellen Sie den ODBC-Wrapper, mit dem die Datenbank auf dem Server für klassische Datenquellen föderiert werden soll.
5. Konfigurieren Sie den DB2-Crawler so, dass er die föderierende Datenbank auf dem Server für nicht klassische Datenquellen durchsucht. Damit wird ermöglicht, dass die Datenbank auf dem Server für klassische Datenquellen über WebSphere Information Integrator Classic Federation durchsucht werden kann.

---

## DB2 Content Manager-Crawler

Wenn Sie IBM DB2 Content Manager-Elementtypen in eine Objektgruppe für die Unternehmenssuche einschließen möchten, müssen Sie einen DB2 Content Manager-Crawler konfigurieren.

### Crawler-Server-Konfiguration

Bevor Sie einen DB2 Content Manager-Server durchsuchen können, müssen Sie ein Script auf dem Crawler-Server ausführen. Dieses Script, das im Lieferumfang von OmniFind Enterprise Edition enthalten ist, ermöglicht es dem DB2 Content Manager-Crawler, mit DB2 Content Manager-Servern zu kommunizieren.

Führen Sie die Ihrer Umgebung entsprechende Task aus, bevor Sie mit der Administrationskonsole für die Unternehmenssuche einen DB2 Content Manager-Crawler konfigurieren:

- „Konfigurieren des Crawler-Servers unter UNIX für DB2 Content Manager-Crawler“ auf Seite 63.
- „Konfigurieren des Crawler-Servers unter Windows für DB2 Content Manager-Crawler“ auf Seite 65.

#### Tipp:

Detaillierte Beispiele für die Konfiguration eines sicheren DB2 Content Manager-Crawlers finden Sie im Szenario für ein großes Unternehmen im IBM Redbook *IBM OmniFind Enterprise Edition Configuration and Implementation Scenarios*.

### Konfigurationsüberblick

Mit dem DB2 Content Manager-Crawler können Sie eine beliebige Anzahl DB2 Content Manager-Server durchsuchen. Wenn Sie den Crawler konfigurieren, geben Sie Optionen dafür an, wie der Crawler alle DB2 Content Manager-Server im Crawlerbereich durchsuchen soll. Sie wählen außerdem die jeweiligen Elementtypen aus, die auf jedem Server durchsucht werden sollen.

Wenn Sie einen DB2 Content Manager-Crawler erstellen oder ändern möchten, müssen Sie sich an der Administrationskonsole für die Unternehmenssuche anmelden. Sie müssen die Rolle eines Administrators für die Unternehmenssuche haben oder ein Objektgruppenadministrator für die Objektgruppe sein, zu der der Crawler gehört.

Wenn Sie den Crawler erstellen, unterstützt Sie ein Assistent bei den folgenden Tasks:

- Angeben von Merkmalen, die steuern, wie der Crawler ausgeführt wird und Systemressourcen verwendet. Die Crawlermerkmale steuern, wie der Crawler sämtliche Elementtypen auf allen DB2 Content Manager-Servern im Crawlerbereich durchsucht.
- Auswählen der DB2 Content Manager-Server, die durchsucht werden sollen.
- Angeben der Benutzer-IDs und Kennwörter, die es dem Crawler ermöglichen, auf den Inhalt von DB2 Content Manager-Servern zuzugreifen.
- Definieren eines Zeitplans für das Durchsuchen der Server.
- Auswählen der Elementtypen, die Sie auf jedem DB2 Content Manager-Server durchsuchen wollen.

- Angeben der Optionen zum Aktivieren der Attribute für Suchvorgänge in einigen Elementtypen. Sie können z. B. bestimmte Dokumenttypen aus dem Crawlerbereich ausschließen und angeben, welche Attribute in den Suchergebnissen zurückgegeben werden können.
- Konfigurieren von Optionen für die Sicherheit auf Dokumentebene. Wenn die Sicherheit bei der Erstellung der Objektgruppe aktiviert war, kann der Crawler Sicherheitsdaten Dokumenten im Index zuordnen. Mithilfe dieser Daten können Suchanwendungen die Zugriffssteuerung auf Basis der gespeicherten Zugriffssteuerungslisten oder Sicherheitstoken erzwingen.

Sie können außerdem eine Option auswählen, um Berechtigungsnachweise eines Benutzers bei der Übergabe einer Abfrage zu prüfen. In diesem Fall vergleicht das System die Berechtigungsnachweise des Benutzers nicht mit indextierten Sicherheitsdaten, sondern mit den aktuellen Zugriffssteuerungslisten, die von den ursprünglichen Datenquellen verwaltet werden.

Klicken Sie **Hilfe** an, während Sie den Crawler erstellen, um Informationen zu den Feldern im Assistenten und zur Vorgehensweise abzurufen, wie Sie dem Crawler die für das Durchsuchen der Daten erforderlichen Informationen zur Verfügung stellen.

## Konfigurieren des Crawler-Servers unter UNIX für DB2 Content Manager-Crawler

Wenn Sie OmniFind Enterprise Edition auf einem Computer installieren, auf dem IBM AIX, Linux oder die Solaris-Betriebsumgebung ausgeführt wird, müssen Sie ein Script ausführen, um den Crawler-Server zu konfigurieren. Mithilfe des Scripts kann der DB2 Content Manager-Crawler mit IBM DB2 Content Manager-Servern kommunizieren.

### Informationen zu dieser Task

Der DB2 Content Manager-Crawler verwendet den Java-Connector für DB2 Content Manager Version 8, um auf DB2 Content Manager-Server zuzugreifen. Sie installieren diesen Connector, indem Sie eines der folgenden Produkte auf Ihrem Crawler-Server installieren:

- IBM DB2 Information Integrator for Content, Version 8.3 für AIX, Solaris oder Linux
- IBM DB2 Information Integrator for Content, Version 8.2 für AIX oder Solaris
- IBM DB2 Content Manager Toolkit, Version 8.2 für Linux

Nach der Installation des Connectors führen Sie auf dem Crawler-Server ein Setup-Script aus, das von OmniFind Enterprise Edition bereitgestellt wird, um sicherzustellen, dass der DB2 Content Manager-Crawler mit DB2 Content Manager zusammenarbeiten kann.

### Vorgehensweise

Gehen Sie wie folgt vor, um den Crawler-Server zu konfigurieren, damit er DB2 Content Manager-Server durchsuchen kann:

1. Installieren Sie den Java-Connector für DB2 Content Manager Version 8 auf dem Crawler-Server:
  - a. Melden Sie sich auf dem Crawler-Server als Root an:  
`su - root`
  - b. Führen Sie die Datei `db2profile` aus. Beispiel:  
`./home/db2inst/sqllib/db2profile`

- c. Exportieren Sie die Umgebungsvariable JAVAHOME. Beispiel:  
export JAVAHOME=/usr/IBMJava2-141
  - d. Fügen Sie das Java-Verzeichnis der Umgebungsvariablen PATH hinzu:  
export PATH=\$PATH:\$JAVAHOME/bin
  - e. Legen Sie die DB2 Information Integrator for Content-Installations-CD ein, und führen Sie den Installationsassistenten aus.
  - f. Führen Sie im **Component Selection** die folgenden Aktionen aus. (Wenn Sie mit Information Integrator for Content Version 8.3 arbeiten, enthält das Fenster **Component Selection** die Installationsoption **Custom**.)
    - 1) Wählen Sie **Local connectors** in der Liste **Components** und dann **Content Manager V8 connector** in der Liste **Subcomponents** aus.
    - 2) Wählen Sie **Connector toolkits and samples** in der Liste **Components** und dann **Content Manager V8 connector** in der Liste **Subcomponents** aus.
  - g. Geben Sie einen Datenbanknamen, einen Benutzernamen und ein Kennwort für die DB2 Content Manager-Bibliothek an, und übernehmen Sie die Standardeinstellungen für die übrigen Fenster.
2. Melden Sie sich am Crawler-Server mit einer Benutzer-ID an, die zur DB2-Verwaltungsgruppe gehört.
  3. Katalogisieren Sie die ferne DB2 Content Manager-Bibliotheksserverdatenbank, und prüfen Sie, ob der Crawler-Server eine Verbindung zum DB2 Content Manager-Server herstellen kann:

```
db2 catalog tcpip node knotenname remote hostname server port
db2 catalog database datenbankname as aliasname at node knotenname
```

Dabei gilt Folgendes:

*knotenname*

Der Kurzname des Hosts für den DB2 Content Manager-Server (wie z. B. ibmes).

*hostname*

Der vollständig qualifizierte Hostname des DB2 Content Manager-Servers (wie z. B. ibmes.ibm.com).

*port*

Die Portnummer des DB2 Content Manager-Servers.

*datenbankname*

Der Name der DB2 Content Manager-Datenbank (wie z. B. ICMNLSDB).

*aliasname*

Der Aliasname der DB2 Content Manager-Datenbank (wie z. B. CMSVR).

4. Optional: Melden Sie sich als Root an, und testen Sie die Datenbankverbindung:

```
.installationsverzeichnis_von_information_integrator_for_content
/bin/cmbenv81.sh cd
installationsverzeichnis_von_information_integrator_for_content/samples/java/icm
javac *.java
java SConnectDisconnectICM icm-datenbankname cm-admin-id cm-admin-kennwort
```

5. Führen Sie das Setup-Script für den DB2 Content Manager-Crawler auf dem Crawler-Server aus:
  - a. Wechseln Sie in das Verzeichnis ES\_INSTALL\_ROOT/bin:

```
cd $ES_INSTALL_ROOT/bin
```

- b. Starten Sie das folgende Script, und beantworten Sie die Eingabeaufforderungen:

```
escrcm.sh
```

6. Stoppen Sie das System für die Unternehmenssuche auf dem Crawler-Server, und starten Sie es erneut:

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

## Konfigurieren des Crawler-Servers unter Windows für DB2 Content Manager-Crawler

Wenn Sie OmniFind Enterprise Edition auf einem Microsoft Windows-Computer installieren, müssen Sie ein Script ausführen, um den Crawler-Server zu konfigurieren. Mithilfe des Scripts kann der DB2 Content Manager-Crawler mit IBM DB2 Content Manager-Servern kommunizieren.

### Informationen zu dieser Task

Der DB2 Content Manager-Crawler verwendet den Java-Connector für DB2 Content Manager Version 8, um auf DB2 Content Manager-Server zuzugreifen. Sie installieren diesen Connector, indem Sie IBM DB2 Information Integrator for Content Version 8.2 oder Version 8.3 für Windows auf dem Crawler-Server installieren. Nach der Installation des Connectors führen Sie auf dem Crawler-Server ein Setup-Script aus, das von OmniFind Enterprise Edition bereitgestellt wird, um sicherzustellen, dass der DB2 Content Manager-Crawler mit DB2 Content Manager zusammenarbeiten kann.

### Vorgehensweise

Gehen Sie wie folgt vor, um den Crawler-Server zu konfigurieren, damit er DB2 Content Manager-Server durchsuchen kann:

1. Installieren Sie den Java-Connector für DB2 Content Manager Version 8 auf dem Crawler-Server:
  - a. Legen Sie die DB2 Information Integrator for Content-Installations-CD ein. Das Installationsprogramm wird automatisch gestartet.  
Der Installationsassistent von DB2 Content Manager Enterprise Information Portal wird geöffnet.
  - b. Führen Sie im **Component Selection** die folgenden Aktionen aus. (Wenn Sie mit Information Integrator for Content Version 8.3 arbeiten, enthält das Fenster **Component Selection** die Installationsoption **Custom**.)
    - 1) Wählen Sie **Local connectors** in der Liste **Components** und dann **Content Manager V8 connector** in der Liste **Subcomponents** aus.
    - 2) Wählen Sie **Connector toolkits and samples** in der Liste **Components** und dann **Content Manager V8 connector** in der Liste **Subcomponents** aus.
  - c. Geben Sie einen Datenbanknamen, einen Benutzernamen und ein Kennwort für die DB2 Content Manager-Bibliothek an, und übernehmen Sie die Standardeinstellungen für die übrigen Fenster.
2. Katalogisieren Sie die ferne DB2 Content Manager-Bibliotheksserverdatenbank, und prüfen Sie, ob der Crawler-Server eine Verbindung zum DB2 Content

Manager-Server herstellen kann. Geben Sie die folgenden Befehle an einer Eingabeaufforderung auf dem Crawler-Server ein:

```
db2 catalog tcpip node knotenname remote hostname server port
db2 catalog database datenbankname as aliasname at node knotenname
```

Dabei gilt Folgendes:

*knotenname*

Der Kurzname des Hosts für den DB2 Content Manager-Server (wie z. B. ibmes).

*hostname*

Der vollständig qualifizierte Hostname des DB2 Content Manager-Servers (wie z. B. ibmes.ibm.com).

*port*

Die Portnummer des DB2 Content Manager-Servers.

*datenbankname*

Der Name der DB2 Content Manager-Datenbank (wie z. B. ICMNLSDB).

*aliasname*

Der Aliasname der DB2 Content Manager-Datenbank (wie z. B. CMSVR).

- Optional: Testen Sie die Datenbankverbindung, indem Sie eine Eingabeaufforderung öffnen und die folgenden Befehle eingeben:

```
cmbenv81.bat
cd installationsverzeichnis_von_information_integrator_for_content
 \samples\java\icm javac *.java
java SConnectDisconnectICM icm-datenbankname cm-admin-id cm-admin-kennwort
```

- Führen Sie das Setup-Skript für den DB2 Content Manager-Crawler auf dem Crawler-Server aus:

- Wechseln Sie in das Verzeichnis ES\_INSTALL\_ROOT\bin:

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

- Starten Sie das folgende Skript, und beantworten Sie die Eingabeaufforderungen:

```
escrcm.vbs
```

- Stoppen Sie das System für die Unternehmenssuche auf dem Crawler-Server, und starten Sie es erneut:

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

---

## Domino Document Manager-Crawler

Wenn Sie Domino Document Manager-Bibliotheken und -Aktenschrankdateien in eine Objektgruppe für die Unternehmenssuche einschließen möchten, müssen Sie einen Domino Document Manager-Crawler konfigurieren.

### Crawler-Server-Konfiguration

Wenn ein Domino Document Manager-Server, den Sie durchsuchen wollen, das NRPC-Protokoll (Notes Remote Procedure Call) verwendet, müssen Sie ein Setup-Skript auf dem Crawler-Server ausführen. Dieses Skript, das im Lieferumfang von OmniFind Enterprise Edition enthalten ist, ermöglicht es dem Domino Document Manager-Crawler, mit den Servern zu kommunizieren, die NRPC verwenden.

Wenn ein Domino Document Manager-Server, den Sie durchsuchen wollen, das DIIOP-Protokoll (Domino Internet Inter-ORB Protocol) verwendet, müssen Sie kein Setup-Script auf dem Crawler-Server ausführen. Sie müssen allerdings den Domino Document Manager-Server so konfigurieren, dass der Domino Document Manager-Crawler auf den Server zugreifen kann.

**Wichtig:** Wenn der Domino Document Manager-Server das DIIOP-Protokoll verwendet und Sie den Crawler für die Verwendung von HTTPS oder DIIOP über SSL (Secure Sockets Layer) so konfigurieren, dass die Übertragungen zwischen Crawler und Server verschlüsselt werden, müssen Sie die Datei `TrustedCerts.class` (z. B. `c:\certs` oder `/data/certs`) vom Domino Document Manager-Server auf den Crawler-Server kopieren. In einer Konfiguration mit zwei oder vier Servern müssen Sie die Datei `TrustedCerts.class` auch auf die Server kopieren, auf denen die Suchkomponente installiert ist. Stellen Sie sicher, dass sich die Datei auf dem Crawler-Server und auf den Suchservern an derselben Speicherposition befindet. Den Verzeichnispfad für die Datei `TrustedCerts.class` geben Sie bei der Konfiguration des Crawlers an.

Wenn OmniFind Enterprise Edition auf einem IBM AIX-System installiert wurde, muss sichergestellt werden, dass das I/O Completion Port-Modul installiert und auf dem Crawler-Server verfügbar ist.

Führen Sie die Ihrer Umgebung entsprechenden Tasks aus, bevor Sie mit der Administrationskonsole für die Unternehmenssuche einen Domino Document Manager-Crawler konfigurieren:

- „Konfigurieren des Crawler-Servers unter UNIX für das Durchsuchen von Lotus Domino-Quellen“ auf Seite 83.
- „Konfigurieren des Crawler-Servers unter Windows für das Durchsuchen von Lotus Domino-Quellen“ auf Seite 85.
- „Konfigurieren von Servern, die das DIIOP-Protokoll verwenden“ auf Seite 86.
- „Konfigurieren des I/O Completion Ports unter AIX für das Durchsuchen von Lotus Domino-Quellen“ auf Seite 88.

## Sicherheit auf Dokumentenebene

Wenn die Objektgruppensicherheit aktiviert ist und ein Server, den Sie durchsuchen wollen, das NRPC-Protokoll verwendet, müssen Sie auf dem Crawler-Server einen Lotus Domino-Trusted Server konfigurieren. Mit dem Trusted Server wird eine Zugriffssteuerung auf Dokumentenebene umgesetzt. Führen Sie die folgenden Tasks aus, bevor Sie den Benutzern die Objektgruppe für die Suche zur Verfügung stellen:

- Konfigurieren der Lotus Domino-Trusted Server, um Benutzerberechtigungen zu prüfen.
- Aktivieren der globalen Sicherheit in WebSphere Application Server und Konfigurieren der Suchanwendung für die Verwendung der Sicherheit. Dieser Schritt stellt sicher, dass Anmeldeberechtigungen geprüft werden, wenn Benutzer versuchen, die Suchanwendung zu verwenden. Die Suchserver verwenden die Berechtigungsnachweise, um die Berechtigung jedes Benutzers zu prüfen, auf Lotus Domino-Dokumente zuzugreifen.

## Konfigurationsüberblick

Mit dem Domino Document Manager-Crawler können Sie eine beliebige Anzahl Domino Document Manager-Bibliotheken durchsuchen. Bei der Erstellung des

Crawlers wählen Sie die zu durchsuchenden Bibliotheken auf einem einzigen Domino Document Manager-Server aus. Wenn Sie später den Crawlerbereich bearbeiten, können Sie Dokumente von einem anderen Domino Document Manager-Server hinzufügen, die Sie in denselben Crawlerbereich einschließen möchten. Bei der Erstellung bzw. Bearbeitung des Crawlers können Sie angeben, ob alle Aktenschrankdateien in den für die Crawleruche ausgewählten Bibliotheken oder nur bestimmte Aktenschrankdateien durchsucht werden sollen.

Wenn Sie einen Domino Document Manager-Crawler erstellen oder ändern möchten, müssen Sie sich an der Administrationskonsole für die Unternehmenssuche anmelden. Sie müssen die Rolle eines Administrators für die Unternehmenssuche haben oder ein Objektgruppenadministrator für die Objektgruppe sein, zu der der Crawler gehört.

Wenn Sie den Crawler erstellen, unterstützt Sie ein Assistent bei den folgenden Tasks:

- Angeben von Merkmalen, die steuern, wie der Crawler ausgeführt wird und Systemressourcen verwendet. Die Crawlermerkmale steuern, wie der Crawler alle Dokumente im Crawlerbereich durchsucht.
- Angeben des Domino Document Manager-Servers und Kommunikationsprotokolls.
- Wenn der Server für die Verwendung des DIIOP-Protokolls konfiguriert ist, können Sie angeben, wie der Crawler eine Verbindung zu Domino-Objekten herstellen soll. Sie können beispielsweise Optionen angeben, um HTTPS oder Secure Sockets Layer (SSL) zu verwenden, um die Kommunikation zu verschlüsseln.
- Auswählen der zu durchsuchenden Bibliotheken.
- Definieren eines Zeitplans für das Durchsuchen der Bibliotheken.
- Auswählen der zu durchsuchenden Dokumente. Der Crawler kann alle Aktenschrankdateien in einer Bibliothek oder nur die Dokumente in den von Ihnen ausgewählten Aktenschrankdateien durchsuchen.
- Angeben von Optionen für das Aktivieren der Felder in verschiedenen Bibliotheken und Aktenschrankdateien für Suchvorgänge. Sie können z. B. bestimmte Felder aus dem Crawlerbereich ausschließen und Optionen für das Durchsuchen von Anlagen angeben.
- Konfigurieren von Optionen für die Sicherheit auf Dokumentebene. Wenn die Sicherheit bei der Erstellung der Objektgruppe aktiviert war, kann der Crawler Sicherheitsdaten Dokumenten im Index zuordnen. Mithilfe dieser Daten können Suchanwendungen die Zugriffssteuerung auf Basis der gespeicherten Zugriffssteuerungslisten oder Sicherheitstoken erzwingen.

Sie können außerdem eine Option auswählen, um Berechtigungsnachweise eines Benutzers bei der Übergabe einer Abfrage zu prüfen. In diesem Fall vergleicht das System die Berechtigungsnachweise des Benutzers nicht mit indexierten Sicherheitsdaten, sondern mit den aktuellen Zugriffssteuerungslisten, die von den ursprünglichen Datenquellen verwaltet werden.

Klicken Sie **Hilfe** an, während Sie den Crawler erstellen, um Informationen zu den Feldern im Assistenten und zur Vorgehensweise abzurufen, wie Sie dem Crawler die für das Durchsuchen der Daten erforderlichen Informationen zur Verfügung stellen.

#### **Zugehörige Konzepte**

„Tipps für das Durchsuchen von Lotus Domino-Datenbanken“ auf Seite 82

„Umsetzung der Sicherheit auf Dokumentebene für Lotus Domino-Dokumente“ auf Seite 307

### Zugehörige Tasks

„Konfigurieren des Crawler-Servers unter UNIX für das Durchsuchen von Lotus Domino-Quellen“ auf Seite 83

„Konfigurieren des Crawler-Servers unter Windows für das Durchsuchen von Lotus Domino-Quellen“ auf Seite 85

„Konfigurieren von Servern, die das DIIOP-Protokoll verwenden“ auf Seite 86

„Konfigurieren des I/O Completion Ports unter AIX für das Durchsuchen von Lotus Domino-Quellen“ auf Seite 88

„Konfigurieren der Lotus Domino-Trusted Server, um Benutzerberechtigungen zu prüfen“ auf Seite 308

---

## Exchange Server-Crawler

Wenn Sie allgemein zugängliche Ordner von Microsoft Exchange Server in eine Objektgruppe für die Unternehmenssuche einschließen möchten, müssen Sie einen Exchange Server-Crawler konfigurieren.

Mit dem Exchange Server-Crawler können Sie eine beliebige Anzahl Ordner und Unterordner auf Exchange Server-Servern mit allgemein zugänglichen Ordnern durchsuchen. Wenn Sie einen Crawler erstellen, wählen Sie den Inhalt aus, der auf einem Server mit allgemein zugänglichen Ordnern durchsucht werden soll. Sie können den Crawlerbereich später bearbeiten, um Inhalt eines weiteren Servers mit allgemein zugänglichen Ordnern hinzuzufügen.

Wenn Sie einen Exchange Server-Crawler erstellen oder ändern möchten, melden Sie sich an der Administrationskonsole für die Unternehmenssuche an. Sie müssen die Rolle eines Administrators für die Unternehmenssuche haben oder ein Objektgruppenadministrator für die Objektgruppe sein, zu der der Crawler gehört.

Wenn Sie den Crawler erstellen, unterstützt Sie ein Assistent bei den folgenden Tasks:

- Angeben von Merkmalen, die steuern, wie der Crawler ausgeführt wird und Systemressourcen verwendet. Die Crawlermerkmale steuern, wie der Crawler sämtliche Unterordner auf allen Servern im Crawlerbereich durchsucht.
- Angeben der Informationen zu dem Exchange Server-Server mit allgemein zugänglichen Ordnern, die durchsucht werden sollen.

Sie müssen eine Benutzer-ID und ein Kennwort angeben, sodass der Crawler auf den Inhalt des Servers zugreifen kann. Wenn der Server das SSL-Protokoll (Secure Sockets Layer) verwendet, können Sie Optionen angeben, die es dem Crawler ermöglichen, auf die Schlüsselspeicherdatei auf dem Crawler-Server zuzugreifen.

- Definieren eines Zeitplans für das Durchsuchen des Servers mit allgemein zugänglichen Ordnern.
- Auswählen der zu durchsuchenden Ordner und Unterordner.
- Angeben von Optionen für das Aktivieren der Dokumente für Suchvorgänge in Unterordnern. Sie können z. B. bestimmte Dokumenttypen aus dem Crawlerbereich ausschließen.
- Konfigurieren von Optionen für die Sicherheit auf Dokumentebene. Wenn die Sicherheit bei der Erstellung der Objektgruppe aktiviert war, kann der Crawler Sicherheitsdaten Dokumenten im Index zuordnen. Mithilfe dieser Daten können Suchanwendungen die Zugriffssteuerung auf Basis der gespeicherten Zugriffssteuerungslisten oder Sicherheitstoken erzwingen.

Klicken Sie **Hilfe** an, während Sie den Crawler erstellen, um Informationen zu den Feldern im Assistenten und zur Vorgehensweise abzurufen, wie Sie dem Crawler die für das Durchsuchen der Daten erforderlichen Informationen zur Verfügung stellen.

#### Zugehörige Tasks

„Zugriffsprüfung zum Schutz von Exchange Server-Dokumenten“ auf Seite 307

---

## JDBC-Datenbankcrawler

Mithilfe des JDBC-Datenbankcrawlers schließen Sie Datenbanken, auf die mit einem JDBC-Protokoll (Java Database Connectivity) zugegriffen werden kann, in eine Objektgruppe für die Unternehmenssuche ein.

Sie müssen für jeden Datenbanksystemtyp, den Sie durchsuchen wollen, einen separaten Crawler konfigurieren. Wenn Sie einen Crawler erstellen, geben Sie Optionen für das Durchsuchen einer einzelnen Datenbank an. Später können Sie dem Crawlerbereich Datenbanken desselben Typs hinzufügen.

Jede Zeile einer Datenbanktabelle wird wie ein Dokument behandelt, und die Werte der Datenbankspalten werden als durchsuchbare Felder syntaktisch analysiert und indiziert. Sie können den Crawler so konfigurieren, dass er mehrere strukturierte Tabellen durchsucht, indem Sie dem Crawler bei der Konfiguration der Crawlermerkmale ein Plug-in zuordnen. Mit dem Plug-in können Zeilen aus mehreren Tabellen einer relationalen Datenbank, die dieselben Schlüsselfelder aufweisen, verknüpft und wie ein einzelnes Dokument behandelt werden. Wenn ein Benutzer die Datenbank durchsucht, werden Daten aus den verknüpften Tabellen als zusätzliche Felder angezeigt, wenn das Dokument in den Suchergebnissen angezeigt wird.

### Unterstützte Datenbanksysteme und Treiber

Auf dem Crawler-Server muss der geeignete JDBC-Treiber vorhanden sein, damit das JDBC-Protokoll für das Durchsuchen von Tabellen in einer Datenbank verwendet werden kann. Der JDBC-Datenbankcrawler unterstützt die folgenden Datenbanksysteme und JDBC-Treiber des Typs 4:

| Datenbanksystem                                                                                                      | Name des JDBC-Treibers des Typs 4                                                                                            | Klassenpfade der JDBC-Standardtreiber                                                                                                                                                  |
|----------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| IBM DB2 Universal Database Version 8.2 und IBM DB2 Enterprise Server Edition Version 9.1 für Linux, UNIX und Windows | com.ibm.db2.jcc.DB2Driver                                                                                                    | <i>db2-installationsroot</i> /java/db2jcc.jar<br><i>db2-installationsroot</i> /java/db2jcc_license_cu.jar                                                                              |
| Oracle 9i und 10g                                                                                                    | Oracle.jdbc.driver.OracleDriver                                                                                              | <i>oracle-ausgangsverzeichnis</i> /jdbc/lib/ojdbc14.jar                                                                                                                                |
| Microsoft SQL Server 2000                                                                                            | com.microsoft.jdbc.sqlserver.SQLServerDriver                                                                                 | <i>mssql-jdbc-ausgangsverzeichnis</i> /lib/mssqlserver.jar<br><i>mssql-jdbc-ausgangsverzeichnis</i> /lib/msbase.jar<br><i>mssql-jdbc-ausgangsverzeichnis</i> /lib/msutil.jar           |
| Microsoft SQL Server 2005                                                                                            | com.microsoft.sqlserver.jdbc.SQLServerDriver (Der JDBC-Treiber für SQL Server 2005 wird auf AIX-Systemen nicht unterstützt.) | <i>installationsverzeichnis</i> /sqljdbc_1.0/loc/sqljdbc.jar<br>Dabei ist <i>loc</i> Ihre Ländereinstellung. Beispiel:<br><i>installationsverzeichnis</i> /sqljdbc_1.0/enu/sqljdbc.jar |

## JDBC-Datenbankcrawler im Vergleich mit DB2-Crawlern

Wenn Sie zurzeit den DB2-Crawler verwenden, möchten Sie möglicherweise weiterhin mit ihm arbeiten. Daten, die für einen DB2-Crawler gespeichert sind, können Sie nicht auf einen JDBC-Datenbankcrawler migrieren.

Verwenden Sie in den folgenden Situationen den DB2-Crawler statt des JDBC-Datenbankcrawlers:

- Sie wollen DB2-Datenbanken mit einem JDBC-Treiber des Typs 2 durchsuchen.
- Sie wollen Oracle- und SQL Server-Datenbanken durchsuchen, die mit einer DB2-Datenbank föderiert sind. Mit dem DB2-Crawler können Sie über einen Kurznamen auf alle diese Datenbanktypen zugreifen.
- Sie wollen DB2 für z/OS-, DB2 für iSeries-, Informix-, Sybase-, VSAM-, IMS-, CA-Datacom- oder Software AG Adabas-Datenbanken durchsuchen. Sie sollten diese Datenbanktypen mit einer DB2-Datenbank föderieren und mit dem DB2-Crawler über einen Kurznamen auf sie zugreifen.
- Sie wollen den Index für die Unternehmenssuche mithilfe der Ereignisveröffentlichung aktualisieren, wenn Aktualisierungen für eine Datenbank veröffentlicht werden.

## Konfigurationsüberblick

Wenn Sie einen JDBC-Datenbankcrawler erstellen oder ändern möchten, müssen Sie sich an der Administrationskonsole für die Unternehmenssuche anmelden. Sie müssen die Rolle eines Administrators für die Unternehmenssuche haben oder ein Objektgruppenadministrator für die Objektgruppe sein, zu der der Crawler gehört.

Wenn Sie den Crawler erstellen, unterstützt Sie ein Assistent bei den folgenden Tasks:

- Angeben von Merkmalen, die steuern, wie der Crawler ausgeführt wird und Systemressourcen verwendet. Die Crawlermerkmale steuern, wie der Crawler alle Datenbanken im Crawlerbereich durchsucht.
- Angeben des zu durchsuchenden Datenbanktyps.
- Angeben der zu durchsuchenden Datenbank sowie gegebenenfalls einer Benutzer-ID und eines Kennworts, die dem Crawler den Zugriff auf die Datenbank ermöglichen.
- Definieren eines Zeitplans für das Durchsuchen der Datenbank.
- Auswählen der zu durchsuchenden Tabellen.

**Achtung:** Zum Optimieren der Leistung der Discoveryprozesse und zum Verhindern des Überschreitens von Zeitlimits für den Crawlerkonfigurationsprozess wählen Sie das Durchsuchen aller Tabellen nur aus, wenn die Datenbank wenige Tabellen enthält bzw. die Tabellen wenige Spalten enthalten. Wenn Sie jetzt nur einige Tabellen zum Durchsuchen auswählen, können Sie den Crawlerbereich später bearbeiten und der Objektgruppe weitere Tabellen hinzufügen.

- Angeben von Optionen für das Aktivieren der Spalten für Suchvorgänge in bestimmten Tabellen. Sie können z. B. bestimmte Spalten für die Verwendung in parametrischen Suchen aktivieren oder angeben, welche Spalten in den Suchergebnissen zurückgegeben werden können.
- Konfigurieren von Optionen für die Sicherheit auf Dokumentenebene. Wenn die Sicherheit bei der Erstellung der Objektgruppe aktiviert war, kann der Crawler Sicherheitsdaten Dokumenten im Index zuordnen. Mithilfe dieser Daten können Suchanwendungen die Zugriffssteuerung auf Basis der gespeicherten Zugriffssteuerungslisten oder Sicherheitstoken erzwingen.

Klicken Sie **Hilfe** an, während Sie den Crawler erstellen, um Informationen zu den Feldern im Assistenten und zur Vorgehensweise abzurufen, wie Sie dem Crawler die für das Durchsuchen der Daten erforderlichen Informationen zur Verfügung stellen.

## Beziehungszuordnungen für JDBC-Datenbanken

Wenn Sie Regeln für ein Plug-in erstellen, das mehrere strukturierte JDBC-Datenbanktabellen durchsucht, geben Sie Informationen zur Stammtabelle an und wie die übergeordneten und untergeordneten Tabellen verknüpft werden.

Ein für die Unternehmenssuche bereitgestelltes Plug-in aktiviert den JDBC-Datenbank-Crawler für die Verknüpfung mehrerer strukturierter Tabellen. Sie erstellen das Plug-in, indem Sie in der Datei `ES_INSTALL_ROOT/default_config/crawler_rdb_plugin.xml` Regeln angeben. Nachdem Sie den Crawler für die Verwendung des Plug-ins konfiguriert haben, werden Zeilen in Tabellen, die dieselben Schlüsselfelder haben, verknüpft und wie ein einzelnes Dokument behandelt. Wenn ein Benutzer die Datenbank durchsucht, werden Daten aus den verknüpften Tabellen als zusätzliche Felder angezeigt, wenn das Dokument in den Suchergebnissen angezeigt wird.

### Verknüpfen von Tabellen durch Schlüsselspalten

Die folgende Abbildung zeigt, wie die Beziehungszuordnung für mehrere Tabellen erstellt wird. Der JDBC-Datenbankcrawler durchsucht eine Stammtabelle in einer Datenbank. Einige Spalten in der Tabelle sind Schlüsselfelder, die verwendet werden können, um die Tabelle mit anderen Tabellen zu verknüpfen. Anschließend können Spalten in den verknüpften Tabellen als Schlüssel zum Verknüpfen weiterer Tabellen verwendet werden. Die Zeilen in den Mehrfachtabellen werden im Crawlerbereich wie ein einzelnes Dokument behandelt. Die Spaltenwerte werden wie Dokumentmetadaten behandelt. Die Stammtabelle ist das *übergeordnete Objekt* in der Beziehung, und eine auf der ersten Ebene verknüpfte Tabelle ist ein *untergeordnetes Objekt*. Untergeordnete Tabellen der ersten Ebene können auch übergeordnete Objekte von Tabellen sein, die auf einer zweiten Ebene verknüpft sind.

In diesem Beispiel sind die Spalten **Schlüssel 1** und **Schlüssel 2** in der Stammtabelle (übergeordnet) Schlüsselfelder, die es ermöglichen, dass die Tabelle mit den untergeordneten Tabellen verknüpft werden kann, die ebenfalls die Spalten **Schlüssel 1** und **Schlüssel 2** aufweisen. Eine auf dieser ersten Ebene verknüpfte Tabelle hat die Schlüsselfelder **Schlüssel 3** und **Schlüssel 4**, aufgrund derer die Tabelle mit weiteren Tabellen verknüpft werden kann.

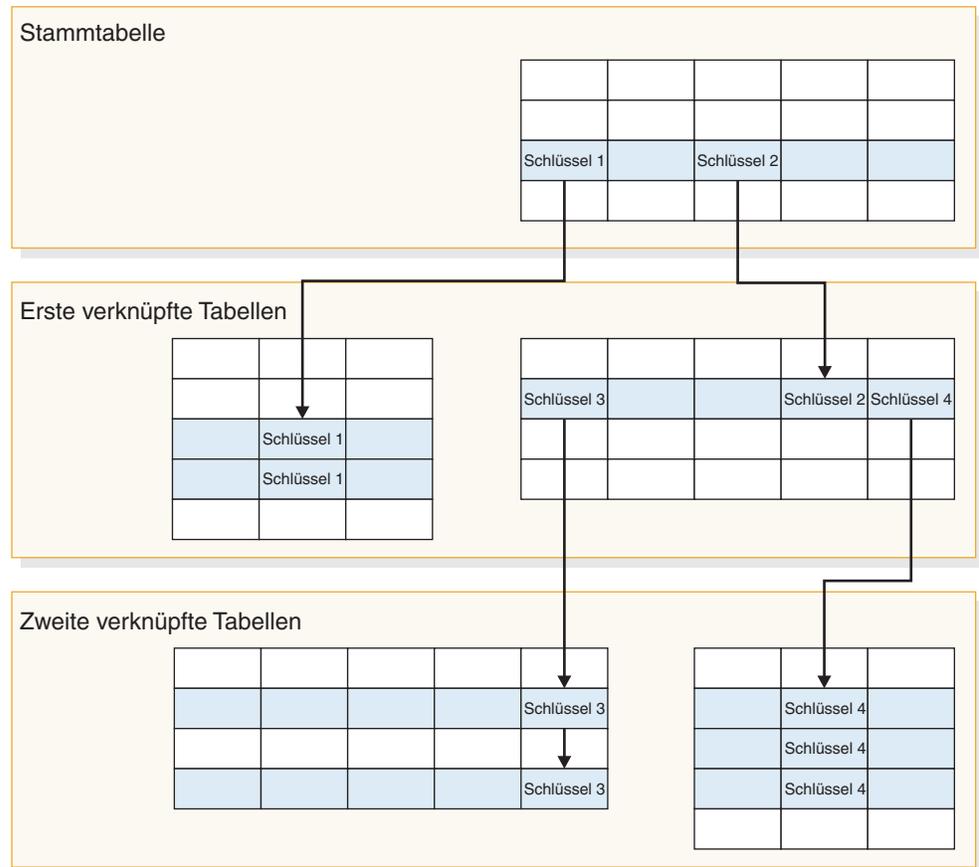


Abbildung 2. Durch Schlüsselfelder verknüpfte JDBC-Datenbanktabellen

## Anzeigen von Suchergebnissen

Die folgende Abbildung zeigt wie Daten aus mehreren strukturierten Tabellen in den Suchergebnissen wie ein einzelnes Dokument angezeigt werden. Einem Benutzer, der die Tabelle EMPLOYEE ohne das Plug-in durchsucht hat, wird möglicherweise eine Zeile der Stammtabelle in den Suchergebnissen angezeigt, sodass er nur die Werte für die Spalten (ID, Name und Office) der Tabelle EMPLOYEE sieht.

Wird das Plug-in verwendet, kann der Crawler die Spalte **Office** als Schlüssel verwenden, um die Tabelle EMPLOYEE mit der Tabelle OFFICE zu verknüpfen. Die Spalte **Country** der Tabelle OFFICE dient als Schlüssel für die Verknüpfung dieser Tabelle mit der Tabelle COUNTRY. Nachdem die Tabellen verknüpft sind, werden Benutzern, die die Tabelle EMPLOYEE durchsuchen, die Werte der Spalten aus den Tabellen OFFICE und COUNTRY als zusätzliche Felder in den Suchergebnissen angezeigt.

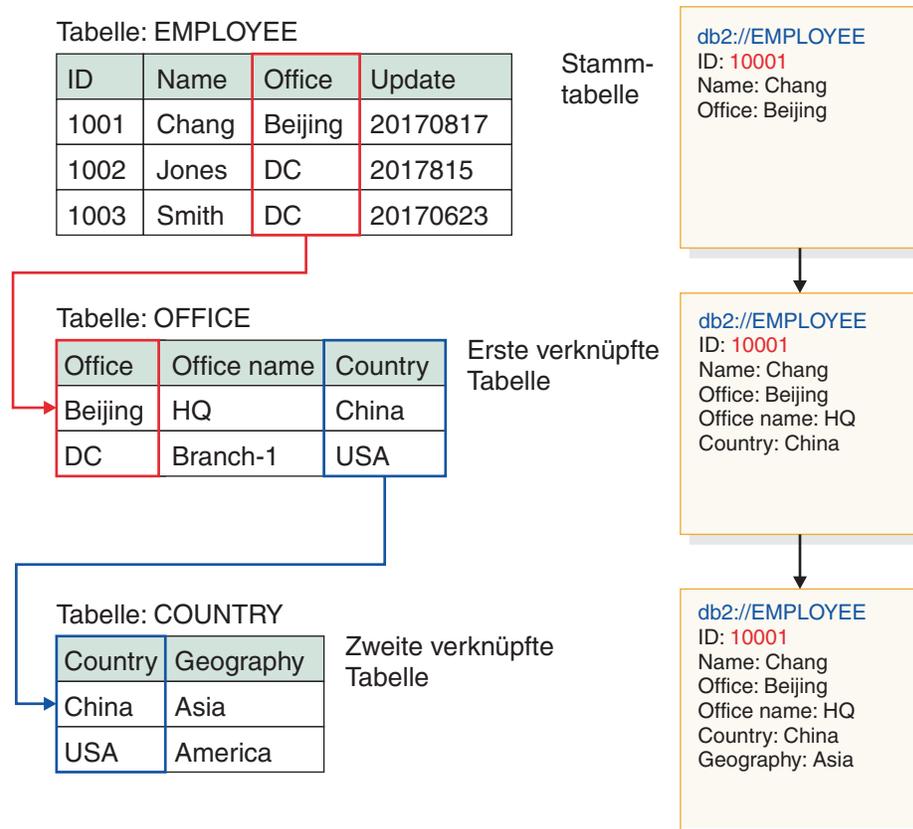


Abbildung 3. Werte aus verknüpften JDBC-Tabellen werden in den Suchergebnissen angezeigt

## Durchsuchen mehrerer strukturierter JDBC-Datenbanktabellen

Sie können den JDBC-Datenbankcrawler so konfigurieren, dass er mehrere strukturierte Tabellen verknüpft, die dieselben Schlüsselfelder enthalten.

### Vorbereitung

Für diese Task müssen Sie berechtigt sein, sich als Administrator für die Unternehmenssuche anzumelden.

### Informationen zu dieser Task

Wenn Sie die Crawlermerkmale für einen JDBC-Datenbankcrawler konfigurieren, können Sie ein Plug-in zum Durchsuchen mehrerer strukturierter Tabellen angeben, die durch Schlüsselfelder miteinander verbunden sind. Ohne das Plug-in werden die Zeilen einer Datenbanktabelle wie einzelne Dokumente behandelt, und die Werte der Datenbankspalten werden als einzelne Felder durchsucht. Mit dem Plug-in werden Zeilen aus mehreren Tabellen einer relationalen Datenbank, die dieselben Schlüsselfelder aufweisen, verknüpft und wie ein einzelnes Dokument behandelt. Der Crawler fügt die Daten, die er aus den verknüpften Tabellen abrufen kann, den Metadaten der ursprünglichen Zeile einer Datenbanktabelle hinzu. Wenn ein Benutzer die Datenbank durchsucht, werden diese Zusatzdaten als zusätzliche Felder angezeigt, wenn das Dokument in den Suchergebnissen angezeigt wird.

## Einschränkungen

### Datentypen, die nicht durchsucht werden können

Der Crawler kann Felder in den verknüpften Tabellen nicht durchsuchen, die die folgenden binären Datentypen enthalten:

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

### Einschränkungen für den Umfang des Crawlerbereichs

Die Tabellen, die verknüpft werden sollen, müssen sich in derselben relationalen Datenbank befinden. Sie können Tabellen nicht datenbankübergreifend verknüpfen.

Ist eine Tabelle in einer Datenbank für die Verknüpfung mit anderen Tabellen konfiguriert, gilt diese Einstellung universell für alle Crawler in einer Objektgruppe, für die die Verwendung des Plug-ins aktiviert wurde. Sie können jedoch mehrere Objektgruppen erstellen und verschiedene Crawler für die Crawlersuche in unterschiedlichen Stammtabellen und für die Verknüpfung verschiedener Tabellen konfigurieren.

### Einschränkungen für die Verwendung anderer Plug-ins

Wenn Sie den Crawler für die Verwendung des Plug-ins zum Durchsuchen von mehreren strukturierten Tabellen konfiguriert haben, können Sie dem Crawler kein anderes Plug-in zuordnen. Sie können z. B. kein benutzerdefiniertes Plug-in für die Anwendung von Geschäfts- und Sicherheitsregeln angeben. Sie können einem Crawler immer nur ein Plug-in zuordnen.

### Einschränkungen der Anzahl Tabellen, Zeilen, Felder und Schlüssel

Die maximale Anzahl verknüpfter Tabellen pro Datenbank ist fünf, und die Summe der Zeilen in diesen Tabellen muss kleiner sein als eine Million. Die maximale Anzahl Felder, die aus einer Tabelle gelesen werden können, ist 10. Für die Verknüpfung von Tabellen wird ein Schlüsselpaar verwendet. Das heißt, es ist nicht möglich, für die Verknüpfung von Tabellen mehrere Schlüssel zu verwenden.

### Sicherstellen, dass Änderungen in verknüpften Tabellen durchsucht werden

Wenn sich die Zeilen einer Stammtabelle zwischen zwei Suchläufen nicht ändern und der Crawler nicht für eine vollständige Crawlersuche konfiguriert ist, ignoriert der Crawler die unveränderten Zeilen. Ändern sich Zeilen einer Tabelle, die mit der Stammtabelle verknüpft ist, auch wenn die Stammtabelle selbst nicht geändert wird, müssen Sie eine der folgenden Aktionen ausführen, um sicherzustellen, dass die Änderungen erkannt und durchsucht werden:

- Eine Stammtabelle in der Zieldatenbank sollte ein Zeitmarkenfeld aufweisen. Konfigurieren Sie die Zieldatenbank so, dass das Zeitmarkenfeld aktualisiert wird, wenn eine Zeile in der Stammtabelle geändert wird oder wenn Zeilen in einer beliebigen verknüpften, untergeordneten Tabelle geändert werden. Geben Sie beim Einrichten des JDBC-Datenbankcrawlers dieses Zeitmarkenfeld als das Feld an, anhand dessen der Crawler ermittelt, ob Änderungen in der Tabelle aufgetreten sind.
- Geben Sie bei der Konfiguration des Crawlerzeitplans an, dass der Crawler eine vollständige Crawlersuche ausführen soll. Diese Option gewährleistet, dass jedes Mal alle Tabellen durchsucht werden, unabhängig davon, ob Änderungen aufgetreten sind oder nicht.

## Vorgehensweise

Gehen Sie wie folgt vor, um den JDBC-Datenbankcrawler zum Durchsuchen mehrerer strukturierter Tabellen zu konfigurieren:

1. Melden Sie sich als Administrator für die Unternehmenssuche am Crawler-Server an, und kopieren Sie die Datei `ES_INSTALL_ROOT/default_config/crawler_rdb_plugin.xml`, um die Datei `ES_NODE_ROOT/master_config/crawler_rdb_plugin.xml` zu erstellen.
2. Bearbeiten Sie die Datei `ES_NODE_ROOT/master_config/crawler_rdb_plugin.xml` mit einem Texteditor, der die UTF-8-Codierung unterstützt.
  - a. Bearbeiten Sie das Element `<Server DBURL="jdbc:db2://db_server_url:50000/SAMPLE">`, und ersetzen Sie `jdbc:db2://db_server_url:50000/SAMPLE` durch die URL der JDBC-Datenbank, die durchsucht werden soll. Geben Sie bei der Konfiguration des Crawlers dieselbe URL für die zu durchsuchende Datenbank an.
  - b. Handelt es sich bei der zu durchsuchenden Datenbank nicht um eine DB2-Datenbank, bearbeiten Sie das Element `<JDBCdriver>com.ibm.db2.jcc.DB2Driver</JDBCdriver>`, und ersetzen Sie `com.ibm.db2.jcc.DB2Driver` durch den entsprechenden JDBC-Treiber. Geben Sie bei der Konfiguration des Crawlers denselben Treiber für die zu durchsuchende Datenbank an.
  - c. Bearbeiten Sie das Element `<User>username</User>`, und ersetzen Sie `username` durch eine Benutzer-ID, die über die Berechtigung verfügt, auf die zu durchsuchende Datenbank zuzugreifen.
  - d. Bearbeiten Sie das Element `<Password Encryption="True">encrypted_password</Password>`, und ersetzen Sie `encrypted_password` durch ein verschlüsseltes Kennwort für die angegebene Benutzer-ID. Sie können das verschlüsselte Kennwort aus der Datei `ES_NODE_ROOT/master_config/col_objektgruppenname.JDBC_crawlername/jdbcrawler.xml` kopieren und hier einfügen. Wenn es nicht erforderlich ist, das Kennwort zu verschlüsseln, ersetzen Sie `Encryption="True"` durch `Encryption="False"`, und ersetzen Sie `encrypted_password` durch ein unverschlüsseltes Kennwort.
  - e. Wenn Sie das Element `<Delimiters Use="True">` unverändert lassen, werden Mehrfachbegriffe in einer Spalte durch Zeichen (,) getrennt, die im Element `<Delimiter>` definiert sind. Die Begriffsgruppen pro Tabelle werden durch Zeichen (;) getrennt, die im Element `<SecondDelimiter>` definiert sind. Wenn Sie `<Delimiters Use="True">` auf `<Delimiters Use="False">` setzen, werden keine Begrenzungszeichen verwendet, und mehrere Metadatenfelder mit demselben Feldnamen werden als Dokumentmetadaten hinzugefügt.
  - f. Wenn Sie das Element `<Delimiters Use="True">` verwenden, bearbeiten Sie die Elemente `<Delimiter>`, `</Delimiter>` und `<SecondDelimiter>`; `</SecondDelimiter>`, um die Zeichen anzugeben, die als Trennzeichen für Werte verwendet werden sollen.
  - g. Bearbeiten Sie das Element `<RelationMap Root="DB2INST1.TABLE_0">`, und ersetzen Sie `DB2INST1.TABLE_0` durch den Namen einer Stammtabelle, die durchsucht werden soll.
  - h. Bearbeiten Sie das Element `<Relation Parent="DB2INST1.TABLE_0" ParentAlias="T0" ParentKey="ID" Child="DB2INST1.TABLE_1" ChildAlias="T1" ChildKey="ID"/>`.
    - Ersetzen Sie `Parent="DB2INST1.TABLE_0"` durch den Namen einer Tabelle, die in der Beziehung übergeordnet ist.

- Ersetzen Sie ParentKey= "T0" durch einen Aliasnamen der übergeordneten Tabelle. Der Aliasname muss eindeutig sein und darf in der Datei crawler\_rdb\_plugin.xml nur einmal vorkommen.
- Ersetzen Sie ParentKey="ID" durch den Namen einer Spalte, die in der Beziehung als Schlüsselfeld verwendet wird.
- Ersetzen Sie Child="DB2INST1.TABLE\_1" ChildAlias="T1" ChildKey="ID" durch Informationen zu einer zu durchsuchenden untergeordneten Tabelle.

Diese Struktur definiert, wie die Tabellen verknüpft werden. Die folgende Beziehungszuordnung gibt z. B. an, dass eine Stammtabelle namens DB2INST1.TABLE\_A durchsucht werden soll. Die Tabellen DB2INST1.TABLE\_B und DB2INST1.TABLE\_C werden mit der Bedingung DB2INST1.TABLE\_A.ID=DB2INST1.TABLE\_B.ID AND DB2INST1.TABLE\_B.ID=DB2INST1.TABLE\_C.ID verknüpft.

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

- Wiederholen Sie Schritt 2h auf Seite 76, um für alle Beziehungen, die Tabellen aus einer Stammtabelle verknüpfen, die Elemente <Relation> zu erstellen.
- Bearbeiten Sie das Element <Target TableAlias="T1">, und ersetzen Sie TableAlias="T1" durch einen Wert ChildAlias, den Sie in Schritt 2h auf Seite 76 definiert haben.
- Bearbeiten Sie das Element <Field Name="ID" FieldName="ID\_1" Enabling="True" Searchable="True" FieldSearchable="True" IsContent="True"/>.
  - Ersetzen Sie Name="ID" durch den Namen einer Spalte in den zu durchsuchenden Dokumenten.
  - Ersetzen Sie FieldName="ID\_1" durch den Namen eines Metadatenfelds in den zu durchsuchenden Dokumenten. Dieser Wert wird als Anzeigename für die Spalte in der Administrationskonsole für die Unternehmenssuche und die Suchergebnisse verwendet.
  - Ersetzen Sie Enabling="True" durch "False", wenn diese Spalte nicht in die Dokumentmetadaten eingeschlossen werden soll.
  - Ersetzen Sie Searchable="True" durch "False", um zu verhindern, dass Benutzer diese Spalte mit einer Freiformatabfrage durchsuchen.
  - Ersetzen Sie FieldSearchable="True" durch "False", um zu verhindern, dass Benutzer diese Spalte nach dem Spaltennamen durchsuchen.
  - Ersetzen Sie IsContent="True" durch "False", um anzugeben, dass die Spalte keinen durchsuchbaren Inhalt enthält. Wenn Sie Searchable="True" und IsContent="True" angeben, wird der Wert der Spalte verwendet, um Kopien eines Dokuments zu erkennen, und er wird Teil der dynamischen Dokumentzusammenfassung in den Suchergebnissen.
- Wiederholen Sie Schritt 2k, um für alle zu durchsuchenden Spalten die Elemente <Field> zu erstellen.
- Wiederholen Sie die Schritte 2j und 2k, um die Elemente <Target> und <Field> für alle untergeordneten Tabellen zu erstellen, auf die in der Beziehungszuordnung (<RelationMap>) verwiesen wird.
- Wiederholen Sie die Schritte 2g auf Seite 76 bis 2m, um mehrere Beziehungszuordnungen für mehrere Stammtabellen zu erstellen.

- o. Wiederholen Sie die Schritte 2a auf Seite 76 bis 2n auf Seite 77, um eine Beziehungszuordnung für eine andere Datenbank zu konfigurieren.
- 3. Konfigurieren Sie den Crawler für die Verwendung des Plug-ins:
  - a. Öffnen Sie die Administrationskonsole für die Unternehmenssuche, bearbeiten Sie eine Objektgruppe, und wählen Sie die Seite **Durchsuchen** aus.
  - b. Erstellen Sie einen JDBC-Datenbankcrawler, oder suchen Sie einen Crawler, den Sie ändern wollen, und klicken Sie  **Crawlermerkmale** an.
  - c. Geben Sie den Namen des Plug-ins zum Durchsuchen mehrerer strukturierter Tabellen in das Feld **Plug-in-Klassenname** ein:
 

```
com.ibm.es.plugin.rdb.RDBPlugin
```
  - d. Geben Sie die vollständig qualifizierten Pfade des Plug-ins und der vom Plug-in verwendeten JDBC-Treiber in das Feld **Plug-in-Klassenpfad** ein. Der Pfad eines JDBC-Treibers für eine DB2-Datenbank in einem Windows-System kann z. B. wie folgt aussehen:
 

```
C:\Programme\IBM\es\lib\plugin_rdb.jar;C:\Programme\IBM\SQLLIB\java\
db2jcc.jar;C:\Programme\IBM\SQLLIB\java\db2jcc_license_cu.jar;
```
  - e. Klicken Sie **Weiter** an, um die Erstellung des Crawlers fortzusetzen, oder klicken Sie **OK** an, um Ihre Änderungen zu speichern.
- 4. Starten Sie das System für die Unternehmenssuche erneut, um die Datei `crawler_rdb_plugin.xml` in der Systemkonfiguration zu implementieren:

```
esadmin system stop
esadmin system start
```

---

## NNTP-Crawler

Wenn Sie Artikel von NNTP-Newsgroups in eine Objektgruppe für die Unternehmenssuche einschließen wollen, müssen Sie einen NNTP-Crawler konfigurieren.

Mit dem NNTP-Crawler können Sie eine beliebige Anzahl NNTP-Server durchsuchen. Wenn Sie den Crawler konfigurieren, wählen Sie die zu durchsuchenden Newsgroups von einem NNTP-Server aus. Wenn Sie später den Crawlerbereich bearbeiten, können Sie weitere NNTP-Server hinzufügen, die vom Crawler durchsucht werden sollen.

Wenn Sie die zu durchsuchenden Newsgroups angeben, können Sie Gruppen auswählen, die in den Crawlerbereich aufgenommen werden sollen, sowie Gruppen, die aus dem Crawlerbereich ausgeschlossen werden sollen. Auf diese Weise können Sie es dem Crawler problemlos ermöglichen, einen Großteil der Newsgroups auf einem Server zu durchsuchen, sowie dem Crawler das Durchsuchen einiger Newsgroups zu untersagen, die nicht von Benutzern durchsucht werden sollen.

Sie können z. B. Regeln festlegen, um alle Newsgroups auf einem bestimmten NNTP-Server einzuschließen. Anschließend geben Sie an, dass Sie Newsgroups auf diesem Server ausschließen wollen, wenn deren Name die Zeichenfolge `private` enthält.

Wenn Sie einen NNTP-Crawler erstellen oder ändern möchten, melden Sie sich an der Administrationskonsole für die Unternehmenssuche an. Sie müssen die Rolle eines Administrators für die Unternehmenssuche haben oder ein Objektgruppenadministrator für die Objektgruppe sein, zu der der Crawler gehört.

Wenn Sie den Crawler erstellen, unterstützt Sie ein Assistent bei den folgenden Tasks:

- Angeben von Merkmalen, die steuern, wie der Crawler ausgeführt wird und Systemressourcen verwendet. Die Crawlermerkmale steuern, wie der Crawler alle Newsgroups im Crawlerbereich durchsucht.
- Angeben des zu durchsuchenden NNTP-Servers. Ist der Server kennwortgeschützt, müssen Sie eine Benutzer-ID und ein Kennwort angeben, mit denen der Crawler auf Newsgroups auf dem Server zugreifen kann.
- Definieren eines Zeitplans für das Durchsuchen des Servers.
- Angeben von Mustern zum Einschließen von Newsgroups in den Crawlerbereich und Angeben von Mustern zum Ausschließen von bestimmten Newsgroups aus dem Crawlerbereich.
- Angeben, ob der Crawler Sprache und Codepage der zu durchsuchenden Artikel automatisch erkennen soll oder ob der Crawler eine bestimmte Sprache und eine bestimmte Codepage verwenden soll.
- Konfigurieren von Optionen für die Sicherheit auf Dokumentenebene. Wenn die Sicherheit bei der Erstellung der Objektgruppe aktiviert war, kann der Crawler Sicherheitsdaten Dokumenten im Index zuordnen. Mithilfe dieser Daten können Suchanwendungen die Zugriffssteuerung auf Basis der gespeicherten Zugriffssteuerungslisten oder Sicherheitstoken erzwingen.

Klicken Sie **Hilfe** an, während Sie den Crawler erstellen, um Informationen zu den Feldern im Assistenten und zur Vorgehensweise abzurufen, wie Sie dem Crawler die für das Durchsuchen der Daten erforderlichen Informationen zur Verfügung stellen.

---

## Notes-Crawler

Wenn Sie IBM Lotus Notes-Datenbanken in eine Objektgruppe für die Unternehmenssuche einschließen möchten, müssen Sie einen Notes-Crawler konfigurieren.

### Tipp:

Detaillierte Beispiele für die Konfiguration eines sicheren Notes-Crawlers finden Sie im Szenario für ein großes Unternehmen im IBM Redbook *IBM OmniFind Enterprise Edition Configuration and Implementation Scenarios*.

## Crawler-Server-Konfiguration

Wenn ein Lotus Notes-Server, den Sie durchsuchen wollen, das NRPC-Protokoll (Notes Remote Procedure Call) verwendet, müssen Sie ein Script auf dem Crawler-Server ausführen. Dieses Script, das im Lieferumfang von OmniFind Enterprise Edition enthalten ist, ermöglicht es dem Notes-Crawler, mit den Servern zu kommunizieren, die NRPC verwenden.

Wenn ein Lotus Notes-Server, den Sie durchsuchen wollen, das DIIOP-Protokoll (Domino Internet Inter-ORB Protocol) verwendet, müssen Sie kein Setup-Script auf dem Crawler-Server ausführen. Sie müssen allerdings den Lotus Notes-Server so konfigurieren, dass der Notes-Crawler auf den Server zugreifen kann.

**Wichtig:** Wenn der Lotus Notes-Server das DIIOP-Protokoll verwendet und Sie den Crawler für die Verwendung von HTTPS oder DIIOP über SSL (Secure Sockets Layer) so konfigurieren, dass die Übertragungen zwischen Crawler und Server verschlüsselt werden, müssen Sie die Datei `TrustedCerts.class` (z. B. `c:\certs` oder

/data/certs) vom Lotus Notes-Server auf den Crawler-Server kopieren. In einer Konfiguration mit zwei oder vier Servern müssen Sie die Datei TrustedCerts.class auch auf die Server kopieren, auf denen die Suchkomponente installiert ist. Stellen Sie sicher, dass sich die Datei auf dem Crawler-Server und auf den Suchservern an derselben Speicherposition befindet. Den Verzeichnispfad für die Datei TrustedCerts.class geben Sie bei der Konfiguration des Crawlers an.

Wenn OmniFind Enterprise Edition auf einem IBM AIX-System installiert wurde, muss sichergestellt werden, dass das I/O Completion Port-Modul installiert und auf dem Crawler-Server verfügbar ist.

Führen Sie die Ihrer Umgebung entsprechenden Tasks aus, bevor Sie mit der Administrationskonsole für die Unternehmenssuche einen Notes-Crawler konfigurieren:

- „Konfigurieren des Crawler-Servers unter UNIX für das Durchsuchen von Lotus Domino-Quellen“ auf Seite 83.
- „Konfigurieren des Crawler-Servers unter Windows für das Durchsuchen von Lotus Domino-Quellen“ auf Seite 85.
- „Konfigurieren von Servern, die das DIIOP-Protokoll verwenden“ auf Seite 86.
- „Konfigurieren des I/O Completion Ports unter AIX für das Durchsuchen von Lotus Domino-Quellen“ auf Seite 88.

## **Sicherheit auf Dokumentebene**

Wenn die Objektgruppensicherheit aktiviert ist und ein Server, den Sie durchsuchen wollen, das NRPC-Protokoll verwendet, müssen Sie auf dem Crawler-Server einen Lotus Domino-Trusted Server konfigurieren. Mit dem Trusted Server wird eine Zugriffssteuerung auf Dokumentebene umgesetzt. Führen Sie die folgenden Tasks aus, bevor Sie den Benutzern die Objektgruppe für die Suche zur Verfügung stellen:

- Konfigurieren der Lotus Domino-Trusted Server, um Benutzerberechtigungen zu prüfen.
- Aktivieren der globalen Sicherheit in WebSphere Application Server und Konfigurieren der Suchanwendung für die Verwendung der Sicherheit. Dieser Schritt stellt sicher, dass Anmeldeberechtigungen geprüft werden, wenn Benutzer versuchen, die Suchanwendung zu verwenden. Die Suchserver verwenden die Berechtigungsnachweise, um die Berechtigung jedes Benutzers zu prüfen, auf Lotus Domino-Dokumente zuzugreifen.

## **Konfigurationsüberblick**

Mit dem Notes-Crawler können Sie eine beliebige Anzahl Lotus Notes-Standarddatenbanken (NSF-Dateien) durchsuchen. Bei der Erstellung des Crawlers wählen Sie die zu durchsuchenden Datenbanken oder Verzeichnisse von einem einzigen Lotus Notes-Server aus. Wenn Sie später den Crawlerbereich bearbeiten, können Sie Dokumente von einem anderen Lotus Notes-Server hinzufügen, die Sie in denselben Crawlerbereich einschließen möchten. Wenn Sie den Crawler erstellen oder bearbeiten, können Sie angeben, ob Sie alle Datenbanken oder Verzeichnisse auf dem Server oder nur bestimmte Datenbanken, Sichten und Ordner durchsuchen möchten.

Wenn Sie einen Notes-Crawler erstellen oder ändern möchten, müssen Sie sich an der Administrationskonsole für die Unternehmenssuche anmelden. Sie müssen die

Rolle eines Administrators für die Unternehmenssuche haben oder ein Objektgruppenadministrator für die Objektgruppe sein, zu der der Crawler gehört.

Wenn Sie den Crawler erstellen, unterstützt Sie ein Assistent bei den folgenden Tasks:

- Angeben von Merkmalen, die steuern, wie der Crawler ausgeführt wird und Systemressourcen verwendet. Die Crawlermerkmale steuern, wie der Crawler alle Dokumente im Crawlerbereich durchsucht.
- Angeben des Hostnamens, des Ports und des Kommunikationsprotokolls für den Lotus Notes-Server.
- Wenn der Server für die Verwendung des DIIOP-Protokolls konfiguriert ist, können Sie angeben, wie der Crawler eine Verbindung zu Domino-Objekten herstellen soll. Sie können beispielsweise Optionen angeben, um HTTPS oder Secure Sockets Layer (SSL) zu verwenden, um die Kommunikation zu verschlüsseln.
- Auswählen der zu durchsuchenden Datenbanken oder Verzeichnisse. Wenn Sie Verzeichnisse durchsuchen, können Sie Muster angeben, um Datenbanken ein- oder auszuschließen, was Ihnen dabei helfen kann, die Task, große Verzeichnisse zu durchsuchen, auf mehrere Crawler zu verteilen.
- Definieren eines Zeitplans für das Durchsuchen der Datenbanken oder Verzeichnisse.
- Auswählen der zu durchsuchenden Dokumente. Sie können alle Dokumente in einem Verzeichnis, alle Dokumente in einer Datenbank oder Dokumente aus ausgewählten Sichten und Ordnern einer Datenbank durchsuchen.
- Angeben von Optionen für das Aktivieren der Felder für Suchvorgänge in verschiedenen Datenbanken, Sichten und Ordnern. Sie können z. B. bestimmte Felder aus dem Crawlerbereich ausschließen und Optionen für das Durchsuchen von Anlagen angeben.
- Konfigurieren von Optionen für die Sicherheit auf Dokumentenebene. Wenn die Sicherheit bei der Erstellung der Objektgruppe aktiviert war, kann der Crawler Sicherheitsdaten Dokumenten im Index zuordnen. Mithilfe dieser Daten können Suchanwendungen die Zugriffssteuerung auf Basis der gespeicherten Zugriffssteuerungslisten oder Sicherheitstoken erzwingen.

Sie können außerdem eine Option auswählen, um Berechtigungsnachweise eines Benutzers bei der Übergabe einer Abfrage zu prüfen. In diesem Fall vergleicht das System die Berechtigungsnachweise des Benutzers nicht mit indexierten Sicherheitsdaten, sondern mit den aktuellen Zugriffssteuerungslisten, die von den ursprünglichen Datenquellen verwaltet werden.

Klicken Sie **Hilfe** an, während Sie den Crawler erstellen, um Informationen zu den Feldern im Assistenten und zur Vorgehensweise abzurufen, wie Sie dem Crawler die für das Durchsuchen der Daten erforderlichen Informationen zur Verfügung stellen.

#### **Zugehörige Konzepte**

„Tipps für das Durchsuchen von Lotus Domino-Datenbanken“ auf Seite 82

„Umsetzung der Sicherheit auf Dokumentenebene für Lotus Domino-Dokumente“ auf Seite 307

#### **Zugehörige Tasks**

„Konfigurieren des Crawler-Servers unter UNIX für das Durchsuchen von Lotus Domino-Quellen“ auf Seite 83

„Konfigurieren des Crawler-Servers unter Windows für das Durchsuchen von Lotus Domino-Quellen“ auf Seite 85

„Konfigurieren von Servern, die das DIIOP-Protokoll verwenden“ auf Seite 86

„Konfigurieren des I/O Completion Ports unter AIX für das Durchsuchen von Lotus Domino-Quellen“ auf Seite 88

„Konfigurieren der Lotus Domino-Trusted Server, um Benutzerberechtigungen zu prüfen“ auf Seite 308

## Tipps für das Durchsuchen von Lotus Domino-Datenbanken

Lesen Sie die Richtlinien zum Durchsuchen von Lotus Domino-Datenbanken, bevor Sie einen Notes-Crawler konfigurieren.

- Notes-Datenbanken, die auf Standardschablonen basieren (z. B. eine Diskussionsdatenbank) eignen sich am besten für die Crawlersuche in Datenbanken.
- Der Notes-Crawler wendet die folgenden Feldzuordnungsregeln an:
  - Zunächst werden die zentralen Feldnamen in den Domino-Standard-schablonen registriert.
  - Werte aus Notes-Feldern, die in der Zuordnungsregeltabelle angegeben sind, werden in den Suchergebnissen als Dokumentzusammenfassung verwendet.
  - Werte aus Notes-Feldern, die nicht in der Zuordnungsregeltabelle angegeben sind, werden nicht in den Dokumentzusammenfassungen verwendet.
  - Werte aus Notes-Feldern, die dem Feld **Title** zugeordnet sind, werden in den Suchergebnissen als Dokumenttitel verwendet.
  - Die Felder in der folgenden Tabelle werden standardmäßig für die Suchfeldnamen zugeordnet:

*Tabelle 2. Standardfeldzuordnungsregeln*

Notes-Datenbankfeldname	Suchfeldname
Title	Title
EventTitle	Title
Subject	Title
Body	Body
Mission	Body
From	Creator
Author	Creator
Keywords	Categories
Categories	Categories
TeamRoomName	Organization
TeamName	Organization
Department	Organization

- Der Notes-Crawler kann alle Feldtypen durchsuchen mit Ausnahme von Feldern des Typs Für die Anzeige berechnet.
- Unveränderlicher Text sowie Bilder in einem Notes-Formular werden nicht durchsucht.
- Wenn Sie den Crawler konfigurieren, wählen Sie das Markierungsfeld **Alle** unter **Crawler** aus, um alle Felder sowie die maximale Anzahl Felddaten zu durchsuchen (mit dem Feld **Alle Felder durchsuchen außer** lässt sich die Anzahl der zu durchsuchenden Felder begrenzen).

Wenn Sie die Anzahl der unnötigerweise durchsuchten Felder minimieren möchten, inaktivieren Sie das Markierungsfeld **Crawler** für alle Felder, mit Ausnahme der Felder, die Suchfeldern zugeordnet sind.

### Zugehörige Konzepte

„Domino Document Manager-Crawler“ auf Seite 66

„Notes-Crawler“ auf Seite 79

„QuickPlace-Crawler“ auf Seite 89

## Konfigurieren des Crawler-Servers unter UNIX für das Durchsuchen von Lotus Domino-Quellen

Wenn Sie OmniFind Enterprise Edition auf einem Computer installieren, auf dem IBM AIX, Linux oder die Solaris-Betriebsumgebung ausgeführt wird, und Server durchsuchen möchten, die das NRPC-Protokoll (Notes Remote Procedure Call) verwenden, müssen Sie ein Script ausführen, um den Crawler-Server zu konfigurieren. Das Script aktiviert die Notes-, QuickPlace- und Domino Document Manager-Crawler für die Kommunikation mit den Datenbankservern.

### Einschränkungen

Ein Domino-Server kann nicht gleichzeitig auf einem Computer mit einem Notes-, QuickPlace- oder Domino Document Manager-Crawler ausgeführt werden, der für die Verwendung von NRPC konfiguriert wurde. Wenn Sie versuchen, einen dieser Crawler zu starten, während der Domino-Server aktiv ist, tritt ein Fehler auf und der Crawler wird gestoppt.

### Informationen zu dieser Task

Crawler, die das NRPC-Protokoll verwenden, nutzen Domino-Bibliotheken als Client. Sie installieren diese Bibliotheken, indem Sie Lotus Domino Server auf dem Crawler-Server installieren. Nach der Installation der Domino-Bibliotheken führen Sie auf dem Crawler-Server ein Setup-Script aus, das von OmniFind Enterprise Edition bereitgestellt wird, um sicherzustellen, dass die Crawler mit Domino-Bibliotheken zusammenarbeiten können.

### Vorgehensweise

Gehen Sie wie folgt vor, um den Crawler-Server zu konfigurieren, damit er Lotus Notes-, Lotus QuickPlace- und Domino Document Manager-Server durchsuchen kann:

1. Erstellen Sie den Benutzer server und die Gruppe notes auf dem Crawler-Server:
  - a. Melden Sie sich als Root an:  
`su - root`
  - b. Fügen Sie einen Benutzer hinzu:  
`useradd server`
  - c. Fügen Sie für diesen Benutzer ein Kennwort hinzu:  
`passwd server`  
Sie werden zur Änderung des Kennworts aufgefordert.
2. Installieren Sie Lotus Domino Server auf dem Crawler-Server:
  - a. Legen Sie die Domino Server-CD ein, und hängen Sie sie an. (Wenn Sie keine CD haben, können Sie das Image herunterladen.)
  - b. Wechseln Sie zum Ordner für Ihr Betriebssystem.

AIX: cd /mnt/cdrom/aix  
Linux: cd /mnt/cdrom/linux  
Solaris: cd /mnt/cdrom/solaris

- c. Starten Sie das Installationsprogramm:  
./install
  - d. Beantworten Sie die Eingabeaufforderungen, und akzeptieren Sie die Standardwerte, oder geben Sie Ihre bevorzugten Installationseinstellungen an (wie z. B. Pfade für das Installationsverzeichnis und das Datenverzeichnis).  
Lesen Sie die Domino-Dokumentation, wenn Sie bei der Installation von Domino Server Unterstützung benötigen.
  - e. Stellen Sie sicher, dass die Administrator-ID für die Unternehmenssuche berechtigt ist, auf das Verzeichnis home/server zuzugreifen. Diese Administrator-ID wurde bei der Installation von OmniFind Enterprise Edition angegeben.
3. Führen Sie das von OmniFind Enterprise Edition bereitgestellte Setup-Script auf dem Crawler-Server aus:
- a. Melden Sie sich als Administrator für die Unternehmenssuche an (diese Benutzer-ID wurde während der Installation von OmniFind Enterprise Edition angegeben).
  - b. Starten Sie das folgende Script, das im Verzeichnis \$ES\_INSTALL\_ROOT/bin installiert wurde:  
escrnote.sh
  - c. Beantworten Sie die Eingabeaufforderungen:
    - Beantworten Sie die folgende Eingabeaufforderung mit Y, falls Domino Server im Standardverzeichnis installiert ist, antworten Sie mit N, falls dies nicht der Fall ist:  
Der Lotus Notes-Verzeichnispfad /opt/lotus/notes/latest/linux wurde gefunden.  
Ist dies der richtige Lotus Notes-Verzeichnispfad?  
  
Der Standardpfad für AIX lautet /opt/lotus/notes/latest/ibmpow.  
Der Standardpfad für Linux lautet /opt/lotus/notes/latest/linux.  
Der Standardpfad für Solaris lautet /opt/lotus/notes/latest/sunspa.
    - Wenn Domino Server auf dem Crawler-Server nicht im Standardverzeichnis installiert ist, geben Sie als Antwort auf die folgende Eingabeaufforderung an, wo Domino installiert ist:  
Geben Sie den Pfad für das Lotus Notes-Verzeichnis ein.  
  
Sie könnten auf einem Linux-Computer z. B. Folgendes angeben: /opt/lotus/notes/latest/linux.
    - Beantworten Sie die folgende Eingabeaufforderung mit Y, falls das Domino Server-Datenverzeichnis im Standardverzeichnis installiert ist, antworten Sie mit N, falls dies nicht der Fall ist:  
Der Lotus Notes-Datenverzeichnispfad /local/notesdata wurde gefunden.  
Ist dies der richtige Lotus Notes-Datenverzeichnispfad?  
  
Der Standardpfad lautet /local/notesdata.
    - Wenn das Domino Server-Datenverzeichnis auf dem Crawler-Server nicht an der Standardposition eingerichtet wurde, geben Sie den Domino-Datenpfad als Antwort auf die folgende Eingabeaufforderung an:  
Geben Sie den Pfad für das Lotus Notes-Datenverzeichnis ein.
4. Stoppen Sie das System für die Unternehmenssuche auf dem Crawler-Server, und starten Sie es erneut:

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

### **Zugehörige Konzepte**

„Domino Document Manager-Crawler“ auf Seite 66

„Notes-Crawler“ auf Seite 79

„QuickPlace-Crawler“ auf Seite 89

## **Konfigurieren des Crawler-Servers unter Windows für das Durchsuchen von Lotus Domino-Quellen**

Wenn Sie OmniFind Enterprise Edition auf einem Microsoft Windows-Computer installieren und Server durchsuchen möchten, die das NRPC-Protokoll (Notes Remote Procedure Call) verwenden, müssen Sie ein Script ausführen, um den Crawler-Server zu konfigurieren. Das Script aktiviert die Notes-, QuickPlace- und Domino Document Manager-Crawler für die Kommunikation mit den Datenbankservern.

### **Einschränkungen**

Lotus Domino Server und der Lotus Notes-Client können nicht gleichzeitig auf einem Computer mit einem Notes-, QuickPlace- oder Domino Document Manager-Crawler ausgeführt werden, der für die Verwendung von NRPC konfiguriert wurde. Wenn Sie versuchen, einen dieser Crawler zu starten, während der Domino-Server aktiv ist, tritt ein Fehler auf und der Crawler wird gestoppt.

### **Informationen zu dieser Task**

Crawler, die das NRPC-Protokoll verwenden, arbeiten mit Lotus Domino-Clientbibliotheken. Sie installieren diese Bibliotheken, indem Sie Lotus Domino Server auf dem Crawler-Server installieren. Nach der Installation der Domino-Bibliotheken führen Sie auf dem Crawler-Server ein Setup-Script aus, das von OmniFind Enterprise Edition bereitgestellt wird, um sicherzustellen, dass die Crawler mit Domino-Bibliotheken zusammenarbeiten können.

### **Vorgehensweise**

Gehen Sie wie folgt vor, um den Crawler-Server zu konfigurieren, damit er Lotus Note-, Lotus QuickPlace- und Domino Document Manager-Server durchsuchen kann:

1. Melden Sie sich am Crawler-Server mit einer Benutzer-ID an, die ein Mitglied der Administratorgruppe ist. Stellen Sie sicher, dass die Benutzer-ID über die Berechtigung zum Installieren von Lotus Notes verfügt.
2. Installieren Sie Lotus Notes:
  - a. Legen Sie die Domino Server-CD ein. (Wenn Sie keine CD haben, können Sie das Image herunterladen.)
  - b. Starten Sie das Installationsprogramm: setup.exe
  - c. Beantworten Sie die Eingabeaufforderungen, und akzeptieren Sie die Standardwerte, oder geben Sie Ihre bevorzugten Installationseinstellungen an (wie z. B. Pfade für das Installationsverzeichnis und das Datenverzeichnis). Lesen Sie die Lotus domino-Dokumentation, wenn Sie Unterstützung benötigen.

- d. Stellen Sie sicher, dass die Administrator-ID für die Unternehmenssuche berechtigt ist, auf das Domino-Datenverzeichnis zuzugreifen. Diese Administrator-ID wurde bei der Installation von OmniFind Enterprise Edition angegeben.
3. Führen Sie das von OmniFind Enterprise Edition bereitgestellte Setup-Script auf dem Crawler-Server aus:
    - a. Melden Sie sich mit der ID des Administrators für die Unternehmenssuche an (diese Benutzer-ID wurde während der Installation von OmniFind Enterprise Edition angegeben).
    - b. Starten Sie das folgende Script, das im Verzeichnis %ES\_INSTALL\_ROOT%\bin installiert wurde:  
escrnote.vbs
    - c. Beantworten Sie die Eingabeaufforderungen:
      - Beantworten Sie die folgende Eingabeaufforderung mit Y, falls Lotus Notes im Standardverzeichnis installiert ist, antworten Sie mit N, falls dies nicht der Fall ist:  
Der Lotus Notes-Verzeichnispfad c:\lotus\notes wurde gefunden.  
Ist dies der richtige Lotus Notes-Verzeichnispfad?  
  
Der Standardinstallationspfad auf einem Windows-Computer lautet c:\lotus\notes bzw. c:\lotus\domino.
      - Wenn Lotus Notes auf dem Crawler-Server nicht im Standardverzeichnis installiert ist, geben Sie als Antwort auf die folgende Eingabeaufforderung an, wo Lotus Notes installiert ist:  
Geben Sie den Pfad für das Lotus Notes-Verzeichnis ein.
      - Beantworten Sie die folgende Eingabeaufforderung mit Y, falls das Lotus Notes-Datenverzeichnis an der Standardposition installiert ist, antworten Sie mit N, falls dies nicht der Fall ist:  
Der Lotus Notes-Datenverzeichnispfad c:\lotus\notes\data wurde gefunden.  
Ist dies der richtige Lotus Notes-Datenverzeichnispfad?  
  
Der Standardpfad auf einem Windows-Computer lautet c:\lotus\notes\data bzw. c:\lotus\domino\data.
      - Wenn das Lotus Notes-Datenverzeichnis auf dem Crawler-Server nicht an der Standardposition eingerichtet wurde, geben Sie den Datenverzeichnispfad als Antwort auf die folgende Eingabeaufforderung an:  
Geben Sie den Pfad für das Lotus Notes-Datenverzeichnis ein.
  4. Stoppen Sie das System für die Unternehmenssuche auf dem Crawler-Server, und starten Sie es erneut:

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

#### Zugehörige Konzepte

„Domino Document Manager-Crawler“ auf Seite 66

„Notes-Crawler“ auf Seite 79

„QuickPlace-Crawler“ auf Seite 89

## Konfigurieren von Servern, die das DIIOP-Protokoll verwenden

Zum Durchsuchen von Servern, die das DIIOP-Protokoll (Domino Internet Inter-ORB Protocol) verwenden, müssen Sie den Server so konfigurieren, dass die Notes-, QuickPlace- und Domino Document Manager-Crawler das Protokoll verwenden können.

## Vorbereitung

Der Server, den Sie durchsuchen wollen, muss die DIIOP- und HTTP-Tasks ausführen.

## Vorgehensweise

Gehen Sie wie folgt vor, um Server zu konfigurieren, die das DIIOP-Protokoll verwenden:

1. Konfigurieren Sie das Serverdokument:
  - a. Öffnen Sie das Dokument `server` auf dem Lotus Notes-, Lotus QuickPlace- oder Domino Document Manager-Server, den Sie durchsuchen wollen. Dieses Dokument ist im Domino-Verzeichnis gespeichert.
  - b. Erweitern Sie auf der Seite **Konfiguration** den Abschnitt **Server**.
  - c. Geben Sie auf der Seite **Sicherheit** im Bereich **Einschränkungen der Programmierbarkeit** die entsprechenden Sicherheitseinschränkungen für Ihre Umgebung in den folgenden Feldern an:
    - **Beschränkte LotusScript/Java-Agenten ausführen**
    - **Beschränktes Java/Javascript/COM ausführen**
    - **Unbeschränktes Java/Javascript/COM ausführen**

Sie könnten z. B. einen Stern (\*) angeben, um unbeschränkten Zugriff durch Lotus Script/Java-Agenten zuzulassen, und Benutzernamen angeben, die im Domino-Verzeichnis für die Java/Javascript/COM-Einschränkungen registriert sind.

**Wichtig:** Der Crawler, den Sie für das Durchsuchen dieses Servers mit dem DIIOP-Protokoll konfigurieren, muss die Benutzernamen, die Sie in diesen Feldern angeben, verwenden können.

- d. Öffnen Sie die Seite **Internet-Protokoll**, dann die Seite **HTTP**, und setzen Sie die Option **HTTP-Clients zum Suchen von Datenbanken zulassen** auf **Ja**.
2. Konfigurieren Sie das Benutzerdokument:
  - a. Öffnen Sie das Dokument `user` auf dem Lotus Notes-, Lotus QuickPlace- oder Domino Document Manager-Server, den Sie durchsuchen wollen. Dieses Dokument ist im Domino-Verzeichnis gespeichert.
  - b. Geben Sie auf der Seite **Allgemein** im Feld **Internet-Kennwort** ein Kennwort an.

Wenn Sie mit der Administrationskonsole für die Unternehmenssuche Optionen für das Durchsuchen dieses Servers konfigurieren, geben Sie diese Benutzer-ID und dieses Kennwort auf der Seite an, auf der der zu durchsuchende Server angegeben wird. Der Crawler greift dann mithilfe dieser Berechtigungsnachweise auf den Server zu.
3. Starten Sie die DIIOP-Task auf dem Server erneut.

### Zugehörige Konzepte

„Domino Document Manager-Crawler“ auf Seite 66

„Notes-Crawler“ auf Seite 79

„QuickPlace-Crawler“ auf Seite 89

## Konfigurieren des I/O Completion Ports unter AIX für das Durchsuchen von Lotus Domino-Quellen

Bevor Sie die Notes-, QuickPlace- oder Domino Document Manager-Crawler auf einem IBM AIX-System verwenden können, müssen Sie das IOCP-Modul (I/O Completion Port) installieren und für die Verwendung durch den Crawler konfigurieren.

### Informationen zu dieser Task

Ohne das IOCP-Modul schlagen die Discoveryprozesse fehl, wenn Sie versuchen, einen Crawler zu erstellen. Die folgende Fehlermeldung wird angezeigt:

```
FFQM0105E Fehler vom Server empfangen -
Nachricht: FFQG0024E Unerwartete Ausnahmebedingung: discover
```

Die folgende Nachricht, die den Fehler ENOEXEC enthält, wird in die Datei `$ES_NODE_ROOT/logs/system_jjjjmmtt.log` geschrieben. (Zur besseren Lesbarkeit wurde der Nachrichtentext teilweise auf mehrere Zeilen verteilt.)

```
5/20/05 18:08:52.423 JST [Error] [ES_ERR_EXCEPTION_DEFAULT_MESSAGE] [] [discovery]
ies10.yamato.ibm.com:0:2108088751:control:ComponentDiscoveryW.java:
com.ibm.es.control.discovery.server.ComponentDiscoveryW.discover:86
FFQ00277E Eine Ausnahmebedingung mit der Zusatzinformation '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'
und dem Stack-Trace '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) wurde abgefangen
```

### Vorgehensweise

Gehen Sie wie folgt vor, um das IOCP-Modul zu installieren und sicherzustellen, dass das Modul ordnungsgemäß auf dem Crawler-Server installiert ist:

Sie müssen folgende Schritte ausführen:

1. Installieren Sie das IOCP-Modul (`bos.iocp.rte`) von der AIX-Produkt-CD auf dem Crawler-Server.

Wenden Sie nach der Installation des IOCP-Moduls, aber noch vor der Erstellung eines Notes-, QuickPlace- oder Domino Document Manager-Crawlers, einen Software-Fix für das Modul an. Anweisungen hierzu finden Sie unter dem folgenden Link:

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

2. Geben Sie den folgenden Befehl ein, um sicherzustellen, dass das IOCP-Modul auf dem Crawler-Server installiert ist:

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

Die Ausgabe des Befehls `ls1pp` sollte dem folgenden Beispiel ähneln:

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. Geben Sie den folgenden Befehl ein, um sicherzustellen, dass der IOCP-Port den Status **Available** hat:

```
$ lsdev -Cc iocp
```

Die Ausgabe des Befehls `lsdev` sollte mit dem folgenden Beispiel übereinstimmen:

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

4. Wenn der IOCP-Port den Status **Defined** hat, ändern Sie diesen zu **Available**:
  - a. Melden Sie sich auf dem Crawler-Server als Root an, und setzen Sie den folgenden Befehl ab:

```
smit iocp
```
  - b. Wählen Sie **Change / Show Characteristics of I/O Completion Ports** aus, und ändern Sie **STATE to be configured at system restart** von **Defined** zu **Available**.
  - c. Führen Sie einen Warmstart des Crawler-Servers durch.
  - d. Geben Sie den Befehl `lsdev` erneut ein, und vergewissern Sie sich, dass der IOCP-Port nun den Status **Available** hat.

#### Zugehörige Konzepte

„Domino Document Manager-Crawler“ auf Seite 66

„Notes-Crawler“ auf Seite 79

„QuickPlace-Crawler“

---

## QuickPlace-Crawler

Wenn Sie Lotus QuickPlace-Bereiche und -Räume in eine Objektgruppe für die Unternehmenssuche einschließen möchten, müssen Sie einen QuickPlace-Crawler konfigurieren.

Sie können den QuickPlace-Crawler auch verwenden, um Bereiche zu durchsuchen, die Sie mit Lotus Quickr-Services für Lotus Domino verwalten. Wenn Sie Lotus Quickr-Services für WebSphere Portal verwenden, verwenden Sie den Crawler für die Einstiegspunktliste, um einer Objektgruppe für die Unternehmenssuche Dokumente hinzuzufügen, die in Lotus Quickr-Bibliotheken gespeichert sind.

#### Tipp:

Detaillierte Beispiele für die Konfiguration eines sicheren QuickPlace-Crawlers finden Sie im Szenario für ein kleines Unternehmen im IBM Redbook *IBM OmniFind Enterprise Edition Configuration and Implementation Scenarios*.

## Crawler-Server-Konfiguration

Wenn ein Server, den Sie durchsuchen wollen, das NRPC-Protokoll (Notes Remote Procedure Call) verwendet, müssen Sie ein Setup-Script auf dem Crawler-Server ausführen. Dieses Script, das im Lieferumfang von OmniFind Enterprise Edition enthalten ist, ermöglicht es dem QuickPlace-Crawler, mit den Servern zu kommunizieren, die NRPC verwenden.

Wenn ein Server, den Sie durchsuchen wollen, das DIIOP-Protokoll (Domino Internet Inter-ORB Protocol) verwendet, müssen Sie kein Setup-Script auf dem Crawler-Server ausführen. Sie müssen allerdings den Server so konfigurieren, dass der QuickPlace-Crawler auf den Server zugreifen kann.

Verwendet ein Server, den Sie durchsuchen wollen, einen LDAP-Server (Lightweight Directory Access Protocol), dann muss für den Zielservers die Verwendung des Protokolls DIIOP konfiguriert werden (der QuickPlace-Crawler kann LDAP-Daten nicht mit dem NRPC-Protokoll durchsuchen). Sie müssen außerdem eine Directory Assistance-Datenbank konfigurieren und die Konfiguration des Zielservers so vornehmen, dass dieser den LDAP-Server als sekundären Domino-Server nutzt.

**Wichtig:** Wenn der Zielservers das DIIOP-Protokoll verwendet und Sie den Crawler für die Verwendung von HTTPS oder DIIOP über SSL (Secure Sockets Layer) so konfigurieren, dass die Übertragungen zwischen Crawler und Server verschlüsselt werden, müssen Sie die Datei `TrustedCerts.class` (z. B. `c:\certs` oder `/data/certs`) vom Zielservers auf den Crawler-Server kopieren. In einer Konfiguration mit zwei oder vier Servern müssen Sie die Datei `TrustedCerts.class` auch auf die Server kopieren, auf denen die Suchkomponente installiert ist. Stellen Sie sicher, dass sich die Datei auf dem Crawler-Server und auf den Suchservern an derselben Speicherposition befindet. Den Verzeichnispfad für die Datei `TrustedCerts.class` geben Sie bei der Konfiguration des Crawlers an.

Wenn Sie den Crawler konfigurieren und eine Benutzer-ID angeben, die der Crawler verwenden soll, stellen Sie sicher, dass die entsprechende ID über die Berechtigung für den Zugriff auf alle QuickPlace-Bereiche in einer Domino-Domäne verfügt. Wenn Sie dies sicherstellen wollen, weisen Sie die Benutzer-ID der reservierten Gruppe mit dem Namen `QuickPlaceAdministratorsSUGroup` zu.

Wenn OmniFind Enterprise Edition auf einem IBM AIX-System installiert wurde, muss sichergestellt werden, dass das I/O Completion Port-Modul installiert und auf dem Crawler-Server verfügbar ist.

Führen Sie die Ihrer Umgebung entsprechenden Tasks aus, bevor Sie mit der Administrationskonsole für die Unternehmenssuche einen QuickPlace-Crawler konfigurieren:

- „Konfigurieren des Crawler-Servers unter UNIX für das Durchsuchen von Lotus Domino-Quellen“ auf Seite 83.
- „Konfigurieren des Crawler-Servers unter Windows für das Durchsuchen von Lotus Domino-Quellen“ auf Seite 85.
- „Konfigurieren von Servern, die das DIIOP-Protokoll verwenden“ auf Seite 86.
- „Konfigurieren des QuickPlace-Servers für die Verwendung der Sicherheit "Local User" (Lokaler Benutzer)" auf Seite 309.
- „Konfigurieren von Directory Assistance auf einem QuickPlace-Server“ auf Seite 310.

- „Konfigurieren des I/O Completion Ports unter AIX für das Durchsuchen von Lotus Domino-Quellen“ auf Seite 88.

## Sicherheit auf Dokumentenebene

Wenn die Objektgruppensicherheit aktiviert ist und ein Server, den Sie durchsuchen wollen, das NRPC-Protokoll verwendet, müssen Sie auf dem Crawler-Server einen Lotus Domino-Trusted Server konfigurieren. Mit dem Trusted Server wird eine Zugriffssteuerung auf Dokumentenebene umgesetzt. Führen Sie die folgenden Tasks aus, bevor Sie den Benutzern die Objektgruppe für die Suche zur Verfügung stellen:

- Konfigurieren der Lotus Domino-Trusted Server, um Benutzerberechtigungen zu prüfen.
- Aktivieren der globalen Sicherheit in WebSphere Application Server und Konfigurieren der Suchanwendung für die Verwendung der Sicherheit. Dieser Schritt stellt sicher, dass Anmeldeberechtigungen geprüft werden, wenn Benutzer versuchen, die Suchanwendung zu verwenden. Die Suchserver verwenden die Berechtigungsnachweise, um die Berechtigung jedes Benutzers zu prüfen, auf Lotus Domino-Dokumente zuzugreifen.

## Durchsuchen von Anhängen

In Lotus QuickPlace können Sie Microsoft Office-Dokumente importieren und bereitstellen (zu den Optionen, die Sie beim Importieren auswählen, gehören **Imported Page**, **Microsoft Word Page**, **Microsoft Excel Page**, **Microsoft PowerPoint Page** und **Multiple Imported Pages**). Der QuickPlace-Crawler kann importierte Dokumente dieser Typen als Anhänge nur unter den folgenden Bedingungen durchsuchen:

- Der Server verwendet das DIIOP-Protokoll.
- Sie aktivieren das Durchsuchen von Anhängen, wenn Sie für den Crawler die Optionen für die Crawler suche konfigurieren.
- Sie konfigurieren den Crawler so, dass er das Feld "\$FILE" oder alle Felder durchsucht.

## Konfigurationsüberblick

Mit dem QuickPlace-Crawler können Sie eine beliebige Anzahl QuickPlace-Bereiche durchsuchen. Bei der Erstellung des Crawlers wählen Sie die zu durchsuchenden Bereiche auf einem einzigen QuickPlace-Server aus. Wenn Sie später den Crawlerbereich bearbeiten, können Sie Dokumente von einem anderen QuickPlace-Server hinzufügen, den Sie in denselben Crawlerbereich einschließen möchten. Bei der Erstellung bzw. Bearbeitung des Crawlers können Sie angeben, ob alle Bereiche des Servers oder nur die von Ihnen angegebenen Bereiche durchsucht werden sollen und ob alle Räume in den für die Crawler suche ausgewählten Bereichen oder nur die von Ihnen angegebenen Räume durchsucht werden sollen.

**Einschränkung:** Wenn Sie eine Benutzer-ID angeben, die der Crawler verwenden soll, stellen Sie sicher, dass die entsprechende ID über die Berechtigung für den Zugriff auf alle QuickPlace-Bereiche in der Domino-Domäne verfügt. Dazu können Sie den QuickPlace-Server konfigurieren und der reservierten Gruppe QuickPlace-AdministratorsSUGroup eine Benutzer-ID hinzufügen. Beachten Sie, dass der Gruppenname keine eingebetteten Leerzeichen enthält und die Groß- und Kleinschreibung eingehalten werden muss.

Wenn Sie einen QuickPlace-Crawler erstellen oder ändern möchten, müssen Sie sich an der Administrationskonsole für die Unternehmenssuche anmelden. Sie müssen die Rolle eines Administrators für die Unternehmenssuche haben oder ein Objektgruppenadministrator für die Objektgruppe sein, zu der der Crawler gehört.

Wenn Sie den Crawler erstellen, unterstützt Sie ein Assistent bei den folgenden Tasks:

- Angeben von Merkmalen, die steuern, wie der Crawler ausgeführt wird und Systemressourcen verwendet. Die Crawlermerkmale steuern, wie der Crawler alle Dokumente im Crawlerbereich durchsucht.
- Angeben des QuickPlace-Servers und des Kommunikationsprotokolls.
- Wenn der Server für die Verwendung des DIIOP-Protokolls konfiguriert ist, können Sie angeben, wie der Crawler eine Verbindung zu Domino-Objekten herstellen soll. Sie können beispielsweise Optionen angeben, um HTTPS oder Secure Sockets Layer (SSL) zu verwenden, um die Kommunikation zu verschlüsseln.
- Angeben von Informationen zu dem Benutzerverzeichnis, das dem Server zugeordnet ist (der Crawler benötigt diese Informationen, um die Zugriffssteuerungen umzusetzen, wenn Benutzer die Objektgruppe durchsuchen).
- Auswählen der zu durchsuchenden Bereiche.
- Konfigurieren eines Zeitplans für das Durchsuchen der Bereiche.
- Auswählen der zu durchsuchenden Dokumente. Der Crawler kann alle Räume in einem Bereich oder nur die Dokumente in den von Ihnen ausgewählten Räumen durchsuchen.
- Angeben von Optionen für das Aktivieren der Felder in verschiedenen Bereichen und Räumen für Suchvorgänge. Sie können z. B. bestimmte Felder aus dem Crawlerbereich ausschließen und Optionen für das Durchsuchen von Anlagen angeben.
- Konfigurieren von Optionen für die Sicherheit auf Dokumentenebene. Wenn die Sicherheit bei der Erstellung der Objektgruppe aktiviert war, kann der Crawler Sicherheitsdaten Dokumenten im Index zuordnen. Mithilfe dieser Daten können Suchanwendungen die Zugriffssteuerung auf Basis der gespeicherten Zugriffssteuerungslisten oder Sicherheitstoken erzwingen.  
Sie können außerdem eine Option auswählen, um Berechtigungsnachweise eines Benutzers bei der Übergabe einer Abfrage zu prüfen. In diesem Fall vergleicht das System die Berechtigungsnachweise des Benutzers nicht mit indexierten Sicherheitsdaten, sondern mit den aktuellen Zugriffssteuerungslisten, die von den ursprünglichen Datenquellen verwaltet werden.

Klicken Sie **Hilfe** an, während Sie den Crawler erstellen, um Informationen zu den Feldern im Assistenten und zur Vorgehensweise abzurufen, wie Sie dem Crawler die für das Durchsuchen der Daten erforderlichen Informationen zur Verfügung stellen.

#### **Zugehörige Konzepte**

„Tipps für das Durchsuchen von Lotus Domino-Datenbanken“ auf Seite 82

„Umsetzung der Sicherheit auf Dokumentenebene für Lotus Domino-Dokumente“ auf Seite 307

#### **Zugehörige Tasks**

„Konfigurieren des Crawler-Servers unter UNIX für das Durchsuchen von Lotus Domino-Quellen“ auf Seite 83

„Konfigurieren des Crawler-Servers unter Windows für das Durchsuchen von Lotus Domino-Quellen“ auf Seite 85

„Konfigurieren von Servern, die das DIIOP-Protokoll verwenden“ auf Seite 86

„Konfigurieren des I/O Completion Ports unter AIX für das Durchsuchen von Lotus Domino-Quellen“ auf Seite 88

„Konfigurieren der Lotus Domino-Trusted Server, um Benutzerberechtigungen zu prüfen“ auf Seite 308

„Konfigurieren des QuickPlace-Servers für die Verwendung der Sicherheit "Local User" (Lokaler Benutzer)" auf Seite 309

„Konfigurieren von Directory Assistance auf einem QuickPlace-Server“ auf Seite 310

---

## Crawler für Einstiegspunktlisten

Wenn Sie IBM Lotus Quickr-Services für WebSphere Portal verwenden, können Sie den Crawler für die Einstiegspunktliste verwenden, um einer Objektgruppe für die Unternehmenssuche Dokumente in Lotus Quickr-Bibliotheken hinzuzufügen.

Eine Lotus Quickr-Bibliothek ist ein Container für Dokumentdateien. Der Crawler für die Einstiegspunktliste unterstützt nicht die Crawlersuche von webbasierten Inhalten wie z. B. Wikis und Blogs.

Wenn Sie Lotus Quickr-Services für Lotus Domino verwenden, verwenden Sie den QuickPlace-Crawler, um einer Objektgruppe Lotus Quickr-Dokumente hinzuzufügen.

### WebSphere Portal-Serverkonfiguration

Wenn Sie Lotus Quickr auf einem WebSphere Portal-Server der Version 6 installieren, können Sie den Crawler für die Einstiegspunktliste verwenden, um Lotus Quickr-Bibliotheksdokumente zu durchsuchen. Optionen für die Crawlersuche in diesen Dokumenten können Sie getrennt von den Optionen konfigurieren, die Sie für Portalsites angeben, die von einem WebSphere Portal-Crawler durchsucht werden.

Bevor Sie einen Crawler für die Einstiegspunktliste erstellen, müssen Sie die Prozeduren für die Einrichtung der Unternehmenssuche in WebSphere Portal befolgen. Sie führen ein Script (`wp6_install.sh` unter AIX, Linux oder Solaris bzw. `wp6_install.bat` unter Windows), das mit OmniFind Enterprise Edition bereitgestellt wird, auf den Suchservern aus, um die Umgebung für die Unternehmenssuche zu konfigurieren.

Eine Benutzeragentenzeichenfolge gibt an, welcher Browser oder Roboter auf einen Server zugreift. Bei der Crawlersuche auf einem Lotus Quickr-Server verwendet der Crawler für die Einstiegspunktliste die Benutzeragentenzeichenfolge `OmniFind SeedlistCrawler/1.0`.

### Konfigurationsüberblick

Mit dem Crawler für die Einstiegspunktliste können Sie eine beliebige Anzahl Lotus Quickr-Dokumente durchsuchen. Wenn Sie den Crawler konfigurieren, geben Sie den zu durchsuchenden Server an. Danach durchsucht der Crawler alle Dokumente in den Lotus Quickr-Bibliotheken des betreffenden Servers.

Die zu durchsuchenden Dokumente müssen über dieselbe Lotus Quickr-Administrator-ID und das zugehörige Kennwort zugänglich sein. Wenn Sie Sites durchsuchen wollen, die unterschiedliche Berechtigungsnachweise verwenden, müssen Sie einen separaten Crawler für die Einstiegspunktliste konfigurieren.

Wenn Sie einen Crawler für die Einstiegspunktliste erstellen oder ändern möchten, müssen Sie sich an der Administrationskonsole für die Unternehmenssuche anmelden. Sie müssen die Rolle eines Administrators für die Unternehmenssuche haben oder ein Objektgruppenadministrator für die Objektgruppe sein, zu der der Crawler gehört.

Wenn Sie den Crawler erstellen, unterstützt Sie ein Assistent bei den folgenden Tasks:

- Angeben von Merkmalen, die steuern, wie der Crawler ausgeführt wird und Systemressourcen verwendet. Die Crawlermerkmale steuern, wie der Crawler die Lotus Quickr-Dokumente durchsucht.
- Angeben des zu durchsuchenden Servers und der Informationen, die der Crawler benötigt, um eine Verbindung zum Server herzustellen.

Wenn Sie den Crawler erstellen oder bearbeiten, können Sie testen, ob er eine Verbindung zu den zu durchsuchenden Dokumenten herstellen kann. Nachrichten zeigen Ihnen an, ob der Crawler auf die zu durchsuchenden Dokumente zugreifen kann, bevor Sie den Crawler starten.

- Angeben von Sicherheitsoptionen auf Dokumentenebene. Wenn die Sicherheit bei der Erstellung der Objektgruppe aktiviert war, kann der Crawler Sicherheitsdaten Dokumenten im Index zuordnen. Mithilfe dieser Daten können Zugriffssteuerungen auf Basis der gespeicherten Zugriffssteuerungslisten oder Sicherheitstoken erzwungen werden.

Sie können außerdem eine Option auswählen, um Berechtigungsnachweise eines Benutzers bei der Übergabe einer Abfrage zu prüfen. In diesem Fall vergleicht das System die Berechtigungsnachweise des Benutzers nicht mit indexierten Sicherheitsdaten, sondern mit den aktuellen Zugriffssteuerungslisten, die von den ursprünglichen Datenquellen verwaltet werden.

**Wichtig:** Wenn Sie sichere Lotus Quickr-Dokumente durchsuchen wollen, müssen Sie Suchen unter Verwendung des Suchportlets für die Unternehmenssuche über WebSphere Portal übergeben. Suchen, die über die Mustersuchanwendung (ESSearchApplication) übergeben werden, verfügen nicht über die geeigneten Berechtigungsnachweise und können die Berechtigung des Benutzers für den Dokumentenzugriff nicht überprüfen.

- Angeben von Informationen, die dem Crawler die Kommunikation mit einem Proxy-Server ermöglichen, wenn ein Proxy-Server für das Bereitstellen von Seiten verwendet wird.
- Angeben von SSO-Berechtigungsnachweisen (Single Sign-on), mit denen der Crawler auf Dokumente auf dem Server zugreifen kann, wenn Sie ein anderes Produkt (wie z. B. IBM Tivoli Access Manager WebSEAL oder CA SiteMinder SSO Agent for PeopleSoft) verwenden, um Ihren WebSphere Portal-Server und Ihre Lotus Quickr-Dokumente zu schützen.
- Angeben von Informationen zu einer Schlüsselspeicherdatei, damit der Crawler das Protokoll SSL (Secure Sockets Layer) verwenden kann, um eine Verbindung zum Server herzustellen.
- Angeben der Sprache und Codepage der zu durchsuchenden Dokumente.
- Angeben von Optionen für die Crawlersuche und die Suche in Metadaten in Lotus Quickr-Dokumenten.
- Angeben von Zeitplänen für die Crawlersuche in Lotus Quickr-Dokumenten.

Klicken Sie **Hilfe** an, während Sie den Crawler erstellen, um Informationen zu den Feldern im Assistenten und zur Vorgehensweise abzurufen, wie Sie dem Crawler die für das Durchsuchen der Daten erforderlichen Informationen zur Verfügung stellen.

### Zugehörige Konzepte

„Integration in WebSphere Portal“ auf Seite 373

### Zugehörige Tasks

„Konfigurieren der Unternehmenssuche in WebSphere Portal Version 6“ auf Seite 381

„Konfigurieren des Portlets für die Unternehmenssuche für Lotus Quickr“ auf Seite 387

---

## Crawler für UNIX-Dateisysteme

Wenn Sie Dokumente, die in AIX-, Linux- oder Solaris-Dateisystemen gespeichert sind, in eine Objektgruppe für die Unternehmenssuche einschließen möchten, müssen Sie einen Crawler für UNIX-Dateisysteme konfigurieren.

Sie können mit dem Crawler für UNIX-Dateisysteme eine beliebige Anzahl Dateisysteme durchsuchen. Wenn Sie den Crawler konfigurieren, wählen Sie die lokalen und fernen Verzeichnisse und Unterverzeichnisse aus, die Sie durchsuchen wollen.

Wenn Sie den Crawler-Server auf einem Windows-Computer installieren, können Sie diesen Server nicht zum Durchsuchen von Quellen in AIX-, Linux- oder Solaris-Dateisystemen verwenden (der Crawler für UNIX-Dateisysteme wird in der Liste der verfügbaren Crawlertypen nicht angezeigt).

Der Crawler für UNIX-Dateisysteme durchsucht Dokumente entsprechend den Leseberechtigungen, die für den Administrator für die Unternehmenssuche angegeben sind.

Wenn Sie einen Crawler für UNIX-Dateisysteme erstellen oder ändern möchten, müssen Sie sich an der Administrationskonsole für die Unternehmenssuche anmelden. Sie müssen die Rolle eines Administrators für die Unternehmenssuche haben oder ein Objektgruppenadministrator für die Objektgruppe sein, zu der der Crawler gehört.

Wenn Sie den Crawler erstellen, unterstützt Sie ein Assistent bei den folgenden Tasks:

- Angeben von Merkmalen, die steuern, wie der Crawler ausgeführt wird und Systemressourcen verwendet. Die Crawlermerkmale steuern, wie der Crawler alle Unterverzeichnisse im Crawlerbereich durchsucht.
- Definieren eines Zeitplans für das Durchsuchen der Dateisysteme.
- Auswählen der Unterverzeichnisse und der Unterverzeichnisebenen, die der Crawler durchsuchen soll.
- Angeben von Optionen für das Aktivieren der Dokumente für Suchvorgänge in Unterverzeichnissen. Sie können z. B. bestimmte Dokumenttypen aus dem Crawlerbereich ausschließen.
- Konfigurieren von Optionen für die Sicherheit auf Dokumentenebene. Wenn die Sicherheit bei der Erstellung der Objektgruppe aktiviert war, kann der Crawler Sicherheitsdaten Dokumenten im Index zuordnen. Mithilfe dieser Daten können Suchanwendungen die Zugriffssteuerung auf Basis der gespeicherten Zugriffssteuerungslisten oder Sicherheitstoken erzwingen.

Klicken Sie **Hilfe** an, während Sie den Crawler erstellen, um Informationen zu den Feldern im Assistenten und zur Vorgehensweise abzurufen, wie Sie dem Crawler die für das Durchsuchen der Daten erforderlichen Informationen zur Verfügung stellen.

---

## Web-Crawler

Sie müssen einen Web-Crawler konfigurieren, um Seiten von Websites in eine Objektgruppe für die Unternehmenssuche einzuschließen.

Mit dem Web-Crawler können Sie eine beliebige Anzahl HTTP-Server (HTTP - Hypertext Transfer Protocol) und HTTPS-Server durchsuchen. Der Crawler ruft eine Website auf und liest die Daten auf der Site. Er durchsucht dann über Links in Dokumenten zusätzliche Dokumente. Der Web-Crawler kann Links auf einzelnen Seiten oder *Framesets* (Seiten, die mit HTML-Frames erstellt wurden) durchsuchen und extrahieren.

Die durchsuchten Daten können eines von vielen Formaten haben und aus verschiedenen Quellen innerhalb Ihres Intranets oder aus dem Internet stammen. Zu den allgemeinen Formaten gehören HTML, PDF, Microsoft Word, Lotus WordPro, XML (Extensible Markup Language) usw.

### Tipp:

Detaillierte Beispiele für die Konfiguration eines Web-Crawlers finden Sie im Szenario für ein mittleres Unternehmen im IBM Redbook *IBM OmniFind Enterprise Edition Configuration and Implementation Scenarios*.

Wenn Sie einen Web-Crawler erstellen oder ändern möchten, melden Sie sich an der Administrationskonsole für die Unternehmenssuche an. Sie müssen zudem die Rolle eines Administrators für die Unternehmenssuche haben oder ein Objektgruppenadministrator für die Objektgruppe sein, zu der der Crawler gehört.

Wenn Sie den Crawler erstellen, unterstützt Sie ein Assistent bei den folgenden Tasks:

- Angeben von Merkmalen, die steuern, wie der Crawler ausgeführt wird und Systemressourcen verwendet. Die Crawlermerkmale steuern, wie der Crawler alle Webseiten im Crawlerbereich durchsucht.
- Angeben von Regeln, um das Aufrufen von Websites zu ermöglichen bzw. zu verhindern. Wenn Sie Regeln für die Crawlersuche angeben, können Sie die Regeln testen und prüfen, ob der Crawler auf die Sites zugreifen kann, die Sie in den Crawlerbereich einschließen wollen.
- Angeben von Optionen, um bestimmte Dateitypen einzuschließen und Dateien mit bestimmten Dateierweiterungen auszuschließen.
- Angeben von Regeln, wie der Web-Crawler Seiten mit detaillierten Fehlerhinweisen handhabt.
- Konfigurieren von Optionen für die Sicherheit auf Dokumentenebene. Wenn die Sicherheit bei der Erstellung der Objektgruppe aktiviert war, kann der Crawler Sicherheitsdaten Dokumenten im Index zuordnen. Mithilfe dieser Daten können Suchanwendungen die Zugriffssteuerung auf Basis der gespeicherten Zugriffssteuerungslisten oder Sicherheitstoken erzwingen.
- Angeben von Optionen für das Durchsuchen von kennwortgeschützten Websites (die zu durchsuchenden Web-Server müssen eine HTTP-Basisauthentifizierung oder HTML-Formulare für die Aufforderung zur Eingabe eines Kennworts verwenden).
- Angeben von Optionen für das Durchsuchen von Websites, die von einem Proxy-Server bereitgestellt werden.

Klicken Sie **Hilfe** an, während Sie den Crawler erstellen, um Informationen zu den Feldern im Assistenten und zur Vorgehensweise abzurufen, wie Sie dem Crawler die für das Durchsuchen der Daten erforderlichen Informationen zur Verfügung stellen.

## Benutzeragentenkonfiguration

Wenn Sie eine Website durchsuchen möchten, die das Protokoll zum Sperren von Websitebereichen für Robots verwendet, sollten Sie sicherstellen, dass die Datei `robots.txt` auf der Website dem Benutzeragentennamen, den Sie für den Web-Crawler konfiguriert haben, den Zugriff auf die Website erlaubt.

Wenn das System für die Unternehmenssuche gestartet wird, lädt der Web-Crawler den Benutzeragentennamen, den Sie für den Web-Crawler konfiguriert haben. Bevor der Crawler eine Seite von einer Website herunterlädt, die noch nicht aufgerufen wurde (oder über einen längeren Zeitraum nicht aufgerufen wurde), versucht der Crawler zuerst, die Datei `robots.txt` herunterzuladen. Diese Datei befindet sich im Stammverzeichnis der Website.

Wenn die Datei `robots.txt` nicht vorhanden ist, kann die Website uneingeschränkt durchsucht werden. Wenn die Datei vorhanden ist, wird darin angegeben, welche Bereiche der Site (Verzeichnisse) von Crawlern nicht durchsucht werden dürfen. Die Datei `robots.txt` gibt Berechtigungen für Crawler an, indem ihre Benutzeragentennamen angegeben werden.

Die Einhaltung des Protokolls zum Sperren von Websitebereichen für Robots ist freiwillig. Der Web-Crawler für die Unternehmenssuche versucht jedoch, es in den folgenden Fällen einzuhalten:

- Wenn die Datei `robots.txt` einen Eintrag für den Benutzeragentennamen enthält, der für den Web-Crawler konfiguriert wurde, hält der Web-Crawler die Einschränkungen für diesen Benutzeragenten ein.
- Wenn der Benutzeragentenname in der Datei `robots.txt` nicht enthalten ist, im letzten Eintrag jedoch `User-agent: *` (also jeder beliebige Benutzeragent) angegeben wird und die Einschränkung `Disallow: /` lautet (was bedeutet, dass jegliches Durchsuchen ab dem Stammverzeichnis der Website nicht zugelassen ist), darf der Web-Crawler diese Site nicht durchsuchen.
- Wenn der Benutzeragentenname in der Datei `robots.txt` nicht enthalten ist, im letzten Eintrag jedoch `User-agent: *` angegeben wird und die Einschränkung `Allow: /` lautet, darf der Web-Crawler diese Site durchsuchen.

Administratoren von Websites geben oft einen letzten Eintrag an, mit dem allen Web-Crawlern der Zugriff verwehrt wird, denen der Zugriff nicht explizit gewährt wurde. Wenn Sie einen neuen Web-Crawler konfigurieren und Sie wissen, dass einige Websites, die Sie durchsuchen möchten, das Protokoll zum Sperren von Websitebereichen für Robots verwenden, bitten Sie die Administratoren der Websites, einen Eintrag für Ihren Crawler deren `robots.txt`-Dateien hinzuzufügen.

Stellen Sie sicher, dass Sie in den Merkmalen des Web-Crawlers und in allen `robots.txt`-Dateien, die zu den entsprechenden Websites gehören, denselben Benutzeragentennamen angeben.

Wenn keine der zu durchsuchenden Websites das Protokoll zum Sperren von Websitebereichen für Robots verwendet, spielt der Wert, den Sie als Benutzeragent angeben, in der Regel keine Rolle. Einige Anwendungsserver, JSPs und Servlets passen ihre Antworten jedoch an den Benutzeragentennamen an. Es gibt z. B. verschiedene Antworten für die Handhabung von Browserinkompatibilitäten. Der von

Ihnen für den Web-Crawler angegebene Benutzeragentenname kann in diesen Fällen eine Rolle spielen, ungeachtet des Protokolls zum Sperren von Websitebereichen für Robots. Wenn Sie solche Sites durchsuchen möchten, wenden Sie sich an die Administratoren der Websites, um sicherzustellen, dass dem Web-Crawler der Zugriff gewährt wird.

## Vorgehensweise des Web-Crawlers bei der Verwendung des Protokolls zum Sperren von Websitebereichen für Robots

Der Web-Crawler versucht, das Protokoll zum Sperren von Websitebereichen für Robots einzuhalten und Websites nicht zu durchsuchen, wenn Regeln in der Datei `robots.txt` des Servers das Durchsuchen nicht zulassen.

Ein erfolgreicher Download findet statt, wenn der Crawler die Datei `robots.txt` von einem Web-Server abrufen oder bestätigen kann, dass keine Datei `robots.txt` vorhanden ist. Der Download gilt als fehlgeschlagen, wenn der Crawler die Regeln nicht abrufen oder nicht bestätigen kann, dass eine Datei `robots.txt` vorhanden ist.

Ein erfolgreicher Download bedeutet nicht, dass der Crawler zum Durchsuchen berechtigt ist, da es möglich ist, dass Regeln in der Datei `robots.txt` das Durchsuchen nicht zulassen. Ein Downloadfehler verhindert vorübergehend das Durchsuchen, weil der Crawler die Regeln nicht ermitteln kann.

Dies sind die Schritte, die der Crawler ausführt, wenn er versucht, die Datei `robots.txt` herunterzuladen:

1. Wenn der Crawler eine neue Site erkennt, versucht er, die IP-Adresse des Servers abzurufen. Schlägt dieser Versuch fehl, ist kein Durchsuchen möglich.
2. Wenn zumindest eine IP-Adresse verfügbar ist, versucht der Crawler, die Datei `robots.txt` mithilfe von **HTTP (oder HTTPS) GET** zu laden.
3. Wenn für die Socketverbindung eine Zeitlimitüberschreitung auftritt, die Verbindung unterbrochen wird oder ein anderer Fehler der unteren Ebene auftritt (wie z. B. ein SSL-Zertifikatsproblem), protokolliert der Crawler das Problem und wiederholt den Versuch für jede bekannte IP-Adresse des Zielservers.
4. Wurde keine Verbindung hergestellt, nachdem der Crawler alle Adressen getestet hat, wartet der Crawler zwei Sekunden, bevor er erneut versucht, alle Adressen zu erreichen.
5. Wurde eine Verbindung hergestellt und wurden HTTP-Header ausgetauscht, wird der Rückgabestatus untersucht. Ist der Statuscode 500 oder höher, interpretiert der Crawler dies als eine schlechte Verbindung und testet weiterhin andere IP-Adressen. Bei jedem anderen Status beendet der Crawler das Testen alternativer IP-Adressen und fährt gemäß dem Statuscode fort.

Nachdem der Crawler einen HTTP-Statuscode unter 500 empfangen oder alle IP-Adressen zweimal getestet hat, fährt er wie folgt fort:

1. Wurde kein HTTP-Status unter 500 empfangen, ist die Site vorerst disqualifiziert.
2. Wurde ein HTTP-Status von 400, 404 oder 410 empfangen, ist die Site für das Durchsuchen ohne Regeln qualifiziert.
3. Wurde ein HTTP-Status im Bereich von 200 bis 299 empfangen, wird die nächste Aktion von den folgenden Bedingungen gesteuert:
  - Wenn der Inhalt abgeschnitten wurde, ist die Site vorerst disqualifiziert.
  - Wenn der Inhalt ohne Fehler syntaktisch analysiert wurde, ist die Site für das Durchsuchen unter Verwendung der gefundenen Regeln qualifiziert.

- Wenn der Inhalt mit Fehlern syntaktisch analysiert wurde, ist die Site für das Durchsuchen ohne Regeln qualifiziert.
4. Wenn ein anderer HTTP-Status zurückgegeben wurde, ist die Site vorerst disqualifiziert.

Wenn der Crawler versucht, die Datei `robots.txt` für eine Site herunterzuladen, aktualisiert er eine persistente Zeitmarke für diese Zeit, die als Datum des Robots bezeichnet wird. Ist eine Site disqualifiziert, weil die Informationen aus der Datei `robots.txt` nicht verfügbar sind, wird der persistente Zähler für Robotsfehler erhöht.

Wenn das Wiederholungsintervall erreicht ist, versucht der Crawler erneut, die Informationen aus der Datei `robots.txt` für die fehlgeschlagene Site abzurufen. Wenn die Anzahl aufeinander folgender Fehler die Anzahl maximal zulässiger Fehler erreicht, stoppt der Crawler seine Versuche, die Datei `robots.txt` für die Site abzurufen, und disqualifiziert die Site für die Crawlersuche.

Wenn eine Site für die Crawlersuche qualifiziert ist (die Überprüfung auf Regeln in der Datei `robots.txt` ist erfolgreich), wird der Zähler für Fehler auf null gesetzt. Der Crawler verwendet die Ergebnisse des Downloads, bis das Intervall für die Überprüfung von Regeln endet. Zu diesem Zeitpunkt muss die Site erneut qualifiziert werden.

#### **Tipp:**

- Wenn ein Server Inhalt zurückgibt, dieser aber Syntaxfehler enthält, oder wenn der Server ein anderes Protokoll für Robots verwendet als die Version von 1994, oder wenn der Inhalt etwas anderes als Regeln für Robots umfasst (z. B. eine Seite mit detaillierten Fehlerhinweisen), dann verhält der Crawler sich so, als ob der Server keine anwendbare Regeldatei hätte, und durchsucht die Site. Diese Vorgehensweise ist normalerweise korrekt, weil Administratoren für Objektgruppen weder den Inhalt von Sites noch das Standardserververhalten steuern. Wenn ein Web-Server-Administrator eine Site von der Suche ausschließen möchte und keine entsprechende Regeldatei installieren will, kann der Administrator für Objektgruppen die Site über den Web-Crawler blockieren, indem er die Domäne, die IP-Adresse oder das HTTP-Präfix der Site in den Regeln des Crawlers angibt.
- Wenn ein Server einen Statuscode 302 oder sonstige Umleitungscode zurückgibt, interpretiert der Crawler den Code so, dass die Site über eine Datei `robots.txt` verfügt, die verwendet werden soll, sich aber nicht an der entsprechenden Speicherposition (dem Stammverzeichnis der Site) befindet. Der Web-Server-Administrator muss die Datei an die korrekte Speicherposition versetzen, damit der Web-Crawler die Regeln in der Datei einhalten kann.
- Wenn Zertifikatsprobleme auftreten (das Zertifikat ist z. B. abgelaufen, die Zertifizierungsstelle ist nicht vertrauenswürdig, oder das Zertifikat ist selbst signiert und der Crawler nicht dafür konfiguriert, selbst signierte Zertifikate zu akzeptieren), interpretiert der Crawler das Problem als Fehler beim Herstellen einer Verbindung mit der Site und disqualifiziert die Site. Wahrscheinlich würden die gleichen Probleme ohnehin verhindern, dass andere Seiten der Site durchsucht werden können. Der Administrator für Objektgruppen muss selbst signierte Zertifikate ermöglichen, der Speicherdatei für vertrauenswürdige Schlüssel die Berechtigung der Site hinzufügen oder den Web-Server-Administrator bitten, ein aktuelles Zertifikat zu beziehen, um die zu durchsuchende Site zu aktivieren.
- Der Web-Crawler ist möglicherweise für die Verwendung der HTTP-Basisauthentifizierung konfiguriert (einschließlich HTTP-Basisauthentifizierung für Proxys). Bei korrekter Konfiguration ist die Authentifizierung auch für Downloads

von robots.txt-Dateien erforderlich. Ein Statuscode 403, 407 oder eine andere Antwort bezüglich einer Authentifizierung zeigt Berechtigungsprobleme an, woraufhin der Crawler die Site disqualifiziert. (Nur die HTTP-Basisauthentifizierung wird unterstützt.)

- Wenn die Datei robots.txt für eine Site die maximale Länge für eine Robotsseite überschreitet, kann der Administrator für Objektgruppen den konfigurierten Maximalwert erhöhen (der Standardwert von einer Million Byte sollte ausreichen).

Zur Fehlerbehebung können Sie einen Sitebericht anfordern, wenn Sie den Crawler überwachen. Wählen Sie Optionen für Folgendes aus: Anzeigen des Inhalts der Datei robots.txt (um zu sehen, ob Regeln dem Web-Crawler den Zugriff auf die Site untersagen), Anzeigen von Datum und Uhrzeit des letzten Versuchs des Crawlers, die Datei robots.txt herunterzuladen (der Crawler unternimmt bis zum Ende des Wiederholungsintervalls keinen neuen Versuch), und Anzeigen, wie viele aufeinander folgende erfolglose Versuche der Crawler unternommen hat, die Datei robots.txt herunterzuladen. Klicken Sie beim Überwachen des Web-Crawlers **Hilfe** an, um weitere Informationen zu diesen Siteberichtsoptionen und zum Interpretieren von Ergebnissen zu erhalten.

Weitere Informationen zum Protokoll zum Sperren von Websitebereichen für Robots finden Sie unter der folgenden URL-Adresse: <http://www.robotstxt.org/wc/exclusion.html>.

## Unterstützung für JavaScript

Der Web-Crawler für die Unternehmenssuche kann einige Links (URLs) finden, die in den JavaScript-Abschnitten von Webdokumenten enthalten sind.

Der Web-Crawler kann sowohl relative als auch absolute Links finden. Wenn ein HTML-Dokument das Element BASE enthält, verwendet der Crawler dieses Element zum Auflösen von relativen Links. Andernfalls verwendet der Crawler die URL des Dokuments.

Die Unterstützung für JavaScript ist auf die Extraktion von Links beschränkt. Der Crawler analysiert JavaScript nicht syntaktisch und erstellt kein DOM (Dokumentobjektmodell). Außerdem werden vom Crawler keine JavaScript-Anweisungen interpretiert oder ausgeführt. Der Crawler sucht nach Zeichenfolgen im Dokumentinhalt (einschließlich, jedoch nicht ausschließlich in JavaScript-Abschnitten), die wahrscheinlich URLs in JavaScript-Anweisungen sind. Dies bedeutet Folgendes:

- Einige URLs werden gefunden, die vom strengeren HTML-Parser ignoriert werden. Der Crawler weist alles außer syntaktisch gültigen URLs zurück. Einige der gültigen URLs, die beim Scannen zurückgegeben werden, sind jedoch eventuell nur von geringem Interesse für das Durchsuchen.
- Dokumentinhalt, der von JavaScript generiert wird, wie z. B. wenn ein Benutzer eine Seite mit einem Browser anzeigt und der Browser JavaScript-Code ausführt, kann vom Web-Crawler nicht erkannt werden und wird somit nicht indiziert.

Da der Web-Crawler JavaScript in HTML-Dateien nicht syntaktisch analysiert, werden URL-Adressen in JavaScript nicht durchsucht. Wenn Sie den Web-Crawler für die Crawlersuche in URL-Adressen in JavaScript aktivieren möchten, können Sie eine der folgenden Aktionen ausführen:

- Bearbeiten Sie den Web-Crawler in der Administrationskonsole für die Unternehmenssuche, fügen Sie auf der Seite **Web-Crawler-Bereich** der Liste mit URL-Adressen die URL-Adressen hinzu, die der Crawler als Ausgangspunkt ver-

wenden soll, um der Objektgruppe URL-Adressen hinzuzufügen (**Start-URL-Adressen**). Damit die Änderungen wirksam werden, starten Sie den Web-Crawler erneut. (Sie brauchen keine vollständige Crawleruche zu starten.)

- Verwenden Sie den Ankertag (<a href="..">), um die URL-Adressen in der HTML-Datei als Hypertext-Links anzugeben.

## Regeln zum Begrenzen des Web-Crawler-Bereichs

Sie geben Regeln an, um den Bereich zu begrenzen, den der Web-Crawler durchsuchen kann. Dadurch stellen Sie sicher, dass Benutzer nur auf die Websites zugreifen können, die sie durchsuchen dürfen.

Wenn ein Web-Crawler eine Webseite durchsucht, spürt er Links zu anderen Seiten auf und stellt diese Links in eine Warteschlange, um sie als Nächstes zu durchsuchen. Crawleruche und Discovery können wiederholt werden, sofern Zeit- und Speicherressourcen dies zulassen. Wenn Sie einen Web-Crawler konfigurieren, geben Sie an, wo der Crawler mit der Crawleruche beginnen soll. Der Web-Crawler kann von diesen ursprünglichen URL-Adressen (*Start-URL-Adressen*) aus jedes beliebige Dokument im Web erreichen, das durch direkte oder indirekte Links verbunden ist.

Zum Begrenzen des Crawlerbereichs konfigurieren Sie den Web-Crawler so, dass er bestimmte URL-Adressen genau durchsucht und Links ignoriert, die vom Interessengebiet wegführen. Da der Crawler standardmäßig jede URL-Adresse akzeptiert, die er aufspürt, müssen Sie Regeln angeben, die festlegen, welche URL-Adressen in die Objektgruppe eingeschlossen werden sollen, und die übrigen Seiten ausschließen.

Sie können auf verschiedene Weise angeben, was der Web-Crawler durchsuchen soll. Sie können Folgendes angeben:

- Eine Liste der Start-URL-Adressen, bei denen der Crawler mit der Crawleruche beginnen soll
- Drei Typen von Crawlerucheregeln: Domäne, IP-Adresse (Internet Protocol), URL-Präfix
- Eine Liste mit MIME-Typen für Dokumente, die Sie einschließen wollen
- Eine Liste mit Dateierweiterungen für Dokumente, die Sie ausschließen wollen
- Die maximale Anzahl Verzeichnisse in einem URL-Adressenpfad

Crawlerucheregeln haben das folgende Format:

```
aktion typ ziel
```

Die Werte für *aktion* sind *forbid* oder *allow*, der Wert für *typ* ist Domäne (*domain*), IP-Adresse (*address*) oder URL-Adressenpräfix (*prefix*, HTTP oder HTTPS). Der Wert für *ziel* hängt vom Wert für *typ* ab. Sie können einen Stern (\*) eingeschränkt als Platzhalterzeichen verwenden, um Ziele anzugeben, die mit einem Muster übereinstimmen.

### Domänenregeln

Das Ziel einer Domänenregel ist ein DNS-Domänenname. Sie können z. B. das Durchsuchen der gesamten Domäne *www.ibm.com* angeben:

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

Sie können einen Stern als Platzhalterzeichen angeben, was dazu führt, dass die Regel für alle Hostnamen gilt, die dem übrigen Muster entsprechen. Sie können z. B. angeben, dass keine Domänen durchsucht werden sollen, die mit 'server' beginnen und mit 'ibm.com' enden:

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

Beim Abgleich von Hostnamen muss die Groß-/Kleinschreibung sowohl bei einem expliziten Domännennamen als auch bei einem Domännennamensmuster beachtet werden. \*.user.ibm.com stimmt z. B. mit joe.user.ibm.com und mary.smith.user.ibm.com überein, jedoch nicht mit joe.user.IBM.com.

Eine Domänenregel, die keine Portnummer angibt, gilt für alle Ports dieser Domäne. In dem folgenden Beispiel sind alle Ports der Domäne sales zulässig:

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

Wenn eine Domänenregel eine Portnummer angibt, gilt die Regel nur für diesen Port. In dem folgenden Beispiel ist nur der Port 443 der Domäne sales zulässig:

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

## Präfixregeln

Eine Präfixregel steuert die Crawlersuche für URL-Adressen, die mit einer angegebenen Zeichenfolge beginnen. Das Ziel ist eine einzelne URL-Adresse, die in der Regel mindestens einen Stern enthält, um ein Muster anzugeben. Ein Stern wird z. B. oft als Schlusszeichen in der Präfixzeichenfolge angegeben.

Mit einer Präfixregel können Sie eine Website ganz oder in Teilen durchsuchen. Sie können einen Verzeichnispfad oder ein Muster angeben, und dann von diesem Punkt in der Verzeichnisbaumstruktur an alles zulassen (allow) oder ausschließen (forbid). Beispiel: Die folgenden Regeln werden zusammen angewendet, um dem Crawler das uneingeschränkte Durchsuchen im Verzeichnis public unter sales.ibm.com zu ermöglichen, aber lassen den Crawler nicht auf andere Seiten auf der Site zugreifen:

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

Wenn Sie Präfixregeln festlegen, können Sie mehrere Sterne verwenden und diese an beliebigen Stellen angeben, nicht nur an der letzten Position in der Zeichenfolge. Beispiel: Die folgende Regel verbietet, dass der Crawler eine Crawlersuche für Dokumente in einem Verzeichnis der höchsten Ebene von der Site sales.ibm.com ausführt, wenn der Verzeichnisname auf fs endet. (Sie könnten z. B. über Dateisystemmounts verfügen, die keine Informationen enthalten, die im Suchindex nützlich wären.)

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

## Adressenregeln

Mit einer Adressenregel können Sie die Crawlersuche gesamter Hosts oder Netze steuern, indem Sie eine IP-Adresse und die Netzmaske als das Ziel angeben. Beispiel:

**IPv4** allow address 9.0.0.0 255.0.0.0

**IPv6** Wenn Sie die Unternehmenssuche auf einem Windows 2003-Server ausfüh-

ren und das System für die Unternehmenssuche für die Verwendung des IP-Protokolls der Version 6 (IPv6) aktiviert ist, müssen Sie die Adresse in eckigen Klammern einschließen.

```
allow address [2001:db8:0:1:0:0:0:1]
```

Mit der Netzmaske können Sie die Mustererkennung angeben. Damit eine Adressenregel auf eine Kandidaten-IP-Adresse angewendet wird, müssen die IP-Adresse in der Regel und die Kandidaten-IP-Adresse übereinstimmen. Mit Ausnahme der Stellen, an denen Nullen in der Netzmaske angegeben sind. Die Adressenregel definiert ein Muster, und die Netzmaske definiert die signifikanten Bit im Adressenmuster. Eine Null in der Netzmaske fungiert als Platzhalterzeichen und bedeutet, dass ein beliebiger Wert, der an derselben Bitposition in der Adresse angegeben ist, als Übereinstimmung gilt.

Im vorherigen Beispiel gilt die Regel `allow` für beliebige IP-Adressen mit 9 im ersten Oktett und beliebigen Werten in den letzten drei Oktetten.

Die folgende Regel ist eine nützliche Regel, die Sie als letzte Adresse in Ihrer Liste mit Adressenregeln einschließen sollten. Diese Regel stimmt mit jeder beliebigen IP-Adresse überein, da die Netzmaske alle Bit irrelevant macht (die Regel schließt alle Adressen aus, die von keiner der vorherigen Regeln in Ihrer Regelliste zugelassen wurden).

```
IPv4 forbid address 0.0.0.0 0.0.0.0
```

**IPv6**

```
forbid address :: ::
```

**Einschränkungen für Proxy-Server:** Geben Sie keine IP-Adressenregeln an, wenn Sie Websites durchsuchen wollen, die von einem Proxy-Server bereitgestellt werden. Ein Proxy-Server wird in der Regel dann verwendet, wenn ein Benutzeragent (ein Browser oder Crawler) keinen Direktzugriff auf die Netze hat, auf denen sich die Web-Server befinden. Ein HTTP-Proxy-Server kann beispielsweise die HTTP-Anforderungen von einem Crawler an einen Web-Server weiterleiten und die Antworten zurück an den Crawler übertragen.

Wenn ein Web-Crawler einen Proxy-Server verwendet, ist die IP-Adresse des Proxy-Servers die einzige IP-Adresse, die der Crawler für einen anderen Host hat. Wenn IP-Adressenregeln verwendet werden, um den Crawler auf ein Teilnetz von IP-Adressen zu begrenzen, führt die Einschränkung dazu, dass fast alle URL-Adressen mit dem Statuscode 760 klassifiziert werden (dadurch wird angegeben, dass sie durch den Web-Space verboten sind).

## Reihenfolge für Crawlerregeln

Der Crawler wendet die Crawlerregeln zu unterschiedlichen Zeiten während des Discoveryprozesses und des Crawlerprozesses für URL-Adressen an. Die Reihenfolge der Regeln ist wichtig, aber nur innerhalb der Regeln des jeweiligen Typs. Es ist ein Unterschied, ob eine Adressenregel vor oder nach einer anderen Adressenregel kommt, aber es macht keinen Unterschied, ob eine Adressenregel vor oder nach einer Präfixregel kommt, da der Crawler die Regeln nicht zur selben Zeit anwendet.

Innerhalb der Gruppe von Regeln eines Typs testet der Crawler eine Kandidatendomäne, -adresse oder -URL-Adresse für jede Regel, von der ersten angegebenen

Regel bis zur letzten, bis er eine entsprechende Regel findet. Die Aktion, die für die erste entsprechende Regel angegeben ist, wird verwendet.

Die Abhängigkeit von der Reihenfolge führt zu einer typischen Struktur für die meisten Crawlersuchregeln:

- Die Gruppe von Domänenregeln beginnt gewöhnlich mit forbid-Regeln, die einzelne Domänen aus dem Crawlerbereich eliminieren. Der Administrator für die Objektgruppe könnte z. B. festlegen, dass bestimmte Domänen keine nützlichen Informationen enthalten.
- Der Liste der forbid-Regeln folgt meist eine Reihe von allow-Regeln (mit Platzhalterzeichen), die dem Crawler das Aufrufen einer beliebigen Domäne ermöglichen, die mit einem der Domännennamen der höchsten Ebene endet, die ein Unternehmensintranet, wie z. B. \*.ibm.com und \*.lotus.com, definieren.

Setzen Sie die folgende Standardregel, die Domänen ausschließt, die von einer vorherigen Regel nicht zugelassen wurden, an das Ende der Gruppe von Domänenregeln:

```
forbid domain *
```

Diese letzte Regel ist entscheidend, da sie verhindert, dass das gesamte Internet in den Crawlerbereich eingeschlossen wird.

- Die Gruppe von Adressenregeln beginnt normalerweise mit einer kleinen Anzahl allow-Regeln, die es dem Crawler ermöglichen, Netze der höchsten Ebene (Klasse-A, Klasse-B oder Klasse-C), die ein Unternehmensintranet umfassen, zu durchsuchen.

In den vorangegangenen Erläuterungen finden Sie Beispieladressenregeln zum Angeben der letzten Regel in Ihrer Liste mit Adressenregeln, die verhindert, dass der Crawler Websites außerhalb des Unternehmensnetzes durchsucht.

- Die Gruppe von Präfixregeln ist für gewöhnlich die größte, da sie beliebig detaillierte Spezifikationen für zugelassene und ausgeschlossene Bereiche enthält, die als Baumstrukturen und Unterbaumstrukturen angegeben sind. Eine empfohlene Methode ist hier, die genauer umrissenen Bereiche zuerst zuzulassen bzw. auszuschließen, und dann die gegenteilige Regel mit einem allgemeineren Muster anzugeben, um alles Übrige zuzulassen bzw. auszuschließen.

Der Präfixabschnitt endet normalerweise nicht mit einer typischen Regel. Die empfohlenen Domänen- und Adressenabschlussregeln können effizienter sicherstellen, dass der Crawler nicht über das Unternehmensnetz hinaus durchsucht, als dies beim Testen von URL-Präfixen der Fall ist.

Der Crawler kann Präfixregeln effizienter anwenden, wenn Sie die Regeln nach Aktion (forbid oder allow) gruppieren. Anstatt also kurze Abfolgen von allow- und forbid-Regeln anzugeben, die einander abwechseln, geben Sie eine lange Abfolge von Regeln an, die die eine Aktion festlegen, und dann geben Sie eine lange Abfolge von Regeln an, die die andere Aktion festlegen. Sie können allow- und forbid-Regeln miteinander verbinden, um die Zielsetzung Ihres Crawlerbereichs zu erreichen. Das Zusammenfassen der allow-Regeln einerseits und der forbid-Regeln andererseits kann jedoch die Crawlerleistung verbessern.

## **Dateierweiterungen, MIME-Typen und maximale Crawlersuchtiefe**

Diese Optionen bieten Ihnen zusätzliche Möglichkeiten, Inhalt für den Crawlerbereich anzugeben. Sie können bestimmte Dokumenttypen auf der Basis der Dateierweiterung eines Dokuments ausschließen, und Sie können bestimmte Dokumenttypen auf der Basis des MIME-Typs eines Dokuments einschließen. Wenn Sie angeben, welche MIME-Typen der Crawler durchsuchen soll, bedenken Sie, dass der MIME-Typ in Webdokumenten oft nicht richtig definiert ist.

Die maximale Crawlersuchtiefe ist die Anzahl Schrägstriche in einer URL-Adresse ab ihrem Site-Root. Mit dieser Option können Sie verhindern, dass der Crawler in rekursive Dateisystemstrukturen mit unendlicher Tiefe gezogen wird. Die Crawlersuchtiefe entspricht nicht den Ebenen, die der Crawler durchquert, wenn er den Links von einem Dokument zum nächsten Dokument folgt.

## Start-URL-Adressen

Start-URL-Adressen sind die URL-Adressen, bei denen der Crawler die Crawlersuche startet. Diese URL-Adressen werden jedesmal in die Crawlersuche eingefügt, wenn der Crawler gestartet wird. Wenn die Start-URL-Adressen bereits aufgespürt sind, werden sie nicht schneller als andere Websites, die Sie in den Crawlersuchregeln zulassen, durchsucht bzw. erneut durchsucht.

Eine Start-URL-Adresse ist wichtig, wenn der Web-Crawler das erste Mal gestartet wird und der Crawlerbereich leer ist. Eine Start-URL-Adresse ist ebenfalls wichtig, wenn Sie eine URL-Adresse hinzufügen, die vorher noch nicht in der Liste der Start-URL-Adressen in einem Crawlerbereich aufgespürt wurde.

Start-URL-Adressen müssen vollständig qualifizierte URL-Adressen sein, nicht nur Domännennamen. Sie müssen das Protokoll und, falls der Port nicht 80 ist, die Portnummer angeben.

Die folgenden URL-Adressen sind gültige Start-URL-Adressen:

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

Die folgende URL-Adresse ist keine gültige Start-URL-Adresse:

```
www.ibm.com
```

Sie müssen die Start-URL-Adressen in Ihre Crawlersuchregeln einschließen. Der Crawler kann z. B. mit der Crawlersuche nicht bei einer angegebenen Start-URL-Adresse beginnen, wenn die Crawlersuchregeln nicht zulassen, dass diese URL-Adresse durchsucht wird.

**Unterstützung für IPv6-Adressen:** Wenn Sie die Unternehmenssuche auf einem Windows 2003-Server ausführen und das System für die Unternehmenssuche für die Verwendung des IP-Protokolls der Version 6 (IPv6) aktiviert ist, müssen Sie die Start-URL-Adressen in eckigen Klammern einschließen. Beispiel:

```
http://[2001:db8:0:1:0:0:0:1]
http://[2001:db8:0:1::1]
```

### Zugehörige Tasks

„Aktivieren der Unterstützung für das IPv6-Protokoll“ auf Seite 29

## Testen von URL-Verbindungen mit dem Web-Crawler

Nach der Angabe von URL-Adressen für den für die Crawlersuche zu verwendenden Web-Crawler, können Sie die Konfiguration der Crawlersuchregeln testen.

Sie können **Testen** anklicken, wenn Sie die zu durchsuchenden Domänen, HTTP-Präfixe oder IP-Adressen angeben. Oder Sie können die Seite **URL-Adressen testen** auswählen, um zu testen, ob der Crawler zusätzlich zu den von Ihnen angegebenen URL-Adressen auch zu den Start-URL-Adressen eine Verbindung herstellen kann.

Die Testergebnisse zeigen an, ob der Crawler mit dem Benutzeragentennamen, der in den Crawlermerkmalen angegeben ist, auf URL-Adressen zugreifen kann. Die Testergebnisse zeigen außerdem an, ob eine URL-Adresse aufgrund der Ausschlussregeln nicht durchsucht werden kann. (Ein Dokument wird möglicherweise nicht durchsucht, weil es eine Dateierweiterung hat, die mit einer vom Crawlerbereich ausgeschlossenen Erweiterung übereinstimmt.)

Nachdem eine Site mindestens einmal durchsucht worden ist, können Sie die URL-Adressen testen, um weitere Informationen erhalten. Der Testbericht kann z. B. den aktuellen HTTP-Statuscode enthalten (der angibt, ob eine Crawleruche der URL-Adresse erfolgreich war), zeigen, wann die URL-Adresse zuletzt durchsucht worden ist und wann sie erneut durchsucht wird, sowie ferner zeigen, ob der Benutzeragent die aktuelle Datei `robots.txt` des Web-Servers verwendet.

## Intervalleinstellungen für erneute Suche im Web-Crawler

Sie geben Optionen in den Web-Crawler-Merkmalen an, um zu beeinflussen, wie oft der Web-Crawler URL-Adressen aufruft.

Die meisten anderen Crawlertypen in einem System für die Unternehmenssuche werden entsprechend der Zeitpläne ausgeführt, die ein Administrator angibt. Im Gegensatz dazu wird ein Web-Crawler, nachdem Sie ihn gestartet haben, in der Regel kontinuierlich ausgeführt. Sie geben die minimalen und maximalen Intervalle für erneute Suche an, um zu steuern, wie oft er URL-Adressen erneut aufruft, die er zuvor durchsucht hat.

Wenn Sie mit der Administrationskonsole für die Unternehmenssuche einen Web-Crawler erstellen oder Web-Crawler-Merkmale bearbeiten, können Sie eine Option auswählen, um erweiterte Merkmale zu konfigurieren. Sie geben auf der Seite **Erweiterte Web-Crawler-Merkmale** die Optionen für das minimale Intervall für erneute Suche und das maximale Intervall für erneute Suche an. Der Web-Crawler verwendet die Werte, die Sie angeben, um ein Intervall für eine erneute Datensuche zu berechnen.

Wenn eine Seite zum ersten Mal durchsucht wird, verwendet der Crawler das Datum und die Uhrzeit des Zeitpunkts, zu dem die Seite durchsucht wird, und den Durchschnitt der angegebenen minimalen und maximalen Intervalle für erneute Suche, um ein Datum für die erneute Suche festzulegen. Die Seite wird vor diesem Datum nicht erneut durchsucht werden. Die Zeit, zu der die Seite nach diesem Datum erneut durchsucht wird, hängt von der Crawlerauslastung und der Ausgewogenheit zwischen neuen und alten URL-Adressen im Crawlerbereich ab.

Jedes Mal, wenn die Seite erneut durchsucht wird, prüft der Crawler, ob sich der Inhalt geändert hat. Wenn sich der Inhalt geändert hat, wird das nächste Intervall für erneute Suche kürzer als das vorherige sein, aber nie kürzer als das angegebene minimale Intervall für erneute Suche. Wenn sich der Inhalt nicht geändert hat, wird das nächste Intervall für erneute Suche länger als das vorherige sein, aber nie länger als das angegebene maximale Intervall für erneute Suche.

## Optionen für das Aufrufen von URL-Adressen mit dem Web-Crawler

Sie können den Web-Crawler zwingen, bestimmte URL-Adressen so schnell wie möglich aufzurufen.

Wenn Sie den Crawlerbereich mit Informationen von bestimmten Websites aktualisieren müssen, können Sie den Crawler überwachen, die Option **Aufzurufende**

**oder erneut aufzurufende URL-Adressen** auswählen und dann die URL-Adressen oder URL-Adressenmuster der Seiten angeben, die durchsucht bzw. erneut durchsucht werden müssen.

Wenn z. B. Ihre Kommunikationsabteilung Ihrem Intranet eine Webseite hinzufügt oder eine Seite überarbeitet, um eine wichtige Richtlinienänderung widerzuspiegeln, können Sie die URL-Adresse der neuen oder geänderten Seite angeben. Wenn der Crawler ausgeführt wird, stellt der Crawler die angegebene URL-Adresse für die Crawler suche in die Warteschlange, bis er das nächste Mal Seiten überprüft, die darauf warten, aufgerufen zu werden (dies ist in der Regel alle zehn Minuten). Wenn der Crawler gerade nicht aktiv ist, stellt er die angegebene URL-Adresse in die Warteschlange, sodass sie beim nächsten Crawlerstart durchsucht werden kann.

Stellen Sie sicher, dass die Crawler suchregeln eine Regel einschließen, die dem Crawler ermöglicht, die von Ihnen angegebenen URL-Adressen aufzurufen. Der Crawler kann die von Ihnen angegebenen URL-Adressen schneller aufrufen, als dies normalerweise der Fall wäre. Damit jedoch eine URL-Adresse überhaupt durchsucht werden kann, muss eine Crawler suchregel vorhanden sein, die zulässt, dass die URL-Adresse durchsucht werden kann.

Die neu durchsuchten Daten werden für die Suche verfügbar, wenn der Hauptindex das nächste Mal erstellt wird.

## **Vorgehensweise des Web-Crawlers bei Seiten mit detaillierten Fehlerhinweisen**

Sie können den Web-Crawler so konfigurieren, dass benutzerdefinierte Seiten bearbeitet werden, die Administratoren von Websites erstellen, wenn sie nicht möchten, dass ein Standardfehlercode als Antwort auf Anforderungen für bestimmte Seiten zurückgegeben wird.

Wenn ein HTTP-Server die von einem Client angeforderte Seite nicht zurückgeben kann, gibt der Server in der Regel eine Antwort zurück, die aus einem Header mit einem Statuscode besteht. Der Statuscode gibt das Problem an (beispielsweise Fehler 404, was bedeutet, dass die Datei nicht gefunden werden konnte). Einige Administratoren von Websites erstellen spezielle Seiten, die das Problem genauer erläutern, und konfigurieren den HTTP-Server so, dass stattdessen diese Seiten zurückgegeben werden. Diese benutzerdefinierten Seiten werden *Seiten mit detaillierten Fehlerhinweisen* genannt.

Seiten mit detaillierten Fehlerhinweisen können die Ergebnisse des Web-Crawlers verzerren. Anstelle eines Headers mit der Angabe des Problems empfängt der Crawler beispielsweise eine Seite mit einem normalen Fehler und den Statuscode 200, der angibt, dass eine gültige HTML-Seite erfolgreich heruntergeladen wurde. Diese heruntergeladene Seite mit einem detaillierten Fehlerhinweis bezieht sich jedoch nicht auf die angeforderte URL, und ihr Inhalt ist jedes Mal fast identisch, wenn sie anstelle einer angeforderten Seite zurückgegeben wird. Diese irrelevanten und fast identischen Seiten verzerren den Index und die Suchergebnisse.

Für diesen Fall können Sie bei der Konfiguration des Web-Crawlers Optionen für die Handhabung von Seiten mit detaillierten Fehlerhinweisen angeben. Der Web-Crawler benötigt die folgenden Informationen zu den einzelnen Websites, die Seiten mit detaillierten Fehlerhinweisen zurückgeben:

- Ein URL-Muster für eine Site, die Seiten mit detaillierten Fehlerhinweisen verwendet. Dieses URL-Muster besteht aus dem Protokoll (HTTP oder HTTPS), dem Hostnamen, der Portnummer (falls keine Standardportnummer) und einem

Pfadnamen. Sie können einen Stern (\*) als Platzhalterzeichen verwenden, das mindestens einem Zeichen bis zum nächsten Vorkommen eines Nicht-Platzhalterzeichens in dem Muster entspricht. Bei dem von Ihnen angegebenen Muster muss die Groß-/Kleinschreibung beachtet werden.

- Ein Titelmuster für Text, der dem Tag <TITLE> eines HTML-Dokuments entspricht. Sie können einen Stern (\*) als Platzhalterzeichen verwenden, um dieses Muster anzugeben. Bei diesem von Ihnen angegebenen Muster muss die Groß-/Kleinschreibung beachtet werden.
- Ein Inhaltsmuster für Text, der dem Inhalt eines HTML-Dokuments entspricht. Der Inhalt ist nicht einfach nur der Inhalt des Tags <BODY>, wenn der Tag <BODY> vorhanden ist. Zum Inhalt gehört alles, was nach dem HTTP-Header in der Datei steht. Sie können einen Stern (\*) als Platzhalterzeichen verwenden, um dieses Muster anzugeben. Bei diesem von Ihnen angegebenen Muster muss die Groß-/Kleinschreibung beachtet werden.
- Eine ganze Zahl, die den Statuscode darstellt, der für Dokumente verwendet werden soll, die den von Ihnen angegebenen URL-, Titel- und Inhaltsmustern entsprechen.

## Beispiel

Bei der folgenden Konfiguration soll der Web-Crawler alle gültigen HTML-Seiten (Statuscode 200), die von der Website `http://www.mysite.com/hr/*` zurückgegeben werden, mit den angegebenen Titel- und Inhaltsmustern vergleichen. Wenn der Tag <TITLE> einer Seite mit "Sorry, the page" beginnt und der Inhalt des Dokuments beliebigen Text (\*) enthält, handhabt der Crawler die Seite so wie den Statuscode 404 (die Seite wurde nicht gefunden).

Tabelle 3. Beispiel einer Seite mit einem normalen Fehler

URL-Muster	Titelmuster	Inhaltsmuster	HTTP-Statuscode
<code>http://www.mysite.com/hr/*</code>	<code>Sorry, the page*</code>	<code>*</code>	404

Sie können mehrere Einträge für dieselbe Website erstellen, um verschiedene Statuscodes zu handhaben. Die einzelnen Statuscodes einer Website erfordern jeweils einen eigenen Eintrag in der Konfiguration des Web-Crawlers.

## Verwenden von Platzhalterzeichen

Die URL-, Titel- und Inhaltsmuster sind keine regulären Ausdrücke. Der Stern entspricht allen Zeichen bis zum nächsten Auftreten eines Nicht-Platzhalterzeichens. Beispiel:

`*404` ist gleich *beliebige Zeichen*404  
`404: *` ist gleich `404: beliebige Zeichen`  
`http://*.mysite.com/*` ist gleich `http://beliebiger Host.mysite.com/beliebige Datei`  
`*` ist gleich *beliebige Zeichen*

## Auswirkungen auf die Leistung

Wenn Sie Optionen für die Handhabung von Seiten mit detaillierten Fehlerhinweisen konfigurieren, erhöhen Sie die Verarbeitungszeit des Crawlers, da alle erfolgreich durchsuchten Seiten überprüft werden müssen. Es ist eine erhöhte Verarbeitungszeit erforderlich, wenn nach Musterentsprechungen gesucht und festgestellt werden muss, ob eine Seite oder ein Ersatzstatuscode zurückgegeben werden soll.

## Unterstützung für die Crawlersuche auf sicheren Websites

Durch die Angabe von Berechtigungsnachweisen in der Administrationskonsole für die Unternehmenssuche kann der Web-Crawler auf eingeschränkt zugänglichen Inhalt zugreifen, beispielsweise auf Dokumente, die für den Zugriff ein Kennwort erfordern.

Wenn ein Web-Server eine HTTP-Basisauthentifizierung oder formularbasierte HTML-Authentifizierung verwendet, um den Zugriff auf Websites einzuschränken, können Sie bei der Konfiguration des Web-Crawlers Berechtigungsnachweise angeben, damit Seiten auf den kennwortgeschützten Websites durchsucht werden können. Sie können auch Optionen für die manuelle Konfiguration von Cookie-Dateien angeben.

### Durch HTTP-Basisauthentifizierung geschützte Websites

Wenn ein Web-Server eine HTTP-Basisauthentifizierung verwendet, um den Zugriff auf Websites einzuschränken, können Sie die Authentifizierungsnachweise angeben, mit denen der Web-Crawler auf kennwortgeschützte Seiten zugreifen kann.

Zur Feststellung, ob ein Benutzer (oder eine Clientanwendung) eine Zugriffsberechtigung für Seiten einer Website hat, verwenden viele Web-Server ein Clientauthentifizierungsschema, die so genannte HTTP-Basisauthentifizierung, um die Identität des Benutzers zu ermitteln. In der Regel ist dieser Vorgang interaktiv:

- Wenn ein HTTP-Benutzeragent (wie z. B. ein Web-Browser) eine Seite anfordert, die durch eine HTTP-Basisauthentifizierung geschützt ist, antwortet der Web-Server mit dem Statuscode 401, der angibt, dass der Anforderer nicht berechtigt ist, auf die angeforderte Seite zuzugreifen.
- Der Web-Server fordert darüber hinaus den Anforderer auf, die Berechtigungsnachweise anzugeben, anhand derer geprüft werden kann, ob der Benutzer Zugriff auf den eingeschränkt zugänglichen Inhalt hat.
- Der Web-Browser fordert den Benutzer in einem Dialog auf, einen Benutzernamen, ein Kennwort und weitere Informationen anzugeben, die als Berechtigungsnachweise des Benutzers erforderlich sind.
- Der Web-Browser codiert die Berechtigungsnachweise und schließt sie bei der Wiederholung der Anforderung der geschützten Seite ein.
- Wenn die Berechtigungsnachweise gültig sind, antwortet der Web-Server mit dem Rückkehrcode 200 und dem Inhalt der angeforderten Seite.
- Nachfolgende Anforderungen von Seiten von demselben Web-Server schließen in der Regel dieselben Berechtigungsnachweise ein, sodass der berechtigte Benutzer auf weitere eingeschränkt zugängliche Inhalte Zugriff hat, ohne zusätzliche Berechtigungsnachweise anzugeben.

Nachdem die Identität eines Benutzers festgestellt wurde, tauschen der Web-Server und der HTTP-Benutzeragent in der Regel Token aus, so genannte *Cookies*, sodass der Anmeldestatus des Benutzers zwischen den HTTP-Anforderungen erhalten bleibt.

Da der Web-Crawler nicht interaktiv ausgeführt wird, müssen die Berechtigungsnachweise, mit denen er kennwortgeschützte Seiten durchsuchen kann, vor der Crawlersuche des Crawlers angegeben werden. Wenn Sie einen Web-Crawler erstellen oder den Crawlerbereich bearbeiten, geben Sie Informationen zu den einzelnen sicheren Websites an, die durchsucht werden sollen.

Zur Angabe dieser Informationen müssen Sie eng mit den Administratoren dieser durch die HTTP-Basisauthentifizierung geschützten Websites oder Web-Server zusammenarbeiten. Sie müssen Ihnen die Sicherheitsanforderungen für die zu

durchsuchenden Websites zur Verfügung stellen, einschließlich aller Informationen, die verwendet werden, um die Identität des Web-Crawlers zu authentifizieren und um festzustellen, ob der Crawler die Berechtigung hat, die eingeschränkt zugänglichen Seiten zu durchsuchen.

Wenn für die Objektgruppe bei ihrer Erstellung die Sicherheit aktiviert wurde, können Sie bei der Konfiguration des Crawlers Sicherheitstoken angeben, wie z. B. Benutzer-IDs, Gruppen-IDs oder Benutzerrollen, um den Zugriff auf Dokumente zu steuern. Der Web-Crawler ordnet diese Sicherheitstoken jedem zu durchsuchenden Dokument im Dateisystembaum für die angegebene Stamm-URL zu. Die Token werden neben den Sicherheitstoken auf Dokumentenebene verwendet, die Sie für den gesamten Web-Crawler-Bereich konfigurieren.

Die Reihenfolge der URLs ist wichtig. Nachdem Sie Informationen zu einer kennwortgeschützten Website hinzugefügt haben, müssen Sie sie in der Reihenfolge anordnen, in der sie vom Crawler verarbeitet werden sollen. Führen Sie die genaueren URLs zuerst auf, und setzen Sie die generischeren URLs in der Liste weiter nach unten. Wenn der Web-Crawler eine Kandidaten-URL auswertet, verwendet er die Authentifizierungsdaten, die für die erste URL in der Liste angegeben sind, die der Kandidaten-URL entspricht.

### **Durch formularbasierte Authentifizierung geschützte Websites**

Wenn ein Web-Server HTML-Formulare verwendet, um den Zugriff auf Websites einzuschränken, können Sie die Authentifizierungsnachweise angeben, mit denen der Web-Crawler auf kennwortgeschützte Seiten zugreifen kann.

Zur Feststellung, ob ein Benutzer (oder eine Clientanwendung) eine Zugriffsberechtigung für Seiten einer Website hat, verwenden viele Web-Server HTML-Formulare, um die Identität des Benutzers zu ermitteln. In der Regel ist dieser Vorgang interaktiv:

- Wenn ein HTTP-Benutzeragent (wie z. B. ein Web-Browser) eine Seite anfordert, die durch eine formularbasierte Authentifizierung geschützt ist, prüft der Web-Server, ob die Anforderung ein Cookie enthält, mit dem die Identität des Benutzers ermittelt werden kann.
- Wenn das Cookie nicht vorhanden ist, wird der Benutzer vom Web-Server aufgefordert, die Sicherheitsdaten in ein Formular einzugeben. Wenn der Benutzer das Formular übergibt, gibt der Web-Server die erforderlichen Cookies zurück, und die Anforderung für die kennwortgeschützte Seite darf fortgesetzt werden.
- Nachfolgende Anforderungen, die die erforderlichen Cookies enthalten, dürfen ebenfalls fortgesetzt werden. Der berechtigte Benutzer kann auf weitere eingeschränkt zugängliche Inhalte zugreifen, ohne bei jeder Anforderung aufgefordert zu werden, ein Formular auszufüllen und Berechtigungsnachweise anzugeben.

Da der Web-Crawler nicht interaktiv ausgeführt wird, müssen die Berechtigungsnachweise, mit denen er kennwortgeschützte Seiten durchsuchen kann, vor der Crawlersuche des Crawlers angegeben werden. Wenn Sie einen Web-Crawler erstellen oder den Crawlerbereich bearbeiten, geben Sie Informationen zu den einzelnen sicheren Websites an, die durchsucht werden sollen.

Die von Ihnen angegebenen Felder entsprechen den Feldern, die ein interaktiver Benutzer auf Anforderung des Web-Browsers ausfüllt, sowie sämtlichen verdeckten oder statischen Feldern, die für eine erfolgreiche Anmeldung erforderlich sind.

Zur Angabe dieser Informationen müssen Sie eng mit den Administratoren dieser durch die formularbasierte Authentifizierung geschützten Websites oder Web-Ser-

ver zusammenarbeiten. Sie müssen Ihnen die Sicherheitsanforderungen für die zu durchsuchenden Websites zur Verfügung stellen, einschließlich aller Informationen, die verwendet werden, um die Identität des Web-Crawlers zu authentifizieren und um festzustellen, ob der Crawler die Berechtigung hat, die eingeschränkt zugänglichen Seiten zu durchsuchen.

Die Reihenfolge der URL-Muster ist wichtig. Nachdem Sie Informationen zu einer kennwortgeschützten Website hinzugefügt haben, müssen Sie sie in der Reihenfolge anordnen, in der sie vom Crawler verarbeitet werden sollen. Führen Sie die genaueren URL-Muster zuerst auf, und setzen Sie die generischeren URL-Muster in der Liste weiter nach unten. Wenn der Web-Crawler eine Kandidaten-URL auswertet, verwendet er die Formulardaten, die für das erste URL-Muster in der Liste angegeben sind, das der Kandidaten-URL entspricht.

### **Verwenden eines Plug-ins, um sichere WebSphere Portal-Sites zu durchsuchen**

Wenn die globale Sicherheit in WebSphere Application Server aktiviert ist und Sie mit dem Web-Crawler sichere WebSphere Portal-Sites durchsuchen wollen, müssen Sie ein Crawler-Plug-in erstellen, um formularbasierte Authentifizierungsanforderungen zu verarbeiten. Eine Diskussion über formularbasierte Authentifizierung sowie ein Musterprogramm, das Sie für Ihr benutzerdefiniertes Web-Crawler-Plug-in anpassen können, finden Sie unter <http://www.ibm.com/developerworks/db2/library/techarticle/dm-0707nishitani>.

Das Plug-in ist erforderlich, wenn Sie den Web-Crawler verwenden, um beliebige Sites über WebSphere Portal zu durchsuchen, einschließlich Workplace Web Content Management-Sites und Lotus Quickr-Sites.

## **Über Proxy-Server bereitgestellte Websites**

Wenn dem Web-Crawler kein direkter Zugriff auf ein Netz gewährt wird, können Sie den Crawler so konfigurieren, dass er über einen HTTP-Proxy-Server auf den Inhalt zugreift, den Sie durchsuchen möchten.

Wenn auf dem Computer, auf dem der Web-Crawler ausgeführt werden soll, kein Zugriff auf ein TCP/IP-Netz verfügbar ist oder wenn der Zugriff auf berechtigte Prozesse eingeschränkt ist, können Sie den Web-Crawler so konfigurieren, dass ein HTTP-Proxy-Server verwendet wird. Ein HTTP-Proxy ist ein Prozess, der an einem angegebenen Port auf einem angegebenen Host empfangsbereit für HTTP-Anforderungen ist. Der Proxy-Server leitet Anforderungen an den Web-Server weiter und leitet Antworten vom Web-Server an den Client (den Web-Crawler) weiter, von dem die Anforderung kommt. Ein Proxy-Server kann auf demselben oder auf einem anderen Computer ausgeführt werden wie der Web-Crawler.

Bei einer Crawlersuche ohne Proxy-Server wird eine URL-Anforderung direkt an den Host gesendet. Bei einer Crawlersuche mit Proxy-Server wird die Anforderung an den Proxy-Server gesendet.

Wenn Sie einen Web-Crawler erstellen oder den Crawlerbereich bearbeiten, geben Sie Informationen zu den Proxy-Servern an, die der Web-Crawler verwendet, wenn Seiten in der Domäne des Proxy-Servers durchsucht werden sollen. Bevor Sie dem Crawlerbereich einen Proxy-Server hinzufügen, ermitteln Sie die Namen der vom Proxy-Server bereitgestellten Domänen, den Hostnamen oder die IP-Adresse des Proxy-Servers sowie die vom Proxy-Server verwendete Portnummer.

Wenn der Proxy-Server eine Authentifizierung erfordert, ermitteln Sie außerdem einen Benutzernamen und ein Kennwort, mit denen der Web-Crawler auf die vom Proxy-Server bereitgestellten Seiten zugreifen kann. Der Web-Crawler unterstützt nur die HTTP-Basisauthentifizierung für Proxys, wie in RFC2616 (<http://rfc.net/rfc2616.html>) beschrieben. Andere Authentifizierungstypen, einschließlich Windows NT LAN Manager (NTLM), werden nicht unterstützt.

Nachdem Sie einen Proxy-Server hinzugefügt haben, müssen Sie ihn auswählen und an der Stelle der Reihenfolge anordnen, an der der Crawler ihn verarbeiten soll. Führen Sie die genaueren Domännennamen zuerst auf, und setzen Sie die generischeren Domännennamen in der Liste weiter nach unten. Wenn der Web-Crawler eine Kandidaten-URL auswertet, verwendet er die Daten des Proxy-Servers, die für die erste Domäne in der Liste angegeben sind, die der Kandidaten-URL entspricht. (Auf URLs, die keiner Proxy-Regel entsprechen, kann der Crawler direkt zugreifen.)

## Cookieverwaltung

In der Regel erfolgt die Cookieverwaltung automatisch, ohne dass Aktionen des Administrators für die Unternehmenssuche erforderlich sind. Falls erforderlich, können Sie Cookies für eine Web-Crawler-Sitzung manuell angeben.

Cookies sind opake Token, die ein Web-Server an einen Benutzeragenten als Teil eines HTTP-Antwortheaders zurückgibt. Sie haben nur für den Web-Server Aussagekraft, der sie ausgibt, und sie werden verwendet, um den Status zwischen HTTP-Anforderungen aufrecht zu erhalten. Während der Clientauthentifizierung beispielsweise kann der Web-Server ein Cookie zurückgeben, mit dem der Server feststellen kann, dass ein authentifizierter Benutzer bereits angemeldet ist. Durch das vorhandene Cookie kann der Benutzer weitere Anforderungen für Seiten auf diesem Web-Server absetzen, ohne aufgefordert zu werden, sich erneut anzumelden.

Der Web-Crawler behält Cookies, die vom Web-Server empfangen werden, und verwendet sie für die Dauer der Crawlerinstanz. Er speichert die Cookies in der Datei `cookies.ini`, die am Ende einer Crawlerinstanz vom Crawler neu geschrieben wird. Wenn der Web-Crawler gestoppt wird, werden alle nicht abgelaufenen Cookies gespeichert und zu Beginn der nächsten Sitzung erneut geladen.

Wenn Sie Cookies manuell angeben, speichern Sie sie in einer separaten Datei, und fügen Sie sie dann bei Bedarf mit den Cookies in der Datei `cookies.ini` zusammen. Nicht abgelaufene Cookies werden vom Crawler nicht gelöscht. Wenn jedoch eine gesamte Cookiegruppe aufgrund eines Problems nicht geschrieben werden kann, sollen die manuell angegebenen Cookies nicht verloren gehen. Sie müssen Ihre Cookies mit den Cookies, die der Crawler automatisch aufbewahrt, vor dem Start einer Crawlerinstanz zusammenfügen.

### Cookieformat

Cookies, die Sie mit der Datei `cookies.ini` der Unternehmenssuche zusammenfügen möchten, müssen ein bestimmtes Format aufweisen.

- Jedes Cookie muss in einer einzigen Zeile stehen. Leere Zeilen und Kommentare sind zulässig, sie werden jedoch in der Datei `cookies.ini` nicht beibehalten.
- Jedes Cookie muss das folgende Format aufweisen:

```
CookieN(cookielänge,URL-länge)cookietext,prüfungs-URL
```

#### Cookie

Ein erforderliches Schlüsselwort, das den Start eines Cookieeintrags angibt.

Das Cookieschlüsselwort darf keine Leerzeichen enthalten, und es muss eine einzelne Ziffer angehängt sein, entweder 0, 1 oder 2. Die Ziffer gibt den Cookietyp an: version-0 (Netscape), version-1 (RFC2109) oder version-2 (RFC2965). Portlisten werden in RFC2965-Cookies nicht unterstützt.

#### *cookielänge*

Die Länge in Zeichen des zugehörigen Cookietextes.

#### *URL-länge*

Die Länge in Zeichen der zugehörigen Prüfungs-URL.

#### *cookietext*

Der Inhalt des Cookies, das an den Ursprungs-Web-Server gesendet werden soll. Diese Zeichenfolge (die die rechte Seite der Anweisung zum Setzen eines Cookies in einem HTTP-Antwortheader darstellt) gibt das Name-Wert-Paar des Cookies und weiteren Inhalt an (wie z. B. Pfad, Sicherheitseinstellungen usw.), die mit dem Cookie gesendet werden sollen. Auf diese Zeichenfolge folgt ein Komma (,) als Trennzeichen.

#### *prüfungs-URL*

Die URL, unter der dieses Cookie aufgespürt wurde. Diese URL wird verwendet, um zu ermitteln, wohin das Cookie gesendet werden soll (z. B. indem ein Domänen- und Pfadname angegeben wird). Die Prüfungs-URL muss den Sicherheits- und Datenschutzbeschränkungen für Cookies des Ursprungs-Web-Servers entsprechen.

Das folgende Beispiel wird zur besseren Lesbarkeit in zwei Zeilen dargestellt; von Ihnen angegebene Cookies müssen in einer einzigen Zeile stehen:

```
Cookie0(53,40)ASPSESSIONIDQSQTACSD=SLNSIDFNLSIDNFLSNDNLSNL;path=/
https://www.ibm.com:443/help/solutions/
```

## **Konfigurieren von Cookies für den Web-Crawler**

Sie können Cookies für eine Web-Crawler-Sitzung manuell angeben und sie mit Cookies, die vom Web-Crawler verwaltet werden, zusammenfügen.

### **Vorbereitung**

Sie können nur als Administrator für die Unternehmenssuche Cookies für die Verwendung durch den Web-Crawler manuell konfigurieren.

### **Vorgehensweise**

Gehen Sie wie folgt vor, um Cookies für einen Web-Crawler zu konfigurieren:

1. Überwachen Sie von der Administrationskonsole für die Unternehmenssuche aus die Objektgruppe, für die Sie Cookies angeben möchten, und stoppen Sie den Web-Crawler.
2. Melden Sie sich als Administrator für die Unternehmenssuche am Crawler-Server an. (Diese Benutzer-ID wurde während der Installation von OmniFind Enterprise Edition angegeben.)
3. Wechseln Sie in das Datenverzeichnis für den Crawler, den Sie konfigurieren möchten, wobei *crawlersitzungs-ID* eine ID ist, die der Crawler-Sitzung vom System für die Unternehmenssuche zugeordnet wurde. Beispiel:  
ES\_NODE\_ROOT/data/col\_56092.WEB\_88534
4. Bearbeiten Sie die Datei *cookies.ini*, hängen Sie die manuell angegebenen Cookieeinträge an die bereits aufgelisteten an, und speichern und schließen Sie anschließend die Datei. Achten Sie darauf, dass Ihre Cookies keine bereits vorhandenen Cookies überschreiben.

5. Starten Sie den gestoppten Web-Crawler über die Administrationskonsole für die Unternehmenssuche erneut.

## Konfiguration eines globalen Web-Crawler-Bereichs

Sie können einen globalen Crawlerbereich für Web-Crawler konfigurieren, mit dem Sie das Entfernen von URLs aus dem Index besser steuern können.

Jeder Web-Crawler wird mit einem Crawlerbereich konfiguriert, in dem die URLs definiert sind, die durchsucht oder nicht durchsucht werden sollen. Aufgespürte URLs, die sich im Crawlerbereich befinden, werden für eine spätere Crawler suche (in einer Datenbank) aufbewahrt. URLs, die sich nicht im Crawlerbereich befinden, werden gelöscht. Wenn der Crawler mit einer leeren Datenbank beginnt, bleiben die Definition und die Datenbank des Crawlerbereichs konsistent, während der Crawler ausgeführt wird.

Es kommt vor, dass ein Crawler gestoppt wird und der Crawlerbereich verkleinert wird (z. B. aufgrund von neuen Regeln, die ein Durchsuchen von Seiten verbieten). Wenn der Crawler erneut gestartet wird, sind die Definition und die Datenbank des Crawlerbereichs nicht mehr konsistent. Die Datenbank enthält URLs (einige durchsuchte und einige nicht durchsuchte), die sich nicht im neuen, kleineren Crawlerbereich befinden.

Wenn eine Objektgruppe nur über einen Web-Crawler verfügt, kann der Web-Crawler die Konsistenz wiederherstellen, indem die HTTP-Statuscodes für diese URLs in 760 geändert werden (dadurch wird angegeben, dass sie ausgeschlossen werden sollen) und das Entfernen der nun ausgeschlossenen Seiten aus dem Index angefordert wird.

Wenn Sie den Crawlerbereich auf zwei oder mehr Web-Crawler aufteilen (z. B. um sicherzustellen, dass einige Seiten häufiger durchsucht werden als die übrigen), verwaltet jeder Web-Crawler unabhängige (ursprünglich leere) Datenbanktabellen und durchsucht jeweils einen anderen Teil des Web-Crawler-Bereichs. Der Crawlerbereich des ursprünglichen Crawlers wird dann auf den verbleibenden Teil verkleinert, nachdem die von anderen Crawlern zu durchsuchenden Teile entfernt wurden. Probleme treten auf, wenn der ursprüngliche Crawler versucht, die Konsistenz wiederherzustellen, indem die versetzten Seiten aus dem Index entfernt werden. Da die versetzten Seiten nun von anderen Crawlern durchsucht werden, sollten die Seiten im Index bleiben.

Durch die Konfiguration eines globalen Crawlerbereichs auf höherer Ebene können Sie URLs angeben, die von dem ursprünglichen Crawler nicht durchsucht werden sollen, aber trotzdem nicht aus dem Index entfernt werden sollen. URLs, die sich nicht mehr im Crawlerbereich eines Crawlers befinden, werden weiterhin für den Ausschluss durch die Discoveryprozesse markiert und werden aus dem Index entfernt, wenn sie erneut durchsucht werden.

Der globale Crawlerbereich wird durch die Konfigurationsdatei `global.rules` definiert, die sich im Konfigurationsverzeichnis des Crawlers befinden muss (durch die Datei `global.rules` wird die Funktion des globalen Crawlerbereichs aktiviert). Wenn diese Datei vorhanden ist, wird sie während der Initialisierung des Crawlers gelesen. Wenn diese Datei nicht vorhanden ist, verwendet der Crawler einen Crawlerbereich auf einer Ebene und entfernt gegebenenfalls Dokumente aus dem Index, um die Konsistenz zwischen der Definition und der Datenbank seines Crawlerbereichs zu bewahren.

Wenn ein globaler Crawlerbereich vorhanden ist, schließt der Crawler URLs wie bisher ein oder aus, fordert jedoch das Entfernen einer URL aus dem Index nur dann an, wenn sich die URL nicht in einem Web-Crawler-Bereich befindet.

Die Datei `global.rules` weist dieselbe Syntax auf wie die lokale Datei `crawl.rules`, mit dem Unterschied, dass sie nur Regeln für Domännennamen enthalten kann. Aufgrund dieser Einschränkung kann ein Crawlerbereich zwischen Crawlern nur auf der Basis von DNS-Hostnamen und nicht auf der Basis von IP-Adressen oder HTTP-Präfixmustern partitioniert werden. URLs, die durch Regeln für URL-Präfixe oder IP-Adressen im lokalen Crawlerbereich ausgeschlossen werden (gemäß der Definition in der Datei `crawl.rules`) sind von dem globalen Crawlerbereich nicht betroffen; diese URLs werden weiterhin ausgeschlossen.

Der globale Crawlerbereich wird nur verwendet, um zu verhindern, dass die URLs, die aufgrund einer lokalen Domänenregel aus dem Crawlerbereich des Crawlers ausgeschlossen werden, aus dem Index entfernt werden. Die folgenden Regeln gelten in der folgenden Reihenfolge:

1. Wenn eine URL aufgrund einer lokalen Präfix- oder Adressenregel aus der Datenbank des Crawlers ausgeschlossen wird, wird der URL der Statuscode 760 zugeordnet, und sie wird aus dem Index entfernt. Die URL wird nicht mehr durchsucht.
2. Wenn eine URL aufgrund einer lokalen Domänenregel aus der Datenbank des Crawlers ausgeschlossen wird und es keinen globalen Crawlerbereich gibt, wird der URL der Statuscode 760 zugeordnet, und sie wird aus dem Index entfernt. Die URL wird nicht mehr durchsucht.
3. Wenn eine URL aufgrund einer lokalen Domänenregel aus der Datenbank des Crawlers ausgeschlossen wird, jedoch explizit durch eine Regel im globalen Crawlerbereich zugelassen ist, wird der URL der Statuscode 761 zugeordnet. Der Crawler wird die URL nicht mehr durchsuchen. Die URL wird jedoch nicht aus dem Index entfernt (es wird davon ausgegangen, dass sie sich im lokalen Crawlerbereich eines anderen Crawlers befindet).
4. Wenn eine URL aufgrund einer lokalen Domänenregel aus der Datenbank des Crawlers ausgeschlossen wird und nicht explizit durch eine Regel im globalen Crawlerbereich zugelassen ist, wird der URL der Statuscode 760 zugeordnet, und sie wird aus dem Index entfernt.

Da der globale Crawlerbereich nur abgefragt wird, um zu verhindern, dass URLs, die bereits durch den lokalen Crawlerbereich ausgeschlossen wurden, gelöscht werden, lautet das Standardergebnis aus dem globalen Crawlerbereich, falls keine Regel für eine Kandidaten-URL gilt, dass diese URL nicht durchsucht werden darf.

Die Datei `global.rules`, muss sich im Verzeichnis `master_config` aller Crawler befinden, die einen globalen Crawlerbereich gemeinsam nutzen. Sie müssen sämtliche Kopien der Datei `global.rules` und der einzelnen lokalen `crawl.rules`-Dateien sorgfältig bearbeiten, um sicherzustellen, dass sie untereinander konsistent sind.

## Die Anweisungen 'no-follow' und 'no-index'

Sie können die Suchqualität verbessern, indem Sie Anweisungen für den Web-Crawler angeben, mit denen gesteuert wird, ob er Links auf Seiten folgen und Seiten indexieren soll.

Einige Webseiten verfügen über die Anweisungen `no-follow` oder `no-index`, die Robots (wie z. B. den Web-Crawler) anweisen, den Links auf diesen Seiten nicht zu folgen und/oder diese Inhalte auf diesen Seiten nicht in den Index einzuschließen.

Durch diese Einstellungen kann die Qualität der Crawlersuche verbessert werden. Einige Verzeichnisseiten können z. B. Tausende von Links, jedoch keinen anderen nützlichen Inhalt enthalten. Diese Seiten sollten durchsucht und ihren Links gefolgt werden, es ist jedoch nicht sinnvoll, die eigentlichen Verzeichnisseiten zu indexieren.

Es kann auch vorkommen, dass Sie nicht möchten, dass der Crawler tiefer in die Hierarchie eindringt, die gewünschten Blattseiten jedoch Links enthalten und keine Anweisungen no-follow. Da einige dieser Seiten automatisch generiert werden, haben sie keine Eigentümer, die die erforderlichen Anweisungen einfügen könnten.

Zur Angabe von Regeln für das Durchsuchen solcher Seiten erstellen oder bearbeiten Sie eine Konfigurationsdatei mit dem Namen `followindex.rules`. Beachten Sie bei der Angabe von Regeln in dieser Datei die folgenden Richtlinien:

- Die von Ihnen konfigurierten Regeln müssen URL-Präfixe angeben (Sie können Websites nicht durch IP-Adressen oder DNS-Hostnamen angeben).
- Die URL-Präfixe können Sterne (\*) als Platzhalterzeichen enthalten, um mehrere Sites mit ähnlichen URLs zuzulassen oder nicht zuzulassen.
- Die Reihenfolge ist wichtig (der Crawler wendet die erste Regel an, die der Kandidaten-URL entspricht).
- Die Regeln, die das Folgen von Links oder das Indexieren explizit zulassen oder nicht zulassen, setzen andere Einstellungen außer Kraft, einschließlich der Einstellungen im Zieldokument.

## Außerkräftsetzen der Anweisungen 'no-follow' und 'no-index' auf Webseiten

Sie können in einer Konfigurationsdatei Regeln angeben, die steuern, ob der Web-Crawler Links zu Seiten folgt oder Seiten indexiert, die die Anweisungen no-follow oder no-index enthalten.

### Vorbereitung

Sie können nur als Administrator für die Unternehmenssuche die Anweisungen no-follow und no-index für den Web-Crawler angeben. Die von Ihnen angegebenen Anweisungen setzen die Anweisungen außer Kraft, die auf den zu durchsuchenden Seiten vorhanden sind.

### Vorgehensweise

Gehen Sie wie folgt vor, um die Anweisungen no-follow und no-index außer Kraft zu setzen:

1. Überwachen Sie von der Administrationskonsole für die Unternehmenssuche aus die Objektgruppe, für die Sie Regeln konfigurieren möchten, und stoppen Sie den Web-Crawler.
2. Melden Sie sich als Administrator für die Unternehmenssuche am Crawler-Server an. Diese Benutzer-ID wurde bei der Installation von OmniFind Enterprise Edition angegeben.
3. Wechseln Sie in das Konfigurationsverzeichnis für den Crawler, den Sie konfigurieren möchten, wobei *crawlersitzungs-ID* eine ID ist, die der Crawler-sitzung vom System für die Unternehmenssuche zugeordnet wurde. Beispiel:  
`ES_NODE_ROOT/master_config/col_56092.WEB_88534`
4. Erstellen oder bearbeiten Sie die Datei `followindex.rules`.

- Geben Sie Regeln für den Crawler im folgenden Format ein, wobei *URLpräfix* die Anfangszeichen für die Websites darstellt, für die ein Folgen von Links oder ein Indexieren zugelassen werden soll oder nicht zugelassen werden soll:

```
forbid follow URLpräfix
allow follow URLpräfix
forbid index URLpräfix
allow index URLpräfix
```

- Speichern und schließen Sie die Datei.
- Starten Sie den gestoppten Web-Crawler über die Administrationskonsole für die Unternehmenssuche erneut.

## Konfigurieren des vom Web-Crawler für durchsuchte Dokumente verwendeten Datums

In einer Konfigurationsdatei können Sie eine Option angeben, die steuert, welches Datum der Web-Crawler als Datum eines durchsuchten Dokuments verwendet.

### Vorbereitung

Standardmäßig setzt der Web-Crawler den Wert des Datumsfelds in durchsuchten Dokumenten auf das Datum, an dem ein Dokument durchsucht wird. Sie können den Web-Crawler aber auch so konfigurieren, dass dieses Datum auf das Datum und die Uhrzeit der letzten Änderung gesetzt wird, die vom Web-Server zurückgegeben werden.

Das Datum der letzten Änderung ist möglicherweise nicht für alle Dokumente verfügbar. Wenn Sie den Crawler für die Verwendung dieses Werts konfigurieren, und der Web-Server kein Datum der letzten Änderung für ein Dokument zurückgibt, setzt der Crawler den Wert des Datumsfelds für die durchsuchten Dokumente auf das Datum und die Uhrzeit, an dem bzw. zu der das Dokument durchsucht wird.

*Tabelle 4. Vorgehensweise des Web-Crawlers beim Konfigurieren des Datumsfelds für durchsuchte Dokumente*

Crawler-konfiguration	Von einem Web-Server zurückgegebene Daten	Datumsmetadatenfeld für durchsuchte Dokumente
Standard-konfiguration	Ein Web-Server gibt einen Wert der letzten Änderung für ein durchsuchtes Dokument zurück.	Datum und Uhrzeit der Crawler-suche
	Ein Web-Server gibt keinen Wert der letzten Änderung für ein durchsuchtes Dokument zurück.	Datum und Uhrzeit der Crawler-suche
Konfiguriert für die Verwendung des Datums der letzten Änderung	Ein Web-Server gibt einen Wert der letzten Änderung für ein durchsuchtes Dokument zurück.	Vom Web-Server zurückgegebener Wert der letzten Änderung
	Ein Web-Server gibt keinen Wert der letzten Änderung für ein durchsuchtes Dokument zurück.	Datum und Uhrzeit der Crawler-suche

Sie können nur als Administrator für die Unternehmenssuche angeben, welches Datum der Web-Crawler für durchsuchte Dokumente verwenden soll.

### Vorgehensweise

Gehen Sie wie folgt vor, um einen Web-Crawler so zu konfigurieren, dass das Datum für durchsuchte Dokumente auf das von den Web-Servern zurückgegebene Datum und die Uhrzeit der letzten Änderung gesetzt wird:

1. Melden Sie sich als Administrator für die Unternehmenssuche am Crawler-Server an. Diese Benutzer-ID wurde bei der Installation von OmniFind Enterprise Edition angegeben.
2. Bearbeiten Sie die Datei `crawl.properties` für den Web-Crawler, den Sie konfigurieren wollen. Dabei ist `crawlersitzungs-id` eine ID, die der Crawlersitzung vom System für die Unternehmenssuche zugeordnet wurde.  
`ES_NODE_ROOT/master_config/crawlersitzungs-id/crawl.properties`

Beispiel:

```
/home/esadmin/master_config/co1_00112.WEB_23344/crawl.properties
C:\Programme\IBM\es\esadmin\master_config\co1_55667.WEB_78899\crawl.properties
```

3. Fügen Sie die folgende Zeile hinzu, und speichern Sie die Datei:  
`which_date=LastModified`
4. Starten Sie den Web-Crawler über die Administrationskonsole für die Unternehmenssuche erneut.

Wenn ein Web-Server das Datum der letzten Änderung für Dokumente zurückgibt, die von diesem Web-Crawler durchsucht werden, verwendet der Crawler das zurückgegebene Datum und die Uhrzeit als Datum der durchsuchten Dokumente.

---

## Web Content Management-Crawler

Wenn Sie IBM Workplace Web Content Management-Dokumente in eine Objektgruppe für die Unternehmenssuche einschließen möchten, müssen Sie einen Web Content Management-Crawler konfigurieren.

### WebSphere Portal-Serverkonfiguration

Wenn Sie Workplace Web Content Management auf einem WebSphere Portal-Server der Version 6 installieren, können Sie den Web Content Management-Crawler verwenden, um Web Content Management-Sites zu durchsuchen. Optionen für die Crawlersuche auf diesen Sites können Sie getrennt von den Optionen konfigurieren, die Sie für Portalsites angeben, die von einem WebSphere Portal-Crawler durchsucht werden.

Bevor Sie einen Web Content Management-Crawler erstellen, müssen Sie die Prozeduren für die Einrichtung der Unternehmenssuche in WebSphere Portal befolgen. Sie führen ein Script (`wp6_install.sh` unter AIX, Linux oder Solaris bzw. `wp6_install.bat` unter Windows), das mit OmniFind Enterprise Edition bereitgestellt wird, auf den Suchservern aus, um die Umgebung für die Unternehmenssuche zu konfigurieren.

### URL-Format

Wenn Sie die zu durchsuchenden URL-Adressen angeben, müssen Sie das folgende Format verwenden:

```
http-protokoll://portalhostname:portnummer/portalpräfix
/wcm-servletpfad_sucheinstiegspunkt/searchseed?site=wcm-sitenam&lib=wcm-bibliotheksname
```

Im folgenden Beispiel sehen Sie die URL-Adresse für eine Site im Standardinstallationspfad von Workplace Web Content Management unter WebSphere Portal:

[http://portal.server.ibm.com:80/wps/wcmsearchseed/  
searchseed?site=SiteTest01&lib=Web+Content](http://portal.server.ibm.com:80/wps/wcmsearchseed/searchseed?site=SiteTest01&lib=Web+Content)

Wenn der Sitename oder der Bibliotheksname Leerzeichen enthält, müssen Sie diese durch ein Pluszeichen (+) ersetzen. Ersetzen Sie z. B. Web Content durch Web+Content.

## Konfigurationsüberblick

Mithilfe des Web Content Management-Crawlers können Sie eine beliebige Anzahl Web Content Management-Sites durchsuchen. Wenn Sie den Crawler konfigurieren, geben Sie die URL-Adressen für die zu durchsuchenden Sites an. Der Crawler lädt dann die Seiten herunter, die über die angegebenen Sites verknüpft sind.

Die zu durchsuchenden Sites müssen über dieselbe WebSphere Portal-Administrator-ID und das zugehörige Kennwort zugänglich sein. Wenn Sie Sites durchsuchen wollen, die unterschiedliche Berechtigungsnachweise verwenden, müssen Sie einen separaten Web Content Management-Crawler konfigurieren.

Wenn Sie einen Web Content Management-Crawler erstellen oder ändern möchten, müssen Sie sich an der Administrationskonsole für die Unternehmenssuche anmelden. Sie müssen die Rolle eines Administrators für die Unternehmenssuche haben oder ein Objektgruppenadministrator für die Objektgruppe sein, zu der der Crawler gehört.

Wenn Sie den Crawler erstellen, unterstützt Sie ein Assistent bei den folgenden Tasks:

- Angeben von Merkmalen, die steuern, wie der Crawler ausgeführt wird und Systemressourcen verwendet. Die Crawlermerkmale steuern, wie der Crawler alle Dokumente im Crawlerbereich durchsucht.
- Angeben der URL-Adressen, für die zu durchsuchenden Sites und der Informationen, die der Crawler benötigt, um eine Verbindung zu den Sites herzustellen. Wenn Sie den Crawler erstellen oder bearbeiten, können Sie testen, ob er eine Verbindung zu den zu durchsuchenden URL-Adressen herstellen kann. Nachrichten zeigen Ihnen an, ob der Crawler auf die zu durchsuchenden Dokumente zugreifen kann, bevor Sie den Crawler starten.
- Angeben von Sicherheitsoptionen auf Dokumentenebene. Wenn die Sicherheit bei der Erstellung der Objektgruppe aktiviert war, kann der Crawler Sicherheitsdaten Dokumenten im Index zuordnen. Mithilfe dieser Daten können Zugriffssteuerungen auf Basis der gespeicherten Zugriffssteuerungslisten oder Sicherheitstoken erzwungen werden.

Sie können außerdem eine Option auswählen, um Berechtigungsnachweise eines Benutzers bei der Übergabe einer Abfrage zu prüfen. In diesem Fall vergleicht das System die Berechtigungsnachweise des Benutzers nicht mit indexierten Sicherheitsdaten, sondern mit den aktuellen Zugriffssteuerungslisten, die von den ursprünglichen Datenquellen verwaltet werden.

**Wichtig:** Wenn Sie sichere Web Content Management-Seiten durchsuchen wollen, müssen Sie Suchen unter Verwendung des Suchportlets für die Unternehmenssuche über WebSphere Portal übergeben. Suchen, die über die Muster-suchanwendung (ESSearchApplication) übergeben werden, verfügen nicht über die geeigneten Berechtigungsnachweise und können die Berechtigung des Benutzers für den Dokumentenzugriff nicht überprüfen.

- Angeben von Informationen, die dem Crawler die Kommunikation mit einem Proxy-Server ermöglichen, wenn die Web Content Management-Sites für das Bereitstellen von Dokumenten einen Proxy-Server verwenden.
- Angeben von SSO-Berechtigungs nachweisen (Single Sign-on), mit denen der Crawler auf Dokumente auf dem Server zugreifen kann, wenn Sie ein anderes Produkt (wie z. B. IBM Tivoli Access Manager WebSEAL oder CA SiteMinder SSO Agent for PeopleSoft) verwenden, um Ihren WebSphere Portal-Server und Ihre Websites zu schützen.
- Angeben von Informationen zu einer Schlüsselspeicherdatei, damit der Crawler das Protokoll SSL (Secure Sockets Layer) verwenden kann, um eine Verbindung zu den Web Content Management-Sites herzustellen.
- Angeben der Sprache und Codepage der zu durchsuchenden Dokumente.
- Angeben von Optionen für die Crawlersuche und die Suche in Metadaten in Web Content Management-Dokumenten.
- Angeben von Zeitplänen für die Crawlersuche auf Web Content Management-Sites.

Klicken Sie **Hilfe** an, während Sie den Crawler erstellen, um Informationen zu den Feldern im Assistenten und zur Vorgehensweise abzurufen, wie Sie dem Crawler die für das Durchsuchen der Daten erforderlichen Informationen zur Verfügung stellen.

#### **Zugehörige Konzepte**

„Integration in WebSphere Portal“ auf Seite 373

#### **Zugehörige Tasks**

„Konfigurieren der Unternehmenssuche in WebSphere Portal Version 6“ auf Seite 381

---

## **WebSphere Portal-Crawler**

Wenn Sie Seiten einer IBM WebSphere Portal-Site in einer Objektgruppe für die Unternehmenssuche einschließen möchten, müssen Sie einen WebSphere Portal-Crawler konfigurieren.

### **WebSphere Portal-Serverkonfiguration**

Bevor Sie einen WebSphere Portal-Crawler erstellen, müssen Sie ein Script ausführen, um die Umgebung für die Unternehmenssuche in WebSphere Portal zu konfigurieren. Für verschiedene Versionen von WebSphere Portal sind verschiedene Scripts erforderlich. Die Scripts werden bei der Installation von OmniFind Enterprise Edition auf den Suchservern installiert.

- Für WebSphere Portal Version 5.1.0 oder später führen Sie das Script `wp5_install.sh` auf AIX-, Linux- und Solaris-Systemen bzw. das Script `wp5_install.bat` auf einem Windows-System aus.
- Für WebSphere Portal Version 6 führen Sie das Script `wp6_install.sh` auf AIX-, Linux- und Solaris-Systemen bzw. das Script `wp6_install.bat` auf einem Windows-System aus.

#### **Tipp:**

Detaillierte Beispiele für die Konfiguration eines sicheren WebSphere Portal-Crawlers finden Sie im Szenario für ein mittleres Unternehmen im IBM Redbook *IBM OmniFind Enterprise Edition Configuration and Implementation Scenarios*.

## Konfigurationsüberblick

Mit dem WebSphere Portal-Crawler können Sie eine einzelne WebSphere Portal-Site durchsuchen. Die URL-Adresse für die zu durchsuchende Portalsite geben Sie bei der Konfiguration des Crawlers an. Der Crawler lädt dann die Portlets und Seiten herunter, die über die angegebene Portal-URL-Adresse verknüpft sind. Wenn Sie eine weitere Portal-Site durchsuchen wollen, erstellen Sie einen weiteren Crawler.

Wenn Sie einen WebSphere Portal-Crawler erstellen oder ändern möchten, müssen Sie sich an der Administrationskonsole für die Unternehmenssuche anmelden. Sie müssen die Rolle eines Administrators für die Unternehmenssuche haben oder ein Objektgruppenadministrator für die Objektgruppe sein, zu der der Crawler gehört.

Wenn Sie den Crawler erstellen, unterstützt Sie ein Assistent bei den folgenden Tasks:

- Angeben von Merkmalen, die steuern, wie der Crawler ausgeführt wird und Systemressourcen verwendet. Die Crawlermerkmale steuern, wie der Crawler sämtliche Seiten auf der Site durchsucht.
- Angeben der URL-Adresse, für die zu durchsuchende Portalsite und der Informationen, die der Crawler benötigt, um eine Verbindung zu der Site herzustellen. Da diese URL-Adressentypen lang sein und zudem codierte Nicht-ASCII-Zeichen enthalten können, ist es empfehlenswert, die URL-Adresse vom WebSphere Portal-Server zu kopieren und in die Administrationskonsole für die Unternehmenssuche einzufügen.

Wenn Sie den Crawler erstellen oder bearbeiten, können Sie testen, ob der Crawler eine Verbindung zu der zu durchsuchenden URL herstellen kann. Nachrichten zeigen Ihnen an, ob der Crawler auf die zu durchsuchenden Dokumente zugreifen kann, bevor Sie den Crawler starten.

- Angeben von Sicherheitsoptionen auf Dokumentenebene. Wenn die Sicherheit bei der Erstellung der Objektgruppe aktiviert war, kann der Crawler Sicherheitsdaten Dokumenten im Index zuordnen. Mithilfe dieser Daten können Zugriffssteuerungen auf Basis der gespeicherten Zugriffssteuerungslisten oder Sicherheitstoken erzwungen werden.

Sie können außerdem eine Option auswählen, um Berechtigungsnachweise eines Benutzers bei der Übergabe einer Abfrage zu prüfen. In diesem Fall vergleicht das System die Berechtigungsnachweise des Benutzers nicht mit indexierten Sicherheitsdaten, sondern mit den aktuellen Zugriffssteuerungslisten, die von den ursprünglichen Datenquellen verwaltet werden.

**Wichtig:** Wenn Sie sichere WebSphere Portal-Seiten durchsuchen wollen, müssen Sie Suchen unter Verwendung des Suchportlets für die Unternehmenssuche über WebSphere Portal übergeben. Suchen, die über die Mustersuchanwendung (ESSearchApplication) übergeben werden, verfügen nicht über die geeigneten Berechtigungsnachweise und können die Berechtigung des Benutzers für den Dokumentenzugriff nicht überprüfen.

- Angeben von Informationen, die dem Crawler die Kommunikation mit einem Proxy-Server ermöglichen, wenn die WebSphere Portal-Site für das Bereitstellen von Seiten einen Proxy-Server verwendet.
- Angeben von SSO-Berechtigungsnachweisen (Single Sign-on), mit denen der Crawler auf Dokumente auf dem Server zugreifen kann, wenn Sie ein anderes Produkt (wie z. B. IBM Tivoli Access Manager WebSEAL oder CA SiteMinder SSO Agent for PeopleSoft) verwenden, um Ihren WebSphere Portal-Server und Ihre Websites zu schützen.

- Angeben von Informationen zu einer Schlüsselspeicherdatei, damit der Crawler das Protokoll SSL (Secure Sockets Layer) verwenden kann, um eine Verbindung zur WebSphere Portal-Site herzustellen.
- Angeben der Sprache und Codepage der zu durchsuchenden Dokumente.
- Angeben von Optionen für die Crawlersuche und die Suche in Metadaten in WebSphere Portal-Dokumenten.
- Angeben von Zeitplänen für die Crawlersuche in WebSphere Portal-Dokumenten.

Klicken Sie **Hilfe** an, während Sie den Crawler erstellen, um Informationen zu den Feldern im Assistenten und zur Vorgehensweise abzurufen, wie Sie dem Crawler die für das Durchsuchen der Daten erforderlichen Informationen zur Verfügung stellen.

#### Zugehörige Konzepte

„Integration in WebSphere Portal“ auf Seite 373

#### Zugehörige Tasks

„Kopieren der zu durchsuchenden URL-Adresse von WebSphere Portal“

„Konfigurieren der Unternehmenssuche in WebSphere Portal Version 6“ auf Seite 381

„Konfigurieren der Unternehmenssuche in WebSphere Portal Version 5.1“ auf Seite 375

## Kopieren der zu durchsuchenden URL-Adresse von WebSphere Portal

Sie können Fehler bei der Eingabe einer URL-Adresse vermeiden, indem Sie bei der Konfiguration eines WebSphere Portal-Crawlers oder eines Web Content Management-Crawlers die URL-Adresse der zu durchsuchenden Site kopieren und in das entsprechende Feld einfügen.

### Informationen zu dieser Task

Beim Erstellen eines WebSphere Portal-Crawlers oder eines Web Content Management-Crawlers geben Sie die URL-Adresse der Site auf dem WebSphere Portal-Server an, den Sie durchsuchen möchten. Da die URL-Adressen lang sind und in der Regel codierte Nicht-ASCII-Zeichen enthalten, ist es empfehlenswert, die URL-Adresse anhand folgender Vorgehensweise vom WebSphere Portal-Server zu kopieren und in die Administrationskonsole für die Unternehmenssuche einzufügen.

### Vorgehensweise

Gehen Sie wie folgt vor, um die URL-Adresse anzugeben, die der Crawler durchsuchen soll:

1. Wenn Sie bereit sind, die zu durchsuchende URL-Adresse bzw. die zu durchsuchenden URL-Adressen in der Administrationskonsole für die Unternehmenssuche anzugeben, stellen Sie sicher, dass der WebSphere Portal-Server gestartet ist, und melden Sie sich dann bei WebSphere Portal als Administrator an.
2. Wenn Sie WebSphere Portal Version 5.1 verwenden, führen Sie die folgenden Schritte auf dem WebSphere Portal-Server aus:
  - a. Klicken Sie **Verwaltung** in der rechten oberen Ecke an.
  - b. Klicken Sie zuerst **Portaleinstellungen** (im Navigationsbereich auf der linken Seite) und dann **Verwaltung der Suchfunktion** an.

- c. Klicken Sie **PortalCollection** im Bereich **Suchsammlungen** der Seite **Suchsammlungen verwalten** an. Sie können auch eine andere Objektgruppe auswählen, sofern welche zur Verfügung stehen.
  - d. Klicken Sie **Inhaltsquelle hinzufügen** im Bereich **Inhaltsquellen in der Objektgruppe** an.
  - e. Klicken Sie **Portalsite** für die Option **Quellentyp durchlaufen und erfassen** an. Die URL-Adresse der Site wird im Feld **Über diese URL verknüpfte Dokumente abrufen** angezeigt.
  - f. Kopieren Sie die URL-Adresse in die Zwischenablage. Heben Sie beispielsweise die URL-Adresse hervor, und halten Sie beim Betätigen der Einfügetaste die Steuertaste gedrückt.
3. Wenn Sie WebSphere Portal Version 6 verwenden, führen Sie die folgenden Schritte auf dem WebSphere Portal-Server aus:
    - a. Klicken Sie **Verwaltung** in der linken unteren Ecke an.
    - b. Klicken Sie **Suchläufe verwalten** im Navigationsbereich auf der linken Seite an.
    - c. Klicken Sie **Suchsammlungen** auf der Seite **Suchläufe verwalten** an.
    - d. Klicken Sie **Standardsuchservice des Portals** in den Suchserviceoptionen auf der Seite **Suchläufe verwalten** an.
    - e. Klicken Sie die Sammlung **Portalinhalte** in der Tabelle **Suchsammlungen** an.
    - f. Klicken Sie in der Tabelle **Inhaltsquellen** das Symbol für das Editieren (Stiftsymbol) auf der rechten Seite neben **Portalinhaltsquelle** an.
    - g. Klicken Sie für **Inhaltsquellentyp** die Option **Portalsite** an. Die URL-Adresse der Site wird im Feld **Über diese URL verknüpfte Dokumente abrufen** angezeigt.
    - h. Kopieren Sie die URL-Adresse in die Zwischenablage. Heben Sie beispielsweise die URL-Adresse hervor, und halten Sie beim Betätigen der Einfügetaste die Steuertaste gedrückt.
  4. Kehren Sie zur Administrationskonsole für die Unternehmenssuche zurück, und fügen Sie die kopierte URL-Adresse in das Feld für die URL-Adresse der Site ein.

#### Zugehörige Konzepte

„WebSphere Portal-Crawler“ auf Seite 120

---

## Crawler für das Windows-Dateisystem

Wenn Sie Dokumente einschließen möchten, die in einem Microsoft Windows-Dateisystem in einer Objektgruppe für die Unternehmenssuche gespeichert sind, müssen Sie einen Crawler für das Windows-Dateisystem konfigurieren.

Mit dem Crawler für das Windows-Dateisystem können Sie eine beliebige Anzahl Windows-Dateisysteme durchsuchen. Wenn Sie den Crawler konfigurieren, wählen Sie die lokalen und fernen Verzeichnisse und Unterverzeichnisse aus, die Sie durchsuchen wollen.

Wenn Sie den Crawler-Server auf AIX-, Linux- oder Solaris-Systemen installieren, können Sie diesen Server nicht zum Durchsuchen von Quellen in Windows-Dateisystemen verwenden (der Crawler für das Windows-Dateisystem wird in der Liste der verfügbaren Crawlertypen nicht angezeigt).

#### Tipp:

Detaillierte Beispiele für die Konfiguration eines sicheren Crawlers für das Windows-Dateisystem finden Sie im Szenario für ein kleines Unternehmen im IBM Redbook *IBM OmniFind Enterprise Edition Configuration and Implementation Scenarios*.

## Durchsuchen gemeinsam genutzter Netzverzeichnisse

Der Crawler für Windows-Dateisysteme durchsucht Dokumente entsprechend den Leseberechtigungen, die für den Administrator für die Unternehmenssuche angegeben sind. Der Administrator ist das Konto der OmniFind Enterprise Edition-Services.

Sie können eine Benutzer-ID und ein Kennwort für die zu durchsuchenden Verzeichnisse angeben. Die Benutzer-ID und das Kennwort werden jedoch nur zum Herstellen einer Verbindung zu den gemeinsam genutzten Netzverzeichnissen verwendet. Der Crawler durchsucht Dateien gemäß den Leseberechtigungen, die für diesen Benutzer für die gemeinsam genutzten Netzverzeichnisse (nicht für lokale Laufwerke) festgelegt sind.

Verbindungen zu Netzverzeichnissen werden erst getrennt, wenn Sie den OmniFind Enterprise Edition-Service erneut starten. Nach dem Herstellen einer Verbindung ist es möglich, mit einer nicht korrekten Benutzer-ID und einem Kennwort auf das Verzeichnis zuzugreifen. Diese Verbindung ist jedoch nur für Windows-Dateisystem-Erkennungs- und Crawler Sitzungen zulässig, die vom System für die Unternehmenssuche gesteuert werden. Stellen Sie sicher, dass die Berechtigungen für das Konto des Administrators für die Unternehmenssuche (unter dem der OmniFind Enterprise Edition-Service ausgeführt wird) korrekt konfiguriert sind, um mögliche Sicherheitsrisiken zu vermeiden.

Geben Sie immer denselben Benutzernamen und dasselbe Kennwort für dasselbe Netzverzeichnis an, um zukünftig Probleme beim Herstellen einer Verbindung zu einem Netzverzeichnis zu vermeiden. Wenn Sie die falsche Benutzer-ID und das falsche Kennwort angeben und den OmniFind Enterprise Edition-Service erneut starten, schlägt das Durchsuchen mit dem Crawler für das Windows-Dateisystem möglicherweise fehl, weil der Crawler versucht, mit den falschen Berechtigungsnachweisen eine Verbindung zu dem Verzeichnis herzustellen. Die Crawler suchung kann erfolgreich sein, wenn die Netzverbindung von einem anderen Crawler für Windows-Dateisysteme hergestellt wird, der die korrekte Benutzer-ID und das korrekte Kennwort verwendet.

## Konfigurationsüberblick

Wenn Sie einen Crawler für Windows-Dateisysteme erstellen oder ändern möchten, müssen Sie sich an der Administrationskonsole für die Unternehmenssuche anmelden. Sie müssen die Rolle eines Administrators für die Unternehmenssuche haben oder ein Objektgruppenadministrator für die Objektgruppe sein, zu der der Crawler gehört.

Wenn Sie den Crawler erstellen, unterstützt Sie ein Assistent bei den folgenden Tasks:

- Angeben von Merkmalen, die steuern, wie der Crawler ausgeführt wird und Systemressourcen verwendet. Die Crawlermerkmale steuern, wie der Crawler alle Unterverzeichnisse im Crawlerbereich durchsucht.
- Definieren eines Zeitplans für das Durchsuchen der Dateisysteme.
- Auswählen der zu durchsuchenden Unterverzeichnisse.

Sie können angeben, wie viele Unterverzeichnisebenen der Crawler durchsuchen soll. Sie geben außerdem eine Benutzer-ID und ein Kennwort an, mit denen der Crawler auf Daten zugreifen kann, um ferne Dateisysteme zu durchsuchen.

- Angeben von Optionen für das Aktivieren der Dokumente für Suchvorgänge in Unterverzeichnissen. Sie können z. B. bestimmte Dokumenttypen aus dem Crawlerbereich ausschließen oder eine Benutzer-ID und ein Kennwort angeben, mit deren Hilfe der Crawler auf Dateien in einem bestimmten Unterverzeichnis zugreifen kann.
- Konfigurieren von Optionen für die Sicherheit auf Dokumentebene. Wenn die Sicherheit bei der Erstellung der Objektgruppe aktiviert war, kann der Crawler Sicherheitsdaten Dokumenten im Index zuordnen. Mithilfe dieser Daten können Suchanwendungen die Zugriffssteuerung auf Basis der gespeicherten Zugriffssteuerungslisten oder Sicherheitstoken erzwingen.

Sie können außerdem eine Option auswählen, um Berechtigungsnachweise eines Benutzers bei der Übergabe einer Abfrage zu prüfen. In diesem Fall vergleicht das System die Berechtigungsnachweise des Benutzers nicht mit indextierten Sicherheitsdaten, sondern mit den aktuellen Zugriffssteuerungslisten, die von den ursprünglichen Datenquellen verwaltet werden.

Zur Umsetzung der Sicherheit auf Dokumentebene müssen Sie sicherstellen, dass die Benutzer- und Domänenkontoinformationen auf dem Crawler-Server ordnungsgemäß konfiguriert sind.

Klicken Sie **Hilfe** an, während Sie den Crawler erstellen, um Informationen zu den Feldern im Assistenten und zur Vorgehensweise abzurufen, wie Sie dem Crawler die für das Durchsuchen der Daten erforderlichen Informationen zur Verfügung stellen.

#### **Zugehörige Konzepte**

„Umsetzung der Sicherheit auf Dokumentebene für Dokumente des Windows-Dateisystems“ auf Seite 311

„Sicheres Suchen in vertrauenswürdigen Windows-Domänen“ auf Seite 313

---

## **Konfigurieren der Unterstützung für Anwendungen der Datenempfangsfunktion**

Sie können die Unternehmenssuche erweitern, indem Sie die API der Datenempfangsfunktion verwenden, um einen externen Crawler zu erstellen. Ihre benutzerdefinierten Anwendungen der Datenempfangsfunktion können einer Objektgruppe Daten hinzufügen, Daten aus einer Objektgruppe entfernen oder einen Web-Crawler anweisen, URLs zu besuchen und nochmals zu bearbeiten.

### **Vorbereitung**

Sie müssen ein Administrator für die Unternehmenssuche sein, um Anwendungen der Datenempfangsfunktion zu konfigurieren.

**Wichtig:** Die Datenempfangsfunktion wird in zukünftigen Releases nicht unterstützt. Verwenden Sie SIAPI (Search and Index API) anstelle der Anwendungsprogrammierschnittstellen der Datenempfangsfunktion, um Clientanwendungen für die Unternehmenssuche zu entwickeln. Die folgenden Informationen werden für Benutzer bereitgestellt, die vorher Anwendungen der Datenempfangsfunktion erstellt haben.

### **Informationen zu dieser Task**

Eine Clientanwendung für die Datenempfangsfunktion aktiviert die Crawlersuche für Datenquellentypen, die von den Standardcrawlern für die Unternehmenssuche nicht durchsucht werden können. Bevor Sie eine Anwendung der Datenempfangsfunktion verwenden können, müssen Sie Berechtigungsnachweise konfigurieren, die es der Anwendung ermöglichen, auf Objektgruppen zuzugreifen und diese zu aktualisieren.

Wenn Ihre Clientanwendung für die Datenempfangsfunktion eine Verbindung zur Datenempfangsfunktion herstellt, muss sie die ID und das Kennwort der Clientanwendung und die ID der Objektgruppe übermitteln, die aktualisiert werden soll. Diese Angaben müssen mit den Angaben übereinstimmen, die Sie für die Anwendung in der Administrationskonsole konfigurieren.

Die Datenempfangsfunktion wird automatisch gestartet, wenn das System für die Unternehmenssuche gestartet wird. Wenn Sie die Portnummer ändern, nachdem Sie die Anwendung in der Administrationskonsole konfiguriert haben, müssen Sie die Datenempfangsfunktion erneut starten.

### Vorgehensweise

Gehen Sie wie folgt vor, um Anwendungen der Datenempfangsfunktion zu konfigurieren:

1. Klicken Sie **System** an, um die Sicht **System** zu öffnen.
2. Klicken Sie  **Bearbeiten** an, um zur Sicht für die Systembearbeitung zu wechseln.
3. Klicken Sie **Anwendungen der Datenempfangsfunktion konfigurieren** auf der Seite **Datenempfangsfunktion** an.
4. Geben Sie auf der Seite **Anwendungen der Datenempfangsfunktion** die Anzahl Threads, die die Datenempfangsfunktion für Verarbeitungsanforderungen von Clientanwendungen erstellen kann, sowie die Portnummer an, für die die Datenempfangsfunktion für Anforderungen empfangsbereit ist. Geben Sie außerdem die maximale Anzahl Dokumente pro Objektgruppe an, die im temporären Speicher gespeichert werden können, bis der Parser sie syntaktisch analysiert.
5. Klicken Sie **Anwendung der Datenempfangsfunktion hinzufügen** an, um Informationen zu einer Clientanwendung hinzuzufügen.
6. Geben Sie auf der Seite **Anwendung der Datenempfangsfunktion hinzufügen** die Authentifizierungsinformationen an, die es Ihren Clientanwendungen für die Datenempfangsfunktion ermöglichen, auf Objektgruppen für die Unternehmenssuche zuzugreifen. Die Client-IDs der Datenempfangsfunktion müssen im System für die Unternehmenssuche eindeutig sein.
7. Wählen Sie die Objektgruppen aus, die die Anwendung für die Datenempfangsfunktion aktualisieren soll:
  - Klicken Sie **Alle Objektgruppen** an, wenn die Anwendung alle Objektgruppen aktualisieren soll.
  - Klicken Sie **Bestimmte Objektgruppen** an, wenn die Anwendung nur die von Ihnen angegebenen Objektgruppen aktualisieren soll.  
Wenn Sie diese Option auswählen, wird eine Liste der Objektgruppennamen angezeigt. Wählen Sie das Markierungsfeld **Auswählen** für jede Objektgruppe aus, die die Anwendung aktualisieren soll.
8. Klicken Sie **OK** an.

9. Wenn Sie die Portnummer der Datenempfangsfunktion oder die Anzahl Dokumente geändert haben, die im temporären Speicher gespeichert werden können, starten Sie die Datenempfangsfunktion erneut:
    - a. Klicken Sie  **Überwachen** an, um zur Sicht für die Systemüberwachung zu wechseln.
    - b. Klicken Sie  **Erneut starten** auf der Seite **Datenempfangsfunktion** an.
- Zugehörige Tasks**  
„Überwachen der Datenempfangsfunktion“ auf Seite 346

---

## Benutzerdefinierte Crawler-Plug-ins

Bei der Konfiguration von Crawlermerkmalen können Sie eine Java-Klasse angeben, die verwendet werden soll, um eine Zugriffssteuerung auf Dokumentenebene umzusetzen. Sie können die Java-Klasse außerdem verwenden, um den Index durch Hinzufügen, Modifizieren oder Entfernen von Metadaten und Dokumentinhalt zu aktualisieren. Durch Schreiben eines Plug-ins können Sie außerdem die Möglichkeiten des Crawlers beim Durchsuchen von Archivdateien erweitern.

Plug-ins enthalten eine Java-Klasse, die für jedes vom Crawler durchsuchte Dokument aufgerufen wird. An die Java-Klasse werden die Dokumentkennung (URI) aus dem Index für die Unternehmenssuche, Sicherheitstoken, Metadaten und Dokumentinhalt übergeben. Die Klasse kann neue oder modifizierte Sicherheitstoken, Metadaten und Inhalt zurückgeben oder Sicherheitstoken, Metadaten und neuen oder modifizierten Inhalt entfernen.

Nachdem alle Dokumente im Crawlerbereich einmal durchsucht wurden, wird das Plug-in nur noch bei neuen oder geänderten Dokumenten aufgerufen. Starten Sie eine vollständige Crawlersuche für alle Dokumente im Crawlerbereich, und erstellen Sie den Index erneut, um die Sicherheitstoken, Metadaten und Inhalte für Dokumente zu ändern, die sich im Index für die Unternehmenssuche befinden, aber in der ursprünglichen Datenquelle nicht aktualisiert wurden.

### Plug-ins zur Umsetzung von Sicherheit

Die Sicherheit auf Dokumentenebene wird umgesetzt, indem jedem Dokument, das von einem Crawler durchsucht wird, mindestens ein Sicherheitstoken (eine durch Kommata begrenzte Zeichenfolge) zugeordnet wird. Als Sicherheitstoken werden in der Regel Gruppen-IDs verwendet.

Standardmäßig wird jedem Dokument ein allgemein zugängliches Token zugeordnet, wodurch das Dokument für jeden zugänglich wird. Das allgemein zugängliche Token kann durch einen Wert ersetzt werden, der vom Administrator bereitgestellt oder der einem Feld im durchsuchten Dokument entnommen wird.

Das Plug-in ermöglicht Ihnen die Anwendung Ihrer eigenen Geschäftsregeln, um den Wert der Sicherheitstoken für durchsuchte Dokumente zu bestimmen. Die jedem Dokument zugeordneten Sicherheitstoken werden im Index gespeichert. Sie werden verwendet, um Dokumente zu filtern, die mit den Sicherheitstoken übereinstimmen, und um sicherzustellen, dass in den Suchergebnissen nur Dokumente zurückgegeben werden, für die der Benutzer eine Anzeigeberechtigung hat.

## Plug-ins zum Hinzufügen, Modifizieren oder Entfernen von Metadaten

Dokumentmetadaten (z. B. das Datum der letzten Änderung eines Dokuments) werden für alle durchsuchten Dokumente erstellt. Das Crawler-Plug-in ermöglicht Ihnen die Anwendung Ihrer eigenen Geschäftsregeln, um den Wert der Metadaten zu bestimmen, die für jedes Dokument indexiert werden sollen.

Die Metadaten werden als Name/Wert-Paar erstellt. Der Benutzer kann die Metadaten über eine Abfrage mit freiem Text oder eine Abfrage durchsuchen, in der der Name des Metadatenfelds angegeben ist.

## Plug-ins zum Hinzufügen, Modifizieren oder Entfernen von Dokumentinhalt

Dokumentinhalt umfasst die Teile eines Dokuments, die durchsuchbaren Inhalt sowie Inhalt enthalten, der Teil der dynamischen Dokumentzusammenfassung in den Suchergebnissen werden kann. Das Crawler-Plug-in ermöglicht Ihnen die Anwendung Ihrer eigenen Geschäftsregeln, um den Inhalt zu bestimmen, der für jedes Dokument indexiert werden soll.

## Web-Crawler-Plug-ins

Mit den Anwendungsprogrammierschnittstellen für den Web-Crawler lässt sich steuern, wie Dokumente durchsucht und für die Syntaxanalyse vorbereitet werden. Sie können beispielsweise dem HTTP-Anforderungsheader Felder hinzufügen, der verwendet wird, wenn der Crawler ein Dokument anfordert. Nachdem ein Dokument durchsucht, aber noch nicht syntaktisch analysiert und aufbereitet wurde, haben Sie die Möglichkeit, den Inhalt sowie die Sicherheitstoken und Metadaten zu ändern. Sie können außerdem verhindern, dass das Dokument an den Parser gesendet wird.

Eine Diskussion über formularbasierte Authentifizierung sowie ein Musterprogramm, das Sie für Ihr benutzerdefiniertes Web-Crawler-Plug-in anpassen können, finden Sie in <http://www.ibm.com/developerworks/db2/library/techarticle/dm-0707nishitani>.

## Plug-ins für Archivdateien

Durch Schreiben eines Plug-ins können Sie die Crawler erweitern und die Unterstützung für das Durchsuchen von Archivdateiformaten in anderen Formaten als ZIP und TAR aktivieren. Sie können beispielsweise ein Plug-in schreiben, das die Crawleruche für Dokumente im LZH-Format unterstützt.

## Nicht abgeschirmter Modus

Wenn Sie einen Nicht-Web-Crawler konfigurieren, können Sie eine Option auswählen, um das Plug-in in einem nicht abgeschirmten Modus auszuführen. In diesem Modus wird der Plug-in-Prozess im Crawlerprozess ausgeführt, was zu einer Verbesserung der Plug-in-Leistung führen kann.

**Wichtig:** Wenn das Plug-in auf ein Problem stößt, das in diesem Modus nicht behebbbar ist, wird der Crawlerprozess möglicherweise beendet.

### Zugehörige Konzepte

 Crawler-Plug-ins

---

## Unterstützung für die Crawlersuche in Archivdateien

Die Crawler für die Unternehmenssuche können Dateien aus einer Archivdatei (wie z. B. einer ZIP- oder TAR-Datei) extrahieren, sodass einzelne Dateien im Archiv indexiert und durchsucht werden können.

### Unterstützte Archivdateiformate

Die folgenden Archivdateiformate werden unterstützt:

*Tabelle 5. Von Crawlern für die Unternehmenssuche unterstützte Archivdateiformate*

Dateierweiterung	MIME-Typ	Datentyp	Anmerkungen
.zip, .ZIP	anwendung/zip	zip	<ul style="list-style-type: none"><li>• Hängt von der Funktionalität des Pakets <code>java.utl.zip</code> ab</li><li>• Unterstützt 'Deflate'-Komprimierung (Methode 8):<ul style="list-style-type: none"><li>– Keine Unterstützung für verschlüsselte Dateien</li><li>– Keine Unterstützung für zip64</li></ul></li></ul>
.tar	anwendung/tar	tar	Unterstützte TAR-Formate: <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	Nicht zutreffend	tgz	Hängt von der Funktionalität des Pakets <code>java.utl.zip</code> ab

### Einschränkungen und Richtlinien

Automatische Codepageerkennung ist nicht für Dateien verfügbar, die aus einer Archivdatei extrahiert werden. Beim Extrahieren der Dateien verwendet der Crawler die Codepageeinstellung, die er gemäß seiner Konfiguration für einfachen Text und unbekannte Dokumenttypen verwenden soll. Wenn Sie die Administrationskonsole für die Unternehmenssuche verwenden, um Sprach- und Codepageeinstellungen für einen Crawler zu konfigurieren, geben Sie die Codepage ein, die der Crawler für Dokumente mit einfachem Text und für Dokumente verwenden soll, deren Codepage nicht automatisch erkannt werden kann.

Der Crawler verwendet das Änderungsdatum in den Headerdaten der Archiveinträge für jede Datei, um festzustellen, wann Dateien in einer Archivdatei erneut durchsucht werden müssen. Wenn Sie einen Crawler überwachen, enthalten die für durchsuchte Dokumente angezeigten Statistikdaten, einschließlich Statistikdaten für eingefügte, aktualisierte und gelöschte Dokumente, Informationen zu Dateien, die aus Archivdateien extrahiert wurden.

Wenn Sie Crawlern das Durchsuchen von Archivdateien in anderen Archivdateiformaten ermöglichen wollen, wie z. B. LZH-Dateien, müssen Sie ein Crawler-Plug-in schreiben und den Crawler anschließend für die Verwendung des Plug-ins konfigurieren.

### Migration

Wenn Sie für Crawler, die vor der Installation von OmniFind Enterprise Edition Version 8.4 vorhanden waren, das Durchsuchen von Archivdateien ermöglichen wollen, müssen Sie den Crawlerbereich des entsprechenden Crawlers bearbeiten.

Stellen Sie sicher, dass die Einstellungen für auszuschließende MIME-Typen und auszuschließende Dateierweiterungen keine Verweise auf .zip-, .tar-, .tgz- oder .gz-Dateien enthalten.

**Wichtig:** Wenn Sie diese Einstellungen für einen Content Edition-Crawler oder einen DB2 Content Manager-Crawler ändern, müssen Sie alle Dokumente erneut durchsuchen, damit die Änderungen angewendet werden können.

---

## URI-Formate in einem Index für die Unternehmenssuche

Der URI (Uniform Resource Identifier) in jedem Dokument eines Index für die Unternehmenssuche zeigt den Typ des Crawlers an, der das Dokument der Objektgruppe hinzugefügt hat.

Sie können URIs oder URI-Muster bei der Konfiguration von Kategorien, Bereichen und Quick Links für eine Objektgruppe angeben. Den URI geben Sie auch an, wenn Sie Dokumente aus dem Index entfernen müssen, oder um ausführliche Statusinformationen zu einem spezifischen URI anzuzeigen.

Durchsuchen Sie die Objektgruppe, um die URIs oder URI-Muster für ein Dokument zu ermitteln. Sie können die URIs in den Suchergebnissen anklicken, um gewünschte Dokumente abzurufen. Sie können einen URI aus den Suchergebnissen kopieren, um ihn in der Administrationskonsole für die Unternehmenssuche zu verwenden. Sie können beispielsweise ein URI-Muster angeben, um Dokumente, die diesem URI-Muster entsprechen, automatisch einer Kategorie der Unternehmenssuche zuzuordnen.

### Archivdateien

Dokumente, die aus einer Archivdatei (wie z. B. einer .zip- oder .tar-Datei) extrahiert und dann durchsucht werden, haben folgenden URI-Format:

```
ursprungs-uri(?|&)ArchiveEntry=eintragsname(&ArchiveEntry=eintragsname)
```

#### Parameter

*ursprungs-uri*

Der Speicherort der Archivdatei an der Datenquelle.

*eintragsname*

Der URL-codierte Name des Archiveintrags in der Archivdatei.

#### Beispiele

```
file:///d:/Archive1.zip
```

```
file:///d:/Archive1.zip?ArchiveEntry=Folder1/PowerPoint.ppt
```

```
file:///d:/Archive1.zip?ArchiveEntry=Folder2/Text.txt
```

### Content Edition-Crawler

Dokumente, die von einem Content Edition-Crawler im Serverzugriffsmodus durchsucht werden, haben folgendes URI-Format:

```
vbr://servername/repository-system-id/persistente_repository-id
/element-id/versions-id
/elementtyp?[Page=seitenzahl&] jndi-merkmale
```

Dokumente, die von einem Content Edition-Crawler im Direktzugriffsmodus durchsucht werden, haben folgendes URI-Format:

vbr://repository-system-id/persistente\_repository-id  
/element-id/versions-id  
/elementtyp/[?Page=seitenzahl]

### Parameter

Die URL-Codierung wird auf alle Felder angewendet.

*servername*

Der Name des Servers von WebSphere Information Integrator Content Edition.

*repository-system-id*

Die System-ID des Repositories.

*persistente\_repository-id*

Die persistente ID des Repositories.

*element-id*

Die ID des Elements.

*versions-id*

Die ID der Version. Ist die Versions-ID leer, weist dieser Wert auf die neueste Version des Dokuments hin.

*elementtyp*

Der Typ des Elements (CONTENT oder FOLDER).

*seitenzahl*

Die Seitenzahl.

*jndi-merkmale*

Die JNDI-Merkmale für den J2EE-Anwendungsclient. Es gibt zwei Merkmalstypen:

**java.naming.factory.initial**

Der Name der Klasse für den Anwendungsserver, mit dem die EJB-Kennung erstellt wird.

**java.naming.provider.url**

Die URL für den Benennungsservice für den Anwendungsserver, mit der die EJB-Kennung angefordert wird.

### Beispiele

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-Crawler

Dokumente, die von einem DB2-Crawler durchsucht werden, haben folgendes URI-Format:

```
db2://datenbankname/tabellenname
/name_der_eindeutigen_id-spalte1/wert_der_eindeutigen_id1
[/name_der_eindeutigen_id-spalte2/wert_der_eindeutigen_id2/...
/name_der_eindeutigen_id-spalteN/wert_der_eindeutigen_idN]
```

### Parameter:

Die URL-Codierung wird auf alle Felder angewendet.

*datenbankname*

Der interne Name oder der Aliasname der Datenbank.

*tabellenname*

Der Name der Zieltabelle, einschließlich des Schemanamens.

*name\_der\_eindeutigen\_id-spalte1*

Der Name der ersten Spalte der eindeutigen Kennung der Tabelle.

*wert\_der\_eindeutigen\_id1*

Der Wert der ersten Spalte der eindeutigen Kennung.

*name\_der\_eindeutigen\_id-spalteN*

Der Name der *n*. Spalte der eindeutigen Kennung der Tabelle.

*wert\_der\_eindeutigen\_idN*

Der Wert der *n*. Spalte der eindeutigen Kennung.

### Beispiele

Lokale, katalogisierte Datenbank:

```
db2://LOCALDB/SCHEMA1.TABLE1/MODEL/ThinkPadA20
```

Ferne, nicht katalogisierte Datenbank:

```
db2://myserver.mycompany.com:50001/REMOVEDB/SCHEMA2.TABLE2/NAME/DAVID
```

## DB2 Content Manager-Crawler

Dokumente, die von einem DB2 Content Manager-Crawler durchsucht werden, haben folgendes URI-Format:

```
cm://servername/elementtypname/pid
```

### Parameter

Die URL-Codierung wird auf den Parameter *pid* angewendet.

*servername*

Der Name des IBM DB2 Content Manager-Bibliotheksservers.

*elementtypname*

Der Name des Zielelementtyps.

*pid* Die persistente ID von DB2 Content Manager.

### Beispiel

```
cm://cmsrvctg/ITEMTYPE1/92+3+ICM8+icm1sdb12+ITEMTYPE159+26+A1001001A
03F27B94411D1831718+A03F27B+94411D183171+14+1018
```

## Domino Document Manager-Crawler

Dokumente, die von einem Domino Document Manager-Crawler durchsucht werden, haben folgendes URI-Format:

```
dominodoc://servername:portnummer/datenbankreplikant-id/pfad_und_name_der_datenbank
/universelle_sicht-id/universelle_dokument-id
/?AttNo=anlagenummer&AttName=anlagename
```

### Parameter

Die URL-Codierung wird auf alle Felder angewendet.

*servername*

Der Name des Domino Document Manager-Servers.

*portnummer*

Optional: Die Portnummer des Domino Document Manager-Servers.

*datenbankreplikant-id*

Die Kennung des Datenbankreplikats.

*pfad\_und\_name\_der\_datenbank*

Der Pfad und der Dateiname der NSF-Dokumentdatenbank auf dem Domino Document Manager-Zielserver.

*universelle\_sicht-id*

Die universelle Sicht-ID, die zum Durchsuchen von Domino Document Manager-Dokumenten verwendet wird.

*universelle\_dokument-id*

Die universelle Dokument-ID, die im durchsuchten Dokument definiert ist.

*anlagenummer*

Optional: Eine bei null beginnende fortlaufende Nummer für jede Anlage.

*anlagedateiname*

Optional: Der ursprüngliche Name der angehängten Datei.

### Beispiele

Ein Domino Document Manager-Dokument:

```
dominodoc://dominodocsvr.ibm.com/49256D3A000A20DE/domdoc%2FADMN-6FAJXL.nsf/8178B1C14B1E9B6B8525624F0062FE9F/0205F44FA3F45A9049256DB20042D226
```

Eine Dokumentanlage:

```
dominodoc://dominodocsvr.ibm.com/49256D3A000A20DE/domdoc%2FADMN-6FAJXL.nsf/8178B1C14B1E9B6B8525624F0062FE9F/0205F44FA3F45A9049256DB20042D226?AttNo=0&AttName=AttachedFile.doc
```

## Exchange Server-Crawler

Dokumente, die von einem Exchange Server-Crawler durchsucht werden, haben folgendes URI-Format:

```
exchange://owa-pfad[?useSSL=true]
```

### Parameter

*owa-pfad*

Der Pfad für Outlook Web Access (OWA) ohne Protokoll.

**useSSL=true**

Wird hinzugefügt, wenn das Protokoll des ursprünglichen OWA-Pfads HTTPS ist.

### Beispiele

Dokumenthauptteil:

```
exchange://exchangesvr.ibm.com/public/RootFolder1/Folder1/Document.EML
```

Dokumentanlage:

```
exchange://exchangesvr.ibm.com/public/RootFolder1/Folder1/Document.EML/AttachedFile.doc
```

Für SSL aktiviert:

```
exchange://exchangesvr.ibm.com/public/TeamRoom/Folder1/Document.EML?useSSL=true
```

## JDBC-Datenbankcrawler

Dokumente, die von einem JDBC-Datenbankcrawler durchsucht werden, haben folgendes URI-Format:

```
jdbc://db-url/tabellenname
 /name_der_eindeutigen_id-spalte1/wert_der_eindeutigen_id1
 /[name_der_eindeutigen_id-spalte2/wert_der_eindeutigen_id2
 /.../name_der_eindeutigen_id-spalteN/wert_der_eindeutigen_idN]
```

### Parameter

Die URL-Codierung wird auf alle Felder angewendet.

*db-url* Die URL-Adresse für die Datenbank.

*tabellenname*

Der Name der Zieltabelle, einschließlich des Schemanamens.

*name\_der\_eindeutigen\_id-spalte1*

Der Name der ersten Spalte der eindeutigen Kennung der Tabelle.

*wert\_der\_eindeutigen\_id1*

Der Wert der ersten Spalte der eindeutige Kennung.

*name\_der\_eindeutigen\_id-spalteN*

Der Name der *n*. Spalte der eindeutigen Kennung der Tabelle.

*wert\_der\_eindeutigen\_idN*

Der Wert der *n*. Spalte der eindeutigen Kennung.

### Beispiele:

DB2-Datenbank

```
jdbc:db2://host01.svl.ibm.com:50000/SAMPLE/DB2INST1.ORG/DEPTNUMB/51
```

Oracle-Datenbank:

```
jdbc:oracle:thin:@/host01.svl.ibm.com:1521:ora/SCOTT.EMP/EMPNO/7934
```

MS SQL Server 2000-Datenbank:

```
jdbc:microsoft:sqlserver://host01.svl.ibm.com:1433;
DatabaseName=Northwind/dbo.Region/RegionID/100
```

MS SQL Server 2005-Datenbank:

```
jdbc:sqlserver://host01.svl.ibm.com:1433;
DatabaseName=Northwind/dbo.Region/RegionID/100
```

## Notes-Crawler

Dokumente, die von einem Notes-Crawler durchsucht werden, haben folgendes URI-Format:

```
domino://servername[:portnummer]/datenbankreplikant-id/pfad_und_name_der_datenbank
/[universelle_sicht-id]/universelle_dokument-id
[?AttNo=anlagenummer&AttName=anlagedateiname]
```

### Parameter

Die URL-Codierung wird auf alle Felder angewendet.

*servername*

Der Name des Lotus Notes-Servers.

*portnummer*

Die Portnummer für den Lotus Notes-Server. Die Portnummer ist optional.

*datenbankreplikant-id*

Die Kennung des Datenbankreplikats.

*pfad\_und\_name\_der\_datenbank*

Der Pfad und der Dateiname der NSF-Datenbank auf dem Lotus Notes-Zielservers.

*universelle\_sicht-id*

Die auf der Zieldatenbank definierte universelle Sicht-ID. Diese ID wird nur angegeben, wenn das Dokument aus einer Sicht oder aus einem Ordner ausgewählt wird. Bei keiner Angabe einer Sicht oder eines Ordners für die Crawlersuche (wenn Sie z. B. angeben, dass alle Dokumente in einer Datenbank durchsucht werden sollen) wird die universelle Sicht-ID nicht angegeben.

*universelle\_dokument-id*

Die universelle Dokument-ID, die in dem vom Crawler durchsuchten Dokument definiert ist.

*anlagenummer*

Eine bei null beginnende fortlaufende Nummer für jede Anlage. Die Anlagennummer ist optional.

*anlagedateiname*

Der ursprüngliche Name der angehängten Datei. Der Name der angehängten Datei ist optional.

### Beispiele

Ein Dokument, das für die Crawlersuche nach Sicht oder Ordner ausgewählt wurde:

```
domino://dominosvr.ibm.com/49256D3A000A20DE/Database.nsf/
8178B1C14B1E9B6B8525624F0062FE9F/0205F44FA3F45A9049256DB20042D226
```

Ein Dokument, das nicht für die Crawlersuche nach Sicht oder Ordner ausgewählt wurde:

```
domino://dominosvr.ibm.com/49256D3A000A20DE/Database.nsf/
0205F44FA3F45A9049256DB20042D226
```

Eine Dokumentanlage:

```
domino://dominosvr.ibm.com/49256D3A000A20DE/Database.nsf/
0205F44FA3F45A9049256DB20042D226?AttNo=0&AttName=AttachedFile.doc
```

### QuickPlace-Crawler

Dokumente, die von einem QuickPlace-Crawler durchsucht werden, haben folgendes URI-Format:

```
quickplace://servername:portnummer/datenbankreplikant-id/pfad_und_name_der_datenbank
/universelle_sicht-id/universelle_dokument-id
/?AttNo=anlagenummer&AttName=anlagedateiname
```

#### Parameter

Die URL-Codierung wird auf alle Felder angewendet.

*servername*

Der Name des Lotus QuickPlace-Servers.

*portnummer*

Optional: Die Portnummer des QuickPlace-Servers.

*datenbankreplikant-id*

Die Kennung des Datenbankreplikats.

*pfad\_und\_name\_der\_datenbank*

Der Pfad und der Dateiname der NSF-Dokumentdatenbank auf dem QuickPlace-Zielserver.

*universelle\_sicht-id*

Die universelle Sicht-ID, die zum Durchsuchen von QuickPlace-Dokumenten verwendet wird.

*universelle\_dokument-id*

Die universelle Dokument-ID, die im durchsuchten Dokument definiert ist.

*anlagenummer*

Optional: Eine bei null beginnende fortlaufende Nummer für jede Anlage.

*anlagedateiname*

Optional: Der ursprüngliche Name der angehängten Datei.

## Beispiele

Ein Dokument:

```
quickplace://1twsvr.ibm.com/49257043000214B3/QuickPlace%5Csamplplace%5CPageLibrary4925704300021490.nsf/A7986FD2A9CD47090525670800167225/2B02B1DE3A82B2CE49257043001C2498
```

Eine Seitenanlage:

```
quickplace://1twsvr.ibm.com/49257043000214B3/QuickPlace%5Csamplplace%5CPageLibrary4925704300021490.nsf/A7986FD2A9CD47090525670800167225/2B02B1DE3A82B2CE49257043001C2498?AttNo=0&AttName==QPCons3.ppt
```

## Crawler für die Einstiegspunktliste

Dokumente, die von einem Crawler für die Einstiegspunktliste durchsucht werden, haben folgendes URI-Format:

```
seedlist://seiten-url?pageID=seiten-id[&useSSL;=true]
```

### Parameter

Die URL-Codierung wird auf alle Felder angewendet.

*seiten-url*

Die URL des Dokuments (für jedes Dokument eindeutig).

*seiten-id*

Die Objektkennung für das Dokument.

**useSSL** Wenn das Protokoll HTTPS verwendet wird, wird dem URI `&useSSL;=true` hinzugefügt. Andernfalls wird `useSSL` weggelassen.

### Beispiel

HTTPS-Protokoll

```
seedlist://quickrserver.ibm.com:10035/lotus/mydoc?uri=dm:bec6090046f1cd52bc5cfcb06e9f4550&verb;=view&pageID;=N1FSZUR1MkJQNjZSMZQMUMwM1FPNjZCQzY2SUw2SUhPNk1RQ0M2Uk80Nk9PNjVCRUM2UUu2TDFDMA==&useSSL;=true
```

## UNIX-Dateisystem-Crawler

Dokumente, die von einem UNIX-Dateisystem-Crawler durchsucht werden, haben folgendes URI-Format:

```
file:///verzeichnisname/dateiname
```

### Parameter

Die URL-Codierung wird auf alle Felder angewendet.

*verzeichnisname*

Der absolute Pfadname des Verzeichnisses.

*dateiname*

Der Name der Datei.

### Beispiel

file:///home/user/test.doc

## Web Content Management-Crawler

WebSphere Content Management-Dokumente, die von einem Web Content Management-Crawler durchsucht werden, haben folgendes URI-Format:

wcm://seiten-url?pageID=seiten-id[&useSSL=true]

### Parameter

Die URL-Codierung wird auf alle Felder angewendet.

*seiten-url*

Die URL des Dokuments (für jedes Dokument eindeutig).

*seiten-id*

Die Seitenkennung.

**useSSL** Wenn das Protokoll HTTPS verwendet wird, wird dem URI useSSL=true hinzugefügt. Andernfalls wird useSSL weggelassen.

### Beispiele

HTTP-Protokoll

wcm://wp6server.ibm.com:9081/wps/wcm/myconnect/Web+Content/Site01/SiteArea01/ContentTest01?pageID=6QReDeJ9DI3R0663E03Q06L1E2MR47MHOC3Q862RD6J0863B0GJS86J9E0

HTTPS-Protokoll

wcm://wp6server.ibm.com:9444/wps/wcm/myconnect/Web+Content/Site01/SiteArea01/ContentTest01?pageID=6QReDeJ9DI3R0663E03Q06L1E2MR47MHOC3Q862RD6J0863B0GJS86J9E0&useSSL=true

## WebSphere Portal-Crawler: WebSphere Portal Version 5

Dokumente von WebSphere Portal Version 5, die von einem WebSphere Portal-Crawler durchsucht werden, haben folgendes URI-Format:

wps://seiten-url?portletDefID=portletdefinitions-id&portletID=portlet-id&pageID=seiten-id[&useSSL=true]

### Parameter

Die URL-Codierung wird auf alle Felder angewendet.

*seiten-url*

Die URL des Dokuments (für jedes Dokument eindeutig).

*portletdefinitions-id*

Die Portletdefinitions-ID.

*portlet-id*

Die Portletkennung.

*seiten-id*

Die Seitenkennung.

**useSSL** Wenn das Protokoll HTTPS verwendet wird, wird dem URI useSSL=true hinzugefügt. Andernfalls wird useSSL weggelassen.

### Beispiele

Dokumenthauptteil:

```
wps://wpserver.ibm.com:9081/wps/myportal!/ut/p/kcxml/04_Sj9SPykssy0x+
LKnPy1vM0Y_QjzKCN4g3cQbJgQi0-pFQAW99X4_83FT9AP2C5IhyR0dFRQD8qHRj/delta
/base64xml/L01DU1kvd0NrQUpORUEvNFBVR0VoQSEvN18wXzZPLzZfMF80RA!!
?portletDefID=3_0_3S&pageID=6_0_6J
```

### Beispiele

Für SSL aktiviert:

```
wps://wpserver.ibm.com:9081/wps/myportal!/ut/p/kcxml/04_Sj9SPykssy0x+
LKnPy1vM0Y_QjzKCN4g3cQbJgQi0-pFQAW99X4_83FT9AP2C5IhyR0dFRQD8qHRj/delta
/base64xml/L01DU1kvd0NrQUpORUEvNFBVR0VoQSEvN18wXzZPLzZfMF80RA!!
?portletDefID=7_0_A4&pageID=6_0_6J&useSSL=true
```

## WebSphere Portal-Crawler: WebSphere Portal Version 6

Dokumente von WebSphere Portal Version 6, die von einem WebSphere Portal-Crawler durchsucht werden, haben folgendes URI-Format:

```
wp6://seiten-url?portletURL=portlet-url?portletDefID=portletdefinitions-id
&pageID=seiten-id[&useSSL=true]
```

### Parameter

Die URL-Codierung wird auf alle Felder angewendet.

*seiten-url*

Die URL des Dokuments (für jedes Dokument eindeutig).

*portlet-url*

Die eindeutige URL-Adresse für das Dokument.

*portletdefinitions-id*

Die Portletdefinitions-ID.

*seiten-id*

Die Seitenkennung.

**useSSL** Wenn das Protokoll HTTPS verwendet wird, wird dem URI useSSL=true hinzugefügt. Andernfalls wird useSSL weggelassen.

### Beispiele

HTTP-Protokoll

```
wp6://wp6server.ibm.com:9081/wps/myportal!/ut/p/c1/04_SB8K8xLLM9MSSzPy
8xBz9CP0os3gjE59gQwMLQ0P_IDMnAyNHA3f3UESTD1NjA6B8pFm8AQ7gaEBAAdjIPrz6_
Tzyc1P1C3IjDHQdFRUBTu-saA!!/d12/d0/Y2BkbGBgY1rDwMDEJ1XAwMggYxZvZOITbGh
gYWjobuhmaGDka0Bu5uHqFRpkaAAAEisaBQ!!
?portletUrl=/wps/myportal!/ut/p/c1/04_SB8K8xLLM9MSSzPy8xBz9CP0os3gjE59
gQwMLQ0P_IDMnAyNHA3f3UESTD1NjA6B8pFm8AQ7gaEBAAdjIPrz6_Tzyc1P1C3IjDHQdF
RUBTu-saA!!/d12/d0/Y2BiUZnBwMqsyBykYGBmS2tcDoDE40MebyRiU-woYGFOaG7oZu
hgZGjgbuZh6uXY4ihAYOMGQ7Z0CDBAwDcXPkM
&portletDefID=6_24LS10811G1F102A0G6HEJUR10
&pageID=3_24LS108110R6B02A0GGU94LN00
```

HTTPS-Protokoll

```
wp6://wp6server.ibm.com:9444/wps/myportal!/ut/p/c1/04_SB8K8xLLM9MSSzPy
8xBz9CP0os3gjE59gQwMLQ0P_IDMnAyNHA3f3UESTD1NjA6B8pFm8AQ7gaEBAAdjIPrz6_
Tzyc1P1C3IjDHQdFRUBTu-saA!!/d12/d0/Y2BkbGBgY1rDwMDEJ1XAwMggYxZvZOITbGh
gYWjobuhmaGDka0Bu5uHq5RtqaAAA50L41Q!!
?portletUrl=/wps/myportal!/ut/p/c1/04_SB8K8xLLM9MSSzPy8xBz9CP0os3gjE59
gQwMLQ0P_IDMnAyNHA3f3UESTD1NjA6B8pFm8AQ7gaEBAAdjIPrz6_Tzyc1P1C3IjDHQdF
RUBTu-saA!!/d12/d0/Y2BiUZnBwMqsyBykYGBmS2tcDoDE40MebyRiU-woYGFOaG7oZu
```

hgZGjgbuZh6uXWZihAYOMGQ5Z31BDaWak73P2  
&portletDefID=6\_24LS10811G1F102A0G6HEJMU10  
&pageID=3\_24LS108110R6B02A0GGU94T410&useSSL=true

## Windows-Dateisystem-Crawler

Dokumente, die von einem Windows-Dateisystem-Crawler durchsucht werden, haben folgende URI-Formate:

*file:///verzeichnisname/dateiname*

*file:///netzordnername/verzeichnisname/dateiname*

### Parameter

Die URL-Codierung wird auf alle Felder angewendet.

*verzeichnisname*

Der absolute Pfadname des Verzeichnisses.

*dateiname*

Der Name der Datei.

*netzordnername*

Nur für Dokumente auf fernen Servern: Der Name des gemeinsam genutzten (freigegebenen) Ordners im Windows-Netz.

### Beispiele

Lokales Dateisystem:

*file:///d:/directory/test.doc*

Netzdateisystem:

*file:///filesvr.ibm.com/directory/file.doc*



---

## Parserverwaltung

Zur Verbesserung der Abrufbarkeit von Dokumenten können Sie Optionen angeben, wie Dokumente und Metadaten syntaktisch analysiert und kategorisiert werden sollen, bevor sie dem Index für die Unternehmenssuche hinzugefügt werden.

Folgende Optionen können Sie für die Syntaxanalyse von Dokumentinhalt und für die Optimierung der Abrufbarkeit von Informationen angeben:

### **Konfigurieren von Optionen für die Syntaxanalyse von Dokumenten in chinesischer, japanischer und koreanischer Sprache**

Sie können Optionen für die Verwendung der N-Gram-Segmentierung angeben, damit Dokumente in chinesischer, japanischer und koreanischer Sprache syntaktisch analysiert werden. Sie können außerdem Zeilenvorschubzeichen aus dem Leerraum in Dokumenten in chinesischer und japanischer Sprache entfernen.

### **Aktivieren einer nativen XML-Suche**

Wenn Ihre Objektgruppe XML-Dokumente umfasst, können Sie sie für das Durchsuchen mit einer nativen XML-Abfragesyntax, wie z. B. XPath- und XML-Fragmenten, aktivieren. Bei einer nativen XML-Suche können Benutzer Abfragen auf der Basis von Abhängigkeiten zwischen verschiedenen XML-Elementen angeben.

### **Konfigurieren von Kategorien**

Sie können Dokumente in Kategorien gruppieren, die ein ähnliches URI-Muster verwenden oder bestimmte Wörter enthalten. Wenn Benutzer die Objektgruppe durchsuchen, können sie die Suchergebnisse auf die Dokumente begrenzen, die zu bestimmten Kategorien gehören.

### **Konfigurieren von Suchfeldern**

Sie können Elemente in XML-Dokumenten Suchfeldern im Index zuordnen. Sie können auch Metadatenelemente in HTML-Dokumenten Suchfeldern zuordnen. Durch die Erstellung von Suchfeldern im Index für die Unternehmenssuche können Benutzer bestimmte Teile von XML- und HTML-Dokumenten abfragen und die Genauigkeit der Suchergebnisse verbessern.

### **Konfigurieren von Optionen für die Textverarbeitung**

Wenn dem System für die Unternehmenssuche benutzerdefinierte Textanalysesteuerkomponenten hinzugefügt wurden, können Sie eine Textanalysesteuerkomponente für die Verwendung mit einer Objektgruppe auswählen. Nachdem Sie eine Analysesteuerkomponente einer Objektgruppe zugeordnet haben, können Sie Optionen für das Zuordnen von Inhalt angeben, sodass dieser linguistisch analysiert und mit Kommentaren versehen werden kann. Sie können auch angeben, wie die Ergebnisse der Analyse dem Index für die Unternehmenssuche oder JDBC-Datenbanktabellen zugeordnet werden sollen.

### **Zuordnen von Feldern zu Boostklassen**

Sie können angeben, dass Dokumente mit Feldern, die mit den Abfragebegriffen übereinstimmen, in den Suchergebnissen in der Rangfolge höher eingestuft werden sollen, als andere mit den Abfragebegriffen übereinstimmende Dokumente. Wenn Sie Felder Boostklassen zuordnen, geben Sie an, welcher Inhalt und welche Metadatenfelder einem Boostvorgang unterzo-

gen werden sollen. Sie können außerdem konfigurieren, mit welchen Bewertungen die einzelnen Boostklassen die Rangfolge der Dokumente festlegen.

### Zugehörige Konzepte

-  Linguistische Unterstützung für semantische Suche
  -  Textanalyse innerhalb der Unternehmenssuche
  -  Grundkonzepte für die Verarbeitung der Textanalyse
  -  Semantische Suchanwendungen
  -  Semantische Suchabfragebegriffe
- „Sprachen- und Codepage-Unterstützung“ auf Seite 181  
„Dokumentformaterkennung“ auf Seite 165

---

## Funktionsweise von Kategorien

Mit Kategorien können Sie Dokumente mit gleichen Merkmalen gruppieren und die Suche nach sowie das Abrufen von Dokumenten auf Mitglieder dieser Gruppe beschränken.

Wenn Sie Kategorien Dokumente zuordnen und Ihre Suchanwendungen diese Funktion unterstützen, können Benutzer eine Untergruppe der Objektgruppe durch Angabe des Kategorienamens durchsuchen. Wenn sie die gesamte Objektgruppe durchsuchen, können Benutzer die Suchergebnisse eingrenzen und nur die Dokumente durchsuchen, die sich in der gleichen Kategorie befinden wie eines der Ergebnisdokumente.

Wenn Sie eine Kategorie konfigurieren, geben Sie Regeln an, anhand derer der Crawler Dokumente mit der Kategorie verknüpft. Sie können Dokumente gruppieren, die ein URI-Muster gemeinsam nutzen oder einen bestimmten Inhalt aufweisen (z. B. Dokumente, die bestimmte Wörter oder Wortfolgen enthalten oder ausschließen).

Gehen Sie wie folgt vor, um über die Administrationskonsole für die Unternehmenssuche Kategorien zu erstellen und zu verwalten:

- Sie wählen den Kategorisierungstyp beim Erstellen einer Objektgruppe aus. Sie können auswählen, ob keine Kategorien oder regelbasierte Kategorien verwendet werden sollen.
- Beim Konfigurieren von Syntaxanalyseregeln für die Objektgruppe können Sie den Kategorisierungstyp ändern, falls erforderlich. Wenn Sie den Kategorisierungstyp nach dem Durchsuchen und Indexieren von Dokumenten ändern, wird die Suchqualität vermindert, bis Sie alle Dokumente erneut durchsuchen und den Hauptindex erneut erstellen.
- Wenn Sie für den Kategorisierungstyp **Regelbasiert** auswählen, verwalten Sie den Kategoriebaum, Kategorien und Kategorisierungsregeln über die Administrationskonsole. Wenn Sie Kategorien oder Kategorisierungsregeln nach dem Durchsuchen und Indexieren von Dokumenten ändern, wird die Suchqualität vermindert, bis Sie alle Dokumente erneut durchsuchen und den Hauptindex erneut erstellen.

## Regelbasierte Kategorien

Sie können Regeln konfigurieren, um zu steuern, welche Dokumente Kategorien in einer Objektgruppe für die Unternehmenssuche zugeordnet sind.

Sie können Kategorisierungsregeln für Objektgruppen erstellen, die Sie in der Unternehmenssuche erstellen, sowie für regelbasierte Kategorien, die Sie aus IBM WebSphere-Objektgruppen migrieren. Sie müssen beim Erstellen der Objektgruppe bzw. beim Angeben der Syntaxanalyseoptionen für die Objektgruppe angeben, dass Sie regelbasierte Kategorien verwenden wollen, um Regeln für das Kategorisieren von Dokumenten konfigurieren zu können.

Der Parser verwendet die von Ihnen angegebenen Regeln, um Dokumente mindestens einer Kategorie zuzuordnen:

- Wenn ein Dokument mindestens einer Regel in einer Kategorie entspricht, ordnet der Parser das Dokument der Kategorie zu.
- Wenn ein Dokument mindestens einer Regel in mehreren Kategorien entspricht, ordnet der Parser das Dokument allen Kategorien zu.
- Wenn ein Dokument keiner Regel für eine Kategorie entspricht, ordnet der Parser das Dokument keiner Kategorie zu. Benutzer können nach diesem Dokument suchen und es beim Durchsuchen der Objektgruppe abrufen, sie können es jedoch nicht beim Durchsuchen einer Kategorie abrufen.

Wenn Sie den Kategoriebaum (d. h. die Taxonomie) für eine Objektgruppe verwalten, können Sie entscheiden, wo in der Kategoriehierarchie eine neue Kategorie hinzugefügt werden soll. Im Kategoriebaum können Sie zudem eine zu bearbeitende Kategorie auswählen und anschließend Regeln für das Kategorisieren von Dokumenten hinzufügen, Regeln löschen oder den Inhalt einzelner Regeln ändern.

Beim Konfigurieren einer Regel für das Kategorisieren von Dokumenten geben Sie an, dass die Unternehmenssuche den URI eines Dokuments oder den Inhalt im Dokument verwenden soll, um die Zugehörigkeit des Dokuments zur Kategorie zu ermitteln.

### Regeln für URI-Muster

Eine URI-Regel gilt für den URI des Dokuments. Sie geben einen Teil eines URI (d. h. ein Muster) an. Dokumente, deren URIs das angegebene Muster aufweisen, entsprechen der Regel.

Wenn Sie z. B. angeben, dass der Regeltext `/hr/` ist, dann entspricht der erste unten stehende URI der Regel und der zweite URI nicht:

```
file:///corporate/hr/medicalform.doc
http://company.com/human_resources/medicalform.htm
```

Da alle URIs wie Muster behandelt werden, ignoriert das System alle Sterne, die Sie als Platzhalterzeichen am Anfang oder am Ende des Musters angeben. Beispiel: `*/hr/*` und `/hr/` entsprechen derselben Gruppe von URIs.

Bei den Regeln für URI-Muster muss die Groß-/Kleinschreibung nicht beachtet werden. Wenn ein URI Leerzeichen enthält, muss das URI-Muster für die Codierung von URIs die Regeln für die Unternehmenssuche einhalten. Das folgende Beispiel zeigt, wie ein URI für einen Windows-Dateisystempfad angegeben werden sollte und wie nicht:

Falscher URI: file:///c:/program files/  
Korrektter URI: file:///c:/program+files/

## Regeln für Dokumentinhalt

Sie können die Regeln für den Dokumentinhalt im gleichen Format wie eine Abfrage formulieren. Wenn das Dokument für die Abfrage gültig ist, entspricht es der Regel. Beim Konfigurieren der Regel geben Sie die Wörter und Ausdrücke an, die Dokumente enthalten müssen bzw. nicht enthalten dürfen. Sie können außerdem eine Sprache für die Anwendung von Wortstammregeln auswählen.

Beispiel: Die folgende Regel gibt an, dass ein Dokument der Regel entspricht, wenn es das Wort `hr` oder den Ausdruck `human resources` enthält:

```
hr "human resources"
```

Weiteres Beispiel: Die folgende Regel gibt an, dass ein Dokument der Regel entspricht, wenn sein Titel das Wort `"health"` enthält und die Wortfolge `"employee benefits"` nicht enthält:

```
+title:health -title:"employee benefits"
```

Für Inhaltsregeln gelten dieselben linguistischen Normalisierungen wie für SI-API-Abfragen (Search and Index API). Die Syntax für Inhaltsregeln unterstützt jedoch eine Untergruppe der Operationen, die in der SI-API-Abfragesyntax verfügbar sind. Nur die folgenden Abfrageoperatoren sind in Regeln für Inhalt zulässig:

- + Setzen Sie vor einen Begriff ein Pluszeichen, um anzugeben, dass der Begriff in dem Dokument enthalten sein muss.
- Setzen Sie vor einen Begriff ein Minuszeichen, um anzugeben, dass der Begriff in dem Dokument nicht enthalten sein darf.
- " " Setzen Sie zwei oder mehr Begriffe in Anführungszeichen, um anzugeben, dass die genaue Wortfolge in dem Dokument enthalten sein muss.

### *feldname:*

Stellen Sie einem Begriff oder einer Wortfolge einen Feldnamen voran, um anzugeben, dass der Begriff oder die Wortfolge im angegebenen Feld vorkommen muss (bzw. nicht vorkommen darf).

Alle Inhalts- und Metadatenfelder, die in der Objektgruppe als durchsuchbare Felder konfiguriert sind, werden unterstützt. Die folgenden SI-API-Feldschlüsselwörter und -Feldtypen werden nicht unterstützt:

- `site:`
- `url:`
- `link:`
- `docID:`
- `samegroupas:`
- Parameterfelder
- Sicherheitstoken
- Attribute (wie z. B. `$source`, `$language`, `$doctype` und so weiter)

## Kategoriebäume

Mit einem Kategoriebaum können Sie alle regelbasierten Kategorien in einer Objektgruppe anzeigen. Im Kategoriebaum können Sie Kategorien erstellen bzw. löschen und die Regeln bearbeiten, die Dokumente Kategorien zuordnen.

Ein Kategoriebaum, der auch als Taxonomie bezeichnet wird, ist hierarchisch angeordnet. Der Baum beginnt mit der Stammkategorie, von der alle anderen Kategorien ausgehen. Sie können eine beliebige Anzahl von Kategorien und Unterkategorien verschachteln, um Benutzern verschiedene Auswahlmöglichkeiten für das Durchsuchen und Abrufen von Dokumenten zu bieten.

Wenn z. B. ein Dokument den Regeln in mehreren Kategorien entspricht, wird es diesen Kategorien zugeordnet. Suchen Benutzer in den Suchergebnissen nach einer Kategorie oder durchsuchen sie Dokumente in einer bestimmten Kategorie, ist die Suche erfolgreicher, wenn ein Dokument zu mehreren Kategorien gehört.

Wenn Sie den Kategoriebaum verwalten, können Sie steuern, welche Dokumente zu mindestens einer Kategorie gehören, indem Sie neue Kategorien unter vorhandenen Kategorien verschachteln. Beim Erstellen einer Kategorie geben Sie an, ob sie auf der Stammebene oder als eine Unterkategorie einer anderen Kategorie erstellt werden soll. Im Kategoriebaum können Sie auch Kategorien aus der Objektgruppe löschen und die Regeln für das Zuordnen von Dokumenten zu Kategorien ändern. Wenn Sie eine Kategorie bearbeiten, können Sie die Kategorie umbenennen, Kategorisierungsregeln hinzufügen bzw. löschen oder den Inhalt einzelner Regeln modifizieren.

Wenn Sie den Kategoriebaum verwalten, verwenden Sie die folgenden Beschreibungen des Suchverhaltens als Richtlinie:

- Wenn ein Benutzer eine übergeordnete Kategorie durchsucht, werden diese Kategorie und ihre gesamten Unterkategorien nach Dokumenten durchsucht, die die Suchbedingungen erfüllen. Wenn ein Benutzer eine Kategorie ohne zusätzliche Unterkategorien durchsucht, wird nur diese Kategorie durchsucht.
- Wenn ein Benutzer Suchergebnisse durchsucht und eine Option zum Durchsuchen von Dokumenten in einer bestimmten Kategorie auswählt, werden nur die Dokumente in dieser Kategorie angezeigt. In den Suchergebnissen werden zusätzlich die Namen von vorhandenen Unterkategorien angezeigt, damit der Benutzer zwischen Kategorien navigieren und jeweils Untergruppen von Dokumenten anzeigen kann.

## Auswählen des Kategorisierungstyps

Wenn Sie einen Kategorisierungstyp auswählen, geben Sie die Methode an, die Sie verwenden wollen, um Dokumente und Kategorien in einer Objektgruppe miteinander zu verknüpfen.

### Vorbereitung

Sie müssen ein Administrator für die Unternehmenssuche oder ein Objektgruppenadministrator der Objektgruppe sein, die Sie ändern wollen, um den Kategorisierungstyp zu ändern.

### Informationen zu dieser Task

Der Kategorisierungstyp wird beim Erstellen der Objektgruppe angegeben. Falls erforderlich, können Sie ändern, wie Dokumente für eine Objektgruppe kategorisiert werden.

siert werden. Sie können speziell für eine Objektgruppe konfigurierte regelbasierte Kategorien oder keine Kategorien verwenden.

**Wichtig:** Wenn Sie den Kategorisierungstyp nach dem Durchsuchen von Daten und nach dem Erstellen eines Index für eine Objektgruppe ändern, wird der Index inkonsistent. Durchsuchen Sie alle Dokumente in der Objektgruppe erneut, und erstellen Sie den Hauptindex erneut, um genaue Suchergebnisse sicherzustellen.

### Vorgehensweise

Gehen Sie wie folgt vor, um den Kategorisierungstyp auszuwählen:

1. Bearbeiten Sie eine Objektgruppe, wählen Sie die Seite **Syntaxanalyse** aus, und klicken Sie **Kategorisierungstyp auswählen** an.
2. Wählen Sie auf der Seite **Kategorisierungstyp auswählen** eine der folgenden Optionen aus:

#### **Keiner**

Wählen Sie diese Option aus, wenn in dieser Objektgruppe keine Dokumente kategorisiert werden sollen.

#### **Regelbasiert**

Wählen Sie diese Option aus, wenn Sie Dokumente anhand von Regeln kategorisieren wollen, die Sie speziell für diese Objektgruppe konfigurieren.

## Konfigurieren von Kategorien

Sie können eine beliebige Anzahl Kategorien für eine Objektgruppe erstellen, und jede Kategorie kann eine beliebige Anzahl Regeln enthalten. Die Regeln bestimmen, welche Dokumente der Kategorie automatisch zugeordnet werden.

### Vorbereitung

Sie müssen ein Administrator für die Unternehmenssuche oder ein Objektgruppenadministrator der Objektgruppe sein, zu der die Kategorien gehören, um Kategorien zu konfigurieren.

Die Option, regelbasierte Kategorien zu verwenden, muss als Kategorisierungstyp ausgewählt sein.

Wenn Sie Beispiele anzeigen wollen, wie Regeln für die Zuordnung von Dokumenten zu Kategorien angegeben werden, klicken Sie **Hilfe** an, während Sie eine Kategorie erstellen oder bearbeiten.

### Informationen zu dieser Task

Wenn Ihre Suchanwendungen die Unterstützung von Kategorien ermöglichen, können die Benutzer eine Untergruppe der Objektgruppe durchsuchen, indem sie den Kategorienamen angeben. Benutzer können auch in den Suchergebnissen eine Kategorie auswählen und nur die Dokumente durchsuchen, die zur ausgewählten Kategorie gehören.

**Wichtig:** Wenn Sie Kategorien oder Kategorisierungsregeln nach dem Durchsuchen von Daten und nach dem Erstellen eines Index für eine Objektgruppe ändern, wird der Index inkonsistent. Durchsuchen Sie alle Dokumente in der Objektgruppe erneut, und erstellen Sie den Hauptindex erneut, um genaue Suchergebnisse sicherzustellen.

## Vorgehensweise

Gehen Sie wie folgt vor, um eine Kategorie zu konfigurieren:

1. Bearbeiten Sie eine Objektgruppe, wählen Sie die Seite **Syntaxanalyse** aus, und klicken Sie **Kategoriebaum konfigurieren** an.
2. Wählen Sie auf der Seite **Kategoriebaum** die Position in der Baumstruktur aus, an der Sie eine Kategorie hinzufügen wollen, und klicken Sie **Kategorie erstellen** an.

Wenn Sie das Stammverzeichnis auswählen, wird die neue Kategorie auf der Stammverzeichnisebene erstellt. Wenn Sie einen Kategorienamen auswählen, wird die neue Kategorie unter der ausgewählten Kategorie im Kategoriebaum verschachtelt.

Ein Assistent wird geöffnet, der Sie bei der Angabe von Regeln für die Zuordnung von Dokumenten zur neuen Kategorie unterstützt:

- a. Geben Sie auf der Seite **Kategorie erstellen** einen beschreibenden Namen für die Kategorie ein, und klicken Sie **Weiter** an.
- b. Klicken Sie **Regel hinzufügen** auf der Seite **Kategorisierungsregeln erstellen** an.
- c. Geben Sie auf der Seite **Kategorisierungsregel erstellen** einen eindeutigen Namen für die Regel in das Feld **Regelname** ein. Dieser Name muss in allen Kategorien der Objektgruppe eindeutig sein.
- d. Geben Sie die Regel an, die Sie verwenden wollen, um dieser Kategorie Dokumente zuzuordnen, und klicken Sie anschließend **OK** an.

- Wenn Sie mithilfe des URI eines Dokuments feststellen wollen, ob das Dokument zu der Kategorie gehört, klicken Sie **URI-Muster** an, und geben Sie dann das entsprechende URI-Muster an.

Ist der von Ihnen angegebene Text im URI vorhanden, wird das Dokument der Kategorie zugeordnet.

Beispiel: `file:///c:/programme/finance`

- Klicken Sie **Dokumentinhalt** an, wählen Sie die Sprache der Dokumente aus, und geben Sie dann die Wörter und Wortfolgen an, die im Dokumentinhalt vorkommen müssen oder nicht vorkommen dürfen, um durch das Abfragen von durchsuchbarem Inhalt festzustellen, ob ein Dokument zur Kategorie gehört.

Die Regel formulieren Sie im gleichen Format wie eine Abfrage. Zulässig sind jedoch nur die Abfrageoperatoren für das Einschließen (+), das Ausschließen (-), für Wortfolgen (" ") und Feldnamen (*feldname*:). Die N-Gram-Segmentierung wird nicht mit Inhaltsregeln unterstützt.

Wenn ein Dokument die von Ihnen angegebenen Wörter ein- oder ausschließt, wird das Dokument der Kategorie zugeordnet.

Beispiel: `+Rechnungswesen -Abrechnung +title:"Geschäftsjahr"`

- e. Klicken Sie **Fertig stellen** an.

Ihre neue Kategorie wird auf der Seite **Kategoriebaum** mit den anderen Kategorien angezeigt, die zu dieser Objektgruppe gehören.

### Zugehörige Verweise

„URI-Formate in einem Index für die Unternehmenssuche“ auf Seite 130

---

## Funktionsweise von XML-Suchfeldern

Ordnen Sie Suchfeldern XML-Elemente zu, wenn Benutzer bestimmte Teile von XML-Dokumenten durchsuchen sollen.

Sie ordnen XML-Elemente Suchfeldern über die Administrationskonsole für die Unternehmenssuche zu.

Normalerweise werden alle Textdaten in einem XML-Dokument indiziert. Mithilfe der Strukturinformationen von XML-Dokumenten lassen sich speziellere Abfragen unterstützen, indem Sie Suchfeldern XML-Elemente zuordnen. Beispielsweise können Sie die Daten in einem XML-Element so einstellen, dass sie nach Feldnamen durchsucht und in den Suchergebnissen zurückgegeben werden können. Hierfür ordnen Sie das XML-Element einem Feldnamen zu. Sie können die Zuordnung präzisieren, indem Sie Attribute des XML-Elements (und die Werte dieser Attribute) als Bedingungen für die Verwendung des Elements als Suchfeld angeben.

## XML-Suchfelder

Mit XML-Suchfeldern können Benutzer bestimmte Teile von XML-Dokumenten abfragen.

XML-Dokumente werden immer geläufiger, weil sie sowohl semistrukturierten als auch unstrukturierten Text enthalten. Die XML-Struktur ist eingebunden und verwendet einen Kontext, der explizit durch XML-Elemente definiert ist, die den Text umgeben. Beispielsweise könnte der Name eines Autors wie folgt angezeigt werden:

```
<author>John Smith</author>
```

In diesem Kontext gibt der Text John Smith den Autor eines XML-Dokuments an.

Mithilfe der Unternehmenssuche können Sie XML-Elemente Suchfeldnamen zuordnen. Beim Konfigurieren von Syntaxanalyseoptionen für eine Objektgruppe geben Sie an, welche XML-Elemente welchen Suchfeldnamen zugeordnet werden sollen. Durch Zuordnen von XML-Elementen zu Suchfeldern ermöglichen Sie es Benutzern, die Werte dieser Elemente zu durchsuchen, indem sie die Feldnamen in Abfragen angeben. Abfragen, die bestimmte Felder durchsuchen, liefern genauere Suchergebnisse als Freiformatabfragen, die den gesamten Dokumentinhalt durchsuchen.

Beispiel: Wenn Ihre Objektgruppe XML-Dokumente enthält und Sie angeben, dass die Elemente `<title>` und `<author>` als Suchfelder im Index markiert werden sollen, können Benutzer diese Elemente direkt durchsuchen. Eine Suche nach `author:Smith` findet XML-Dokumente, die Smith in Elementen enthalten, die einem Feld mit dem Namen `author` zugeordnet sind.

Ein weiteres Beispiel: Ein XML-Element mit dem Namen `<summary>` enthält möglicherweise Informationen, deren Anzeige in den Suchergebnissen von Nutzen ist. Wenn Sie das Element `<summary>` einem Suchfeld zuordnen und angeben, dass der Wert dieses Elements in den Suchergebnissen angezeigt werden soll, ist der Inhalt des Elements Teil des Ergebnisdokuments.

Wenn Sie ein XML-Element einem Feldnamen zuordnen, kann der Text in dem Element unter dem von Ihnen angegebenen Feldnamen durchsucht werden. Wenn das XML-Element jedoch Attribute einschließt, werden die Attributwerte nicht indiziert und können nicht direkt durchsucht werden. Sie müssen Optionen für die Syntaxanalyse in der Administrationskonsole für die Unternehmenssuche konfigurieren und die native XML-Suche aktivieren, um die Attributwerte abfragen zu können. Nach dem Aktivieren der nativen XML-Suche ist die Struktur des XML-

Dokuments im Index verfügbar, und sie können das Dokument abfragen, indem Sie XPath-Abfrageeinschränkungen angeben. Beispiel:

```
@xmlp::'//elementName[@attributeName="attributeValue"]
```

### Zugehörige Konzepte

 Semantischer Suchabfragebegriff

### Zugehörige Tasks

„Aktivieren der Unterstützung für eine native XML-Suche“ auf Seite 163

## Zuordnen von XML-Elementen zu Suchfeldern

Wenn Sie ein XML-Element einem Suchfeld zuordnen, geben Sie an, nach welchen XML-Elementen ein Benutzer durch Angeben eines Feldnamens in einer Abfrage suchen kann.

### Vorbereitung

Sie müssen ein Administrator für die Unternehmenssuche oder ein Objektgruppen-administrator der Objektgruppe sein, zu der die XML-Dokumente gehören, um Suchfeldern XML-Elemente zuzuordnen.

### Einschränkungen

Pro Objektgruppe können mehrere XML-Feldzuordnungen vorhanden sein, aber nur eine XML-Rootelementzuordnung. Das Rootelement eines XML-Dokuments wendet die Zuordnungen entsprechend an.

### Informationen zu dieser Task

Wenn Sie eine XML-Feldzuordnung erstellen oder Felder in einer vorhandenen XML-Feldzuordnung hinzufügen, ändern oder löschen, werden die Änderungen wirksam, nachdem der Parser erneut gestartet wurde. Die neuen und geänderten Zuordnungen werden auf neue Daten angewendet, die nach dem Neustart des Parsers syntaktisch analysiert werden. Sie haben keine Auswirkungen auf Daten, die sich bereits im Index befinden. Sie müssen die Dokumente erneut durchsuchen und indexieren, um Dokumente zu aktualisieren, die sich bereits im Index befinden.

Diese Task verwendet das folgende XML-Musterdokument, um zu zeigen, wie Sie Personaldatensätze zuordnen und es Benutzern ermöglichen können, bestimmte Elemente direkt abzufragen.

```
<?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>
```

### Vorgehensweise

Gehen Sie wie folgt vor, um XML-Elemente in diesem Beispiel Suchfeldern zuzuordnen:

1. Bearbeiten Sie eine Objektgruppe, wählen Sie die Seite **Syntaxanalyse** aus, und klicken Sie **Feldern XML-Elemente zuordnen** an.
2. Klicken Sie **XML-Zuordnung erstellen** auf der Seite **XML-Feldzuordnungen** an. Die Seite **XML-Feldzuordnung erstellen** wird geöffnet.
3. Geben Sie in das Feld **XML-Stammelementname** den Namen des Stammelements ein: `personnel`.

Stellen Sie sicher, dass der von Ihnen in diesem Feld angegebene Name mit dem Stammelement in den XML-Dokumenten übereinstimmt, die Sie durchsuchen wollen. Bei der Syntaxanalyse und dem Indexieren von XML-Dokumenten, wählt die Unternehmenssuche die zu verwendende Zuordnung anhand des Stammelementnamens aus.

4. Geben Sie in das Feld **XML-Zuordnungsname** einen Namen für diese Gruppe XML-Feldzuordnungsregeln ein.

Nachdem Sie die Gruppe XML-Zuordnungsregeln erstellt haben, wird dieser Name auf der Seite **XML-Feldzuordnungen** angezeigt, sodass Sie den Namen auswählen können, wenn Sie Zuordnungsregeln hinzufügen, löschen oder ändern wollen.

5. Ordnen Sie das XML-Element `jobrole` dem Suchfeld `jobrole` zu:
  - a. Geben Sie `jobrole` in das Feld **Feldname** ein.

**Tipp:** Klicken Sie **Hilfe** an, um Informationen zu ASCII-Zeichen und Metazeichen zu erhalten, die in Feldnamen nicht unterstützt werden.

- b. Geben Sie `jobrole` in das Feld **XML-Elementname** ein.
- c. Wählen Sie das Markierungsfeld **Feldspezifische Suche** aus, damit Benutzer das Feld `jobrole` abfragen können.
- d. Wenn Sie sicherstellen wollen, dass nur dann eine Übereinstimmung auftritt, wenn die Suchbegriffe mit dem gesamten Wert des Felds `jobrole` übereinstimmen (d. h., dass keine anderen Wörter im Feld vorhanden sind), wählen Sie das Markierungsfeld **Vollständige Übereinstimmung** aus.
- e. Wählen Sie das Markierungsfeld **Sortierbar** aus, um es Benutzern zu ermöglichen, die Suchergebnisse nach den Werten im Feld `jobrole` zu sortieren.

Das Feld wird nur dann als sortierbares Feld angezeigt, wenn die Suchanwendung diese Funktion unterstützt. Die Mustersuchanwendung für die Unternehmenssuche enthält keine Felder, die über HTML-Elemente in der Liste der Felder zugeordnet sind, die Sie zum Sortieren von Suchergebnissen auswählen können.

- f. Wählen Sie das Markierungsfeld **Suchergebnisse** aus, um es Benutzern zu ermöglichen, die Werte des Felds `jobrole` in den Suchergebnissen anzuzeigen.
6. Ordnen Sie das XML-Element `jobroles` demselben Suchfeld zu:
    - a. Klicken Sie **Feld hinzufügen** an, um der Liste mit Feldzuordnungsregeln eine Leerzeile hinzuzufügen.
    - b. Geben Sie `jobrole` in das Feld **Feldname** ein.
    - c. Geben Sie `jobroles` in das Feld **XML-Elementname** ein.

**Tipp:** Die XML-Elementnamen müssen nicht mit den Namen der Suchfelder übereinstimmen, und es dürfen mehrere XML-Elemente demselben Suchfeld zugeordnet werden.

- d. Damit Benutzer das Feld `jobrole` abfragen können, geben Sie an, dass die Suchbegriffe vollständig mit dem Feldwert übereinstimmen müssen, verwenden Sie das Feld zum Sortieren der Suchergebnisse, zeigen Sie das Feld in den Suchergebnissen an, und wählen Sie die entsprechenden Markierungsfelder aus.
7. Ordnen Sie das XML-Element `section`, das ein Attribut `ID` mit dem Wert `expertise` aufweist, einem Suchfeld mit dem Namen `expertise` zu:
  - a. Klicken Sie **Feld hinzufügen** an, um der Liste mit Feldzuordnungsregeln eine Leerzeile hinzuzufügen.
  - b. Geben Sie `expertise` in das Feld **Feldname** ein.
  - c. Geben Sie `section` in das Feld **XML-Elementname** ein.
  - d. Geben Sie `id` in das Feld **XML-Attributname** ein.
  - e. Geben Sie `expertise` in das Feld **XML-Attributwert** ein.
  - f. Wählen Sie das Markierungsfeld **Feldspezifische Suche** aus, damit Benutzer das Feld `expertise` abfragen können.
  - g. Wenn Sie sicherstellen wollen, dass nur dann eine Übereinstimmung auftritt, wenn die Suchbegriffe mit dem gesamten Wert des Felds `expertise` übereinstimmen (d. h., dass keine anderen Wörter im Feld vorhanden sind), wählen Sie das Markierungsfeld **Vollständige Übereinstimmung** aus.
  - h. Wählen Sie das Markierungsfeld **Sortierbar** aus, um es Benutzern zu ermöglichen, die Suchergebnisse nach den Werten im Feld `expertise` zu sortieren, falls die Suchanwendung diese Funktion unterstützt.
  - i. Wählen Sie das Markierungsfeld **Suchergebnisse** aus, um es Benutzern zu ermöglichen, die Werte des Felds `expertise` in den Suchergebnissen anzuzeigen.

### Beispiele:

Wenn Sie nach allen Mitarbeitern eines Unternehmens suchen, die mit Suchprodukten arbeiten, führen Sie die folgende Abfrage durch:

```
jobrole:search
```

Wenn Sie alle Mitarbeiter eines Unternehmens suchen, die linguistisches Know-how haben, führen Sie die folgende Abfrage durch:

```
expertise:linguistics
```

---

## Funktionsweise von HTML-Suchfeldern

Ordnen Sie HTML-Metadatenelemente Suchfeldern zu, wenn Sie möchten, dass Benutzer bestimmte Metadatenbereiche von HTML-Dokumenten durchsuchen können.

Sie ordnen HTML-Metadatenelemente Suchfeldern über die Administrationskonsole für die Unternehmenssuche zu.

Wenn Sie HTML-Metadatenelemente Suchfeldern zuordnen, ermöglichen Sie Benutzern, HTML-Dokumente mit genaueren Abfragen zu durchsuchen.

### HTML-Suchfelder

Mit HTML-Suchfeldern können Benutzer Attribute von HTML-Dokumenten abfragen.

Metadatenelemente in HTML-Dokumenten ähneln Dokumentattributen insofern, als sie Informationen zum Dokument enthalten und angeben, wie dieses formatiert ist und wie über das Web darauf zugegriffen werden kann. Beispiel:

```
<meta http-equiv="Content-Type" content="text/html; charset=utf-8" />
<meta name="copyright" content="(C) Copyright IBM Corporation 2005" />
<meta name="content.owner" content="(C) Copyright IBM Corporation 2005" />
<meta name="security" content="public" />
<meta name="abstract" content="This topic describes an IBM product." />
<meta name="format" content="XHTML" />
```

Bei der Unternehmenssuche können die Namen von HTML-Metadatenelementen Suchfeldnamen zugeordnet werden. Bei der Konfiguration von Syntaxanalyseoptionen für eine Objektgruppe geben Sie an, welche HTML-Metadatenelemente welchen Suchfeldnamen zugeordnet werden sollen. Wenn Sie HTML-Metadatenelemente Suchfeldern zuordnen, können Benutzer Dokumente mit diesen Elementen finden, indem sie die Suchfeldnamen in den Abfragen angeben. Abfragen, die bestimmte Felder durchsuchen, liefern präzisere Suchergebnisse als Freiformatabfragen, die den gesamten Dokumentinhalt durchsuchen.

Beispiel: Wenn Ihre Objektgruppe HTML-Dokumente enthält und Sie angeben, dass die Metadatenelemente 'copyright' und 'abstract' als Suchfelder indiziert werden sollen, können Benutzer diese Elemente abfragen. Bei einer Suche nach `copyright:IBM` werden HTML-Dokumente gefunden, die IBM in den Copyright-Metadaten enthalten.

Wenn Sie HTML-Metadatenelemente Suchfeldern zuordnen, geben Sie an, ob alle HTML-Metadatenelemente, nur die zur Gruppe der Dublin Core-Metadatenelemente gehörenden Elemente oder nur die von Ihnen angegebenen HTML-Metadatenelemente zugeordnet werden sollen. Eine Beschreibung der Elemente in der Gruppe der Dublin Core-Metadatenelemente finden Sie auf der Website der Dublin Core Metadata Initiative:

<http://dublincore.org/documents/dcmi-terms/#H2>

Wenn Sie auswählen, dass für alle HTML-Metadatenelemente oder für alle Dublin Core-Metadatenelemente Zuordnungen erstellt werden sollen, stimmt der Standardsuchfeldname mit dem Metadatenelementnamen überein. Sie können den Standardsuchfeldnamen überschreiben und andere Suchoptionen für bestimmte Elemente angeben, indem Sie die Elemente der Liste von Elementen hinzufügen, die Sie einzeln konfigurieren wollen.

## Zuordnen von HTML-Metadatenelementen zu Suchfeldern

Wenn Sie ein HTML-Metadatenelement einem Suchfeld zuordnen, geben Sie an, welche HTML-Metadatenelemente ein Benutzer durch Angeben eines Feldnamens in einer Abfrage durchsuchen kann.

### Vorbereitung

Sie müssen ein Administrator für die Unternehmenssuche oder ein Objektgruppenadministrator der Objektgruppe sein, zu der die HTML-Dokumente gehören, um Suchfeldern HTML-Metadatenelemente zuzuordnen.

### Informationen zu dieser Task

Wenn Sie eine HTML-Feldzuordnung erstellen oder Felder in einer vorhandenen HTML-Feldzuordnung hinzufügen, ändern oder löschen, werden die Änderungen

wirksam, nachdem der Parser erneut gestartet wurde. Die neuen und geänderten Zuordnungen werden auf neue Daten angewendet, die nach dem Neustart des Parsers syntaktisch analysiert werden. Sie haben keine Auswirkungen auf Daten, die sich bereits im Index befinden. Sie müssen die Dokumente erneut durchsuchen und indexieren, um Dokumente zu aktualisieren, die sich bereits im Index befinden.

### Vorgehensweise

Gehen Sie wie folgt vor, um Suchfeldern HTML-Metadatenelemente zuzuordnen:

1. Bearbeiten Sie eine Objektgruppe, wählen Sie die Seite **Syntaxanalyse** aus, und klicken Sie **Feldern HTML-Metadaten zuordnen** an.
2. Geben Sie auf der Seite **HTML-Feldzuordnungen** an, welche HTML-Metadatenelemente Suchfeldern zugeordnet werden sollen:
  - Wenn Sie alle HTML-Elemente oder alle zur Gruppe der Dublin Core-Elemente gehörenden Elemente zuordnen, geben Sie an, ob Benutzer die Möglichkeit haben sollen, die Felder nach Feldnamen zu durchsuchen, und ob die Felder in den Suchergebnissen angezeigt werden sollen. Sie geben außerdem an, ob die Suchbegriffe des Benutzers vollständig mit dem gesamten Metadatenelementwert übereinstimmen müssen (d. h., die einzigen Wörter in dem Feld sind Wörter, die mit den Suchbegriffen des Benutzers übereinstimmen) und ob die Benutzer die Suchergebnisse nach diesem Feld sortieren können.

Das System verwendet die Namen der Metadatenelemente automatisch als Feldnamen. Wenn Sie die Standardfeldnamen für bestimmte Elemente überschreiben oder verschiedene Suchoptionen für bestimmte Elemente verwenden wollen, können Sie einen Elementnamen einem eindeutigen Feldnamen zuordnen und dann die Suchoptionen angeben, die für dieses Element verwendet werden sollen.

- Gehen Sie wie folgt vor, um bestimmte Metadatenelemente Suchfeldern zuzuordnen:
  - a. Klicken Sie **Feld hinzufügen** an, um der Liste mit Feldzuordnungsregeln eine Leerzeile hinzuzufügen.
  - b. Geben Sie einen Namen ein, der mit dem HTML-Metadatenelement verknüpft werden soll, das Sie zuordnen. Benutzer können diesen Feldnamen bei der Abfrage von HTML-Dokumenten in dieser Objektgruppe angeben.

**Tipp:** Klicken Sie **Hilfe** an, um Informationen zu ASCII-Zeichen und Metazeichen zu erhalten, die in Feldnamen nicht unterstützt werden.

- c. Geben Sie den Namen des Metadatenelements ein, das zugeordnet werden soll.
- d. Wählen Sie das Markierungsfeld **Feldspezifische Suche** aus, um Benutzern das Abfragen dieses Felds zu ermöglichen.

**Tipp:** Ist dieses Markierungsfeld nicht ausgewählt, kann das Feld nicht mit einer feldspezifischen Abfrage oder einer Freiformatabfrage durchsucht werden. Diese Maßnahme ist beispielsweise nützlich, wenn Sie die Option zum Einschließen aller HTML-Metadatenelemente oder der Dublin Core-Metadatenelemente ausgewählt haben, aber verhindern wollen, dass bestimmte Felder durchsucht werden.

- e. Wenn Sie Benutzern ermöglichen wollen, dieses Feld zu durchsuchen, und angeben, dass ein Dokument nur übereinstimmt, wenn die Suchbe-

griffe mit dem gesamten Feldwert übereinstimmen (d. h., dass keine anderen Wörter im Feld vorhanden sind), wählen Sie das Markierungsfeld **Vollständige Übereinstimmung** aus.

- f. Wählen Sie das Markierungsfeld **Sortierbar** aus, um es Benutzern zu ermöglichen, Suchergebnisse alphabetisch (Zeichenfolgesortierung) nach diesem Feld zu sortieren.

Das Feld wird nur dann als sortierbares Feld angezeigt, wenn die Suchanwendung diese Funktion unterstützt. Die Mustersuchanwendung für die Unternehmenssuche enthält keine Felder, die über HTML-Metadaten-elemente in der Liste der Felder zugeordnet sind, die Sie zum Sortieren von Suchergebnissen auswählen können.

- g. Wenn der Datentyp des Felds DECIMAL, DOUBLE, INTEGER, SHORT, TIME oder TIMESTAMP ist und Sie den Benutzern die Möglichkeit geben wollen, beim Durchsuchen dieses Felds die parametrische Suche zu verwenden oder Ergebnisse numerisch anhand des Werts dieses Felds zu sortieren, wählen Sie das Markierungsfeld **Parametrische Suche** aus.
- h. Wählen Sie das Markierungsfeld **Suchergebnisse** aus, um es Benutzern zu ermöglichen, dieses Feld in den Suchergebnissen anzuzeigen.

### Beispiel:

Benutzer können die zugeordneten Feldnamen abfragen, um nach HTML-Dokumenten mit bestimmten Metadaten zu suchen. Wenn Sie beispielsweise ein HTML-Metadaten-element namens `description` zugeordnet haben, um nach einem Feld namens `abstract` zu suchen, könnte ein Benutzer eine Abfrage wie die folgende eingeben, wenn er nach HTML-Dokumenten zum Thema ThinkPad-Computer sucht:

```
abstract:ThinkPad
```

---

## Benutzerdefinierte Textverarbeitung

Sie können die Qualität und Genauigkeit von Suchergebnissen verbessern, indem Sie benutzerdefinierte Algorithmen für die Textverarbeitung in die Objektgruppen für die Unternehmenssuche integrieren.

OmniFind Enterprise Edition unterstützt IBM UIMA (Unstructured Information Management Architecture), ein Framework zum Erstellen, Aufspüren und Einsetzen von Textanalysefunktionen. Anwendungsentwickler erstellen und testen Analysealgorithmen für den zu durchsuchenden Inhalt und erstellen anschließend ein Verarbeitungssteuerkomponentenarchiv (PEAR-Datei), das alle Ressourcen umfasst, die für die Verwendung des Archivs für die Unternehmenssuche erforderlich sind. Damit Sie Objektgruppen mit Ihren benutzerdefinierten Analysealgorithmen durchsuchen können, müssen Sie das Archiv (das die Textanalysesteuerkomponente enthält) dem System für die Unternehmenssuche hinzufügen.

Die Analyselogikkomponente in einer Textanalysesteuerkomponente wird *Annotator* (Kommentatorfunktion) genannt. Jeder Annotator führt bestimmte Tasks bei der linguistischen Analyse aus. Eine Textverarbeitungssteuerkomponente kann eine beliebige Anzahl Annotatoren enthalten, oder sie kann aus mehreren Textanalysesteuerkomponenten bestehen, von denen jede eigene benutzerdefinierte Annotatoren enthält.

Die von den Annotatoren erzeugten Informationen werden als *Analyseergebnisse* bezeichnet. Analyseergebnisse, die den von Ihnen gesuchten Informationen entsprechen, werden in eine Datenstruktur geschrieben, die *allgemeine Analysestruktur* genannt wird.

Bei der Konfiguration von Textverarbeitungsoptionen für eine Objektgruppe führen Sie die folgenden Tasks aus:

- Wählen Sie die Textanalysesteuerkomponente aus, die Sie verwenden möchten, um Dokumente in der Objektgruppe mit Kommentaren zu versehen.
- Wenn Ihre Objektgruppe XML-Dokumente mit einer aussagekräftigen Markup enthält und Sie diese Markup in Ihrer benutzerdefinierten Textanalyse verwenden möchten, können Sie der Objektgruppe Zuordnungsdateien zuordnen und die Ausgabe der XML-Zuordnung der allgemeinen Analysestruktur zuordnen. Sie können z. B. den Inhalt der Elemente <empfänger> und <kunde> den Kommentaren zu Personen in der allgemeinen Analysestruktur zuordnen. Auf diese Kommentare kann dann von Ihren benutzerdefinierten Annotatoren zugegriffen werden, die eventuell zusätzliche Informationen finden (z. B. das Geschlecht der Person). Sie können Kommentare zu Personen auch dem Index für die Unternehmenssuche zuordnen, sodass Benutzer nach Personen suchen können, ohne die ursprünglichen XML-Elemente kennen zu müssen.

Wenn Sie es Benutzern ermöglichen möchten, die ursprünglichen XML-Elemente in Abfragen anzugeben, müssen Sie keine XML-Zuordnungen definieren. Stattdessen können Sie Optionen für die Syntaxanalyse konfigurieren und eine native XML-Suche für die Objektgruppe aktivieren.

- Ordnen Sie die allgemeine Analysestruktur dem Index für die Unternehmenssuche zu, sodass Dokumente mit Kommentaren mit einer semantischen Suche durchsucht werden können.

Abhängig von den Entitäten und Abhängigkeiten, die von den Annotatoren gefunden werden, können Benutzer beispielsweise nach Konzepten suchen, die in demselben Satz auftreten (wie z. B. eine bestimmte Person und der Name eines Mitbewerbers), oder nach einem Schlüsselwort und einem Konzept (wie z. B. dem Namen "Alex" und einer Telefonnummer).

- Ordnen Sie die allgemeine Analysestruktur einer relationalen Datenbank zu. Daten können Sie IBM DB2-Tabellen oder Oracle-Tabellen zuordnen. Bei diesem Typ der Zuordnung lassen sich die Analyseergebnisse in Datenbankanwendungen (z. B. Data-Mining) verwenden. Er gibt Ihnen zudem die Möglichkeit, Daten außerhalb der Unternehmenssuche mithilfe von SQL-Abfragen zu durchsuchen.

#### **Zugehörige Konzepte**

-  Integration der benutzerdefinierten Textanalyse
-  Grundkonzepte für die Verarbeitung der Textanalyse
-  Workflow für die Integration der benutzerdefinierten Analyse
-  Textanalysealgorithmen
-  Semantische Suchanwendungen
-  Semantischer Suchabfragebegriff

## Hinzufügen von Textanalysesteuerkomponenten zum System

Wenn Sie eine benutzerdefinierte Textanalysesteuerkomponente erstellen, müssen Sie diese dem System hinzufügen, bevor Sie sie für die Unternehmenssuche verwenden können. Objektgruppen können die Steuerkomponente verwenden, um Dokumente zu analysieren und zu kommentieren sowie um die Genauigkeit der Suchergebnisse zu verbessern.

### Vorbereitung

Sie müssen ein Administrator für die Unternehmenssuche sein, um dem System Textanalysesteuerkomponenten hinzuzufügen.

### Informationen zu dieser Task

Anwendungsentwickler können ein Archiv für Verarbeitungseines (.pear) erstellen, das dem UIMA-Framework für die Textanalyse entspricht. Das Archiv enthält alle Ressourcen, die für das Durchsuchen von Objektgruppen für die Unternehmenssuche erforderlich sind. Damit Sie Objektgruppen mit Ihren benutzerdefinierten Analysealgorithmen durchsuchen können, müssen Sie das Archiv (das die Textanalysesteuerkomponente enthält) dem System für die Unternehmenssuche hinzufügen.

Nachdem Sie dem System eine Textanalysesteuerkomponenten hinzugefügt haben, können Sie deren Anzeigenamen ändern und eine Option zum Anzeigen der XML-Quelle auswählen. Die XML-Quelle zeigt an, welche Informationen von dieser Steuerkomponente generiert werden.

Ist eine Textanalysesteuerkomponente einer Objektgruppe zugeordnet, kann die Textanalysesteuerkomponente nicht vom System entfernt werden.

### Vorgehensweise

Gehen Sie wie folgt vor, um dem System für die Unternehmenssuche eine benutzerdefinierte Textanalysesteuerkomponente hinzuzufügen:

1. Klicken Sie **System** an, um die Sicht **System** zu öffnen.
2. Klicken Sie  **Bearbeiten** an, um zur Sicht für die Systembearbeitung zu wechseln.
3. Klicken Sie **Textanalysesteuerkomponenten konfigurieren** auf der Seite **Syntaxanalyse** an.
4. Klicken Sie **Textanalysesteuerkomponente hinzufügen** auf der Seite **Textanalysesteuerkomponenten** an.
5. Geben Sie auf der Seite **Textanalysesteuerkomponente hinzufügen** einen beschreibenden Namen für die neue Steuerkomponente ein. Das System verwendet diesen Anzeigenamen, um die Textanalysesteuerkomponente in der gesamten Administrationskonsole anzugeben.
6. Geben Sie die Speicherposition der .pear-Datei an. Wenn die Datei kleiner als 8 MB ist, kann sie sich auf Ihrem lokalen Computer befinden, und Sie können die Datei suchen. Wenn die Datei größer als 8 MB ist, muss sie sich auf dem Indexserver befinden, und Sie müssen den vollständig qualifizierten Pfad für die Datei eingeben.
7. Klicken Sie **OK** an. Ihre Textanalysesteuerkomponente wird auf der Seite **Textanalysesteuerkomponenten** aufgelistet.

### Zugehörige Konzepte

 Workflow für die Integration der benutzerdefinierten Analyse

 Integration der benutzerdefinierten Textanalyse

 Grundkonzepte für die Verarbeitung der Textanalyse

 XML-Markupdateien bei der Analyse und Suche

#### Zugehörige Tasks

 Erstellen eines XML-Elements für die Zuordnungsdatei der allgemeinen Analysestruktur

## Zuordnen einer Textanalysesteuerkomponente zu einer Objektgruppe

Wenn dem System für die Unternehmenssuche benutzerdefinierte Textanalysesteuerkomponenten zugeordnet wurden, können Sie eine für die Verwendung mit einer Objektgruppe auswählen. Danach können Benutzer beim Durchsuchen der Objektgruppe semantische Abfragen angeben und die Qualität und Genauigkeit der Suchergebnisse verbessern.

### Vorbereitung

Sie müssen ein Administrator für die Unternehmenssuche oder ein Objektgruppenadministrator der Objektgruppe sein, um einer Objektgruppe eine Textanalysesteuerkomponente zuzuordnen.

### Informationen zu dieser Task

Wenn dieser Objektgruppe bereits eine Textanalysesteuerkomponente zugeordnet ist, werden beim Zuordnen einer anderen Steuerkomponente die folgenden Aktionen ausgeführt:

- Wenn Sie **Keine benutzerdefinierte Analyse** auswählen, werden alle vorher definierten Textanalysezuordnungen für die Objektgruppe zurückgesetzt. Die Objektgruppe verwendet nun die Systemstandardwerte.
- Wenn Sie den Namen einer anderen benutzerdefinierten Textanalysesteuerkomponente auswählen, werden alle vorher definierten Textanalysezuordnungen für die Objektgruppe beibehalten. Wenn Sie z. B. von der Analysesteuerkomponente `steuerkomponente1` zu `steuerkomponente2` wechseln, werden in `steuerkomponente2` die für `steuerkomponente1` konfigurierten XML-Zuordnungsdateien übernommen.

### Vorgehensweise

Gehen Sie wie folgt vor, um einer Objektgruppe eine Textanalysesteuerkomponente zuzuordnen:

1. Bearbeiten Sie eine Objektgruppe, wählen Sie die Seite **Syntaxanalyse** aus, und klicken Sie **Textverarbeitungsoptionen konfigurieren** an.
2. Klicken Sie **Textanalysesteuerkomponente auswählen** an. Wenn dem System für die Unternehmenssuche keine benutzerdefinierten Textanalysesteuerkomponenten hinzugefügt wurden oder wenn die Objektgruppe derzeit die Standardanalysealgorithmen verwendet, ist der Name der Steuerkomponente **Standard**.

3. Wählen Sie auf der Seite **Textanalysesteuerkomponente für diese Objektgruppe auswählen** den Namen der Steuerkomponente aus, die Sie für die betreffende Objektgruppe verwenden wollen. Wenn keine Textanalysesteuerkomponenten verfügbar sind oder wenn Sie **Keine benutzerdefinierte Analyse** auswählen, wendet der Parser die Standardregeln für die Textanalyse an, wenn er Dokumente kommentiert und sie für den Index vorbereitet.

#### Zugehörige Konzepte

-  Workflow für die Integration der benutzerdefinierten Analyse
-  Integration der benutzerdefinierten Textanalyse
-  Grundkonzepte für die Verarbeitung der Textanalyse

## Zuordnen von XML-Elementen zur allgemeinen Analysestruktur

Wenn Ihre Objektgruppe XML-Dokumente mit einer aussagekräftigen Markup enthält und Sie diese Markup verwenden möchten, um es Benutzern zu ermöglichen, den Index für die Unternehmenssuche oder Tabellen relationaler Datenbanken über eine semantische Suche zu durchsuchen, können Sie die XML-Elemente der allgemeinen Analysestruktur zuordnen.

#### Vorbereitung

Sie müssen ein Administrator für die Unternehmenssuche oder ein Objektgruppenadministrator der Objektgruppe sein, um der allgemeinen Analysestruktur XML-Elemente zuzuordnen.

#### Einschränkungen

Die maximale Größe einer Zuordnungsdatei ist 8 MB.

#### Informationen zu dieser Task

Sie können benutzerdefinierte Zuordnungsdateien erstellen, um es den benutzerdefinierten Textanalyseprozessen zu ermöglichen, auf bestimmte Elemente in XML-Dokumenten zuzugreifen oder einige XML-Elemente einem gemeinsamen Typ für die Verwendung bei der semantischen Suche zuzuordnen. Die Zuordnungsdateien müssen dem UIMA-Framework für die Textanalyse entsprechen.

Wenn Sie einer Objektgruppe, die eine benutzerdefinierte Textanalysesteuerkomponente verwendet, Zuordnungsdateien hinzufügen, ermöglichen Sie es, dass XML-Elemente in Quelldokumenten Kommentaren in der allgemeinen Analysestruktur zugeordnet werden. Diese Kommentare können anschließend von Ihrer benutzerdefinierten Textanalysesteuerkomponente verwendet werden. Sie können die allgemeine Analysestruktur dem Index zuordnen und es Benutzern ermöglichen, die Kommentare abzufragen, wenn sie die Objektgruppe mit einer semantischen Suche durchsuchen.

Sie können z. B. den Inhalt der Elemente **Empfänger** und **Kunde** dem Kommentar **Person** in der allgemeinen Analysestruktur zuordnen. Auf diese Kommentare kann dann von Ihren benutzerdefinierten Annotatoren zugegriffen werden, die eventuell zusätzliche Informationen finden (z. B. das Geschlecht der Person). Sie können den

Kommentar **Person** auch dem Index für die Unternehmenssuche zuordnen, damit Benutzer nach Personen suchen können, ohne die ursprünglichen XML-Elemente kennen zu müssen.

Wenn Sie es Benutzern ermöglichen möchten, die ursprünglichen XML-Elemente in Abfragen anzugeben, müssen Sie keine Zuordnungsdateien definieren. Sie können statt dessen Syntaxanalyseoptionen konfigurieren und die native XML-Suche für die Objektgruppe aktivieren.

### Vorgehensweise

Gehen Sie wie folgt vor, um der allgemeinen Analysestruktur XML-Elemente zuzuordnen:

1. Bearbeiten Sie eine Objektgruppe, wählen Sie die Seite **Syntaxanalyse** aus, und klicken Sie **Textverarbeitungsoptionen konfigurieren** an.
2. Klicken Sie **Zuordnung hinzufügen** im Bereich **XML-Elemente der allgemeinen Analysestruktur zuordnen** an.
3. Geben Sie auf der Seite **XML-Elemente der allgemeinen Analysestruktur zuordnen** einen beschreibenden Anzeigenamen für die Zuordnungsdatei ein.
4. Geben Sie die Speicherposition der Datei an. Wenn sich die Zuordnungsdatei auf Ihrem lokalen System befindet, können Sie nach dem Pfad suchen. Wenn sich die Zuordnungsdatei auf dem Indexserver befindet, müssen Sie den vollständig qualifizierten Pfad eingeben.
5. Klicken Sie **OK** an. Ihre neue Zuordnungsdatei wird der Seite **Textverarbeitungsoptionen** hinzugefügt.

#### Zugehörige Konzepte

 Workflow für die Integration der benutzerdefinierten Analyse

 Integration der benutzerdefinierten Textanalyse

 Grundkonzepte für die Verarbeitung der Textanalyse

 XML-Markupdateien bei der Analyse und Suche

#### Zugehörige Tasks

 Erstellen eines XML-Elements für die Zuordnungsdatei der allgemeinen Analysestruktur

## Zuordnen der allgemeinen Analysestruktur zum Index

Sie können angeben, welche Textanalyseergebnisse aus einer allgemeinen Analysestruktur dem Index zugeordnet werden sollen und für Benutzer verfügbar sein sollen, die eine Objektgruppe mit einer semantischen Suche abfragen.

### Vorbereitung

Sie müssen ein Administrator für die Unternehmenssuche oder ein Objektgruppenadministrator der Objektgruppe sein, um dem Index die allgemeine Analysestruktur zuzuordnen.

### Einschränkungen

Die maximale Größe einer Zuordnungsdatei ist 8 MB.

## Informationen zu dieser Task

Indem Sie dem Index für die Unternehmenssuche die allgemeine Analysestruktur zuordnen, ermöglichen Sie es Benutzern, semantisch präzise Abfragen anzugeben, und verbessern damit die Qualität der Suchergebnisse.

Abhängig von den Elementen und Abhängigkeiten, die durch die Annotatoren entdeckt wurden, können Benutzer z. B. nach Konzepten, die im selben Satz vorkommen (wie z. B. eine bestimmte Person und der Name eines Mitbewerbers), oder nach einem Schlüsselwort und einem Konzept (wie z. B. dem Namen "Alex" und eine Telefonnummer) suchen.

## Vorgehensweise

Gehen Sie wie folgt vor, um dem Index die allgemeine Analysestruktur zuzuordnen:

1. Bearbeiten Sie eine Objektgruppe, wählen Sie die Seite **Syntaxanalyse** aus, und klicken Sie **Textverarbeitungsoptionen konfigurieren** an.
2. Klicken Sie **Eine Zuordnungsdatei auswählen** im Bereich **Allgemeine Analysestruktur dem Index zuordnen** an.
3. Wählen Sie auf der Seite **Zuordnungsdatei für diese Objektgruppe auswählen** die Zuordnungsdatei aus, die Sie für den Index für die Unternehmenssuche verwenden wollen.
  - Wählen Sie **Standard** aus, um die Standardzuordnungsregeln für den Index für die Unternehmenssuche zu verwenden.
  - Geben Sie die Speicherposition der Zuordnungsdatei an, um dem Index eine allgemeine Analysestruktur zuzuordnen. Wenn sich die Datei auf Ihrem lokalen System befindet, können Sie nach der Datei suchen. Wenn sich die Datei auf dem Indexserver befindet, geben Sie den vollständig qualifizierten Pfad ein.
4. Klicken Sie **OK** an. Die von Ihnen angegebene Zuordnungsdatei wird auf der Seite **Textverarbeitungsoptionen** angezeigt.

### Zugehörige Konzepte

 Workflow für die Integration der benutzerdefinierten Analyse

 Integration der benutzerdefinierten Textanalyse

 Grundkonzepte für die Verarbeitung der Textanalyse

 Indexzuordnung für angepasste Analyseergebnisse

### Zugehörige Tasks

 Erstellen der allgemeinen Analysestruktur für die Indexzuordnungsdatei

## Zuordnen der allgemeinen Analysestruktur zu einer relationalen Datenbank

Sie können angeben, welche Textanalyseergebnisse aus einer allgemeinen Analysestruktur einer relationalen Datenbank zugeordnet werden sollen, damit sie in Datenbankanwendungen verwendet werden können.

## Vorbereitung

Sie müssen ein Administrator für die Unternehmenssuche oder ein Objektgruppenadministrator der Objektgruppe sein, um die allgemeine Analysestruktur einer relationalen Datenbank zuzuordnen.

### Einschränkungen

Die maximale Größe einer Zuordnungsdatei ist 8 MB.

### Informationen zu dieser Task

Durch Zuordnen der allgemeinen Analysestruktur zu Tabellen relationaler Datenbanken werden die Daten für Datenbank Anwendungen nutzbar. Benutzer können beispielsweise SQL-Abfragen außerhalb der Unternehmenssuche angeben, um die Kommentare zu durchsuchen, die von der Textanalyse hinzugefügt wurden. Die Informationen können Sie darüber hinaus zur weiteren Datenverarbeitung verwenden, z. B. zur Verwendung in Data-Mining-Anwendungen.

Sie können jeweils eine einzelne Zuordnungsdatei verwenden, um einer relationalen Datenbank eine allgemeine Analysestruktur zuzuordnen. Wenn Sie eine neue Zuordnungsdatei hochladen, wird die aktuelle Zuordnungsdatei überschrieben.

### Vorgehensweise

Gehen Sie wie folgt vor, um die allgemeine Analysestruktur einer relationalen Datenbank zuzuordnen:

1. Bearbeiten Sie eine Objektgruppe, wählen Sie die Seite **Syntaxanalyse** aus, und klicken Sie **Textverarbeitungsoptionen konfigurieren** an.
2. Klicken Sie **Zuordnung hinzufügen** im Bereich **Allgemeine Analysestruktur einer relationalen Datenbank zuordnen** an.
3. Geben Sie auf der Seite **Allgemeine Analysestruktur einer relationalen Datenbank zuordnen** einen beschreibenden Anzeigenamen für die Zuordnungsdatei ein, mit der Sie einer relationalen Datenbank Informationen zuordnen wollen.
4. Geben Sie die Speicherposition der Zuordnungsdatei an. Wenn sich die Datei auf Ihrem lokalen System befindet, können Sie nach der Datei suchen. Wenn sich die Datei auf dem Indexserver befindet, geben Sie den vollständig qualifizierten Pfad ein.
5. Klicken Sie **OK** an. Der Anzeigename für die Zuordnungsdatei wird auf der Seite **Textverarbeitungsoptionen** angezeigt.

#### Zugehörige Konzepte

-  Workflow für die Integration der benutzerdefinierten Analyse
-  Integration der benutzerdefinierten Textanalyse
-  Grundkonzepte für die Verarbeitung der Textanalyse
-  Datenbankzuordnung für ausgewählte Analyseergebnisse

#### Zugehörige Tasks

-  Erstellen der allgemeinen Analysestruktur für die Datenbankzuordnungsdatei

---

## Konfigurieren von Threads für den Parser-Service

Wenn Sie über ausreichend Speicherressourcen verfügen, können Sie die Anzahl Threads erhöhen, die dem Parser für die Syntaxanalyse von Dokumenten zur Verfügung stehen.

### Vorbereitung

Wenn die Anzahl Ihrer Objektgruppen groß ist, ist es empfehlenswert, die Anzahl Parser-Treads zu erhöhen. Stellen Sie sicher, dass Ihr System über ausreichend Speicher verfügt, um weitere Threads zu unterstützen. Für einen Parser mit einem Thread sind 200 MB Speicher erforderlich. Für jeden weiteren Thread werden zusätzliche 50 MB Speicher benötigt.

Sie müssen ein Administrator für die Unternehmenssuche oder ein Objektgruppenadministrator der Objektgruppe sein, um die Anzahl Threads zu konfigurieren, die für den Parser gestartet wird.

### Vorgehensweise

Gehen Sie wie folgt vor, um die Anzahl Parser-Threads zu konfigurieren:

1. Bearbeiten Sie eine Objektgruppe, wählen Sie die Seite **Syntaxanalyse** aus, und klicken Sie **Syntaxanalyseoptionen konfigurieren** an.
2. Geben Sie die maximale Anzahl Parser-Threads an, die beim Starten des Parsers gestartet werden soll, und klicken Sie **OK** an.
3. Starten Sie den Parser erneut.

---

## Aktivieren der erweiterten Analyse für zusammengesetzte Begriffe

Sie können die Suchqualität verbessern, indem Sie für den Parser die Nutzung der erweiterten Analyse für zusammengesetzte Begriffe aktivieren. Bei der erweiterten Analyse werden zusammengesetzte Begriffe zerlegt, sodass jedes Teilstück als separater Begriff behandelt werden kann.

### Vorbereitung

Sie müssen ein Administrator für die Unternehmenssuche oder ein Objektgruppenadministrator der Objektgruppe sein, um Optionen für die Syntaxanalyse zusammengesetzter Begriffe anzugeben.

### Informationen zu dieser Task

In manchen Sprachen werden Begriffe aus mehreren Wörtern zu einem einzigen Wort ohne Leerzeichen zusammengefügt (*zusammengesetzte* Begriffe). So erweist sich die erweiterte Analyse und Zerlegung von zusammengesetzten Begriffen beim Durchsuchen von Sprachen wie Deutsch als hilfreich und beim Durchsuchen von Sprachen wie Koreanisch als unerlässlich.

Wenn Sie die erweiterte Analyse für zusammengesetzte Begriffe aktivieren, müssen Benutzer bei der Suche nach Begriffen keine Platzhalterzeichen verwenden, um zusammengesetzte Formen der Abfragebegriffe zu finden. So werden beispielsweise bei der Suche nach dem Begriff *Organ* möglicherweise Dokumente zurückgegeben, die den Begriff *Organspender* enthalten, während Dokumente mit dem Begriff *Organisation* nicht zurückgegeben werden. Während bei Abfragen mit Platzhalterzeichen wie z. B. bei der Abfrage *Organ\** bei der Zurückgabe alle nach

dem Abfragebegriff Organ möglicherweise vorkommenden Zeichenfolgen berücksichtigt werden, muss bei dieser Art der Suche das gesamte linguistische Teilwort in dem zusammengesetzten Begriff vorkommen.

Genauso wird z. B. der zusammengesetzte Begriff Mustermann in zwei Token aufgeteilt (Muster und Mann), die separat im Index gespeichert werden. Wird die Platzhalterabfrage Musterma\* eingegeben, kann der Suchvorgang Musterma nicht als ein Präfix eines aufgeteilten Worts erkennen. Das heißt, Dokumente, die den Begriff Mustermann enthalten, werden nicht gefunden. Wenn Sie möchten, dass Benutzer Abfragen mit Platzhalterzeichen für zusammengesetzte Begriffe verwenden können, dürfen Sie die erweiterte Analyse für zusammengesetzte Begriffe nicht aktivieren.

Benutzerdefinierte Vokabeln, wie Synonyme und Boostwörter, werden ebenfalls auf zusammengesetzte Wörter angewendet, die in der Abfrage in Form einzelner Wörter verwendet werden.

### Vorgehensweise

Gehen Sie wie folgt vor, um die erweiterte Analyse für zusammengesetzte Begriffe zu aktivieren:

1. Bearbeiten Sie eine Objektgruppe, wählen Sie **Syntaxanalyse** aus, und klicken Sie **Syntaxanalyseoptionen konfigurieren** an.
2. Wählen Sie das Markierungsfeld **Erweiterte Analyse für zusammengesetzte Begriffe aktivieren** aus, und klicken Sie **OK** an.

#### Zugehörige Konzepte

„Platzhalterzeichen in Abfragen“ auf Seite 199

 Linguistische Unterstützung für semantische Suche

 Textanalyse innerhalb der Unternehmenssuche

---

## Aktivieren der Unterstützung für eine native XML-Suche

Wenn eine Objektgruppe XML-Dokumente enthält, können Sie es Benutzern ermöglichen, beim Suchen nach Dokumenten die XML-Markup zu verwenden, indem Sie für die Objektgruppe die native XML-Suche aktivieren.

### Vorbereitung

Sie müssen die Rolle eines Administrators für die Unternehmenssuche haben oder ein Objektgruppenadministrator für die Objektgruppe sein, um die Unterstützung für das Durchsuchen von XML-Dokumenten mit einer nativen XML-Suche zu aktivieren.

### Informationen zu dieser Task

Eine native XML-Suche, wie z. B. XPath oder XML-Fragmente, kann genauere Suchergebnisse erzielen, indem die XML-Markup der Dokumente innerhalb der Abfrage genutzt wird. Benutzer können angeben, dass ein Abfragebegriff innerhalb eines bestimmten XML-Elements oder Attributs vorkommen muss.

Beispiel: Rechnungen eines Computerhändlers im XML-Format können Einträge <auftrag> enthalten, die die Elemente <unternehmen> und <computertyp> umfassen. Wenn Rechnungen abgerufen werden sollen, die Aufträge für IBM Notebooks

enthalten, kann eine Schlüsselwortsuche nach 'IBM' und 'Notebook' Dokumente abrufen, die Dell-Notebook-Computer sowie IBM Tischmodelle enthalten. Bei der XML-Suche können Sie angeben, dass 'IBM' im Element <unternehmen> und 'Notebook' im Element <computertyp> vorhanden sein muss und dass sich beide Elemente unter demselben Element <auftrag> befinden müssen. Auf diese Weise werden Rechnungen abgerufen, in denen es speziell um IBM Notebooks geht.

### Vorgehensweise

Gehen Sie wie folgt vor, um es Benutzern zu ermöglichen, eine Objektgruppe mit einer nativen XML-Suche zu durchsuchen:

1. Bearbeiten Sie eine Objektgruppe, wählen Sie die Seite **Syntaxanalyse** aus, und klicken Sie **Syntaxanalyseoptionen konfigurieren** an.
2. Wählen Sie **Benutzern das Durchsuchen von XML-Dokumenten mit nativer XML-Suche erlauben** aus.
3. Klicken Sie **OK** an.

#### Zugehörige Konzepte

 Linguistische Unterstützung für semantische Suche

 Textanalyse innerhalb der Unternehmenssuche

 Semantische Suchanwendungen

 Semantischer Suchabfragebegriff

„XML-Suchfelder“ auf Seite 148

---

## Dokumentformaterkennung

Für die Unternehmenssuche wird eine Standardzuordnung von URL-Erweiterung und MIME-Typ verwendet, um die Dokumenttypen und den Parsertyp für das jeweilige Dokument zu bestimmen.

Durch Bearbeiten der Konfigurationsdatei `parserTypes.cfg` können Sie die Standardzuordnung von URL-Erweiterungen und MIME-Typen zu Parsertypen außer Kraft setzen und erweitern. Die Datei `parserTypes.cfg` definiert Regeln für die Zuordnung von Dateierweiterungen oder MIME-Typen zu Parsertypen. Sie können beispielsweise eine Dateierweiterung wie `.content` zuordnen und angeben, dass Dokumente dieses Typs vom HTML-Parser syntaktisch analysiert werden sollen.

Unterschiedliche Dokumentformate haben unterschiedliche interne Darstellungen. Ein System für die Unternehmenssuche verwendet interne Filter und Filter anderer Anbieter für die syntaktische Analyse von Dokumenten. Darüber hinaus werden viele Dokumente mit Parser-Services syntaktisch analysiert, die auf ein bestimmtes Format spezialisiert sind.

Die Erkennung des Dokumentformats und die Zuordnung des Parsers erfolgt auf folgende Weise:

1. Der Algorithmus für die Erkennung des Dokumentformats überprüft die Erweiterung der URL-Adresse des verarbeiteten Dokuments.
2. Das System überprüft den MIME-Typ des Dokuments, der Teil der vom Crawler festgelegten Metadaten ist.
3. Das System versucht jedem Dokument den entsprechenden Parsertyp zuzuordnen. Für HTML-, Text- (TXT) und XML-Dokumente ordnet das System einen für jedes Dokumentformat spezifischen Parsertyp zu.

Für alle anderen Dokumentformate verwendet das System den Stellent-Parser. Berücksichtigen Sie, dass die Stellent-Dokumentfiltertechnologie nun Oracle gehört. Verweise auf Stellent in dieser Dokumentation sind synonym mit Verweisen auf die Technologie Oracle Outside In Content Access.

Der Stellent-Parser unterstützt mehrere hundert Dokumentformate, aber es ist nur ein Teil der Dokumentfilter für die Unternehmenssuche aktiviert. Sie können aber Konfigurationsdateien bearbeiten, damit andere Dokumenttypen vom Stellent-Parser syntaktisch analysiert werden können.

**Wichtig:** Von Ihnen hinzugefügte Dokumentfilter, die nicht zur Untergruppe der Dokumentfilter gehören, die für die Unternehmenssuche in der Standard-systemkonfiguration aktiviert wurden, wurden nicht getestet und werden nicht unterstützt.

4. Wenn keine Datei `parserTypes.cfg` zur Verfügung steht, wird die Standardzuordnung verwendet, um den Dokumenttyp und den zu verwendenden Parser zu bestimmen. Zur Bestimmung des Dokumenttyps führt das System die folgenden Schritte aus:
  - a. Vergleichen der URL-Erweiterung mit kundendefinierten Erweiterungsregeln in der Datei `parserTypes.cfg`.
  - b. Vergleichen des MIME-Typs mit kundendefinierten MIME-Typregeln in der Datei `parserTypes.cfg`.

- c. Vergleichen der URL-Erweiterung mit den Standardregeln für die Unternehmenssuche.
  - d. Vergleichen des MIME-Typs mit standardmäßigen MIME-Typregeln für die Unternehmenssuche.
5. Wenn das System das Dokumentformat eines Dokuments nicht identifizieren kann, wird das Dokument zurückgewiesen. Es wird möglicherweise eine Fehlermeldung angezeigt, die besagt, dass der Dokumenttyp nicht unterstützt wird.

Wenn "Stellent" als Parsertyp zugeordnet ist, kann eine Fehlermeldung angezeigt werden, wenn Stellent das Dokumentformat nicht erkennen kann. Der Fehler kann in folgenden Fällen auftreten:

- Das Dokument ist beschädigt.
- Das Dokument hat kein von Stellent unterstütztes Format. Zur Behebung dieses Problems müssen Sie die Dateiformate der zurückgewiesenen Dateien der Datei `stellentTypes.cfg` hinzufügen. Darüber hinaus müssen Sie die Datei `parserTypes.cfg` aktualisieren, um anzugeben, dass der MIME-Typ oder die Erweiterung der zurückgewiesenen Dateiformate dem Stellent-Parser zugeordnet werden soll.

---

## Standardmäßig unterstützte Dokumenttypen

Bei der Erkennung des Dokumentformats werden nur bestimmte Dokumenttypen ausgewertet.

Die folgenden Dokumentformate sind native Typen, die automatisch von integrierten Objektgruppen-Parser-Services gefunden und syntaktisch analysiert werden:

HTML  
Einfacher Text  
XML

Die folgenden Dokumentformate werden standardmäßig vom Stellent-Parser syntaktisch analysiert:

Adobe Portable Document Format (PDF)  
Lotus 1-2-3  
Lotus Freelance Graphics  
Lotus Word Pro  
Just System Ichitaro  
Microsoft Excel (Versionen bis 2007)  
Microsoft PowerPoint (Versionen bis 2007)  
Microsoft Visio  
Microsoft Word (Versionen bis 2007)  
Rich-Text-Format (RTF)  
StarOffice/OpenOffice Calc  
StarOffice/OpenOffice Impress  
StarOffice/OpenOffice Draw  
StarOffice/OpenOffice Writer

Open XML-Dateiformate in Microsoft Office und OpenDocument-Formate in OpenOffice werden so behandelt, dass es nicht erforderlich ist, Änderungen an der Konfigurationsdatei vorzunehmen.

Für die Syntaxanalyse in anderen Dokumenttypen müssen Sie die Konfigurationsdateien (`parserTypes.cfg` und `stellentypes.cfg`) aktualisieren, um Regeln für die Zuordnung bestimmter Dokumenttypen zu einem Objektgruppenparserservice oder Stellent-Filter anzugeben.

**Einschränkung:** Die Verarbeitung von bidirektionalem Text in PDF-Dateien zum Abgleich der logischen Lesereihenfolge des Texts liegt außerhalb des Bereichs der Stellent-Viewer-Technologie. Der Stellent-Parser übernimmt keine Garantie für die Reihenfolge von Text, der aus PDF-Dateien extrahiert wird. Bei bidirektionalem Text in PDF-Dateien ist es möglich, dass die Reihenfolge, in der der Text syntaktisch analysiert wird, nicht mit der logischen Lesereihenfolge des Texts übereinstimmt. Diese Einschränkung stellt bei der Verarbeitung von PDF-Dateien ein Problem dar, die in den Sprachen des mittleren Ostens verfasst sind, wie z. B. Hebräisch oder Arabisch, die überwiegend von rechts nach links (bidirektional) geschrieben werden.

---

## Dokumenttypen für Objektgruppenparser und Stellent-Parser

Erstellen Sie Konfigurationsdateien, um sicherzustellen, dass Dokumente in einem Crawlerbereich genau und effizient syntaktisch analysiert werden. In diesen Dateien geben Sie an, welche Dokumenttypen der Objektgruppenparser syntaktisch analysieren soll und welche Typen von Stellent-Dokumentfiltern syntaktisch analysiert werden sollen.

In einer Objektgruppe für die Unternehmenssuche werden die meisten Dokumentformate mit integrierten HTML- oder XML-Parsern verarbeitet. Bestimmte Dokumenttypen werden in der Regel nicht syntaktisch analysiert (z. B. Postscript-Dokumente), andere Dokumenttypen werden von Stellent-Parsingfunktionen übernommen (z. B. Dokumenttypen von Microsoft Word, Microsoft Excel, Microsoft PowerPoint, Lotus Freelance, Lotus 123, PDF, RTF und Ichitaro).

Da Metadaten irreführend sein können, werden Text- und HTML-Dokumente eventuell irrtümlich an den Stellent-Parser gesendet und dann an einen der integrierten Parser zurückgesendet. Dies kann die Leistung beeinträchtigen. Bei anderen Dokumenten ist es eventuell nicht möglich, den Dokumenttyp zu erkennen, sodass diese Dokumente übersprungen werden. Diese Situation können Sie vermeiden, indem Sie Konfigurationsdateien erstellen, die steuern, wo und wie verschiedene Dokumenttypen syntaktisch analysiert werden.

Das Zuordnen der Dokumenttypen zum Objektgruppenparser und zum Stellent-Parser umfasst die folgenden Tasks:

1. Konfigurieren von Dokumenttypen für den Objektgruppenparser. Dieser Schritt umfasst das Erstellen einer Konfigurationsdatei, die Dokumenttypen dem Parser zuordnet, der von einer Objektgruppe verwendet wird. Pro Objektgruppe können Sie eine dieser Konfigurationsdateien erstellen.
2. Konfigurieren von Dokumenttypen für den Stellent-Parser. Dieser Schritt umfasst das Erstellen einer Konfigurationsdatei, die Dokumenttypen den Stellent-Dokumentfiltern zuordnet, die von einer Objektgruppe verwendet werden. Pro Objektgruppe können Sie eine dieser Konfigurationsdateien erstellen.
3. Stoppen und erneutes Starten des Parsers. Damit die Änderungen wirksam werden, müssen Sie mit der Administrationskonsole für die Unternehmenssuche die Objektgruppe überwachen, für die Sie Dokumenttypen konfiguriert haben, und dann den Parser stoppen und erneut starten.

---

## Zuordnen von Dokumenttypen zu einem Objektgruppenparser

Wenn Sie bestimmte Dokumenttypen einem Objektgruppenparser zuordnen möchten, müssen Sie eine Konfigurationsdatei `parserTypes.cfg` erstellen. Die Administrationskonsole für die Unternehmenssuche bietet keine Unterstützung für diese Task.

### Vorbereitung

Sie müssen sich als Administrator für die Unternehmenssuche anmelden, um diese Task auszuführen.

### Informationen zu dieser Task

Wenn keine Konfigurationsdatei vorhanden ist, verwendet der Objektgruppenparser die Standardregeln für Parser-Services. Wenn die Konfigurationsdatei bereits vorhanden ist, geben Regeln in der Datei Folgendes an:

- Welche URL-Erweiterung und welcher MIME-Typ welchem Parsertyp zugeordnet ist.
- Vorgehensweise für die syntaktische Analyse von Dokumenten, deren Typ aufgrund fehlender Metadaten unbekannt ist.

Das Format der Datei `parserTypes.cfg` besteht aus einer Folge von Zeilen, bei der jede Zeile eine der folgenden Regeln umfasst:

#### **EXTENSION** *erweiterung parser*

Alle Dokumente, deren URL mit der angegebenen Erweiterung enden, werden von dem angegebenen Parser verarbeitet. Geben sie keinen Punkt in der Erweiterung an. Beim Vergleich wird die Groß-/Kleinschreibung nicht beachtet.

#### **CONTENTTYPE** *typ/subtyp parser*

Alle Dokumente, deren Typ mit dem angegebenen Typ/Subtyp übereinstimmen, werden von dem angegebenen Parser verarbeitet. Bei einem Inhaltstyp `t/s` eines Dokuments ist eine Übereinstimmung vorhanden, wenn `t` gleich dem Typ ist und entweder `s` gleich dem Subtyp oder der Subtyp ein Platzhalterzeichen (Stern, `*`) ist.

#### **UNKNOWN** *parser*

Alle Dokumente, deren Erweiterung und Inhaltstyp nicht bekannt sind (d. h., sie wurden nicht durch den Crawler verfügbar gemacht), werden von dem angegebenen Parser verarbeitet.

#### **DEFAULT** *parser*

Alle Dokumente, die unter keine der anderen Regeln fallen, werden vom angegebenen Parser verarbeitet.

In jedem Fall muss *parser* Folgendes angeben: `text`, `html`, `xml`, `stellent` oder `none`. Dabei bedeutet `none`, dass Dokumente dieses Typs nicht syntaktisch analysiert werden sollen.

Wenn mehrere Regeln auf ein Dokument zutreffen, hat die genaueste Regel den Vorrang, unabhängig von der Reihenfolge der Regeln:

- Eine Regel **EXTENSION** ist genauer als eine Regel **CONTENTTYPE**.
- Eine Regel **CONTENTTYPE** mit einem Subtyp ist genauer als eine mit einem Platzhalterzeichen. So hat z. B. eine Regel für den Inhaltstyp `application/postscript` Vorrang vor einer Regel `application/*`.

- Für dieselbe Erweiterung oder denselben Inhaltstyp dürfen nicht zwei Regeln gelten. In einem solchen Fall entscheidet die Implementierung, welcher Regel Vorrang gewährt wird.

### Vorgehensweise

Gehen Sie wie folgt vor, um Dokumenttypen dem Objektgruppenparser zuzuordnen:

1. Melden Sie sich am Indexserver als Administrator für die Unternehmenssuche an. Diese Benutzer-ID wurde bei der Installation von OmniFind Enterprise Edition angegeben.
2. Erstellen Sie die Konfigurationsdatei wie folgt. Dabei gibt *objektgruppen-id* die Objektgruppe an, die Sie konfigurieren möchten:  
`ES_NODE_ROOT/master_config/objektgruppen-id.parserdriver/parserTypes.cfg`
3. Bearbeiten Sie die Datei mit einem Texteditor und geben Sie die Parser-Service-Regeln an. Speichern Sie dann die Datei, und beenden Sie sie.
4. Damit die Änderungen wirksam werden, müssen Sie mithilfe der Administrationskonsole für die Unternehmenssuche den Parser für die Objektgruppe überwachen und dann den Parser stoppen und erneut starten.

### Beispiel

In diesem Beispiel verarbeitet der integrierte HTML-Parser alle Dokumente mit der Erweiterung `txt`, `htm` oder `html`, mit einem Inhaltstyp, der mit `text/` beginnt, oder Dokumente mit unbekannter Erweiterung und unbekanntem Inhaltstyp. Der integrierte XML-Parser verarbeitet alle Dokumente mit der Erweiterung `xml` oder mit dem Inhaltstyp `text/xml`. Alle anderen Dokumente, einschließlich der Dokumente mit einem Inhaltstyp, der mit `application/` beginnt, werden an den Stellent-Parser gesendet.

```
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
```

#### Zugehörige Tasks

„Ändern der Ersetzungsregeln für einige HTML-Tags“ auf Seite 171

„Zuordnen von Dokumententypen zu einem Stellent-Parser“ auf Seite 173

#### Zugehörige Verweise

„Standardserviceregeln für Objektgruppenparser“

## Standardserviceregeln für Objektgruppenparser

Wenn Sie keine Konfigurationsdatei erstellen, um dem Parser für eine Objektgruppe Dateitypen und Inhaltstypen zuzuordnen, werden für die syntaktische Analyse der Dokumente Standardregeln verwendet.

Der Objektgruppenparser verwendet folgende Standardregeln:

```
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 jttc stellent
EXTENSION jtdc stellent
EXTENSION jtdx 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

```

### Zugehörige Tasks

„Zuordnen von Dokumenttypen zu einem Objektgruppenparser“ auf Seite 168

## Syntaxanalyse für unbekannte Dokumenttypen

Wenn ein Dokumenttyp unbekannt ist (z. B. wenn ein Dokument keine Dateierweiterung hat oder dem Dokument kein MIME-Typ zugeordnet ist), können Sie Regeln konfigurieren, um zu verhindern, dass der Parser das Dokument löscht.

### Informationen zu dieser Task

Wenn der Parser ein Dateiformat nicht erkennt, versucht er, die Syntaxanalyse des Dokuments mit dem Standard-HTML-Parser auszuführen. Liegt der Inhalt nicht im HTML-Format vor, weist der Parser das Dokument zurück. Ein Datensatz mit allen zurückgewiesenen Dokumenten wird in die folgende Speicherposition auf dem Indexserver geschrieben:

```

ES_NODE_ROOT/data/objektgruppen-id/dropped_doc_logs/
dropped_docs_pd_date.log

```

### Vorgehensweise

Wenn Sie diese Situation vermeiden wollen, konfigurieren Sie den ASCII-Parser für unbekannte Dokumenttypen:

1. Melden Sie sich am Indexserver als Administrator für die Unternehmenssuche an. Diese Benutzer-ID wurde bei der Installation von OmniFind Enterprise Edition angegeben.
2. Erstellen oder bearbeiten Sie die folgende Parserkonfigurationsdatei, wobei *objektgruppen-id* die Objektgruppe angibt, die Sie konfigurieren möchten:  
ES\_NODE\_ROOT/master\_config/*objektgruppen-id*.parserdriver/parserTypes.cfg
3. Fügen Sie der Datei die folgende Regel hinzu:  
UNKNOWN text
4. Damit die Änderungen wirksam werden, müssen Sie mithilfe der Administrationskonsole für die Unternehmenssuche den Parser für die Objektgruppe überwachen und dann den Parser stoppen und erneut starten.

---

## Ändern der Ersetzungsregeln für einige HTML-Tags

Sie können die Ersetzungsregeln für HTML-Tags ändern, die der Parser für die Syntaxanalyse von HTML-Dokumenten verwendet. Die Administrationskonsole für die Unternehmenssuche bietet keine Unterstützung für diese Task.

### Vorbereitung

Sie müssen sich als Administrator für die Unternehmenssuche anmelden, um diese Task auszuführen.

### Informationen zu dieser Task

Damit er dem Index für die Unternehmenssuche Textdaten zur Verfügung stellen kann, ersetzt der HTML-Parser die Markup-Informationen (HTML-Tags) durch andere Zeichen, die die Bedeutung der Tags genauso modellieren, wie sie von einem Web-Browser gehandhabt werden. So wird z. B. aus einem Absatz-Tag (<p>) ein Absatzzeichen in der allgemeinen Analysestruktur und im Index. Die Ersetzungsregeln beeinflussen, welche Teile der Textdaten in einem Absatz vorkommen.

Sie können die Konfigurationsdatei des Parsers für eine Objektgruppe aktualisieren und einige der Ersetzungsregeln ändern, um die Bedeutung bestimmter HTML-Tags enger an den HTML-Standard zu binden.

**Tipp:** Weitere Informationen dazu, wie Sie die Handhabung von HTML-Tags in HTML-Dokumenten steuern können, finden Sie unter <http://www.ibm.com/support/docview.wss?rs=63&uid=swg27011251>.

### Vorgehensweise

Gehen Sie wie folgt vor, um die Ersetzungsregeln für HTML-Tags zu ändern:

1. Melden Sie sich am Indexserver als Administrator für die Unternehmenssuche an. Diese Benutzer-ID wurde bei der Installation von OmniFind Enterprise Edition angegeben.
2. Bearbeiten Sie die Datei ES\_NODE\_ROOT/master\_config/*objektgruppen-id*.parserdriver/collection.properties. Dabei gibt *objektgruppen-id* die Objektgruppe an, die Sie konfigurieren möchten.
3. Setzen Sie den folgenden Parameter auf true:  
trevi.tokenizer.newHtmlTagReplacement=true

Die Ersetzungsregeln für die folgenden HTML-Tags werden in die nachfolgenden Werte geändert. Alle anderen Ersetzungsregeln für HTML-Tags verwenden weiterhin die Standardregeln.

```
dfn EMPTY
div NEW LINE
q BLANK
span EMPTY
```

4. Damit die Änderungen wirksam werden, müssen Sie mithilfe der Administrationskonsole für die Unternehmenssuche den Parser für die Objektgruppe überwachen und dann den Parser stoppen und erneut starten.

#### **Zugehörige Tasks**

„Zuordnen von Dokumenttypen zu einem Objektgruppenparser“ auf Seite 168

#### **Zugehörige Verweise**

„Standard-HTML-Ersetzungsregeln“

## **Standard-HTML-Ersetzungsregeln**

Der HTML-Parser interpretiert und ersetzt Markupinformationen in HTML-Dokumenten, um Textdaten für den Index für die Unternehmenssuche bereitzustellen.

Bei der Syntexanalyse von HTML-Tags verwendet der HTML-Parser für die Unternehmenssuche die folgenden Ersetzungsregeln. Die erste Spalte enthält den Namen des HTML-Tags. Die zweite Spalte enthält den Ersetzungswert.

```
comment BLANK
a EMPTY
applet EMPTY
area EMPTY
b EMPTY
base EMPTY
big EMPTY
body EMPTY
br NEW LINE
caption EMPTY
center EMPTY
del EMPTY
dfn PARAGRAPH
dir PARAGRAPH
div BLANK
dl PARAGRAPH
em EMPTY
form EMPTY
frame PARAGRAPH
h1 PARAGRAPH
h2 PARAGRAPH
h3 PARAGRAPH
h4 PARAGRAPH
h5 PARAGRAPH
h6 PARAGRAPH
hr PARAGRAPH
iframe EMPTY
img NEW LINE
li BLANK
meta EMPTY
object EMPTY
ol BLANK
option EMPTY
p PARAGRAPH
q PARAGRAPH
samp PARAGRAPH
script EMPTY
select EMPTY
spacer BLANK
```

span	NEW LINE
strike	EMPTY
strong	EMPTY
style	EMPTY
table	PARAGRAPH
td	NEW LINE
title	PARAGRAPH
tr	NEW LINE
ul	BLANK
xmp	PARAGRAPH

### Zugehörige Tasks

„Ändern der Ersetzungsregeln für einige HTML-Tags“ auf Seite 171

---

## Zuordnen von Dokumententypen zu einem Stellent-Parser

Erstellen Sie eine Konfigurationsdatei `stellenttypes.cfg`, um anzugeben, welche Dokumenttypen von Stellent-Dokumentfiltern analysiert werden sollen. Die Administrationskonsole für die Unternehmenssuche bietet keine Unterstützung für diese Task.

### Oracle-Outside-In-Technologie:

Die Stellent-Dokumentfiltertechnologie gehört nun Oracle. Verweise auf Stellent in dieser Dokumentation sind synonym mit Verweisen auf die Technologie Oracle Outside In Content Access.

### Einschränkungen

OmniFind Enterprise Edition unterstützt Stellent-Filter für die folgenden Dokumenttypen:

- Adobe Portable Document Format (PDF)
- Lotus 1-2-3
- Lotus Freelance Graphics
- Lotus Word Pro
- Ichitaro
- Microsoft Excel
- Microsoft PowerPoint
- Microsoft Visio
- Microsoft Word
- Rich-Text-Format (RTF)

Wenn Sie einem Index für die Unternehmenssuche weitere Dokumenttypen hinzufügen wollen und die Dokumentformate von einem Stellent-Filter unterstützt werden, können Sie die Syntaxanalyseregeln in der Konfigurationsdatei `stellenttypes.cfg` konfigurieren.

**Wichtig:** Von Ihnen hinzugefügte Dokumentformate, die nicht mit den in der obigen Liste genannten Dokumenttypen übereinstimmen, wurden nicht für die Unternehmenssuche getestet und werden nicht unterstützt.

Eine vollständige Liste aller Stellent-Dokumentformate finden Sie in *Outside In Technology: Supported File Formats* unter [http://www.oracle.com/technology/products/content-management/oit/ds\\_oitFiles.pdf](http://www.oracle.com/technology/products/content-management/oit/ds_oitFiles.pdf).

### Vorbereitung

Sie müssen sich als Administrator für die Unternehmenssuche anmelden, um diese Task auszuführen.

### Informationen zu dieser Task

Die Konfigurationsdatei `stellenttypes.cfg` gibt Folgendes an:

- Akzeptanzregeln für die Dateitypen, die vom Stellent-Parser syntaktisch analysiert werden sollen. Ein Dateityp entspricht einem der Dateitypen, die von der Stellent-Bibliothek erkannt werden.
- Native Regeln für die Dateitypen, die für die Verarbeitung mit einem der integrierten Parser an den Objektgruppenparser zurückgesendet werden müssen. Diese Aktion ist erforderlich, weil der Objektgruppenparser aufgrund irreführender Metadaten eventuell irrtümlich ein Dokument an den Stellent-Parser sendet.
- Zurückweisungsregeln für die Dateitypen, die zurückgewiesen werden sollen, weil sie in der Unternehmenssuche nicht unterstützt werden.

Wenn die Konfigurationsdatei angegeben wurde, aber nicht vorhanden ist, kann der Parser nicht starten. Wenn keine Konfigurationsdatei für das Merkmal `OutsideInSupportedTypes` in der Datei `stellent.properties` angegeben wurde, werden die Standardparserregeln für Stellent-Parser verwendet.

Die Konfigurationsdatei listet Dokumenttypen auf und enthält Informationen dazu, wie sie verarbeitet werden sollen. Das Dateiformat ist eine Folge von Zeilen, wobei jede Zeile eine Regel darstellt, die mit einem der folgenden Formate übereinstimmt:

```
accept DEFAULT
accept ALL dokumenttyp
accept stellent-typ dokumenttyp
native DEFAULT
native stellent-typ dokumenttyp
reject stellent-typ
```

Dabei gilt Folgendes:

*dokumenttyp*

Der Wert für das Abfragetoken für den Dokumenttyp. Dokumente können nach dem Dokumenttyp durchsucht werden. Ein Benutzer gibt z. B. `$doctype::pdf` an, um PDF-Dokumente zu durchsuchen.

*stellent-typ*

Einer der Filtertypwerte in der Stellent-Bibliothek, wie `FI_123R1`.

#### DEFAULT

Bedeutet, dass die Liste der akzeptierten oder nativen Typen abhängig vom Regeltyp alle Standardregeln einschließt. Diese Option ermöglicht Ihnen, die Standardkonfiguration zu erweitern, statt sie zu ersetzen.

**A11** Bedeutet, dass alle Typen, die nicht explizit aufgelistet sind, mit dem angegebenen Dokumenttyp token akzeptiert werden.

Die Regeln in der Konfigurationsdatei werden wie folgt verarbeitet:

1. Wenn eine Regel `reject` für *stellent-typ* vorhanden ist, wird das Dokument nicht akzeptiert.
2. Wenn eine Regel `native` für *stellent-typ* vorhanden ist (einschließlich der Standardparserregeln, falls `native DEFAULT` angegeben wurde), wird das Dokument an den integrierten Parser zurückgesendet, und zwar zusätzlich zu dem Wert für das Token *dokumenttyp*, das durch diese Regel angegeben wird. Der

Wert von *dokumenttyp* muss entweder txt, htm oder xml sein. Diese Werte stehen für einfachen Text, HTML und XML.

3. Wenn eine Regel accept für *stellent-typ* vorhanden ist (einschließlich der Standardliste, falls accept DEFAULT angegeben wurde), wird das Dokument akzeptiert.
4. Sonst gilt: Wenn accept ALL angegeben wurde, wird das Dokument akzeptiert.
5. Andernfalls wird das Dokument zurückgewiesen und nicht syntaktisch analysiert.

Wenn der Dokumenttyp akzeptiert wird, wird der Wert für *dokumenttyp* verwendet, der in der angewendeten Regel angegeben wurde. Dieser Wert wird zusätzlich zum syntaktisch analysierten Inhalt an den Objektgruppenparser zurückgesendet.

### Vorgehensweise

Gehen Sie wie folgt vor, um dem Stellent-Parser Dokumenttypen zuzuordnen:

1. Melden Sie sich am Indexserver als Administrator für die Unternehmenssuche an. Diese Benutzer-ID wurde bei der Installation von OmniFind Enterprise Edition angegeben.
2. Bearbeiten Sie die Datei `ES_NODE_ROOT/master_config/objektgruppen-id.stellent/stellent.properties`. Dabei gibt *objektgruppen-id* die Objektgruppe an, die Sie konfigurieren möchten.
3. Geben Sie für das Merkmal `OutsideInSupportedTypes` den absoluten Pfad der Konfigurationsdatei an, die Sie erstellen.

Beispiel: Sie erstellen die folgende Konfigurationsdatei für eine einzelne Objektgruppe und speichern sie zusammen mit anderen objektgruppenspezifischen Dateien:

```
ES_NODE_ROOT/master_config/objektgruppen-id.stellent/stellenttypes.cfg
```

Oder Sie erstellen beispielsweise die folgende Konfigurationsdatei, sodass Sie sie für alle Objektgruppen die gleichen Einstellungen verwenden können, und speichern diese Datei zusammen mit anderen Dateien auf Systemebene. Achten Sie bei dieser Methode darauf, dass Sie für jede Objektgruppe diesen Pfad in der Datei `stellent.properties` angeben, wie in Schritt 2 beschrieben.

```
ES_INSTALL_ROOT/default_config/stellent/stellenttypes.cfg
```

4. Verwenden Sie einen Texteditor, um die Konfigurationsdatei zu erstellen und Stellent-Parserregeln anzugeben. Speichern Sie dann die Datei, und beenden Sie sie.
5. Damit die Änderungen wirksam werden, müssen Sie mithilfe der Administrationskonsole für die Unternehmenssuche den Parser für die Objektgruppe überwachen und dann den Parser stoppen und erneut starten.

### Beispiele

In der folgenden Konfigurationsdatei akzeptiert die Stellent-Sitzung neben der Standardliste unterstützter Dokumenttypen auch Dokumente im Microsoft Visio-Format.

```
accept DEFAULT
accept FI_VISI03 visio
accept FI_VISI04 visio
accept FI_VISI05 visio
accept FI_VISI06 visio
```

In der folgenden Konfigurationsdatei werden PostScript-Dokumente akzeptiert und können als Dokumenttyp "ps" durchsucht werden, Dokumente im X-Pixmap-Format (XPM) werden an den integrierten Textparser zurückgegeben, Dokumente im PNG-Bildformat werden zurückgewiesen, und alle anderen Dateitypen werden akzeptiert und können mit dem Dokumenttyp "other" durchsucht werden.

```
accept DEFAULT
accept FI_POSTSCRIPT ps
native FI_XPIXMAP txt
reject FI_PNG
accept ALL other
```

#### **Zugehörige Tasks**

„Zuordnen von Dokumenttypen zu einem Objektgruppenparser“ auf Seite 168

#### **Zugehörige Verweise**

„Standardsyntaxanalyseregeln für Stellent-Parser“

## **Standardsyntaxanalyseregeln für Stellent-Parser**

Wenn Sie keine Konfigurationsdatei erstellen, um Stellent-Dokumentfiltern Dateitypen zuzuordnen, verwendet der Parser Standardregeln für die syntaktische Analyse der Dokumente.

Der Stellent-Parser akzeptiert die folgenden Dokumenttypen und führt eine Syntaxanalyse für sie aus:

ACCEPT FI_123R1	123
ACCEPT FI_123R2	123
ACCEPT FI_123R3	123
ACCEPT FI_123R4	123
ACCEPT FI_123R6	123
ACCEPT FI_123R9	123
ACCEPT FI_EXCEL	xls
ACCEPT FI_EXCEL2000	xls
ACCEPT FI_EXCEL2002	xls
ACCEPT FI_EXCEL2003	xls
ACCEPT FI_EXCEL2007	xlsx
ACCEPT FI_EXCEL3	xls
ACCEPT FI_EXCEL4	xls
ACCEPT FI_EXCEL5	xls
ACCEPT FI_EXCEL97	xls
ACCEPT FI_EXTPOWERPOINT4	ppt
ACCEPT FI_EXTPOWERPOINTMAC4	ppt
ACCEPT FI_FREELANCE	prz
ACCEPT FI_FREELANCE3	prz
ACCEPT FI_ICHITAR03	jxw
ACCEPT FI_ICHITAR04	jsw
ACCEPT FI_ICHITAR08	jtd
ACCEPT FI_PDF	pdf
ACCEPT FI_PDFMACBIN	pdf
ACCEPT FI_POWERPOINT2	ppt
ACCEPT FI_POWERPOINT2000	ppt
ACCEPT FI_POWERPOINT2007	pptx
ACCEPT FI_POWERPOINT3	ppt
ACCEPT FI_POWERPOINT4	ppt
ACCEPT FI_POWERPOINT7	ppt
ACCEPT FI_POWERPOINT9597	ppt
ACCEPT FI_POWERPOINT97	ppt
ACCEPT FI_POWERPOINTMAC3	ppt
ACCEPT FI_POWERPOINTMAC4	ppt
ACCEPT FI_POWERPOINTMACB3	ppt
ACCEPT FI_POWERPOINTMACB4	ppt
ACCEPT FI_RTF	rtf
ACCEPT FI_RTFJ	rtf
ACCEPT FI_STAROFFICEWRITER8	odt

ACCEPT FI_STAROFFICEDRAW8	odg
ACCEPT FI_STAROFFICEIMPRESS8	odp
ACCEPT FI_STAROFFICECALC8	ods
ACCEPT FI_STAROFFICECALC6	sxc
ACCEPT FI_STAROFFICEDRAW6	sxd
ACCEPT FI_STAROFFICEIMPRESS6	sxi
ACCEPT FI_STAROFFICEWRITER6	sxw
ACCEPT FI_STAROFFICECALC52	sdc
ACCEPT FI_STAROFFICEIMPRESS52	sdd
ACCEPT FI_STAROFFICEWRITER52	sdw
ACCEPT FI_VISIO3	vsd
ACCEPT FI_VISIO4	vsd
ACCEPT FI_VISIO5	vsd
ACCEPT FI_VISIO6	vsd
ACCEPT FI_VISIO2003	vsd
ACCEPT FI_WINWORD1	doc
ACCEPT FI_WINWORD1COMPLEX	doc
ACCEPT FI_WINWORD1J	doc
ACCEPT FI_WINWORD2	doc
ACCEPT FI_WINWORD2000	doc
ACCEPT FI_WINWORD2002	doc
ACCEPT FI_WINWORD2003	doc
ACCEPT FI_WINWORD2007	docx
ACCEPT FI_WINWORD5J	doc
ACCEPT FI_WINWORD6	doc
ACCEPT FI_WINWORD7	doc
ACCEPT FI_WINWORD97	doc
ACCEPT FI_WORD4	doc
ACCEPT FI_WORD5	doc
ACCEPT FI_WORD6	doc
ACCEPT FI_WORDPRO	lwp
ACCEPT FI_WORDPRO97	lwp

Der Stellant-Parser sendet die folgenden Dokumenttypen für die Verarbeitung mit einem der integrierten Parser an den Objektgruppenparser zurück:

NATIVE FI_7BITTEXT	txt
NATIVE FI_ANSI	txt
NATIVE FI_ANSI8	txt
NATIVE FI_ARABIC_710	txt
NATIVE FI_ARABIC_720	txt
NATIVE FI_ARABIC_WINDOWS	txt
NATIVE FI_ASCII	txt
NATIVE FI_ASCII8	txt
NATIVE FI_CENTRALEU_1250	txt
NATIVE FI_CHINESEBIG5	txt
NATIVE FI_CHINESEGB	txt
NATIVE FI_CYRILLIC1251	txt
NATIVE FI_CYRILLICKO18	txt
NATIVE FI_EBCDIC_1026	txt
NATIVE FI_EBCDIC_273	txt
NATIVE FI_EBCDIC_277	txt
NATIVE FI_EBCDIC_278	txt
NATIVE FI_EBCDIC_280	txt
NATIVE FI_EBCDIC_284	txt
NATIVE FI_EBCDIC_285	txt
NATIVE FI_EBCDIC_297	txt
NATIVE FI_EBCDIC_37	txt
NATIVE FI_EBCDIC_500	txt
NATIVE FI_EBCDIC_870	txt
NATIVE FI_EBCDIC_871	txt
NATIVE FI_HANGEUL	txt
NATIVE FI_HEBREW_E0	txt
NATIVE FI_HEBREW_OLDPCODE	txt
NATIVE FI_HEBREW_PC8	txt
NATIVE FI_HEBREW_WINDOWS	txt
NATIVE FI_HTML	htm

NATIVE FI_HTML_ARABIC_ASMO708	htm
NATIVE FI_HTML_ARABIC_DOS	htm
NATIVE FI_HTML_ARABIC_ISO	htm
NATIVE FI_HTML_ARABIC_MAC	htm
NATIVE FI_HTML_ARABIC_WINDOWS	htm
NATIVE FI_HTML_BALTIC_ISO	htm
NATIVE FI_HTML_BALTIC_WINDOWS	htm
NATIVE FI_HTML_CENTRALEUROPEAN_DOS	htm
NATIVE FI_HTML_CENTRALEUROPEAN_ISO	htm
NATIVE FI_HTML_CENTRALEUROPEAN_MAC	htm
NATIVE FI_HTML_CENTRALEUROPEAN_WINDOWS	htm
NATIVE FI_HTML_CHINESEBIG5	htm
NATIVE FI_HTML_CHINESEEUC	htm
NATIVE FI_HTML_CHINESEGB	htm
NATIVE FI_HTML_CHINESESIMPLIFIED_EUC	htm
NATIVE FI_HTML_CHINESESIMPLIFIED_WINDOWS	htm
NATIVE FI_HTML_CHINESETRADITIONAL_WINDOWS	htm
NATIVE FI_HTML_CYRILLIC_DOS	htm
NATIVE FI_HTML_CYRILLIC_ISO	htm
NATIVE FI_HTML_CYRILLIC_KOI8R	htm
NATIVE FI_HTML_CYRILLIC_MAC	htm
NATIVE FI_HTML_CYRILLIC_WINDOWS	htm
NATIVE FI_HTML_CYRILLIC1251	htm
NATIVE FI_HTML_CYRILLICKOI8	htm
NATIVE FI_HTML_EBCDIC_1026	htm
NATIVE FI_HTML_EBCDIC_273	htm
NATIVE FI_HTML_EBCDIC_277	htm
NATIVE FI_HTML_EBCDIC_278	htm
NATIVE FI_HTML_EBCDIC_280	htm
NATIVE FI_HTML_EBCDIC_284	htm
NATIVE FI_HTML_EBCDIC_285	htm
NATIVE FI_HTML_EBCDIC_297	htm
NATIVE FI_HTML_EBCDIC_37	htm
NATIVE FI_HTML_EBCDIC_500	htm
NATIVE FI_HTML_EBCDIC_870	htm
NATIVE FI_HTML_EBCDIC_871	htm
NATIVE FI_HTML_GREEK_ISO	htm
NATIVE FI_HTML_GREEK_MAC	htm
NATIVE FI_HTML_GREEK_WINDOWS	htm
NATIVE FI_HTML_HEBREW_DOS	htm
NATIVE FI_HTML_HEBREW_ISO_VISUAL	htm
NATIVE FI_HTML_HEBREW_WINDOWS	htm
NATIVE FI_HTML_JAPANESE_MAC	htm
NATIVE FI_HTML_JAPANESE_SHIFTJIS	htm
NATIVE FI_HTML_JAPANESEEUC	htm
NATIVE FI_HTML_JAPANESEJIS	htm
NATIVE FI_HTML_JAPANESESJIS	htm
NATIVE FI_HTML_KOREAN_JOHAB	htm
NATIVE FI_HTML_KOREAN_WINDOWS	htm
NATIVE FI_HTML_KOREANHANGUL	htm
NATIVE FI_HTML_LATIN2	htm
NATIVE FI_HTML_RUSSIAN_DOS	htm
NATIVE FI_HTML_THAI_WINDOWS	htm
NATIVE FI_HTML_TURKISH_DOS	htm
NATIVE FI_HTML_TURKISH_ISO	htm
NATIVE FI_HTML_TURKISH_MAC	htm
NATIVE FI_HTML_TURKISH_WINDOWS	htm
NATIVE FI_HTML_VIETNAMESE_WINDOWS	htm
NATIVE FI_HTML_WESTERNEUROPEAN_ISO	htm
NATIVE FI_HTML_WESTERNEUROPEAN_MAC	htm
NATIVE FI_HTML_WESTERNEUROPEAN_WINDOWS	htm
NATIVE FI_HTMLUNICODE	htm
NATIVE FI_JAPANESE_EUC	txt
NATIVE FI_JAPANESE_JIS	txt
NATIVE FI_LATIN2	txt
NATIVE FI_MAC	txt
NATIVE FI_MAC8	txt

NATIVE FI_PP2KHTML	htm
NATIVE FI_SHIFTJIS	txt
NATIVE FI_UNICODE	txt
NATIVE FI_UTF8	txt
NATIVE FI_W2KHTML	htm
NATIVE FI_WML	xml
NATIVE FI_WML_CHINESEBIG5	xml
NATIVE FI_WML_CHINESEEUC	xml
NATIVE FI_WML_CHINESEGB	xml
NATIVE FI_WML_CYRILLIC1251	xml
NATIVE FI_WML_CYRILLICKO18	xml
NATIVE FI_WML_JAPANESEEUC	xml
NATIVE FI_WML_JAPANESEJIS	xml
NATIVE FI_WML_JAPANESESJIS	xml
NATIVE FI_WML_KOREANHANGUL	xml
NATIVE FI_WML_LATIN2	xml
NATIVE FI_XHTML	htm
NATIVE FI_XL2KHTML	htm
NATIVE FI_XML	xml
NATIVE FI_XML_DOCTYPE_HTML	htm

### **Zugehörige Tasks**

„Zuordnen von Dokumententypen zu einem Stellent-Parser“ auf Seite 173



---

## Sprachen- und Codepage-Unterstützung

Die Verarbeitung auf linguistischer Basis für die Unternehmenssuche wird vom Parser und von den Suchservern unterschiedlich gehandhabt.

Bei der Verarbeitung auf linguistischer Basis unterscheidet der Parser nicht zwischen Sprachen und Ländereinstellungen. Wenn ein Benutzer eine Objektgruppe durchsucht, die Dokumente in mehreren Sprachen enthält, lassen es die Suchserver aber zu, dass die Suchergebnisse auf eine bestimmte Sprache oder Ländereinstellung begrenzt wird.

Wenn beispielsweise die Metadaten in einem englischen Dokument en\_US als Ländereinstellung des Dokuments angeben, wird das Dokument sowohl als englischsprachiges Dokument (en) als auch als Dokument, das die Ländereinstellung für die Vereinigten Staaten für Englisch (en\_US) verwendet, indiziert. Durch diese Art der Indexierung können länderspezifische Informationen wie z. B. Zahlen, Datumsangaben und Uhrzeiten korrekt wiedergegeben werden. Wenn Benutzer die Objektgruppe durchsuchen, kann das Objekt unabhängig davon gefunden werden, ob nach en- oder en\_US-Dokumenten gesucht wird.

Wenn ein Dokument nur nach dem Sprachencode (z. B. en) indiziert wird, wird das Dokument nur nach dem Sprachencode und nicht nach der Ländereinstellung indiziert. Wenn Benutzer die Objektgruppe beispielsweise nach en\_US-Dokumenten durchsuchen, wird das Dokument nicht gefunden.

Ein System für die Unternehmenssuche liefert linguistische Unterstützung für die folgenden Sprachen und zweibuchstabigen Sprachencodes, die in der Norm ISO 639 dokumentiert werden:

### Alphabetische Schriften:

- en=Englisch
- sq=Albanisch
- az=Aserbaidtschänisch (Lateinisch)
- bg=Bulgarisch
- be=Weißrussisch
- ca=Katalanisch
- hr=Kroatisch
- cs=Tschechisch
- da=Dänisch
- nl=Niederländisch
- et=Estnisch
- fi=Finnisch
- fr=Französisch
- de=Deutsch
- el=Griechisch
- hu=Ungarisch
- is=Isländisch
- id=Indonesisch
- in=Indonesisch
- it=Italienisch
- kk=Kasachisch
- lv=Lettisch
- lt=Litauisch
- lo=Lao
- mk=Mazedonisch
- ms=Malaiisch
- mt=Maltesisch
- no=Norwegisch

nb=Norwegisch (Bokmal)  
pl=Polnisch  
pt=Portugiesisch  
ro=RumänischRomanian  
ru=Russisch  
sr=Serbisch (Kyrillisch)  
sh=Serbisch (Lateinisch)  
sk=Slowakisch  
sl=Slowenisch  
es=Spanisch  
sv=Schwedisch  
tr=Türkisch  
uk=Ukrainisch  
cy=Walisisch

#### **Ideographische Schriften:**

Für vereinfachtes und traditionelles Chinesisch werden erweiterte Sprachencodes statt zweibuchstabiger Codes verwendet.

zh-CN=Chinesisch (vereinfacht)  
zh-TW=Chinesisch (traditionell)  
ja=Japanisch  
ko=Koreanisch

#### **Komplexe alphabetische Schriften:**

ar=Arabisch  
as=Assamesisch  
bn=Bengalisch  
gu=Gujaratisch  
iw=Hebräisch  
he=Hebräisch  
hi=Hindi  
kn=Kannada  
ml=Malajalam  
mr=Marathi  
or=Orija  
pa=Pundjabisch  
ta=Tamilisch  
te=Telugu  
th=Thailändisch  
ur=Urdu  
vi=Vietnamesisch

Das System für die Unternehmenssuche kann viele dieser Sprachen automatisch erkennen. Darüber hinaus kann es bei Dokumenten mit einfachem Text automatisch die verwendete Codepage erkennen. Beim Konfigurieren eines Crawlers können Sie die automatische Sprach- und Codepageerkennung inaktivieren, wenn Sie explizit angeben wollen, dass eine bestimmte Sprache oder Codepage verwendet werden soll.

---

## **Automatische Spracherkennung**

Ein System für die Unternehmenssuche kann Dokumente in nahezu allen Sprachen verarbeiten.

Wenn in einem Dokument eine der nachfolgend aufgeführten Sprachen verwendet wird, kann das System die Sprache automatisch erkennen. Wenn Sie die Sprache Ihrer Dokumente kennen, können Sie bei der Konfiguration eines Crawlers die zu verwendende Sprache angeben, statt zuzulassen, dass das System die Sprache automatisch erkennt.

Arabisch  
Bulgarisch  
Tschechisch  
Chinesisch, vereinfacht

Chinesisch, traditionell  
Dänisch  
Niederländisch  
Englisch  
Finnisch  
Französisch, Kanada  
Französisch, Frankreich  
Deutsch, Deutschland  
Deutsch, Schweiz  
Griechisch  
Hebräisch  
Ungarisch  
Isländisch  
Italienisch  
Japanisch  
Koreanisch  
Norwegisch, Bokmal  
Polnisch  
Portugiesisch, Brasilien  
Portugiesisch, Portugal  
Rumänisch  
Russisch  
Spanisch  
Schwedisch  
Thailändisch  
Türkisch

---

## Automatische Codepage-Erkennung

Ein System für die Unternehmenssuche unterstützt Dokumente in einer Vielzahl von Codepages.

Bei Textdateien kann das System die nachfolgend aufgeführten Codepages automatisch erkennen. Für andere Dokumentformate verwendet das System die Metadaten im Dokument (z. B. HTML-Metadatenelemente) zur Erkennung der Codepage. Wenn Sie die Codepage Ihrer Dokumente kennen, können Sie bei der Konfiguration eines Crawlers die zu verwendende Codepage angeben, statt zuzulassen, dass das System die Codepage automatisch erkennt.

### Unicode-Codierungsformate:

UTF-8  
UTF-16BE  
UTF-16LE

### Mehrbyte-Codierungsformate:

Shift-JIS  
ISO-2022-CN  
ISO-2022-JP  
ISO-2022-KR  
GB18030  
EUC-JP  
EUC-KR

### Einzelbyte-Codierungsformate:

ISO-8859-1: Dänisch, Niederländisch, Deutsch, Englisch, Französisch, Italienisch, Norwegisch, Portugiesisch, Schwedisch, Spanisch  
ISO-8859-2: Tschechisch, Ungarisch, Polnisch, Rumänisch  
ISO-8859-5: Russisch  
ISO-8859-6: Arabisch  
ISO-8859-7: Griechisch  
ISO-8859-8: Hebräisch, Hebräisch in visueller Reihenfolge  
ISO-8859-9: Türkisch  
Windows-1250: Tschechisch, Ungarisch, Polnisch, Rumänisch  
Windows-1251: Russisch  
Windows-1252: Dänisch, Niederländisch, Deutsch, Englisch, Französisch, Italienisch,

Norwegisch, Portugiesisch, Schwedisch, Spanisch  
Windows-1253: Griechisch  
Windows-1254: Türkisch  
Windows-1255: Hebräisch  
Windows-1256: Arabisch  
KOI8-R: Russisch

Die Erkennung von Zeichensätzen ist ein ungenauer Vorgang. Bei der Erkennung der Codepage wird versucht, den Zeichensatz zu identifizieren, der am besten zu den Merkmalen der Bytedaten passt. Da dieser Prozess aber zum Teil auf Statistik beruht, kann die Richtigkeit der Ergebnisse nicht garantiert werden.

Die größte Genauigkeit wird erzielt, wenn die Eingabedaten hauptsächlich eine einzelne Sprache aufweisen. Darüber hinaus werden mindestens mehrere hundert Byte reiner Text in dieser Sprache benötigt.

Falls die erkannte Codierung nicht unterstützt wird, verwendet das System die Standardcodepage für die Objektgruppe.

---

## Linguistische Analyse von Dokumenten in chinesischer, japanischer und koreanischer Sprache

Zur Verbesserung der Abrufbarkeit von Dokumenten in chinesischer, japanischer und koreanischer Sprache können Sie Optionen für die linguistische Analyse angeben.

Bei Dokumenten in chinesischer, japanischer und koreanischer Sprache können Sie angeben, dass der Parser die N-Gram-Segmentierungsmethode für die lexikalische Analyse verwendet. Für Dokumente in chinesischer und japanischer Sprache können Sie den Parser auch so konfigurieren, dass Zeilenvorschubzeichen aus Leerraum entfernt werden.

### N-Gram-Segmentierung

Bei der Erstellung einer Objektgruppe wählen Sie den Typ der lexikalischen Analyse aus, den Sie für die Syntaxanalyse von Dokumenten verwenden möchten, die in Sprachen geschrieben sind, in denen kein Leerraum zum Begrenzen von Wörtern verwendet werden.

Bei der Unicode-basierten Segmentierung mit Leerzeichen wird Leerraum als Begrenzer zwischen Wörtern verwendet. Bei der N-Gram-Segmentierung werden sich überlappende Sequenzen einer beliebigen Anzahl von Zeichen als ein Wort angesehen. Bei Sprachen wie Chinesisch, Japanisch und Koreanisch, die keinen Leerraum als Begrenzer verwenden, können aufgrund der N-Gram-Segmentierung bessere Suchergebnisse zurückgegeben werden als bei der Unicode-basierten Segmentierung mit Leerzeichen.

Sie wählen die Segmentierungsmethode, die Sie für die Syntaxanalyse von Dokumenten verwenden möchten, bei der Erstellung einer Objektgruppe aus. Nach der Erstellung der Objektgruppe können Sie die Einstellung anzeigen, indem Sie die Optionen für die Syntaxanalyse anzeigen, Sie können sie jedoch nicht ändern.

Informationen zur Konfiguration der Unterstützung der vollständigen N-Gram-Syntaxanalyse und der Aufbereitung in Objektgruppen für die Unternehmenssuche sowie zur Handhabung von Zeichen in Objektgruppen, die für die vollständige N-Gram-Unterstützung konfiguriert sind, finden Sie unter <http://www.ibm.com/support/docview.wss?rs=63&uid=swg27011088>.

## Entfernen von Leerraum aus Text

Sie können den Parser so konfigurieren, dass er Leerraumzeichen aus Text entfernt.

### Vorbereitung

Sie müssen sich als Administrator für die Unternehmenssuche anmelden, um diese Task auszuführen.

### Informationen zu dieser Task

Wenn Sie diese Option für eine Objektgruppe aktivieren, entfernt der Parser jede Folge von Leerraumzeichen zwischen zwei Buchstabenzeichen. Das Entfernen von Leerraumzeichen ist beispielsweise sinnvoll, wenn Ihre Dokumente in einer Sprache vorliegen, die keinen Leerraum als Wortgrenze verwendet (z. B. Chinesisch oder Japanisch).

Wenn Sie den Parser für das Entfernen von Leerraumzeichen konfigurieren, können Sie angeben, ob der gesamte Leerraum nur zwischen DBCS-Zeichen (Double-Byte Character Set, Doppelbytezeichensatz) oder unabhängig vom Zeichenkontext entfernt werden soll. Die letztere Option ist beispielsweise sinnvoll, wenn Sie englischen Text in ein japanisches Dokument einfügen und den Leerraum auch aus dem englischen Text entfernen wollen.

Der Parser entfernt die folgenden Zeichen:

- Tabulatorzeichen (0x09)
- Zeilenvorschubzeichen (0x0A)
- Absatzzeichen (0x0D)

Damit die Änderung wirksam wird, stoppen Sie den Parser, und starten Sie ihn erneut. Wenn die Änderung auch für Dokumente gelten soll, die bereits im Index gespeichert wurden, müssen diese Dokumente erneut durchsucht werden. Anschließend muss der Hauptindex erneut erstellt werden.

### Vorgehensweise

Gehen Sie wie folgt vor, um Leerraum aus Text zu entfernen:

1. Melden Sie sich am Indexserver als Administrator für die Unternehmenssuche an. Diese Benutzer-ID wurde bei der Installation von OmniFind Enterprise Edition angegeben.
2. Verwenden Sie einen Texteditor, um die folgende Datei zu bearbeiten, wobei *objektgruppen-id* die ID darstellt, die bei der Erstellung der Objektgruppe für die Objektgruppe angegeben wurde (oder vom System zugeordnet wurde):  
ES\_NODE\_ROOT/master\_config/*objektgruppen-id*.parserdriver/collection.properties
3. Geben Sie an, wie Leerraum entfernt werden soll:
  - Wenn der Leerraum zwischen DBCS-Zeichen entfernt werden soll, setzen Sie den Wert des Merkmals `removeCjNewlineChars` auf `true`:

```
removeCjNewlineChars=true
```

- Wenn der Leerraum überall in den Dokumenten entfernt werden soll, setzen Sie den Wert des Merkmals `removeCjNewlineChars` auf `true`, und setzen Sie den Wert des Merkmals `removeCjNewlineCharsMode` auf `all`:

```
removeCjNewlineChars=true
removeCjNewlineCharsMode=all
```



---

## Indexverwaltung

Die Unternehmenssuche erstellt für jede Objektgruppe einen Index und pflegt ihn durch regelmäßige Aktualisierung des Inhalts, um sicherzustellen, dass Benutzer immer Zugriff auf die neuesten Informationen haben.

Sie müssen Indizes erstellen, damit die durch Crawler gesammelten Daten durchsucht werden können. Bei der Ersterstellung einer Objektgruppe erstellt die Unternehmenssuche für alle anfänglich durchsuchten Daten einen Index. Wenn die Crawler neue und geänderte Datenquellen durchsuchen, wird für den neuen Inhalt eine Aktualisierung (der sogenannte *Deltaindex*) erstellt. Die Aktualisierungen müssen schließlich in den Basisindex aufgenommen werden. Dieser Vorgang wird als *Hauptindexerstellung* bezeichnet. Immer wenn ein Deltaindex oder ein Hauptindex erstellt wird, wird der neue Inhalt auf die Suchserver kopiert und für Suchvorgänge zur Verfügung gestellt.

Crawler sammeln Daten kontinuierlich oder zu einem regelmäßig terminierten Zeitpunkt. Wenn Sie Indizes häufig aktualisieren, können Benutzer die neuesten Daten durchsuchen. Ein kontinuierlich aktualisierter Index muss schließlich erneut erstellt werden. Da ein Index immer größer wird, benötigt er immer mehr Systemressourcen. Erstellen Sie den Hauptindex regelmäßig, um eine optimale Leistung zu gewährleisten.

Wie oft Sie den Hauptindex erstellen, hängt von folgenden Faktoren ab:

- Systemressourcen (Speicherplatz des Dateisystems, Prozessorgeschwindigkeit und Hauptspeicher)
- Anzahl zu durchsuchender und erneut zu durchsuchender Dokumente
- Typ der zu durchsuchenden Daten
- Häufigkeit der Änderung von Kategorisierungsregeln (die Änderungen werden erst nach der Hauptindexerstellung wirksam)
- Häufigkeit eines erzwungenen Crawlerstarts statt der Ausführung zu einem terminierten Zeitpunkt

Bei Objektgruppen mit mehreren Millionen Dokumenten, die hauptsächlich aus Webdokumenten erzeugt werden, sollte der Hauptindex ungefähr einmal täglich erstellt und im Abstand von einer Stunde oder zwei Stunden aktualisiert werden.

Führen Sie die folgenden Tasks aus, um einen aktuellen, durchsuchbaren Index zu gewährleisten:

- Angeben von Zeitplänen für die Indexerstellung
- Ändern des Indexzeitplans
- Aktivieren und Inaktivieren des Indexzeitplans
- Konfigurieren gleichzeitig ablaufender Indexstellungen

Sie können auch die folgenden Tasks ausführen, um Optionen anzugeben, die die Indexsicht des Benutzers beeinflussen:

- Konfigurieren der Unterstützung von Platzhalterzeichen in Abfragen
- Konfigurieren von Gültigkeitsbereichen, um den Dokumentbereich einzugrenzen, den Benutzer durchsuchen können

- Ausblenden von Dokumenten in den Suchergebnissen, die von derselben Quelle stammen
- Entfernen von URIs aus dem Index

#### Zugehörige Tasks

„Überwachen der Indexaktivität für eine Objektgruppe“ auf Seite 341

„Überwachen der Indexierungswarteschlange für die Unternehmenssuche“ auf Seite 342

---

## Terminieren von Indexerstellungen

Sie können Zeitpläne für die Hauptindexerstellung und für die Aktualisierung des Index mit neuem Inhalt angeben.

### Vorbereitung

Sie müssen ein Administrator für die Unternehmenssuche oder ein Objektgruppenadministrator der Objektgruppe sein, um eine Indexerstellung zu terminieren.

### Informationen zu dieser Task

Terminieren Sie eine regelmäßige Indexerstellung, um sicherzustellen, dass Benutzer immer auf die aktuellen Informationen in den durchsuchten Quellen zugreifen können. Bei der Hauptindexerstellung wird der gesamte Index erneut erstellt. Die Indexierungsprozesse lesen alle Daten, die von Crawlern erfasst und vom Parser syntaktisch analysiert wurden. Bei der Deltaindexerstellung werden Informationen, die seit der letzten Hauptindexerstellung durchsucht wurden, zu durchsuchbaren Informationen.

Standardmäßig ist die Option zum Terminieren von Indexerstellungen ausgewählt. Diese Option weist den Terminierungsprozess an, Tasks zum Erstellen von Haupt- und Deltaindices zu terminieren, wenn das System für die Unternehmenssuche gestartet wird. Sie können die Auswahl des Markierungsfelds **Beim Systemstart aktivieren** jederzeit zurücknehmen, wenn Sie eine terminierte Indexerstellung nicht ausführen wollen. Sie müssen den Zeitplan z. B. inaktivieren, um Probleme zu beheben.

Um Systemressourcen zu konservieren und die Leistung zu verbessern, prüft das System automatisch, ob Änderungen aufgetreten sind, die auf den Index angewendet werden müssen. Sind keine auf den Index anzuwendenden Änderungen vorhanden, wird die terminierte Anforderung der Indexerstellung verworfen.

### Vorgehensweise

Gehen Sie wie folgt vor, um Indexerstellungen zu terminieren:

1. Bearbeiten Sie eine Objektgruppe, wählen Sie die Seite **Index** aus, und klicken Sie **Indexerstellungen terminieren** an.
2. Geben Sie auf der Seite **Indexerstellungen terminieren** im Bereich **Zeitplan für eine Deltaindexerstellung angeben** die folgenden Optionen an, um anzugeben, wie oft der Index mit neuem Inhalt aktualisiert wird:
  - a. Geben Sie im Bereich **Starten am** in den Feldern **Jahr, Monat, Tag, Stunde** und **Minute** an, wann der erste Deltaindex erstellt werden soll.
  - b. Geben Sie im Bereich **Aktualisierungsintervall** in den Feldern **Tage, Stunden** und **Minuten** an, wie oft Deltaindices erstellt werden sollen.

In der Regel sollten Deltaindizes häufig erstellt werden, zum Beispiel stündlich oder alle zwei Stunden. Geben Sie je nach der Häufigkeit von Quelleninhaltsänderungen ein niedrigeres bzw. höheres Intervall an. Sie können z. B. jede Stunde (0 Tage, 1 Stunde) oder alle 12 Stunden (0 Tage, 12 Stunden) angeben.

3. Im Bereich **Zeitplan für die Hauptindexerstellung angeben** geben Sie an, wie oft der Index vollständig neu erstellt wird. Verwenden Sie hierfür die folgenden Optionen:
  - a. Geben Sie im Bereich **Starten am** in den Feldern **Jahr, Monat, Tag, Stunde** und **Minute** an, wann der Hauptindex zum ersten Mal erstellt werden soll.
  - b. Geben Sie im Bereich **Aktualisierungsintervall** in den Feldern **Tage, Stunden** und **Minuten** an, wie oft der Hauptindex erstellt werden soll.

In der Regel sollte der vollständige Index häufig erstellt werden, zum Beispiel alle 24 Stunden. Geben Sie je nach der Häufigkeit von Quelleninhaltsänderungen ein niedrigeres bzw. höheres Intervall an. Sie können z. B. alle 12 Stunden (0 Tage und 12 Stunden) oder alle zweieinhalb Tage (2 Tage und 12 Stunden) angeben.
4. Klicken Sie **OK** an.

## Ändern des Indexzeitplans

Sie können den Zeitplan für die Indexerstellung ändern.

### Vorbereitung

Sie müssen die Rolle eines Administrators für die Unternehmenssuche haben oder ein Objektgruppenadministrator für diese Objektgruppe sein, um einen Indexzeitplan ändern zu können.

### Vorgehensweise

Gehen Sie wie folgt vor, um den Indexzeitplan zu ändern:

1. Bearbeiten Sie eine Objektgruppe, wählen Sie die Seite **Index** aus, und ändern Sie die entsprechenden Werte in den Feldern **Monat, Tag, Jahr** und **Stunde**. Geben Sie an, wie oft der Index aktualisiert und wie oft der Hauptindex erstellt werden soll.
2. Klicken Sie **Anwenden** an.

## Aktivieren und Inaktivieren der Indexzeitpläne

Sie können die Zeitpläne für die Indexerstellung aktivieren und inaktivieren.

### Vorbereitung

Sie müssen die Rolle eines Administrators für die Unternehmenssuche haben oder ein Objektgruppenadministrator für diese Objektgruppe sein, um einen Indexzeitplan aktivieren bzw. inaktivieren zu können.

### Informationen zu dieser Task

Sie können einen Indexzeitplan inaktivieren, wenn die Ausführung einer terminierten Indexerstellung verhindert werden soll. Die Inaktivierung des Zeitplans kann z. B. dann notwendig sein, wenn Sie statt der Indexerstellung am terminierten Datum und Zeitpunkt Fehler beheben müssen.

Sie können den Zeitplan beim Bearbeiten oder Überwachen einer Objektgruppe aktivieren bzw. inaktivieren.

### Vorgehensweise

1. Führen Sie die folgenden Schritte aus, um den Zeitplan für einen Index durch das Bearbeiten einer Objektgruppe zu aktivieren bzw. inaktivieren:
  - a. Bearbeiten Sie die zu ändernde Objektgruppe.
  - b. Wählen Sie auf der Seite **Index** das Markierungsfeld **Beim Systemstart aktivieren** aus bzw. nehmen Sie seine Auswahl zurück, um den Zeitplan für die Indexaktualisierung zu aktivieren bzw. inaktivieren.
  - c. Wählen Sie das Markierungsfeld **Beim Systemstart aktivieren** aus bzw. nehmen Sie seine Auswahl zurück, um den Zeitplan für die Hauptindexerstellung zu aktivieren oder zu inaktivieren.
  - d. Klicken Sie **Anwenden** an.
2. Führen Sie die folgenden Schritte aus, um den Zeitplan für einen Index durch das Überwachen einer Objektgruppe zu aktivieren bzw. inaktivieren:
  - a. Überwachen Sie die zu ändernde Objektgruppe.
  - b. Wenn ein Index terminiert ist, aber nicht zum terminierten Datum und Zeitpunkt erstellt werden soll, klicken Sie  **Zeitplan inaktivieren** auf der Seite **Index** an. Der Index wird erst erstellt, wenn Sie den Zeitplan wieder aktivieren oder  **Start** anklicken, um die Indexerstellung zu starten.
  - c. Wenn ein Index terminiert ist, aber der Zeitplan für seine Erstellung inaktiviert ist, klicken Sie  **Zeitplan aktivieren** an.  
Der Index wird bis zur Erstellung am im Indexzeitplan angegebenen Datum und Zeitpunkt in eine Warteschlange gestellt.

## Konfigurieren von gleichzeitig ablaufenden Indexerstellungen

Sie können die Verwendung von Indexierungsressourcen steuern, indem Sie angeben, von wie vielen Objektgruppen Indexerstellungsanforderungen gleichzeitig verarbeitet werden können. Wenn genügend Systemressourcen zur Verfügung stehen, können Sie die Suchqualität verbessern, indem Sie angeben, dass der Index zur gleichen Zeit aktualisiert werden soll, zu der der Hauptindex erstellt wird.

### Vorbereitung

Sie müssen ein Administrator für die Unternehmenssuche sein, um Indexerstellungsoptionen für das System anzugeben.

### Informationen zu dieser Task

Die Unternehmenssuche kann mehrere Indizes gleichzeitig erstellen, indem Ressourcen von mehreren Objektgruppen gemeinsam genutzt werden. Dadurch wird ermöglicht, dass Indexerstellungsanforderungen für mehrere Objektgruppen parallel verarbeitet werden. Durch gemeinsame Nutzung der Prozesse können Sie sicherstellen, dass die Erstellung eines sehr großen Index nicht die Verfügbarkeit anderer Indizes blockiert, die in der Warteschlange auf ihre Erstellung warten.

Wird eine Indexerstellung angefordert oder terminiert, wird sie in die Indexierungswarteschlange gestellt und wartet auf ihre Verarbeitung. Da jede Objektgruppe über einen eigenen Index verfügt, können sich mehrere Indexerstellungsanforderungen von verschiedenen Objektgruppen gleichzeitig in der Indexierungswarteschlange befinden. Beim Konfigurieren der Indexierungs-

optionen für das System geben Sie an, wie viele Objektgruppen Indexierungsressourcen gemeinsam nutzen und ihre Anforderungen parallel verarbeiten lassen können.

Ferner können Sie angeben, dass die Anforderungen zur Aktualisierung eines Index zeitgleich mit der Hauptindexerstellung für die Objektgruppe verarbeitet werden sollen. Wenn Sie diese Option aktivieren, werden die Suchserver während der länger dauernden Hauptindexerstellung mit den neuesten Dokumenten aktualisiert (mittels des Deltaindex). Die Indexerstellung ist jedoch ein ressourcenintensiver Prozess. Während einer Indexerstellung wird eine große Menge System Speicher und Plattenspeicherplatz belegt. Wenn Sie diese Option aktivieren und nicht genügend Plattenspeicherplatz oder System Speicher zur Verfügung steht, kann sich das negativ auf die Gesamtleistung des Systems auswirken.

Wenn Sie die Anzahl gleichzeitig ablaufender Indexerstellung erhöhen, werden Anforderungen zur Indexerstellung, die sich bereits in der Warteschlange befinden, nicht automatisch gestartet. Die Änderung wirkt sich auf neue Indexerstellung aus, die nach der Änderung dieses Werts in die Warteschlange eingereiht werden.

Wenn Sie die Anzahl gleichzeitig ablaufender Indexerstellung verringern, werden aktuelle Indexerstellung nicht automatisch gestoppt. Ihre Änderung wird nach Beendigung der aktuellen Indexerstellung wirksam, wodurch der Start der in die Warteschlange eingereihten Indexerstellung möglich wird.

### Vorgehensweise

Gehen Sie wie folgt vor, um Indexerstellungsoptionen für das System anzugeben:

1. Klicken Sie **System** an, um die Sicht **System** zu öffnen.
2. Klicken Sie  **Bearbeiten** an, um zur Sicht für die Systembearbeitung zu wechseln.
3. Klicken Sie **Indexierungsoptionen konfigurieren** auf der Seite **Index** an.
4. Geben Sie auf der Seite **Indexierungsoptionen auf Systemebene** die Anzahl Objektgruppen ein, die Systemressourcen gemeinsam nutzen und deren Indexerstellungsanforderungen parallel verarbeitet werden sollen.

Die Anzahl Objektgruppen, die Indexierungsressourcen gemeinsam nutzen, darf die Anzahl Objektgruppen Ihres Systems für die Unternehmenssuche nicht übersteigen. Wenn Sie z. B. fünf Objektgruppen haben, muss die angegebene Anzahl kleiner-gleich fünf sein.

5. Wenn genügend Systemressourcen zur Unterstützung mehrerer gleichzeitig ablaufender Indexerstellung für einzelne Objektgruppen vorhanden sind, können Sie die Option auswählen, die das gleichzeitige Ausführen von Deltaindexerstellung und Hauptindexerstellung ermöglicht.

**Einschränkung:** Wenn Sie diese Option auswählen, wenn gerade eine Indexerstellung für eine beliebige Objektgruppe ausgeführt wird, kann die Indexerstellung fehlschlagen. Bevor Sie gleichzeitig ablaufende Delta- und Hauptindexerstellung aktivieren, müssen Sie alle terminierten Indexerstellung inaktivieren und warten, bis zurzeit ausgeführte Indexerstellung gestoppt werden (bzw. diese gegebenenfalls vor der Fertigstellung beenden). Nachdem Sie diese Option aktiviert haben, aktivieren Sie die terminierten Indexerstellung, die Sie inaktiviert hatten.

---

## Indexerstellung nur nach der Erkennung von Änderungen

Eine terminierte Indexerstellung wird nur ausgeführt, wenn Änderungen erkannt wurden, die auf den Index angewendet werden müssen. Sie können den Befehl **startIndexBuild** verwenden, um eine Indexerstellung zu starten, und auswählen, ob das System nach Änderungen suchen soll.

### Vorbereitung

Wenn Sie manuell angeben wollen, dass eine Indexerstellung nur ausgeführt werden soll, wenn Änderungen erkannt wurden, müssen Sie sich als Administrator für die Unternehmenssuche anmelden.

### Einschränkungen

Das System sucht erst nach Änderungen, die sich möglicherweise auf den Index auswirken, wenn der Hauptindex mindestens zweimal erstellt wurde. Diese Implementierung erfasst Änderungen, die erst erkannt werden können, nachdem der Hauptindex mindestens einmal erstellt wurde, wie z. B. Änderungen an Bereichsdefinitionen oder Regeln für das Ausblenden von Suchergebnissen.

### Informationen zu dieser Task

Indexerstellung verbrauchen Systemressourcen, wie CPU, Speicher und Ein-/Ausgabe-Bandbreite. Um Systemressourcen zu konservieren und die Leistung zu verbessern, kann das System für die Unternehmenssuche ermitteln, ob Änderungen aufgetreten sind, die auf den Index angewendet werden müssen, sodass der Index nur erstellt wird, wenn es erforderlich ist, Änderungen anzuwenden. Das System kann die folgenden Ereignisse für eine Indexänderung erkennen:

#### Nur Hauptindexerstellung

- Entfernen von URIs nach Mustern, wodurch Dokumente aus dem Index entfernt werden.
- Entfernen eines Crawlers aus dem System, wodurch Dokumente aus dem Index entfernt werden.

#### Haupt- und Deltaindexerstellung

- Modifizieren der Definitionen des Bereichs oder der Regeln zum Ausblenden von Suchergebnissen.
- Direktes Hinzufügen oder Entfernen von Dokumenten durch Crawlerinstellungen oder die Push-API.
- Modifizieren von Boostfaktoren, die die statische Rangfolge der Dokumentbewertung im Index beeinflussen.
- Modifizieren der Regeln, wie Platzhalterzeichen in Abfragen verwendet werden können.

Alle terminierten Anforderungen für die Indexerstellung unterliegen der Erkennung von Änderungen. Die Indexerstellung wird nur ausgeführt, wenn anstehende Änderungen vorhanden sind, die auf den Index angewendet werden müssen. Werden keine Änderungen erkannt, wird die Anforderung für die Indexerstellung verworfen.

Wenn Sie unabhängig von aufgetretenen Änderungen eine Indexerstellung erzwingen wollen, können Sie die Administrationskonsole für die Unternehmenssuche verwenden, um die Indexerstellung zu starten. Sie können auch den Befehl **startIn-**

**dexBuild** verwenden, um die Indexerstellung manuell zu starten, und angeben, ob das System nach Änderungen suchen soll, ehe es mit der Verarbeitung der Anforderung fortfährt.

### Vorgehensweise

Gehen Sie wie folgt vor, um eine Indexerstellung über die Befehlszeile zu starten und anzugeben, dass die Erstellung nur ausgeführt werden soll, wenn Änderungen ermittelt wurden:

1. Melden Sie sich als Administrator für die Unternehmenssuche an. Melden Sie sich in einer Konfiguration mit mehreren Servern auf dem Indexserver an.
2. Geben Sie den folgenden Befehl ein:

```
esadmin controller startIndexBuild -optionen
```

Optionen:

**-cid** *objektgruppen-id*

Die Objektgruppen-ID der Objektgruppe, für die Sie die Indexerstellung starten wollen.

**-buildType** *erstellungstyp*

Gibt den Typ des zu erstellenden Index an. Gültige Werte sind main und delta.

**-detectChanges**

Optional. Veranlasst, dass die Anforderung für die Indexerstellung prüft, ob Änderungen vorliegen. Die Indexerstellung wird nur ausgeführt, wenn Änderungen ermittelt werden, die auf den Index angewendet werden müssen.

Im folgenden Beispiel wird die Anforderung zum Starten einer Hauptindexerstellung für die Objektgruppe col\_1 nur ausgeführt, wenn Änderungen ermittelt werden, die auf den Index angewendet werden müssen:

```
esadmin controller startIndexBuild -cid col_1 -buildType main
-detectChanges
```

---

## Stoppen von Indexstellungen

Mithilfe des Befehls **stopIndex** können Sie die Erstellung eines Haupt- und eines Deltaindex stoppen, ohne die Administrationskonsole für die Unternehmenssuche verwenden zu müssen.

### Vorgehensweise

Gehen Sie wie folgt vor, um eine Indexerstellung mithilfe eines Befehls anstatt über die Administrationskonsole zu stoppen:

1. Melden Sie sich als Administrator für die Unternehmenssuche an. Melden Sie sich bei einer Installation mit mehreren Servern auf dem Indexserver an.
2. Geben Sie den folgenden Befehl ein:

```
esadmin stopIndex -cid objektgruppen-id -buildType typDabei gilt Folgendes:
```

**-cid** *objektgruppen-id*

Gibt die Objektgruppen-ID der Objektgruppe an, die den Index enthält.

*typ*

Gibt den Typ der zu stoppenden Indexerstellung an. Zulässige Werte sind main oder delta.

### Beispiel

```
esadmin stopIndex -cid coll -buildType delta
```

---

## Optionen zum Beeinflussen der durchsuchbaren Sicht des Index

Nach der Indexierung der Dokumente können Sie Optionen für das Durchsuchen der Dokumente und die Anzeige der Dokumente in den Suchergebnissen festlegen.

Sie können die folgenden Tasks ausführen, um Optionen anzugeben, die die Indexsicht des Benutzers beeinflussen:

- Konfigurieren der Unterstützung für Platzhalterzeichen in Abfragebegriffen. Sie können Unterstützung für Platzhalterzeichenabfragen in den Index einbauen, oder Sie können Optionen festlegen, um die Abfragebegriffe während der Abfrageverarbeitung zu erweitern.
- Konfigurieren von Gültigkeitsbereichen, um den Dokumentbereich einzugrenzen, den Benutzer durchsuchen können. Wenn Benutzer die Objektgruppe durchsuchen, werden nur die Dokumente durchsucht, die zum Gültigkeitsbereich gehören, und nicht der gesamte Index.
- Ausblenden von Dokumenten von derselben Quelle in den Suchergebnissen. Dokumente, die einem URI oder einem URI-Muster im Index entsprechen, können Sie gruppieren und nur die obersten Ergebnisdokumente als Suchergebnis anzeigen (Benutzer können Optionen festlegen, mit denen sie die ausgeblendeten Ergebnisdokumente anzeigen können).
- Entfernen von URIs aus dem Index. Sie müssen Benutzer möglicherweise vorübergehend daran hindern, bestimmte Dokumente im Index zu durchsuchen.

Für einige Crawlertypen und Objektgruppen, die die Sicherheit nicht aktivieren, wird die Erkennung doppelter Dokumente verwendet, um zu verhindern, dass Benutzern mehrere identische oder fast identische Dokumente in den Suchergebnissen angezeigt werden.

## Indexierte Optionen für das Durchsuchen von Dokumenten

Wenn Sie Optionen für das Durchsuchen von mit dem Crawler durchsuchten Daten konfigurieren oder XML- und HTML-Metadatenelemente zum Durchsuchen von Feldern zuordnen, geben Sie an, wie die Dokumente durchsucht und in den Suchergebnissen angezeigt werden.

Die angegebenen Suchoptionen werden mit den Dokumenten im Index gespeichert. Mit diesen Optionen können Sie einschränken, welche Abfragen ein Benutzer ausführen kann und welche Suchergebnisse angezeigt werden.

### Crawleroptionen:

Wenn Sie einen Crawler für das Durchsuchen von Datenquellen konfigurieren, die Felder enthalten, können Sie mit den folgenden Optionen angeben, ob ein Feld durchsucht werden kann, wie es durchsucht werden kann und ob es in den Suchergebnissen zurückgegeben wird:

- Freiformatsuche
- Feldspezifische Suche
- Vollständige Übereinstimmung

- Sortierbar
- Parametrische Suche
- Suchergebnisse
- Dokumentinhalt

#### **Optionen für das Zuordnen von XML- und HTML-Feldern:**

Wenn Sie den Parser konfigurieren und angeben, dass XML-Elemente und HTML-Metadatenelemente durchsuchbaren Feldern im Index zugeordnet werden sollen, geben Sie die folgenden Optionen an:

- Feldspezifische Suche
- Vollständige Übereinstimmung
- Sortierbar
- Suchergebnisse

Wenn Sie Suchoptionen für bestimmte HTML-Metadatenelemente konfigurieren (und nicht für alle Elemente oder Elemente, die sich in der Gruppe der Dublin Core-Metadatenelemente befinden), können Sie außerdem angeben, dass Felder, die numerische Werte enthalten, mit einer parametrischen Suche durchsucht werden können.

### **Freiformatsuche**

Der Index für die Unternehmenssuche ist ein Volltextindex mit Inhalt aus verschiedenen Datenquellen. Sie können den Inhalt durchsuchen, indem Sie eine einfache Abfrage in natürlicher Sprache angeben. Die Suchvorgänge durchsuchen die Felder und den Dokumentinhalt, um Dokumente zu finden, die für die Abfrage relevant sind.

Wählen Sie beim Konfigurieren eines Crawlers das Markierungsfeld **Freiformatsuche** aus, damit Felder mit einer Freiformatsuche durchsucht werden können. Wählen Sie dieses Markierungsfeld und das Markierungsfeld **Feldspezifische Suche** aus, um Titel-, Schlüsselwort- und Beschreibungsfelder zu durchsuchen.

#### **Beispiel 1:**

Eine Freiformatsuche kann so einfach wie die folgende Abfrage sein:

interessanter Text

Sie können in die Abfrage besondere Notationen aufnehmen, um anzugeben, welche Wörter in einem Dokument vorkommen müssen bzw. nicht vorkommen dürfen. Sie können z. B. einem Wort ein Pluszeichen (+) voranstellen. Damit geben Sie an, dass ein Dokument dieses Wort enthalten muss, um eine Übereinstimmung zu finden. Stellen Sie einem Wort ein Minuszeichen (-) voran, um Dokumente mit diesem Wort aus den Suchergebnissen auszuschließen. Umschließen Sie mindestens zwei Wörter durch Anführungszeichen ("), um nach einem genauen Ausdruck zu suchen.

#### **Beispiel 2:**

In der folgenden Freiformatabfrage tritt nur dann eine Übereinstimmung auf, wenn ein Dokument den genauen Ausdruck science fiction enthält und das Wort robot ausschließt:

+"science fiction" -robot

## Feldspezifische Suche

Mit einer feldspezifischen Suche können Sie das Abfrageobjekt auf bestimmte Datenfelder und Metadatenfelder eines Dokuments beschränken. Sie können z. B. angeben, dass bestimmte Wörter im Titel eines Dokuments vorhanden sein müssen.

Wenn Sie den Crawler oder Feldzuordnungsoptionen für XML- und HTML-Elemente konfigurieren, wählen Sie das Markierungsfeld **Feldspezifische Suche** aus, damit Felder nach Feldname durchsucht werden können. Wählen Sie dieses Markierungsfeld und das Markierungsfeld **Freiformatsuche** aus, um Titel-, Schlüsselwort- und Beschreibungsfelder zu durchsuchen.

### Beispiel:

Nehmen Sie bei der Angabe einer feldspezifischen Suche in der Unternehmenssuche in Ihre Abfrage den Feldnamen und das Wort bzw. den Ausdruck auf, der in diesem Feld vorhanden sein muss.

Die folgende Abfrage sucht nach Dokumenten, die das Wort `ibm` und den Ausdruck `unternehmenssuche` im Titelfeld enthalten müssen:

```
title:ibm title:"unternehmenssuche"
```

## Vollständige Übereinstimmung

Eine Suche nach vollständiger Übereinstimmung kann die Qualität der Suchergebnisse verbessern, da sie Ihnen ermöglicht, genaue Abfragen anzugeben. Mit einer Suche nach vollständiger Übereinstimmung, können Sie Felder und XML-Elemente abfragen und nur die Dokumente abrufen, in denen der gesamte Feldwert oder XML-Elementwert mit den Abfragebegriffen übereinstimmt. Wenn der Wert des Felds oder des Elements weniger Inhalt oder zusätzlichen Inhalt aufweist, gilt dies nicht als Übereinstimmung.

**Tip:** Bei der Untersuchung, ob eine vollständige Übereinstimmung vorliegt, konvertiert das System die Abfragebegriffe in Kleinbuchstaben, entfernt zusätzliche Leerzeichen in der Abfragezeichenfolge und führt eine Mustererkennung für Platzhalterzeichen aus. Es findet jedoch keine Lemmatisierung oder Synonymsuche statt, und Stoppwörter werden nicht entfernt. Für eine vollständige Übereinstimmung müssen bei einem XML-Element der Elementname ohne verschachtelte Elemente und der Gesamtwert des Elements genau mit den Abfragebegriffen übereinstimmen.

Sie können Suchfelder nach vollständigen Übereinstimmungen durchsuchen, indem Sie die Abfragesyntax für die Unternehmenssuche (SI-API) oder die XML-Frag2-Abfragesyntax verwenden (XPath-Abfragen werden nicht unterstützt). Ein Gleichheitszeichen (=) vor den Abfragebegriffen gibt an, dass eine Suche nach vollständiger Übereinstimmung ausgeführt werden soll.

Wenn Sie einen Crawler oder Feldzuordnungsoptionen für XML- und HTML-Elemente konfigurieren, wählen Sie das Markierungsfeld **Vollständige Übereinstimmung** aus, damit Felder nach vollständigen Übereinstimmungen durchsucht werden können. Wenn Sie beim Konfigurieren von Syntaxanalyseoptionen für eine Objektgruppe Benutzern ermöglichen, XML-Dokumente mit einer nativen XML-Suche zu durchsuchen, können alle XML-Elemente nach vollständigen Übereinstimmungen mit den Suchbegriffen durchsucht werden.

### Beispiel 1:

Ein Feld mit dem Namen `color` enthält den Wert `dark blue`.

- Die folgende Abfrage nach vollständiger Übereinstimmung findet eine Übereinstimmung, weil die Abfrage keine weiteren Begriffe enthält:  
`color:"dark blue"`
- Die folgende Abfrage nach vollständiger Übereinstimmung findet keine Übereinstimmung, weil die Abfrage zusätzlich zu `dark blue` das Wort `skirt` enthält:  
`color:"dark blue skirt"`
- Die folgende Abfrage nach vollständiger Übereinstimmung findet keine Übereinstimmung, weil das Feld `color` außerdem das Wort `dark` enthält:  
`color:=blue`

### Beispiel 2:

Ohne vollständige Übereinstimmung gibt die folgende XMLFrag2-Abfrage möglicherweise Dokumente zurück, die `<diagnosis>intraductal carcinoma comedo type</diagnosis>` oder andere Begriffe enthalten, die nicht genau den Abfragebegriffen entsprechen:

```
@xmlf2::'<diagnosis>intraductal carcinoma</diagnosis>'
```

Mit vollständiger Übereinstimmung stellt die folgende XMLFrag2-Abfrage sicher, dass nur die Dokumente zurückgegeben werden, in denen der gesamte Inhalt des XML-Elementwerts mit den Abfragebegriffen übereinstimmt:

```
@xmlf2::'=<diagnosis>intraductal carcinoma</diagnosis>'
```

## Sortierbar

Wenn eine Datenquelle Felder enthält oder wenn Benutzer XML- oder HTML-Dokumente durchsuchen, können Sie ermöglichen, dass die Ergebnisse nach den Werten in einem bestimmten Feld sortiert werden. In der Mustersuchanwendung für die Unternehmenssuche werden die Namen aller Felder aufgelistet, die als sortierbare Felder konfiguriert wurden. Benutzer können auswählen, dass Ergebnisse alphabetisch (nach einer Zeichenfolgesortierung) statt nach Relevanz oder Dokumentdatum sortiert werden sollen, indem sie eines der aufgelisteten Felder auswählen. Benutzer können außerdem auswählen, ob die Dokumente in aufsteigender oder absteigender Reihenfolge sortiert werden sollen.

Ergebnisdokumente, die das Sortierfeld nicht enthalten, werden am Ende der Suchergebnisse aufgelistet. Ergebnisdokumente, die das Sortierfeld enthalten, aber indiziert wurden, bevor das Feld als sortierbares Feld konfiguriert wurde, werden ebenfalls am Ende der Suchergebnisse aufgelistet.

Damit Benutzer Suchergebnisse alphabetisch nach den Werten in einem Feld sortieren können, wählen Sie beim Konfigurieren des Crawlers oder von Feldzuordnungsoptionen für XML- und HTML-Elemente das Markierungsfeld **Sortierbar** aus. Wenn das Feld numerische Werte enthält, wählen Sie das Markierungsfeld **Parametrische Suche** aus, um anzugeben, dass die Feldwerte für die numerische Sortierung der Suchergebnisse verwendet werden können.

## Parametrische Suche

Eine parametrische Suche ist eine Art von feldspezifischer Suche, mit der Sie vergleichende oder auswertende Abfragen für numerische Felder und Datumsfelder

sowie Metadaten ausführen können. Sie können z. B. nach Dokumenten suchen, die eine bestimmte Größe haben oder nach einem bestimmten Datum geschrieben wurden. Sie können auch mit Attributen, die größer als, kleiner als oder gleich einem angegebenen Wert sind, nach Dokumenten suchen.

Damit Ergebnisse numerisch nach dem Wert eines Felds sortiert werden können, müssen Sie das Feld für die parametrische Suche aktivieren.

Damit ein Feld mit einer parametrischen Suche durchsucht werden kann oder damit Ergebnisse numerisch sortiert werden können, wählen Sie beim Konfigurieren des Crawlers oder von Feldzuordnungsoptionen für bestimmte HTML-Metadatenelemente das Markierungsfeld **Parametrische Suche** aus.

### Beispiel 1:

Die folgende Abfrage sucht nach Elementen, die genau 50 Euro (bzw. die für das Feld price indexierte Währungseinheit) kosten:

```
#price::=50
```

### Beispiel 2:

Die folgende Abfrage sucht nach Dokumenten, deren Dateigröße größer als 1024, aber kleiner-gleich 2048 ist:

```
#filesize::>1024<=2048
```

## Suchergebnisse

Vielleicht möchten Sie bestimmte Felder durchsuchen, diese in den Suchergebnissen aber nicht anzeigen, oder vielleicht möchten Sie ein Feld in den Suchergebnissen anzeigen, obwohl Sie für dieses Feld keine Abfrage ausgeführt haben. Sie könnten zur Erstellung eines aussagekräftigen Berichts beispielsweise Finanzdaten abfragen, wollen die Mitarbeitergehälter aber nicht in den Ergebnissen anzeigen, in denen auch die Mitarbeiternamen aufgeführt sind.

Damit ein Feld in den Suchergebnissen angezeigt werden kann, wählen Sie beim Konfigurieren des Crawlers oder von Feldzuordnungsoptionen für XML- und HTML-Elemente das Markierungsfeld **Suchergebnisse** aus.

## Dokumentinhalt

Bei bestimmten Dokumenttypen, wie z. B. Webdokumenten, wird das gesamte Dokument als Inhalt interpretiert. Bei anderen Dokumenttypen, wie z. B. Dokumenten mit Feldern, können Sie angeben, welche Felder nützlichen Inhalt statt Metadaten enthalten.

Zum Angeben, dass ein Feld Dokumentinhalt darstellt, wählen Sie beim Konfigurieren des Crawlers das Markierungsfeld **Dokumentinhalt** aus. Wenn die Markierungsfelder **Dokumentinhalt** und **Freiformatsuche** beide ausgewählt sind, wird der Wert des Felds verwendet, um doppelte Dokumente zu erkennen, und er wird Teil des dynamischen Dokumentzusammenfassungsbereichs der Suchergebnisse.

### Zugehörige Konzepte

 Abfragesyntax

## Erkennung von doppelten Dokumenten

Die Erkennung von doppelten Dokumenten ist ein Verfahren, das verwendet wird, um zu verhindern, dass Suchergebnisse mehrere Dokumente mit denselben oder ähnlichen Inhalten anzeigen.

Die Suchqualität kann sich verschlechtern, wenn mehrere Kopien desselben (oder fast desselben) Dokuments in den Suchergebnissen aufgelistet werden. Die Analyse auf doppelte Dokumente wird nur ausgeführt, wenn die folgenden zwei Bedingungen gleichzeitig zutreffen:

- Die Objektgruppe verwendet das linkbasierte Rangfolgemodell. Dieses Modell wird auf Crawler angewendet, die Websites durchsuchen, wie z. B. der Web-Crawler oder der WebSphere Portal-Crawler.
- Die Objektgruppensicherheit ist inaktiviert.

Während der globalen Analyse erkennt der Indexierungsprozess doppelte Dokumente, indem er den Dokumentinhalt jedes Dokuments durchsucht. Weisen zwei Dokumente denselben Dokumentinhalt auf, werden sie wie Duplikate behandelt.

Wenn Sie möchten, dass bei der Erkennung von doppelten Dokumenten auch die Dokumentmetadaten berücksichtigt werden, müssen Sie bei der Konfiguration der Crawler für die Objektgruppe das Markierungsfeld **Dokumentinhalt** auswählen und Optionen für die Crawlersuche in Metadaten angeben. In diesem Fall durchsucht der Crawler die Metadatenfelder wie Dokumentinhalt und schließt die Metadaten in die Inhaltsanalyse auf doppelte Dokumente ein. Eine ähnliche Analyse wird ausgeführt, wenn Sie Optionen für die Syntaxanalyse von HTML- und XML-Dokumenten konfigurieren und das Markierungsfeld **Dokumentinhalt** auswählen.

Wenn Sie angeben, dass ein Feld oder Metadatenfeld Dokumentinhalt darstellt, wird der Inhalt dieser Felder der dynamischen Zusammenfassung des Dokuments in den Suchergebnissen hinzugefügt, was sich darauf auswirken kann, ob ein Dokument in den Suchergebnissen angezeigt wird. Wenn die Erkennung von Beinaheduplikaten in der Suchanwendung aktiviert ist (das Merkmal `NearDuplicateDetection` in der Methode `setProperty` ist auf `Yes` gesetzt), werden Dokumente mit ähnlichen Titeln und Zusammenfassungen unterdrückt, wenn ein Benutzer die Suchergebnisse anzeigt. Die Benutzer können jedoch einen Link anklicken, um die unterdrückten ähnlichen Dokumente anzuzeigen.

In einer Gruppe mit Kopien eines Dokuments ist ein Dokument das Hauptdokument und die anderen sind die Duplikate. Alle Dokumente der Gruppe mit Kopien haben dieselbe kanonische Darstellung des Inhalts. Während der Indexierung wird der Inhalt (die Token) des Hauptdokuments indiziert. Für die Kopien werden nur die Metadaten token indiziert. Wird das Hauptdokument aus dem Index gelöscht, wird die nächste Kopie zum neuen Hauptdokument. Durchsucht ein Benutzer die Objektgruppe, wird nur das Hauptdokument zurückgegeben.

### Zugehörige Konzepte

„Analyse von doppelten Dokumenten und Objektgruppensicherheit“ auf Seite 282

## Platzhalterzeichen in Abfragen

Sie können Benutzern die Verwendung von Platzhalterzeichen in Abfragebegriffen und die Suche nach Wörtern, die mit einem bestimmten Muster übereinstimmen, ermöglichen.

Ein Abfragebegriff mit Platzhalterzeichen ist ein Begriff, der einen Stern (\*) enthält. Wenn ein Benutzer eine Abfrage mit einem Platzhalterzeichen übergibt, umfassen die Suchergebnisse alle Dokumente im Index, die mit dem Abfragebegriff übereinstimmen, sowie alle Dokumente im Index, die dem durch das Platzhalterzeichen dargestellten Muster entsprechen. Das abschließende Platzhalterzeichen im Abfragebegriff `zei*` kann z. B. mit `zeichnen`, `zeigen` und `zeitweise` übereinstimmen.

Beim Konfigurieren der Platzhalterzeichenoptionen für einen Index können Sie wählen, ob Benutzer Platzhalterzeichen in Abfragen verwenden dürfen, und falls ja, wie diese Unterstützung bereitgestellt wird:

- Sie können alle Teile eines Dokuments nach Wörtern durchsuchen lassen, die mit dem Platzhalterzeichenmuster übereinstimmen, oder Sie können die Mustererkennung auf Felder begrenzen.
- Sie können alle Felder für die Unterstützung von Abfragen mit Platzhalterzeichen aktivieren oder die Mustererkennung auf bestimmte Felder begrenzen.
- Sie können das Platzhalterzeichen auf das letzte Zeichen in einem Abfragebegriff begrenzen (abschließendes Platzhalterzeichen), oder Sie können angeben, dass das Platzhalterzeichen an beliebigen Stellen des Abfragebegriffs verwendet werden kann. (Das Platzhalterzeichen kann in Feldnamen nicht verwendet werden.)
- Abhängig von der Position des Platzhalterzeichens können Sie festlegen, wie Abfragebegriffe erweitert werden können (Abfragebegriffe mit Platzhalterzeichen werden auf alle übereinstimmenden Begriffe im Index erweitert). Im Index können alle möglichen Begriffserweiterungen gespeichert werden, oder Suchvorgänge können Begriffe während der Abfrageverarbeitung erweitern.

Alle Änderungen an den Einstellungen für Platzhalterzeichen werden nach der nächsten Hauptindexerstellung wirksam.

## Indexerweiterung

Wenn Sie Begriffserweiterungen in den Index aufnehmen möchten, legen Sie fest, wie viele führende Zeichen in einem Wort mit dem Platzhalterzeichenmuster in einem Abfragebegriff übereinstimmen müssen, damit das Wort als Übereinstimmung gewertet wird. Nur Abfragebegriffe mit mindestens dieser Zeichenanzahl (ohne \*) geben Ergebnisse zurück. Wenn Sie z. B. 4 angeben, muss der Abfragebegriff mindestens 4 Zeichen enthalten, damit das Wort als Übereinstimmung gewertet wird.

Wenn Sie 4 angeben, stimmt das Wort `technisch` mit dem Abfragebegriff `tech*` überein, nicht jedoch mit dem Abfragebegriff `te*`.

Während einer Delta- oder Hauptindexerstellung werden alle möglichen Erweiterungen für jeden Begriff in einem Dokument zusätzlich zu den ursprünglichen Begriffen indexiert. Ein Vorteil dieser Methode liegt darin, dass keine zusätzliche Zeit erforderlich ist, um die Begriffe während der Abfrageverarbeitung zu erweitern. Bei dieser Methode wird jedoch der Index vergrößert. Das bedeutet, es müssen genügend Systemressourcen verfügbar sein, um diesen größeren Index speichern zu können.

Diese Methode ist am besten geeignet, wenn die Objektgruppe relativ klein ist oder wenn Platz und Zeit für die Indexerstellung nicht so wichtig sind wie die Abfrageantwortzeit. Zum Beispiel könnten Sie diese Methode wählen, um einen Katalog oder ein Mitarbeiterverzeichnis zu durchsuchen.

Diese Methode steht jedoch nur dann zur Verfügung, wenn Sie die Unterstützung für abschließende Platzhalterzeichen aktivieren. Wenn Sie die Unterstützung für Platzhalterzeichen aktivieren, die an beliebigen Stellen in Abfragebegriffen auftreten können, ist die Auswahl der Option zur Aufnahme von Begriffserweiterungen in den Index nicht möglich.

## **Abfrageerweiterung**

Zum Erweitern von Abfragen und zum Anwenden von Mustererkennungsregeln, wenn Benutzer Abfragen mit Platzhalterzeichen übergeben, geben Sie an, wie viele Variationen eines Abfragebegriffs eine Übereinstimmung darstellen. Wenn Sie z. B. 50 angeben, können bis zu 50 Variationen eines Abfragebegriffs als Übereinstimmung mit dem Abfragebegriff gewertet werden.

Zur Verdeutlichung dieses Beispiels: Der Abfragebegriff Tech\* stimmt mit den Wörtern Technik, Techniker, Technologie und bis zu 50 verschiedenen Wörtern überein, die mit den Zeichen Tech beginnen.

Obwohl die Abfrageerweiterung nur einen geringen Einfluss auf der Größe des Index hat, kann sie die Abfrageleistung verringern. Die Suchvorgänge müssen alle möglichen Erweiterungen eines Abfragebegriffs mit Platzhalterzeichen bis zu dem von Ihnen in den Platzhalterzeicheneinstellungen angegebenen Grenzwert iterieren.

Diese Methode ist am besten geeignet, wenn die Objektgruppe relativ groß ist und Platz und Zeit für die Indexerstellung minimiert werden müssen. Sie könnten diese Methode z. B. für E-Mail-Repositorys wählen, bei denen der Index mit den sich schnell ändernden Dokumenten mithalten muss, die Abfrageantwortzeit jedoch nicht so wichtig ist.

Diese Methode steht unabhängig davon zur Verfügung, ob Sie die Unterstützung für abschließende Platzhalterzeichen oder für Platzhalterzeichen aktivieren, die an beliebigen Stellen des Abfragebegriffs auftreten können.

## **Unterstützung für Platzhalterzeichen in Abfragen**

Die für einen Abfragebegriff mit Platzhalterzeichen zur Verfügung stehende Gruppe von Erweiterungen enthält alle Begriffe im Index, die durch Ersetzen des Platzhalterzeichens durch beliebige Zeichenfolgen abgerufen werden können. Diese Gruppe wird wie folgt definiert:

- Wenn eine Objektgruppe Platzhalterzeichen unterstützt, die an beliebigen Stellen des Abfragebegriffs auftreten können, werden alle Abfragebegriffe mit einem Stern als Begriffe mit Platzhalterzeichen interpretiert.
- Die Gruppe kann höchstens die maximale Anzahl Erweiterungen enthalten, die vom Administrator für die Unternehmenssuche konfiguriert werden. Wenn der Index eine größere Anzahl Erweiterungen enthält, werden diese ignoriert. (Die Suchergebnisse zeigen an, ob Platzhalterweiterungen ignoriert wurden.)
- Wenn die Unterstützung für Platzhalterzeichen auf eine bestimmte Feldgruppe begrenzt ist, enthält die Gruppe nur Begriffe, die in einem der angegebenen Felder auftreten. Ein Begriff muss nur in einem der Felder in mindestens einem Dokument im Index auftreten.
- Wenn der Abfragebegriff ein feldspezifischer Begriff ist, muss das Platzhalterzeichen nach der Feldkennung verwendet werden (z. B. `feldname:*sphere`). Der Feldname darf keinen Doppelpunkt (:) enthalten.
- Wenn die Unterstützung für Platzhalterzeichen auf eine bestimmte Feldgruppe begrenzt ist, muss der Feldname im Abfragebegriff mit Platzhalterzeichen mit

einem der Felder übereinstimmen, die in der Administrationskonsole für die Unternehmenssuche angegeben sind. Ansonsten werden keine Erweiterungen für diesen Begriff gefunden.

- Platzhalterzeichen werden nur für Klartextbegriffe und nicht für XML-Elementnamen, Attributnamen oder Attributwerte unterstützt. Ein nur aus einem Platzhalterzeichen bestehender Begriff wird nicht unterstützt.

### **Auswirkung von Platzhalterzeichen auf den Index**

Die Unterstützung für Platzhalterzeichen, die auf einer Indexerweiterung basiert, steigert die Größe des Index und die für die Indexerstellung erforderliche Zeit.

Bei einer Indexerweiterung wird jedes Präfix eines Begriffs und der Begriff selbst indiziert. Beispielsweise werden die folgenden Begriffe für den Begriff support indiziert:

```
s su sup supp suppo suppor support
```

Die Anzahl der im Index gespeicherten Begriffe wächst mit der durchschnittlichen Wortlänge. Eine Indexkomprimierung reduziert zwar die Größe des Index, aber nicht entscheidend. Die für die Indexerstellung erforderliche Zeit nimmt mit der durchschnittlichen Wortlänge zu.

Ein Index englischer Dokumente wächst um einen Faktor von ungefähr vier, da die durchschnittliche Wortlänge in der englischen Sprache fünf bis sechs Zeichen beträgt. Ein Index mit N-Gram-Aufbereitung wächst um ca. das Doppelte, da jedes N-Gram zwei Zeichen enthält.

Die Indexerweiterung für die Unterstützung von Platzhalterzeichen wird in den folgenden Situationen empfohlen:

- Die Objektgruppe ist klein genug, dass der für den erweiterten Index zusätzlich erforderliche Speicherplatz und Zeitaufwand nicht zu einem Leistungsproblem führt.
- Alle möglichen Erweiterungen von Platzhalterzeichen werden in die Suchergebnisse aufgenommen, um den Anforderungen der Benutzer (bzw. des Unternehmens) zu entsprechen.

Sie können die Auswirkung auf den Index reduzieren, indem Sie eine Mindestpräfixlänge für die Platzhalterzeichenerweiterung festlegen. Wenn beispielsweise die Mindestpräfixlänge auf drei gesetzt wird, werden die Präfixe s und su für das Wort support nicht indiziert und ein englischer Index wächst nur um den Faktor drei statt vier.

Mit der Methode der Abfrageerweiterung für die Unterstützung von Platzhalterzeichen werden keine Präfixe in den Index geschrieben. Begriffe werden erweitert, wenn die Abfrage übergeben wird, und der Index wächst nur um eine kleine Datenstruktur, die zur Unterstützung dieser Erweiterung erforderlich ist. In der Regel ist ein Index mit Abfrageerweiterung zwischen 10 % und 20 % größer als ein Index ohne Unterstützung von Platzhalterzeichen, und es wird weniger als 10 % mehr Zeit für die Indexerstellung benötigt. Die Konfiguration der maximalen Anzahl von Erweiterungen hat keine Auswirkung auf die Größe des Index oder die Zeit für die Indexerstellung.

### **Konfigurieren von Optionen für Platzhalterzeichen in Abfragen**

Wenn Sie die Indexierungsoptionen für eine Objektgruppe für die Unternehmenssuche konfigurieren, können Sie angeben, ob Sie es den Benutzern ermöglichen wollen, Platzhalterzeichen in Abfragebegriffen zu verwenden.

## Vorbereitung

Sie müssen ein Administrator für die Unternehmenssuche oder ein Objektgruppenadministrator der Objektgruppe sein, zu der der Index gehört, um die Optionen für Platzhalterzeichen zu konfigurieren.

### Informationen zu dieser Task

Wenn Sie Optionen für Platzhalterzeichen angeben, werden Ihre Änderungen bei der nächsten Hauptindexerstellung wirksam.

### Vorgehensweise

Gehen Sie wie folgt vor, um die Unterstützung von Platzhalterzeichen in Abfragen zu konfigurieren:

1. Bearbeiten Sie eine Objektgruppe, wählen Sie die Seite **Index** aus, und klicken Sie **Optionen für Platzhalterzeichen konfigurieren** an.
2. Wählen Sie auf der Seite **Optionen für Platzhalterzeichen** das Markierungsfeld **Platzhalterzeichen in Abfragen unterstützen** aus.
3. Optional: Sie können angeben, dass Platzhalterzeichen in Abfragen unterstützt werden sollen, mit denen Text mit freiem Format durchsucht wird. Zum Beispiel gibt die Freiformatabfrage `Tech*`, die keine benannten Felder durchsucht, nur dann erweiterte Ergebnisse (z. B. Technik oder Techniker) zurück, wenn dieses Markierungsfeld ausgewählt ist.
4. Geben Sie an, welche Felder Platzhalterzeichen unterstützen:
  - Wenn Sie angeben wollen, dass Platzhalterzeichen nicht in Abfragen verarbeitet werden können, mit denen Felder durchsucht werden, wählen Sie **Keine Felder** aus.
  - Wenn Sie alle Felder in einem Dokument für Abfragen mit Platzhalterzeichen aktivieren möchten, wählen Sie **Alle Felder** aus.
  - Wenn Sie die Unterstützung für Platzhalterzeichen auf bestimmte Felder einschränken möchten, wählen Sie **Bestimmte Felder** aus, und geben Sie die Feldnamen ein. Erweiterte Ergebnisse werden nur für die angegebenen Felder zurückgegeben. Zum Beispiel gibt die Abfrage `author:john*` nur dann erweiterte Ergebnisse zurück, wenn Sie für das Feld `author` Unterstützung für Platzhalterzeichen aktivieren.
5. Geben Sie an, ob das Platzhalterzeichen an der letzten Stelle des Abfragebegriffs (abschließendes Platzhalterzeichen) auftreten soll oder ob das Platzhalterzeichen uneingeschränkt an beliebigen Stellen des Abfragebegriffs verwendet werden kann.

Bei der Auswahl von Position und Typ des Platzhalterzeichens müssen Sie ferner angeben, wie die Unterstützung für Platzhalterzeichen aktiviert werden soll. Klicken Sie **Hilfe** in der Administrationskonsole an, wenn Sie nähere Informationen wünschen.

## Bereiche

Konfigurieren Sie einen Bereich, wenn Benutzern eine begrenzte Sicht einer Objektgruppe dargestellt werden soll.

Ein Bereich ist eine Gruppe zusammengehöriger URIs in einem Index. Wenn Sie einen Bereich konfigurieren, begrenzen Sie die Dokumente, die Benutzer in der Objektgruppe einsehen können. Wenn Benutzer die Objektgruppe durchsuchen,

werden nur die Dokumente im Bereich durchsucht und nicht der gesamte Index. Damit Sie diese Funktionalität nutzen können, müssen Ihre Suchanwendungen das Durchsuchen von Bereichen unterstützen.

Beim Erstellen eines Bereichs können Sie einen URI-Bereich im Index angeben, den Benutzer durchsuchen können. Durch das Begrenzen durchsuchbarer Dokumente können Sie sicherstellen, dass Dokumente in den Suchergebnissen auf die von den Benutzern gesuchten Informationen zugeschnitten sind.

Sie können z. B. einen Bereich erstellen, der die URIs für Ihre Abteilung "Technische Unterstützung" enthält, und einen weiteren Bereich, der die URIs für Ihre Personalabteilung enthält. Wenn Ihre Suchanwendung Bereiche unterstützt, können Benutzer in der Abteilung "Technische Unterstützung" Dokumente aus dem Bereich "Technische Unterstützung" abrufen, und Benutzer in der Personalabteilung können Dokumente aus dem Bereich "Personalwesen" abrufen.

Sie können beliebig viele Bereiche erstellen. Das Erstellen zu vieler Bereiche kann jedoch die Leistung beeinträchtigen. Konfigurieren Sie Bereiche so, dass die meisten Suchanforderungen nur ein oder zwei Bereiche zu filtern brauchen. Da Bereiche gesamte URIs oder URI-Muster enthalten können, kann das gleiche Dokument zu mehreren Bereichen gehören.

Beim Konfigurieren von Bereichen ist es unter Umständen erforderlich, den Hauptindex zweimal zu erstellen, bevor die Änderungen wirksam werden. Wenn Sie vor der ersten Erstellung des Hauptindex für die Objektgruppe Bereiche konfigurieren, können Benutzer zwar die Objektgruppe durchsuchen, jedoch nicht die Bereichsdaten in den Suchergebnissen anzeigen. Erstellen Sie den Hauptindex erneut, um sicherzustellen, dass die Suchergebnisse den URI-Bereich des Bereichs wiedergeben.

Wenn der Hauptindex erstellt wurde, bevor Sie Bereiche konfigurieren, werden die Änderungen nach der nächsten Hauptindexerstellung wirksam.

## Konfigurieren von Bereichen

Wenn Sie einen Bereich für eine Objektgruppe für die Unternehmenssuche konfigurieren, geben Sie die URIs oder URI-Muster eines Dokumentbereichs im Index an, den ein Benutzer durchsuchen darf.

### Vorbereitung

Sie müssen ein Administrator für die Unternehmenssuche oder ein Objektgruppenadministrator der Objektgruppe sein, zu der die Bereiche gehören, um Bereiche zu konfigurieren.

### Informationen zu dieser Task

Wenn Ihre Suchanwendungen die Unterstützung von Bereichen aktivieren, können Benutzer beim Durchsuchen der Objektgruppe nur die Dokumente durchsuchen, die mit den URIs übereinstimmen, die die Bereichsgrenzen definieren.

Beim Konfigurieren von Bereichen ist es unter Umständen erforderlich, den Hauptindex zweimal zu erstellen, bevor die Änderungen wirksam werden. Wenn Sie vor der ersten Indexerstellung für die Objektgruppe Bereiche konfigurieren, können Benutzer zwar die Objektgruppe durchsuchen, jedoch nicht die Bereichsdaten in

den Suchergebnissen anzeigen. Erstellen Sie den Hauptindex erneut, um sicherzustellen, dass die Suchergebnisse den URI-Bereich des Bereichs wiedergeben.

Wenn der Hauptindex erstellt wurde, bevor Sie Bereiche konfigurieren, werden die Änderungen nach der nächsten Hauptindexerstellung wirksam.

### Vorgehensweise

Gehen Sie wie folgt vor, um einen Bereich zu konfigurieren:

1. Bearbeiten Sie eine Objektgruppe, wählen Sie die Seite **Index** aus, und klicken Sie **Bereiche konfigurieren** an.
2. Klicken Sie **Bereich erstellen** auf der Seite **Bereiche** an.
3. Geben Sie einen Namen für den Bereich und die URIs und URI-Muster an, die die Grenzen des Bereichs definieren. Sie können auch URIs und URI-Muster angeben, die Sie aus dem Bereich ausschließen wollen.
4. Klicken Sie **OK** an.

Ihr neuer Bereich wird mit den anderen Bereichen dieser Objektgruppe auf der Seite **Bereiche** angezeigt.

#### Zugehörige Verweise

„URI-Formate in einem Index für die Unternehmenssuche“ auf Seite 130

## Ausgeblendete URIs

Die Unternehmenssuche kann die Suchergebnisse so organisieren, dass Dokumente aus Quellen, die dasselbe URI-Präfix aufweisen, in den Suchergebnissen ausgeblendet werden.

Wenn Ergebnisse ausgeblendet werden, wird das oberste Ergebnis in der Regel linksbündig ausgerichtet. Darunter wird mindestens ein Ergebnis mit einer niedrigeren Rangfolge gruppiert und eingerückt.

Wenn Sie Ergebnisdokumente mit verschiedenen URI-Präfixen als einzelne Gruppe ausblenden möchten, können Sie den URI-Präfixen einen von Ihnen erstellten Gruppennamen zuordnen. Beispiel: Wenn Sie drei Server für das Verwalten von Finanzdaten haben, können Sie Dokumente von all diesen drei Servern in den Suchergebnissen gruppieren und die Ergebnisse mit der niedrigeren Rangfolge unterhalb der obersten Ergebnisdokumente ausblenden.

Suchanwendungen können das URI-Präfix oder den Gruppennamen verwenden, um Dokumente in den Suchergebnissen auszublenden. In der Mustersuchanwendung für die Unternehmenssuche werden die zwei obersten Suchergebnisdokumente angezeigt. Wenn mehr als zwei Ergebnisdokumente mit demselben URI-Präfix zurückgegeben werden oder mehr als zwei Dokumente zu derselben URI-Gruppe gehören, können Sie eine Option für das Anzeigen der ausgeblendeten Ergebnisse auswählen.

Benutzer können die Abfragesyntax für die Unternehmenssuche (`samegroupas:uri-präfix`) verwenden, um alle Dokumente zu durchsuchen, die sich in derselben Gruppe wie das in der Abfrage angegebene URI-Präfix befinden.

## Vorgehensweise für das Organisieren von URI-Präfixen und Gruppennamen

Wenn Sie über die Administrationskonsole Regeln für das Ausblenden von Suchergebnissen konfigurieren, geben Sie die URI-Präfixe der Dokumente an, die Sie ausblenden möchten, und ordnen den URI-Präfixen wahlweise einen Gruppennamen zu.

Dabei ist die Reihenfolge der konfigurierten URI-Präfixe wichtig. Der Indexserver verwendet die Reihenfolge der URI-Präfixe bei der Berechnung des Werts für jeden URI in einer Objektgruppe. Für jeden URI gilt Folgendes:

1. Der Indexserver durchsucht die URI-Präfixe in den Regeln für das Ausblenden von Suchergebnissen nacheinander.
2. Wenn der Indexserver das erste URI-Präfix findet, das mit dem Präfix eines Dokuments im Index übereinstimmt, ordnet er den Gruppennamen (oder das URI-Präfix, falls die Regel keinen Gruppennamen angibt) als zusätzlichen Suchbegriff für das Dokument zu.

Falls ein Webdokument mit keinem URI-Präfix übereinstimmt, verwendet der Indexserver den Hostnamen der URL als URI-Präfix. Falls ein NNTP-Dokument mit keinem URI-Präfix übereinstimmt, verwendet der Indexserver die erste Nachrichten-ID im Wert des Referenzheaders als URI-Präfix.

Nachdem Sie ein URI-Präfix der Liste der URI-Präfixe hinzugefügt haben, die in den Suchergebnissen ausgeblendet werden sollen, müssen Sie das URI-Präfix in der Reihenfolge anordnen, in der es der Indexserver durchsuchen soll, und es möglicherweise als zusätzlichen Suchbegriff Dokumenten im Index zuordnen:

- Wenn Sie ein URI-Präfix hinzufügen und es keinem Gruppennamen zuordnen, können Sie das einzelne URI-Präfix auswählen und in der Liste nach oben oder unten versetzen.
- Wenn Sie ein URI-Präfix hinzufügen und einem Gruppennamen zuordnen, versetzen Sie immer die gesamte Gruppe der URI-Präfixe, die zu derselben Gruppe gehören, sobald Sie ein URI-Präfix in der Liste nach oben oder unten versetzen. Die Reihenfolge der URI-Präfixe in der Gruppe spielt dabei keine Rolle. Durch das Auswählen eines einzelnen URI-Präfixes wird automatisch die gesamte Gruppe ausgewählt.

## Ausblenden von URIs in den Suchergebnissen

Sie können Optionen für das Gruppieren und Ausblenden von Ergebnisdokumenten aus Quellen angeben, die über dasselbe URI-Präfix verfügen. Sie können auch einen Gruppennamen erstellen, mit dem Ergebnisdokumente mit unterschiedlichen URI-Präfixen zusammen ausgeblendet werden können.

### Vorbereitung

Sie müssen ein Administrator für die Unternehmenssuche oder ein Objektgruppenadministrator der Objektgruppe sein, um Optionen zum Ausblenden von Suchergebnissen anzugeben.

### Informationen zu dieser Task

Die von Ihnen vorgenommenen Änderungen zum Ausblenden von Suchergebnissen werden erst nach der nächsten Hauptindexerstellung wirksam.

## Vorgehensweise

Gehen Sie wie folgt vor, um Optionen zum Ausblenden von Suchergebnissen anzugeben:

1. Bearbeiten Sie eine Objektgruppe, wählen Sie die Seite **Index** aus, und klicken Sie **Suchergebnisse ausblenden** an.
2. Klicken Sie **URI-Präfix hinzufügen** auf der Seite **Suchergebnisse ausblenden** an.
3. Geben Sie auf der Seite **URI-Präfix zum Ausblenden von Ergebnissen hinzufügen** das URI-Präfix der Dokumente ein, die Sie in den Suchergebnissen ausblenden wollen. Beispiel:

```
http://finance/ROI/
http://server1.com/finance/
db2://LOCALDB/SCHEMA1.TABLE1/
exchange://exchangesvr.ibm.com/public/TeamRoom/Folder1/
```

4. Sie können einen beschreibenden Gruppennamen eingeben, den Sie diesem URI-Präfix zuordnen wollen. Wenn Sie mehrere Ergebnisdokumente von unterschiedlichen Quellen als eine Gruppe ausblenden wollen, geben Sie denselben Gruppennamen ein, wenn Sie die einzelnen URI-Präfixe hinzufügen.
5. Klicken Sie **OK** an.
6. Positionieren Sie die neue Regel auf der Seite **Suchergebnisse ausblenden** in der Reihenfolge, in der sie vom Indexserver überprüft werden soll:
  - Wenn Sie ein URI-Präfix hinzugefügt haben, ohne dieses einem Gruppennamen zuzuordnen, wird das neue URI-Präfix ganz unten in der Liste angezeigt. Mithilfe der Pfeiltasten können Sie es an die richtige Position versetzen.
  - Wenn Sie das neue URI-Präfix einem Gruppennamen zugeordnet haben, wird das neue URI-Präfix ganz unten in der Gruppe der URI-Präfixe angezeigt, zu der es gehört. Mithilfe der Pfeiltasten können Sie die gesamte URI-Präfixgruppe an ihre gewünschte Position versetzen.
7. Wenn Sie ein URI-Präfix oder den Gruppennamen ändern wollen, wählen Sie das URI-Präfix aus, und klicken Sie  **Bearbeiten** an.
8. Wenn Sie ein URI-Präfix aus der Liste entfernen wollen, wählen Sie das URI-Präfix aus, und klicken Sie  **Entfernen** an.

## Entfernen von URIs aus dem Index

Wenn Sie verhindern wollen, dass Benutzer Dokumente einer Objektgruppe durchsuchen, können Sie die URIs der betreffenden Dokumente aus dem Index entfernen.

### Vorbereitung

Sie müssen ein Administrator für die Unternehmenssuche oder ein Objektgruppenadministrator dieser Objektgruppe sein, um URIs aus dem Index zu entfernen.

### Informationen zu dieser Task

Wenn Sie einen vollständig qualifizierten URI angeben, wird Benutzern der URI nicht mehr in den Suchergebnissen angezeigt. Wenn ein Benutzer jedoch die gleiche Abfrage übergibt und sich Ergebnisdokumente für diese Abfrage im Suchcache befinden, wird die in den Cache gestellte Ergebnisseite für den entfernten URI wei-

terhin in den Suchergebnissen zurückgegeben. Der Suchcache wird nicht aktualisiert, und der URI wird erst bei der nächsten Delta- oder Hauptindexerstellung aus dem Index entfernt.

Wenn Sie ein URI-Muster zum Entfernen mehrerer URIs angeben, werden den Benutzern in den Suchergebnissen weiterhin URIs angezeigt, die mit dem Muster übereinstimmen, bis die nächste Delta- oder Hauptindexerstellung ausgeführt wird.

Wenn Sie einen URI aus dem Index entfernen, entfernen Sie ihn nicht aus dem Crawlerbereich. Wenn der Crawler das Dokument das nächste Mal durchsucht, wird der URI in den Index eingetragen und steht erneut für die Suche zur Verfügung. Sie müssen die Crawlersuchregeln aktualisieren, um das Dokument auszuschließen, und anschließend den Crawler stoppen und erneut starten, um einen URI aus dem Crawlerbereich zu entfernen.

### Vorgehensweise

Gehen Sie wie folgt vor, um die URIs bestimmter Dokumente aus dem Index zu entfernen:

1. Bearbeiten Sie eine Objektgruppe, wählen Sie die Seite **Index** aus, und klicken Sie **URIs aus dem Index entfernen** an.
2. Geben Sie auf der Seite **URIs aus dem Index entfernen** die URIs (oder die URI-Muster) ein, die Sie aus dem Index entfernen wollen.

Beispiel:

```
http://domain.org/hr/*
db2://knowledgeManagement/ROI*
cm://enterprise/finance*
```

### Zugehörige Verweise

„URI-Formate in einem Index für die Unternehmenssuche“ auf Seite 130

---

## Verwaltung des Suchservers

Die Optionen, die Sie für die Suchserver angeben können, umfassen Folgendes: Verwendung von Cachespeicherplatz für die Rückgabe von Suchergebnissen, Steuern der maximalen Länge von Dokumentzusammenfassungen in den Suchergebnissen, Zuordnen benutzerdefinierter Verzeichnisse zur Verbesserung der Suchqualität, sowie die Rückgabe vordefinierter URIs in den Suchergebnissen, sobald bestimmte Begriffe in der Abfrage vorkommen.

Wenn ein Benutzer eine Abfrage übergibt, lokalisieren die Suchserver über den Index schnell die relevanten Dokumente. Die Suchserver verwenden den Datenspeicher für die Unternehmenssuche, der die syntaktisch analysierten und aufbereiteten Daten enthält, um Metadaten für die relevanten Dokumente abzurufen. Metadaten können unter anderem Dokumentangaben wie URI, Titel, Beschreibung, Datum, Datentyp usw. enthalten.

Wenn Sie die Suchserver für eine Objektgruppe konfigurieren, geben Sie Optionen an, die die Verarbeitung von Abfragen beeinflussen. Dazu gehören auch Optionen, die sich auf die Abfrageleistung auswirken können:

### **Konfigurieren eines Suchcaches**

Zur Optimierung der Abfrageleistung können Sie angeben, dass Suchergebnisse (die Antworten auf Suchanforderungen) in einen Cache gestellt werden sollen, und Sie können die Cachespeichergröße angeben, die für diese Suchergebnisse zugeordnet werden soll.

### **Konfigurieren der maximalen Anzeigelänge für Dokumentzusammenfassungen**

Die meisten Ergebnisdokumente zeigen eine Zusammenfassung des Dokumentinhalts, damit der Benutzer einfacher entscheiden kann, ob er dieses Dokument abrufen will. Sie können angeben, wie viel Platz in den Suchergebnissen für die Anzeige dieser Übersichtsdaten verwendet werden soll.

### **Angaben einer anderen voreingestellten Sprache**

Die Sprache für das Durchsuchen der Dokumente in einer Objektgruppe wird bei der Erstellung der Objektgruppe voreingestellt. Sie können nach Bedarf aber auch eine andere Sprache angeben.

### **Zuordnen von benutzerdefinierten Wörterverzeichnissen**

Wenn die Anwendungsentwickler benutzerdefinierte Verzeichnisse für Synonyme, Stopp- oder Boostwörter erstellt haben, können Sie die Wörterverzeichnisse angeben, die der Benutzer beim Durchsuchen einer Objektgruppe verwenden soll.

### **Konfigurieren von Quick Links**

Sie können vordefinieren, dass URIs für bestimmte Schlüsselwörter und Ausdrücke zurückgegeben werden. Wenn Benutzer die Schlüsselwörter bzw. Ausdrücke in einer Abfrage angeben, wird der vordefinierte URI als Suchergebnis zurückgegeben. Die Quick Link-URIs werden zusätzlich zu den URIs zurückgegeben, die die Suchserver beim Durchsuchen des Index zurückgeben.

### **Zugehörige Konzepte**

„Dokumentrangfolge“ auf Seite 223

„Benutzerdefinierte Verzeichnisse von Boostwörtern“ auf Seite 227

---

## Suchcaches

Wenn die Auslastung der Suchserver hoch ist, können Sie die Leistung verbessern, indem Sie die Suchergebnisse zwischenspeichern.

Wenn die Suchserver Suchanforderungen verarbeiten, überprüfen Sie zunächst, ob bereits Ergebnisse für dieselbe Anforderung im Cache vorhanden sind. Wenn die Suchserver die entsprechende Antwort auf die Suchanforderung finden, können sie die Suchergebnisse schnell an den Benutzer zurückgeben. Wenn die Suchserver die entsprechende Antwort auf die Suchanforderung nicht finden, durchsuchen sie den Index.

Die ältesten Suchergebnisse und Ergebnisse für seltene Suchanforderungen werden bei zunehmender Cachefüllung aus dem Suchcache entfernt, um Platz für neue Suchergebnisse zu schaffen.

Über die Administrationskonsole für die Unternehmenssuche können Sie das Zwischenspeichern von Suchergebnissen aktivieren und außerdem die Kapazität des Cache angeben (die Anzahl Antworten auf Suchanforderungen, die gleichzeitig zwischengespeichert werden kann).

Wenn Sie die Einstellungen für den Suchcache ändern, müssen Sie die Suchserver erneut starten, damit die Änderungen wirksam werden.

---

## Konfigurieren eines Suchcaches

Sie können den Suchcache für eine Objektgruppe aktivieren oder inaktivieren. Sie können auch Optionen angeben, um die Größe des Suchcaches zu steuern.

### Vorbereitung

Sie müssen ein Administrator für die Unternehmenssuche oder ein Objektgruppenadministrator der Objektgruppe sein, um einen Suchcache für eine Objektgruppe zu konfigurieren.

### Vorgehensweise

Gehen Sie wie folgt vor, um den Suchcache zu konfigurieren:

1. Bearbeiten Sie eine Objektgruppe, wählen Sie die Seite **Suche** aus, und klicken Sie **Suchserveroptionen konfigurieren** an.
2. Wählen Sie auf der Seite **Suchserveroptionen** das Markierungsfeld **Suchcache verwenden** aus.
3. Geben Sie in das Feld **Maximale Anzahl Einträge im Cache** die maximale Anzahl Antworten auf Suchanforderungen ein, die der Suchcache enthalten darf.
4. Klicken Sie **OK** an.
5. Damit die Änderungen wirksam werden, überwachen Sie die Suchserver, und starten Sie die Suchserverprozesse erneut.

---

## Benutzerdefinierte Synonymverzeichnisse

Sie können die Qualität der Suchergebnisse verbessern, indem die Benutzer beim Durchsuchen der Objektgruppe auch nach Synonymen ihrer Abfragebegriffe suchen können.

Wenn Sie ein Synonymverzeichnis erstellen, dieses dem System für die Unternehmenssuche hinzufügen und es einer Objektgruppe zuordnen, können Benutzer beim Durchsuchen der Objektgruppe nach Dokumenten suchen, die Synonyme ihrer Abfragebegriffe enthalten. Durch diese Art der Abfrageerweiterung finden Benutzer eher alle Dokumente, die für sie von Interesse sind, und nicht nur die Dokumente, in denen der genaue Abfragebegriff enthalten ist. Dadurch, dass Sie bei der Erstellung des Synonymverzeichnisses festlegen, welche Wörter Synonyme sind, helfen Sie sicherzustellen, dass Benutzer relevante Dokumente auch ohne Angabe aller Variationen des Abfragebegriffs finden.

Beispiel: Ihr Unternehmen verwendet Akronyme und Abkürzungen für die einzelnen Abteilungen, Geräte usw., oder die Dokumente in Ihren Objektgruppen enthalten branchenspezifische Begriffe. Durch das Erstellen eines Synonymverzeichnisses stellen Sie sicher, dass Abfragen, die ein Akronym (z. B. ACL) verwenden, Dokumente als Suchergebnis zurückgeben, die auch erweiterte Formen des Akronyms behandeln (z. B. ACLs, Access Control Lists, Access Controls usw.).

Die Abfragesprache für die Unternehmenssuche unterstützt Synonyme, indem Benutzer dem Abfragebegriff eine Tilde voranstellen können. So kann beispielsweise die Abfrage `~WAS` Dokumente als Suchergebnis zurückgeben, die sich mit WebSphere Application Server befassen. Anwendungsentwickler können Synonymunterstützung auch über die Abfragemerkmale, die keine besondere Syntax erfordern, verfügbar machen.

Synonymverzeichnisse enthalten Wortvarianten und haben folgende Merkmale:

- Die Wörter sind nicht auf eine Sprache beschränkt, sondern sie können in verschiedenen Sprachen verwendet werden. Pro Objektgruppe gibt es jeweils nur ein Synonymverzeichnis.
- Die Wörter sind nicht flektiert. Alle möglichen Beugungen des Worts müssen der Synonymliste hinzugefügt werden. Eine Beugung kann beispielsweise der Singular oder der Plural eines Worts sein (ACL und ACLs).

Die meisten Begriffe, die Sie einem Synonymverzeichnis hinzufügen, sind streng semantische Entsprechungen, das heißt, wenn Begriff A ein Synonym von Begriff B ist, dann ist B auch ein Synonym von Begriff A. Jedes Mal, wenn A in einer Abfrage verwendet wird, kann auch B verwendet werden und umgekehrt.

Sie können jedoch auch Begriffe hinzufügen, die den verschiedenen Verwendungen eines Begriffs entsprechen, einschließlich generischer und spezifischerer Varianten des Begriffs. So können Sie z. B. eine Synonymgruppe haben, die sowohl Gebäude und Haus umfasst, und eine andere Gruppe, die Bank, Sitzgelegenheit und Kreditanstalt umfasst.

Je weniger streng die Beziehung zwischen den Begriffen ist, desto größer das Suchergebnis, wobei einige der Suchergebnisse möglicherweise für die Abfrage gar nicht relevant sind. Search and Index API stellt Methoden bereit, mit denen Benutzer bei der Übergabe einer Suchanforderung die entsprechenden Synonyme auswählen können, sowie Methoden, die Benutzern zeigen, welche Abfragebegriffe durch welche Synonyme erweitert wurden.

Zum Erstellen eines Synonymverzeichnisses muss ein Fachmann auf dem Sachgebiet der Objektgruppe eine Synonymliste im XML-Format erstellen oder mit einem Anwendungsentwickler die XML-Datei erstellen. Zum Umwandeln der XML-Datei in eine binäre Datei (.dic) muss ein Tool für die Unternehmenssuche (**essyndict-builder**) verwendet werden.

Ein Administrator für die Unternehmenssuche lädt die Binärdatei auf das System hoch und ordnet ihr einen Anzeigenamen zu. Objektgruppenadministratoren können ein Synonymverzeichnis für das Durchsuchen von Dokumenten in einer Objektgruppe auswählen, wenn sie die Suchserveroptionen für eine Objektgruppe konfigurieren.

**Einschränkung:** Sie können ein benutzerdefiniertes Synonymverzeichnis nicht mehr ändern, wenn Sie es dem System hinzugefügt haben. Zum Überarbeiten der Synonyme, die für eine Objektgruppe verfügbar sind, müssen Sie wie folgt vorgehen:

1. Aktualisieren Sie die XML-Quellendatei.
2. Wandeln Sie die XML-Quelle in eine neue Verzeichnisdatei um.
3. Entfernen Sie das alte Synonymverzeichnis aus allen Objektgruppen, die es verwenden.
4. Löschen Sie das alte Synonymverzeichnis vom System.
5. Fügen Sie dem System das neue Synonymverzeichnis hinzu.
6. Ordnen Sie das neue Synonymverzeichnis den Objektgruppen zu, die es verwenden sollen.

Sie können ein Script schreiben, das diese Schritte einschließt, und das Script anschließend verwenden, um das Wörterverzeichnis erneut in Ihrem System für die Unternehmenssuche zu implementieren.

#### **Zugehörige Konzepte**

 Synonymunterstützung in Suchanwendungen

#### **Zugehörige Tasks**

 Erstellen einer XML-Datei für Synonyme

 Erstellen eines Synonymverzeichnisses

„Erneutes Implementieren von benutzerdefinierten Wörterverzeichnissen“ auf Seite 216

## **Hinzufügen von Synonymverzeichnissen zum System**

Wenn Sie benutzerdefinierte Synonymverzeichnisse erstellen, um die Dokumente einer Objektgruppe zu durchsuchen, müssen Sie die Verzeichnisse dem System für die Unternehmenssuche zuordnen. Sie können später entscheiden, welches Synonymverzeichnis Sie zum Durchsuchen einer Objektgruppe verwenden wollen.

### **Vorbereitung**

Sie müssen ein Administrator für die Unternehmenssuche sein, um Ihre benutzerdefinierten Synonymverzeichnisse für die Verwendung in Unternehmenssuchabfragen hinzuzufügen.

### **Einschränkungen**

Die maximale Größe eines Synonymverzeichnisses ist 8 MB.

### **Vorgehensweise**

Gehen Sie wie folgt vor, um dem System für die Unternehmenssuche Synonyme zuzuordnen:

1. Klicken Sie **System** an, um die Sicht **System** zu öffnen.
2. Klicken Sie  **Bearbeiten** an, um zur Sicht für die Systembearbeitung zu wechseln.
3. Klicken Sie **Synonymverzeichnisse konfigurieren** auf der Seite **Suche** an.
4. Klicken Sie **Synonymverzeichnis hinzufügen** auf der Seite **Synonymverzeichnisse konfigurieren** an.
5. Geben Sie auf der Seite **Synonymverzeichnis hinzufügen** einen eindeutigen Anzeigenamen für das Synonymverzeichnis ein. Optional können Sie auch eine Beschreibung eingeben.
6. Geben Sie die Speicherposition der .dic-Datei an. Wenn sich die Datei auf Ihrem lokalen System befindet, können Sie nach der Datei suchen. Wenn sich die Datei auf dem Indexserver befindet, geben Sie den vollständig qualifizierten Pfad ein.
7. Klicken Sie **OK** an. Ihr benutzerdefiniertes Synonymverzeichnis wird dem System für die Unternehmenssuche hinzugefügt und ist für das Durchsuchen von Objektgruppen verfügbar.

## Zuordnen eines Synonymverzeichnisses zu einer Objektgruppe

Wenn dem System für die Unternehmenssuche Synonymverzeichnisse zugeordnet wurden, können Sie eines für die Verwendung beim Suchen in einer Objektgruppe auswählen. Wenn ein Abfragebegriff mit einem Begriff des Verzeichnisses übereinstimmt, werden Ergebnisdokumente, die Synonyme des Begriffs enthalten, ebenfalls in den Suchergebnissen angezeigt.

### Vorbereitung

Sie müssen ein Administrator für die Unternehmenssuche oder ein Objektgruppenadministrator der Objektgruppe sein, um ein Synonymverzeichnis für eine Objektgruppe auszuwählen.

### Vorgehensweise

Gehen Sie wie folgt vor, um einer Objektgruppe ein Synonymverzeichnis zuzuordnen:

1. Bearbeiten Sie eine Objektgruppe, wählen Sie die Seite **Suche** aus, und klicken Sie **Suchserveroptionen konfigurieren** an.
2. Wählen Sie im Feld **Synonymverzeichnisname** auf der Seite **Suchserveroptionen** das Synonymverzeichnis aus, das verwendet werden soll, wenn Benutzer diese Objektgruppe abfragen.  
Die Liste der verfügbaren Synonymverzeichnisse enthält alle Synonymverzeichnisse, die dem System für die Unternehmenssuche hinzugefügt wurden.
3. Klicken Sie **OK** an.

---

## Benutzerdefinierte Verzeichnisse von Stoppwörtern

Zur Verbesserung der Qualität der Suchergebnisse können Sie angeben, dass bestimmte Wörter während der Abfrageverarbeitung automatisch aus den Abfragebegriffen entfernt werden.

Ein Verzeichnis von Stoppwörtern enthält bestimmte unternehmensspezifische Begriffe, die häufig verwendet werden und deshalb als Abfragebegriffe ungeeignet

sind. Durch Ausschließen dieser Wörter aus den Abfragen können Sie sicherstellen, dass der Benutzer nicht mit wenig relevanten Ergebnisdokumenten überschwemmt wird (nur Dokumente, die mit anderen Begriffen in der Abfrage übereinstimmen, werden zurückgegeben). Während der Abfrageverarbeitung entfernen die Suchserver die Stoppwörter aus den Abfragen. Es werden Wörter entfernt, die in benutzerdefinierten Wörterverzeichnissen enthalten sind, sowie Wörter, die für die Unternehmenssuche vordefiniert sind (z. B. häufige Präpositionen und Artikel).

Bei der Unternehmenssuche wird standardmäßig eine sprachspezifische Stoppworterkennung ausgeführt. Bei diesem Prozess werden häufig auftretende Wörter wie ein oder der aus der Abfrage entfernt. Ein benutzerdefiniertes Verzeichnis von Stoppwörtern muss nur für unternehmens- oder domänenspezifische Stoppwörter definiert werden.

Wenn eine Abfrage verarbeitet wird, werden Stoppwörter entfernt, bevor Rechtsschreibvorschläge gemacht werden. Wenn alle Wörter in einer Abfrage Stoppwörter sind, werden während der Abfrageverarbeitung keine Stoppwörter entfernt. Um sicherzustellen, dass Ergebnisse zurückgegeben werden, wird die Entfernung von Stoppwörtern inaktiviert, wenn alle Abfragebegriffe Stoppwörter sind. Wenn beispielsweise das Wort Auto ein Stoppwort ist und Sie nach Auto suchen, enthalten die Suchergebnisse Dokumente, die mit dem Wort Auto übereinstimmen. Wenn Sie nach Auto Volvo suchen, enthalten die Suchergebnisse nur Dokumente, die mit dem Wort Volvo übereinstimmen.

Zum Erstellen eines Verzeichnisses von Stoppwörtern muss ein Fachmann auf dem Sachgebiet der Objektgruppe eine Liste mit Stoppwörtern im XML-Format erstellen oder mit einem Anwendungsentwickler die XML-Datei erstellen. Zum Umwandeln der XML-Datei in eine binäre Datei (.dic ) muss ein Tool für die Unternehmenssuche (**esstopworddictbuilder**) verwendet werden.

Ein Administrator für die Unternehmenssuche lädt die Binärdatei auf das System hoch und ordnet ihr einen Anzeigenamen zu. Objektgruppenadministratoren können ein Verzeichnis von Stoppwörtern für das Durchsuchen von Dokumenten in einer Objektgruppe auswählen, wenn sie die Suchserveroptionen für eine Objektgruppe konfigurieren.

**Einschränkung:** Sie können ein benutzerdefiniertes Verzeichnis von Stoppwörtern nicht mehr ändern, wenn Sie es dem System hinzugefügt haben. Gehen Sie wie folgt vor, um die Stoppwörter für die Abfrageverarbeitung zu überarbeiten:

1. Aktualisieren Sie die XML-Quellendatei.
2. Wandeln Sie die XML-Quelle in eine neue Verzeichnisdatei um.
3. Entfernen Sie das alte Verzeichnis von Stoppwörtern aus allen Objektgruppen, die es verwenden.
4. Löschen Sie das alte Verzeichnis von Stoppwörtern vom System.
5. Fügen Sie dem System das neue Verzeichnis von Stoppwörtern hinzu.
6. Ordnen Sie das neue Verzeichnis von Stoppwörtern den Objektgruppen zu, die es verwenden sollen.

Sie können ein Script schreiben, das diese Schritte einschließt, und das Script anschließend verwenden, um das Wörterverzeichnis erneut in Ihrem System für die Unternehmenssuche zu implementieren.

### Zugehörige Konzepte



Benutzerdefinierte Verzeichnisse von Stoppwörtern

### Zugehörige Tasks

 Erstellen einer XML-Datei für Stoppwörter

 Erstellen eines Verzeichnisses von Stoppwörtern

„Erneutes Implementieren von benutzerdefinierten Wörterverzeichnissen“ auf Seite 216

## Hinzufügen von Verzeichnissen von Stoppwörtern zum System

Wenn Sie benutzerdefinierte Verzeichnisse von Stoppwörtern erstellen, um Wörter aus Abfragen zu entfernen, müssen Sie die Verzeichnisse dem System für die Unternehmenssuche hinzufügen. Sie können später entscheiden, welches Verzeichnis von Stoppwörtern Sie zum Durchsuchen einer Objektgruppe verwenden wollen.

### Vorbereitung

Sie müssen ein Administrator für die Unternehmenssuche sein, um dem System benutzerdefinierte Verzeichnisse von Stoppwörtern hinzuzufügen.

### Einschränkungen

Die maximale Größe eines Verzeichnisses von Stoppwörtern ist 8 MB.

### Vorgehensweise

Gehen Sie wie folgt vor, um dem System für die Unternehmenssuche benutzerdefinierte Stoppwörter zuzuordnen:

1. Klicken Sie **System** an, um die Sicht **System** zu öffnen.
2. Klicken Sie  **Bearbeiten** an, um zur Sicht für die Systembearbeitung zu wechseln.
3. Klicken Sie **Verzeichnisse von Stoppwörtern konfigurieren** auf der Seite **Suche** an.
4. Klicken Sie **Verzeichnis von Stoppwörtern hinzufügen** auf der Seite **Verzeichnisse von Stoppwörtern konfigurieren** an.
5. Geben Sie auf der Seite **Verzeichnis von Stoppwörtern hinzufügen** einen eindeutigen Anzeigenamen für das Verzeichnis ein.
6. Geben Sie die Speicherposition der .dic-Datei an. Wenn sich die Datei auf Ihrem lokalen System befindet, können Sie nach der Datei suchen. Wenn sich die Datei auf dem Indexserver befindet, geben Sie den vollständig qualifizierten Pfad ein.
7. Klicken Sie **OK** an. Ihr benutzerdefiniertes Verzeichnis von Stoppwörtern wird dem System für die Unternehmenssuche hinzugefügt und ist für das Durchsuchen von Objektgruppen verfügbar.

## Zuordnen eines Verzeichnisses von Stoppwörtern zu einer Objektgruppe

Wenn dem System für die Unternehmenssuche Verzeichnisse von Stoppwörtern zugeordnet wurden, können Sie eines für die Suche in einer Objektgruppe auswählen. Wenn ein Abfragebegriff mit einem Begriff im Verzeichnis übereinstimmt, wird dieser Begriff vor der Verarbeitung aus der Abfrage entfernt.

## Vorbereitung

Sie müssen ein Administrator für die Unternehmenssuche oder ein Objektgruppenadministrator der Objektgruppe sein, um ein Verzeichnis von Stoppwörtern für eine Objektgruppe auszuwählen.

## Vorgehensweise

Gehen Sie wie folgt vor, um einer Objektgruppe ein Verzeichnis von Stoppwörtern zuzuordnen:

1. Bearbeiten Sie eine Objektgruppe, wählen Sie die Seite **Suche** aus, und klicken Sie **Suchserveroptionen konfigurieren** an.
2. Wählen Sie im Feld **Name des Verzeichnisses von Stoppwörtern** auf der Seite **Suchserveroptionen** das Verzeichnis von Stoppwörtern aus, das verwendet werden soll, wenn Benutzer diese Objektgruppe abfragen.

Die Liste der verfügbaren Verzeichnisse enthält alle Verzeichnisse von Stoppwörtern, die dem System für die Unternehmenssuche hinzugefügt wurden.

3. Klicken Sie **OK** an.

---

## Erneutes Implementieren von benutzerdefinierten Wörterverzeichnissen

Sie können die Administrationskonsole für die Unternehmenssuche nicht verwenden, um Änderungen an Verzeichnissen von Synonymen, Stoppwörtern oder Boostwörtern vorzunehmen. Sie können jedoch Schritte in einem Script einschließen und dann dieses Script für die erneute Implementierung des Wörterverzeichnisses verwenden.

## Vorbereitung

Sie müssen sich als Administrator für die Unternehmenssuche anmelden, um ein benutzerdefiniertes Wörterverzeichnis erneut zu implementieren.

## Vorgehensweise

Gehen Sie wie folgt vor, um ein benutzerdefiniertes Wörterverzeichnis für die Unternehmenssuche erneut zu implementieren:

1. Öffnen Sie auf dem Indexserver eine der folgenden Dateien, um nach dem Namen des Wörterverzeichnisses zu suchen, das Sie erneut implementieren möchten:
  - Öffnen Sie `ES_NODE_ROOT/master_config/SynonymConfiguration.xml`, um nach dem Namen eines Synonymverzeichnisses zu suchen.
  - Öffnen Sie `ES_NODE_ROOT/master_config/StopWordDictionaryConfiguration.xml`, um nach dem Namen eines Verzeichnisses von Stoppwörtern zu suchen.
  - Öffnen Sie `ES_NODE_ROOT/master_config/BoostingWordDictionaryConfiguration.xml`, um nach dem Namen eines Verzeichnisses von Boostwörtern zu suchen.

Das folgende Beispiel zeigt eine Datei `SynonymConfiguration.xml` mit zwei Synonymverzeichnissen, die mit den Namen `hello` und `goodbye` hochgeladen wurden:

```
% cat $ES_NODE_ROOT/master_config/SynonymConfiguration.xml
```

```
<SynonymConfiguration>
<Synonyms>
<Synonym Name="hello" ID="SynonymId_1">
<Filename>synonym_hello1.dic</Filename>
<Timestamp>1169766691776</Timestamp>
</Synonym>
<Synonym Name="goodbye" ID="SynonymId_2">
<Filename>synonym_goodbye2.dic</Filename>
<Timestamp>1169767224839</Timestamp>
</Synonym>
</Synonyms>
</SynonymConfiguration>
```

Beim Hochladen von Wörterverzeichnissen vergibt das System eindeutige Datei-IDs. Im obigen Beispiel sind diese IDs `synonym_hello1.dic` und `synonym_goodbye2.dic`. Der absolute Pfad zu diesen Dateien auf dem Indexserver ist der folgende:

- Der Pfad zu einem Synonymverzeichnis ist `ES_NODE_ROOT/data/custom_dictionary/synonym_*.dic`.
- Der Pfad zu einem Verzeichnis von Stoppwörtern ist `ES_NODE_ROOT/data/custom_dictionary/stopwordDictionary_*.dic`.
- Der Pfad zu einem Verzeichnis von Boostwörtern ist `ES_NODE_ROOT/data/custom_dictionary/boostingwordDictionary_*.dic`.

Unter AIX, Linux und Solaris können Sie den Befehl `ls` verwenden, um die verfügbaren Wörterverzeichnisse aufzulisten. Beispiel:

```
% ls -l $ES_NODE_ROOT/data/custom_dictionary/synonym_*.dic
-rw-rw-r-- 1 esuser users 9 Jan 25 15:11 /home/esuser/node/data/custom_dictionary/synonym_hello1.dic
-rw-rw-r-- 1 esuser users 9 Jan 25 15:11 /home/esuser/node/data/custom_dictionary/synonym_goodbye2.dic
```

2. Wenn Sie das Wörterverzeichnis gefunden haben, das Sie aktualisieren möchten, stoppen Sie das System für die Unternehmenssuche.
3. Überschreiben Sie das zu aktualisierende Wörterverzeichnis mit Ihrer neuen Wörterverzeichnisdatei. Sie können beispielsweise `synonym_hello1.dic` oder `synonym_goodbye2.dic` überschreiben.
4. Wird Ihr System für die Unternehmenssuche auf zwei oder vier Servern ausgeführt, müssen Sie die Wörterverzeichnisdatei auf den Suchservern manuell überschreiben. Die Wörterverzeichnisdateien befinden sich in demselben Pfad wie auf dem Indexserver (`ES_NODE_ROOT/data/custom_dictionary/`).
5. Starten Sie das System für die Unternehmenssuche erneut.

#### Zugehörige Konzepte

„Benutzerdefinierte Synonymverzeichnisse“ auf Seite 210

„Benutzerdefinierte Verzeichnisse von Stoppwörtern“ auf Seite 213

„Benutzerdefinierte Verzeichnisse von Boostwörtern“ auf Seite 227

---

## Dynamische Zusammenfassung

Dynamische Zusammenfassung ist ein Verfahren, das ermittelt, welche Ausdrücke eines Ergebnisdokuments die vom Benutzer gesuchten Begriffe am besten darstellt.

Bei der Unternehmenssuche versucht die dynamische Zusammenfassung, Sätze in Dokumenten zu erfassen, die eine große Vielfalt der Suchbegriffe enthalten. Es werden ein paar Sätze oder Teile von Sätzen ausgewählt und in den Suchergebnissen angezeigt. Die Suchbegriffe werden in der HTML-Darstellung der Suchergebnisse hervorgehoben.

Bei der Konfiguration der Suchserveroptionen für eine Objektgruppe können Sie die maximale Anzeigelänge für Dokumentzusammenfassungen in den Suchergebnissen festlegen. Da die Zusammenfassung auch Hervorhebungszeichen einschließt, ist der Puffer, der an die Suchanwendung zurückgegeben wird, größer als der angegebene Maximalwert. Die Anzeigelänge überschreitet jedoch nicht den angegebenen Maximalwert, obwohl die Zusammenfassung kürzer sein kann (abhängig von den aus dem Quelldokument extrahierten Daten).

## Anpassen von Dokumentzusammenfassungen in der Administrationskonsole

Sie können selbst bestimmen, wie viele Informationen eine Dokumentzusammenfassung enthält, indem Sie in der Administrationskonsole für die Unternehmenssuche Optionen für den Suchserver angeben.

### Vorbereitung

Sie müssen die Rolle eines Administrators für die Unternehmenssuche haben oder ein Objektgruppenadministrator für diese Objektgruppe sein, um die Anzeigelänge von Zusammenfassungen für eine Objektgruppe steuern zu können.

### Informationen zu dieser Task

Der Wert, den Sie für die maximale Anzeigelänge von Dokumentzusammenfassungen angeben, hängt mit dem Wert zusammen, den Sie für die Anzahl Sätze in jeder Zusammenfassung angeben. Der Wert, der die kürzeste Dokumentzusammenfassung ergibt, hat Vorrang.

Wenn Sie z. B. eine Begrenzung von vier Sätzen angeben, dann enthält die Dokumentzusammenfassung nur vier Sätze, selbst wenn die Anzeigelänge mehr Zeichen zulässt als die Gesamtzahl Zeichen in diesen Sätzen. In einem anderen Beispiel könnte eine Begrenzung auf 10 Sätze in Kombination mit einer 500 Zeichen umfassenden Begrenzung für die Anzeigelänge zu einer Dokumentzusammenfassung führen, die weniger als 10 Sätze enthält.

### Vorgehensweise

Gehen Sie wie folgt vor, um eine Anzeigelänge für Dokumentzusammenfassungen zu konfigurieren:

1. Bearbeiten Sie eine Objektgruppe, wählen Sie die Seite **Suche** aus, und klicken Sie **Suchserveroptionen konfigurieren** an.
2. Geben Sie auf der Seite **Suchserveroptionen** die maximale Anzeigelänge für Dokumentzusammenfassungen an. Wenn ein Benutzer Suchergebnisse anzeigt, überschreiten die Dokumentzusammenfassungen den von Ihnen angegebenen Wert nicht.
3. Geben Sie an, wie viele Sätze eine Dokumentzusammenfassung enthalten kann (Zusammenfassungen können bis zu 10 Sätze enthalten).
4. Klicken Sie **OK** an.
5. Damit die Änderungen wirksam werden, überwachen Sie die Suchserver, und starten Sie die Suchserverprozesse erneut.

## Anpassen von Dokumentzusammenfassungen durch Bearbeitung von Merkmalen

Jedes Ergebnisdokument für eine Unternehmenssuchabfrage enthält eine Zusammenfassung. Sie können selbst bestimmen, wie viele Informationen jede Zusammenfassung enthält, indem Sie eine Merkmaldatei bearbeiten.

### Informationen zu dieser Task

Sie können die Beschreibungen der Suchergebnisse anpassen, indem Sie die Werte für die folgenden Merkmale in der Datei `ES_NODE_ROOT/master_config/objektgruppen-id.runtime.node1/runtime-generic.properties` ändern:

#### **MinWordsPerSentence**

Die minimale Anzahl Wörter in jedem Satz der Zusammenfassung. In die Zusammenfassung werden kürzere Sätze eingeschlossen, wenn nicht genug Sätze vorhanden sind, die mehr Wörter enthalten, als im `MinWordsPerSentence`-Wert angegeben sind. Der Standardwert ist 4.

#### **MaxWordsPerSentence**

Die maximale Anzahl Wörter in jedem Satz, der in die Zusammenfassung eingeschlossen wird. Wenn ein Satz mehr Wörter enthält als angegeben, wird nur ein Teil des Satzes (der Teil, der den Abfragebegriff enthält, bis zum angegebenen Maximalwert von `MaxWordsPerSentence`) in die Zusammenfassung aufgenommen. Der Rest des Satzes wird abgeschnitten. Der Standardwert ist 20.

Die Sätze für die Dokumentzusammenfassung werden entsprechend eines proprietären, internen Algorithmus ausgewählt, der die Relevanz aller Sätze ermittelt, die die Suchbegriffe enthalten. Die Auswahl nach der Relevanz erfolgt, bevor die Sätze nach der Satzlänge gefiltert werden.

#### **NumberOfReturnedSentences**

Die Anzahl Sätze, die die Beschreibung eines Dokuments bilden. Der Standardwert ist 5.

#### **MaxSentencesPerDocument**

Die maximale Anzahl Sätze in einem Dokument, die als Kandidaten bei der Beschreibungserstellung in Frage kommen. Der Standardwert ist 1000.

### Vorgehensweise

Gehen Sie wie folgt vor, um die Dokumentzusammenfassungen in den Suchergebnissen anzupassen:

1. Melden Sie sich an den Suchservern als Administrator für die Unternehmenssuche an. Diese Benutzer-ID wurde bei der Installation von OmniFind Enterprise Edition angegeben.
2. Bearbeiten Sie die folgende Datei mit einem Texteditor. Dabei ist *objektgruppen-id* die ID, die bei der Erstellung der Objektgruppe für die Objektgruppe angegeben wurde (oder vom System zugeordnet wurde):  
`ES_NODE_ROOT/master_config/objektgruppen-id.runtime.node1/runtime-generic.properties`

**Tip:** Informationen zum Ermitteln der Zuordnung zwischen einem Objektgruppennamen und der zugehörigen ID finden Sie in der Datei `ES_NODE_ROOT/master_config/collections.ini`.

3. Ändern Sie die Merkmale, die Sie anpassen möchten. Dann speichern und beenden Sie die Datei.

4. Stoppen Sie die Suchserver, und starten Sie sie erneut, um die Änderungen auszuführen.

---

## Funktionsweise von Quick Links

Quick Links sind Dokumente, die als Suchergebnisse zurückgegeben werden, wenn ein Benutzer eine Abfrage mit bestimmten Wörtern oder Ausdrücken übergibt.

Die Quick Links für eine Objektgruppe konfigurieren Sie über die Administrationskonsole für die Unternehmenssuche.

### Quick Links

Mit Quick Links können Sie Benutzern Links zu Dokumenten bereitstellen, für die bereits festgelegt ist, dass sie für die Abfragebegriffe relevant sind.

Ein Quick Link ist ein URI, den die Unternehmenssuche automatisch in die Suchergebnisse aufnimmt, wenn eine Abfrage bestimmte Wörter oder Ausdrücke enthält. In der Regel werden die URIs von Quick Links oben in der Ergebnisliste angezeigt, um die Aufmerksamkeit der Benutzer auf die Dokumente zu lenken, die Sie als für die Abfrage relevant vorbestimmt haben.

Quick Links werden zusätzlich zu anderen Suchergebnissen zurückgegeben. Die Suchvorgänge suchen im Index nach Dokumenten, die mit den Abfragebegriffen übereinstimmen, und geben URIs für diese Dokumente zusätzlich zu den URIs von Quick Links zurück.

Beim Konfigurieren eines Quick Links können Sie einen beschreibenden Titel und eine Zusammenfassung für den URI angeben. Dadurch helfen Sie Benutzern, das Dokument einfach zu erkennen und schnell zu ermitteln, ob es abgerufen werden soll.

Beispielsweise könnten Sie für den URI <http://www.ibm.com/education/us/> den Titel IBM Schulung in den USA verwenden und die Zusammenfassung Lösungen, Produkte und Ressourcen für professionelle Anwender, Schulungsbeauftragte und Schulungsteilnehmer in den USA angeben.

Die Option für das Anzeigen von Quick Links muss in der Suchanwendung verfügbar sein, um Quick Links in Objektgruppen für die Unternehmenssuche verwenden zu können. In einigen Suchanwendungen können Benutzer eventuell die Rückgabe von Quick Links beim Durchsuchen der Objektgruppe aktivieren und inaktivieren.

### Konfigurieren von Quick Links

Beim Erstellen eines Quick Links für eine Objektgruppe für die Unternehmenssuche ordnen Sie dem URI eines Dokuments die Schlüsselwörter zu, die dessen Aufnahme in die Suchergebnisse auslösen.

#### Vorbereitung

Sie müssen ein Administrator für die Unternehmenssuche oder ein Objektgruppenadministrator der Objektgruppe sein, zu der der Quick Link gehört, um Quick Links zu konfigurieren.

#### Informationen zu dieser Task

Wenn Sie Beispiele anzeigen wollen, wie Schlüsselwörter und URIs für Quick Links angegeben werden, klicken Sie **Hilfe** an, während Sie einen Quick Link erstellen oder bearbeiten.

Es ist nicht erforderlich, die Suchserver erneut zu starten, damit Ihre Änderungen wirksam werden.

### **Vorgehensweise**

Gehen Sie wie folgt vor, um einen Quick Link zu konfigurieren:

1. Bearbeiten Sie eine Objektgruppe, wählen Sie die Seite **Suche** aus, und klicken Sie **Quick Links konfigurieren** an.
2. Klicken Sie **Quick Link erstellen** auf der Seite **Quick Links** an.
3. Geben Sie die Schlüsselwörter und Wortfolgen, die die Rückgabe dieses Quick Links in den Suchergebnissen auslösen, den URI des Dokuments, das Ihrer Meinung nach für die Abfrage relevant ist, und andere Optionen für den Quick Link an.

Sie können ein einzelnes Schlüsselwort, mehrere Schlüsselwörter oder eine Wortfolge (mindestens zwei Wörter in Anführungszeichen) pro Zeile angeben. Trennen Sie die Schlüsselwörter mit Leerzeichen (Sie können zur Trennung von Schlüsselwörtern keine Kommata verwenden). Drücken Sie die Eingabetaste, um eine neue Zeile zu beginnen.

4. Klicken Sie **OK** an.

Ihr neuer Quick Link wird auf der Seite **Quick Links** mit den anderen zu dieser Objektgruppe gehörenden Quick Links angezeigt.

### **Zugehörige Verweise**

„URI-Formate in einem Index für die Unternehmenssuche“ auf Seite 130



---

## Dokumentrangfolge

Wenn ein Benutzer eine Objektgruppe durchsucht, geben die Suchvorgänge die relevantesten Ergebnisse für die Abfragebedingungen zurück.

Die Suchserver unterstützen eine umfangreiche Abfragesyntax und verwenden mehrere Verfahren zum Erzielen der relevantesten Suchergebnisse, wie beispielsweise textbasierte Bewertung und statische Rangfolge. Sie können das Standardverhalten für Rangfolgen durch das Konfigurieren von Optionen erweitern, die den Rang von Dokumenten in den Suchergebnissen beeinflussen:

- Sie können benutzerdefinierte Verzeichnisse von Boostwörtern erstellen, um die Rangfolge der Dokumente mit den angegebenen Boostwörtern in den Suchergebnissen zu beeinflussen.
- Sie können die Bewertungen von Dokumenten beeinflussen, die mit einem angegebenen URI-Muster übereinstimmen.
- Sie können die Bewertungen von Dokumenten beeinflussen, die Felder enthalten, die Boostklassen zugeordnet werden.

### Zugehörige Konzepte

„Auf Boostklassen basierende Dokumentrangfolge“ auf Seite 232

„Benutzerdefinierte Verzeichnisse von Boostwörtern“ auf Seite 227

---

## Textbasierte Bewertung

Die Unternehmenssuche berechnet dynamisch eine Bewertung für jedes Dokument, das mit den Begriffen in einer Abfrage übereinstimmt.

Die Textbewertung eines Dokuments stellt die Bedeutung der Abfragebegriffe in einem Dokument dar. Beim Berechnen der Textbewertung für jedes Dokument, das mit einer Abfrage übereinstimmt, zieht die Unternehmenssuche viele Faktoren in Betracht, wie z. B. Folgendes:

- Ob die Begriffe ein Dokument von anderen Dokumenten unterscheiden. Wenn z. B. in einem Dokument Abfragebegriffe vorhanden sind, die in den anderen Dokumenten nicht vorhanden sind, bedeutet das, dass diese Begriffe für das Dokument wichtig sind und dieses Dokument von den anderen Dokumenten unterscheiden. Abfragebegriffe, die in den meisten Dokumenten vorhanden sind, tragen weniger zur Dokumentbewertung bei als Abfragebegriffe, die nur in einer kleinen Anzahl Dokumente vorhanden sind.
- Die Häufigkeit von Abfragebegriffen in einem Dokument. Die Bewertung eines Begriffs wird auf der Basis jedes Vorkommens des betreffenden Begriffs in einem Dokument berechnet. Je häufiger ein Abfragebegriff in einem Dokument vorhanden ist, desto höher wird das Dokument bewertet. Wenn Sie z. B. nach `thinkpad` suchen, wird ein Dokument, das diesen Begriff häufiger enthält, in den Suchergebnissen höher bewertet, als andere Dokumente, in denen der Begriff seltener vorhanden ist.
- Bei HTML-Dokumenten werden Attribute des Abfragebegriffs (das heißt, der Kontext des Begriffs) berücksichtigt. Die Bewertung eines Begriffs wird auf der Basis der Begriffsattribute (wie z. B. Speicherposition, fett, kursiv, Anker usw.) berechnet. Im Allgemeinen werden Begriffe, die im Dokumenttitel vorkommen, höher bewertet als Begriffe, die in einem Textabschnitt vorkommen. Hervorgeho-

bene Begriffe (wie z. B. fetter Text) werden höher bewertet als Begriffe, die in einfachem Text dargestellt sind. Sie können die Bedeutung von Attributen in den Parsereinstellungen konfigurieren.

- Die Wertigkeit von Abfragebegriffen. Wenn Sie die Bedeutung von Begriffen in einem Dokument anpassen wollen, können Sie Boostwerte für Begriffe konfigurieren. In diesem Fall ordnen Sie der Objektgruppe ein Verzeichnis von Boostwörtern zu (das Begriffe und ihre Boostwerte enthält). Während der Suche wird dieses Wörterverzeichnis verwendet, und die Boostwerte der Begriffe tragen zur Bewertung des Dokuments bei. Je höher der Boostwert, desto höher die Bewertung des Dokuments aufgrund des Begriffs.
- Die räumliche Nähe von Abfragebegriffen in einem Dokument. Wenn die Abfragebegriffe in einem Dokument nahe beieinander sind, wird ihre lexikalische Affinität für die Berechnung der Textbewertung verwendet. Angenommen, Sie haben zwei Dokumente. In einem Dokument geht es um einen 'car park' (Parkplatz) in der Stadtmitte (car und park sind dicht nebeneinander). In dem anderen Dokument geht es um einen 'car showroom near to the park' (Autohaus in der Nähe des Parks) in einer Stadt (car und park sind nicht so dicht beieinander). Wenn Sie nun nach car park suchen, wird das erste Dokument aufgrund der räumlichen Nähe der Begriffe höher bewertet als das zweite Dokument.
- Die Länge der Dokumente und die Vokabelvielfalt (wie z. B. die Anzahl einmalig auftretender Wörter) sind weitere Faktoren für die Ermittlung der Dokumentbewertung.

---

## Statische Rangfolge

Bei bestimmten Dokumenttypen können Sie einen statischen Rangfolgefaktor zuordnen, der den Rang dieser Dokumente in den Suchergebnissen erhöht.

Wenn Sie eine Objektgruppe erstellen, geben Sie die Optionen **Dokumentrang** an. Der Dokumentbedeutungstyp, den Sie auswählen, legt fest, ob den Dokumenten in der Objektgruppe ein statischer Rangfolgefaktor zugeordnet wird. Wenn Benutzer eine Objektgruppe durchsuchen, die eine statische Rangfolge verwendet, beeinflusst der statische Rangfolgefaktor wie die zurückgegebenen Dokumente in den Suchergebnissen sortiert werden.

Bei Webinhalten basiert der statische Rangfolgefaktor auf Links. Die Anzahl Links, die von anderen Dokumenten auf ein Dokument verweisen, und der Ursprung dieser Links kann die Relevanz des betreffenden Dokuments in den Suchergebnissen erhöhen.

Bei Dokumenten mit Datumfeldern oder Datumsmetadaten basiert der statische Rangfolgefaktor auf dem Dokumentdatum. Das Datumfeld des Dokuments, das vom Crawler bereitgestellt wird, kann das Datum der letzten Änderung des Dokuments oder der letzten Crawleruche im Dokument sein, je nachdem, wie Sie die Crawlerkonfiguration einrichten.

Das Datum eines Dokuments kann seine Relevanz erhöhen. Beispielsweise sind neuere Artikel in NNTP-Newsgruppen wahrscheinlich relevanter als ältere Artikel. Wenn eine Datenquelle mehrere Datumswerte enthält, können Sie beim Konfigurieren des Crawlers auswählen, welches das wichtigste für das Ermitteln der Relevanz von Dokumenten ist.

Wenn Sie für eine Objektgruppe eine statische Rangfolge verwenden, stellen Sie sicher, dass Sie nicht Datenquellen mischen, die verschiedene Rangfolgetypen in der gleichen Objektgruppe verwenden. Wenn Sie z. B. die Links zu einem Doku-

ment als den statischen Rangfolgefaktor verwenden wollen, stellen Sie sicher, dass die Objektgruppe nur Webdokumente enthält. Die Dokumentrangfolge wird ungenauer, wenn Quellen mit verschiedenen Rangfolgemodellen in der gleichen Objektgruppe kombiniert werden. Außerdem kann die Reihenfolge der Suchergebnisse von der erwarteten Reihenfolge abweichen.

Sie müssen zudem sicherstellen, dass die Dokumente in der Objektgruppe Felder und Werte enthalten, auf die eine statische Rangfolge angewendet werden kann. Beispiel: Eine Objektgruppe ist für die Verwendung der statischen Rangfolge auf der Basis des Dokumentdatums konfiguriert, und ein Crawler der Objektgruppe ist für die Verwendung eines bestimmten Felds für das Dokumentdatum konfiguriert. Wenn ein Dokument das betreffende Feld nicht enthält, kann die Bedeutung des Dokuments nicht richtig eingestuft werden und die Reihenfolge der Suchergebnisse weicht möglicherweise von der erwarteten Reihenfolge ab.

### **Auswirkungen der linkbasierten Rangfolge**

Die statische Rangfolge trägt zusammen mit Faktoren wie z. B. der Zuordnung einer Bewertung zur Verbesserung von URI-Mustern zur statischen Bewertung eines Dokuments bei und beeinflusst die Bedeutung dieses Dokuments. Das linkbasierte Rangfolgemodell wird normalerweise auf Webobjektgruppen angewendet, weil dieses Modell die statische Rangfolge anhand der Links berechnet, die auf das Dokument verweisen. Ein Dokument, auf das viele Links von anderen Dokumenten verweisen, wird mit einer höheren Relevanz eingestuft.

Das heißt, wenn Sie dieses Modell für eine Nicht-Webobjektgruppe oder eine gemischte Objektgruppe (eine die Webdokumente und Nicht-Webdokumente enthält) konfigurieren, wird die Suchqualität möglicherweise vermindert, weil Nicht-Webdokumente kein Linkkonzept haben.

Wenn die linkbasierte Rangfolge aktiviert ist, ist auch die Erkennung doppelter Dokumente aktiviert. Doppelte Dokumente haben dieselben statische Rangfolge wie das Hauptdokument. Wenn für keines der doppelten Dokumente ein Boostfaktor für URI-Muster konfiguriert ist, haben alle doppelten Dokumente dieselbe statische Bewertung.

---

## **Wiederherstellung der Standardwerte für die statische Dokumentrangfolge**

Wenn Sie bei der Erstellung einer Objektgruppe eine Option für die statische Dokumentreihenfolge konfigurieren, können Sie die Merkmale auf die Standardwerte zurücksetzen, indem Sie die Dateien `runtime.properties` für die Objektgruppe bearbeiten.

### **Vorbereitung**

Sie müssen ein Administrator für die Unternehmenssuche sein, um die Standardwerte für die Dokumentrangfolge wiederherstellen zu können.

### **Informationen zu dieser Task**

Zum Wiederherstellen der Standardwerte für die Dokumentrangfolge für eine Objektgruppe müssen Sie die Dateien `runtime.properties` für diese Objektgruppe und alle Suchserver in Ihrem System für die Unternehmenssuche aktualisieren. In einer Konfiguration mit mehreren Servern befindet sich die Datei

runtime.properties auf dem Indexserver im Verzeichnis ES\_NODE\_ROOT/master\_config/objektgruppen-id.runtime.knoten-id. Dabei ist *objektgruppen-id* die ID für die Objektgruppe und *knoten-id* die ID für die Suchserver.

Wenn Sie beispielsweise die Objektgruppe coll in einem System für die Unternehmenssuche mit mehreren Servern aktualisieren wollen, müssen Sie die Datei runtime.properties für diese Objektgruppe und für beide Suchserver (node3 und node4) aktualisieren:

```
ES_NODE_ROOT/master_config/coll.runtime.node3/runtime.properties
ES_NODE_ROOT/master_config/coll.runtime.node4/runtime.properties
```

### Vorgehensweise

Gehen Sie wie folgt vor, um die Werte für die Dokumentrangfolge für eine Objektgruppe auf die Standardwerte zurückzusetzen:

1. Melden Sie sich als Administrator für die Unternehmenssuche am Indexserver an.
2. Geben Sie die Objektgruppen-ID für die Objektgruppe an, für die Sie die Standardwerte für die Dokumentrangfolge wiederherstellen wollen. Die Objektgruppen-ID befindet sich in der Datei ES\_NODE\_ROOT/master\_config/collections.ini. Sortieren Sie diese Datei, damit Sie sie einfacher betrachten können. Im folgenden Beispiel ist coll die Objektgruppen-ID:

```
% sort $ES_NODE_ROOT/master_config/collections.ini | more
collection1.configfile=coll_config.ini
collection1.datadir=/home/eseach/node/data/coll
collection1.description=
collection1.displayname=Collection1
collection1.flags=0
collection1.id=coll
collection1.sectiontype=collection
collection1.type=1
...
```

3. Bearbeiten Sie die Datei runtime.properties für die Objektgruppe, die Sie wiederherstellen wollen, und nehmen Sie folgende Änderungen vor:

- a. Löschen Sie die folgenden Merkmale:

```
trevi.autorank.dfthreshold1
trevi.autorank.dfthreshold2
trevi.autorank.dfthreshold3
trevi.autorank.rc0.*
trevi.autorank.rc1.*
```

- b. Wenn die Datei runtime.properties die Angabe trevi.sourcetype=1 enthält, womit angegeben wird, dass Dokumente nach Links eingestuft werden, müssen Sie die Datei ES\_INSTALL\_ROOT/default\_config/runtime.1/runtime.properties bearbeiten und die folgenden Standardmerkmale kopieren und in die Datei runtime.properties einfügen:

```
trevi.autorank.dfthreshold1
trevi.autorank.dfthreshold2
trevi.autorank.dfthreshold3
trevi.autorank.rc0.*
trevi.autorank.rc1.*
```

- c. Wenn die Datei runtime.properties die Angabe trevi.sourcetype=2 enthält, womit angegeben wird, dass Dokumente nach Datum eingestuft werden, müssen Sie die Datei ES\_INSTALL\_ROOT/default\_config/runtime.2/runtime.properties bearbeiten und die folgenden Standardmerkmale kopieren und in die Datei runtime.properties einfügen:

```
trevi.autorank.dfthreshold1
trevi.autorank.dfthreshold2
trevi.autorank.dfthreshold3
trevi.autorank.rc0.*
trevi.autorank.rc1.*
```

- d. Wenn die Datei `runtime.properties` die Angabe `trevi.sourcetype=3` enthält, womit angegeben wird, dass kein statischer Rangfolgefaktor für die Einstufung von Dokumenten in der Objektgruppe verwendet wird, müssen Sie die Datei `ES_INSTALL_ROOT/default_config/runtime.0/runtime.properties` bearbeiten und die folgenden Standardmerkmale kopieren und in die Datei `runtime.properties` einfügen:

```
trevi.autorank.dfthreshold1
trevi.autorank.dfthreshold2
trevi.autorank.dfthreshold3
trevi.autorank.rc0.*
trevi.autorank.rc1.*
```

4. In einer Konfiguration mit mehreren Servern müssen Sie Schritt 3 auf Seite 226 wiederholen, um die Datei `runtime.properties` für dieselbe Objektgruppe und den zweiten Suchserver zu aktualisieren.
5. Überwachen Sie in der Administrationskonsole die Seite **Suche**, und starten Sie die Suchvorgänge für diese Objektgruppe neu.

Wiederholen Sie diese Schritte für jede Objektgruppe, für die Sie die Standardwerte für die Dokumentrangfolge wiederherstellen wollen.

---

## Benutzerdefinierte Verzeichnisse von Boostwörtern

Zur Verbesserung der Qualität der Suchergebnisse können Sie die Rangfolge von Dokumenten in den Suchergebnissen beeinflussen, indem Sie ein benutzerdefiniertes Verzeichnis von Boostwörtern erstellen.

Wenn eine Abfrage ein Wort enthält, das sich im Verzeichnis von Boostwörtern befindet, wird der Rang von Dokumenten mit diesem Wort erhöht oder herabgesetzt, je nachdem, welcher Boostfaktor für dieses Wort im Wörterverzeichnis konfiguriert ist.

Sie können ein Verzeichnis von Boostwörtern verwenden, um sicherzustellen, dass bestimmte Dokumente zurückgegeben werden, wenn ein Benutzer bestimmte Abfragebegriffe angibt. Angenommen, Sie haben eine Objektgruppe, die viele Dokumente enthält, in denen es um Autos geht. Für diese Dokumente halten Sie vielleicht bestimmte Schlüsselwörter für wichtig, die mit Fahrzeugmodellen, den Herstellernamen und so weiter zu tun haben. Sie können die Rangfolge der Suchergebnisse beeinflussen, indem Sie den Schlüsselwörtern (`Modell`, `Hersteller` und so weiter) in einem Verzeichnis von Boostwörtern einen Boostwert und damit eine bestimmte Gewichtung zuordnen. Wenn Benutzer die Objektgruppe durchsuchen und eine Abfrage angeben, die eines der Schlüsselwörter enthält, werden die Dokumente über Autos in den Suchergebnissen weiter oben angezeigt als andere Dokumente.

Der Bereich der Boostfaktoren liegt zwischen -10 und 10. Während der Abfrageverarbeitung erhöhen die Suchserver den Rang der Dokumente, die Wörter mit einem positiven Boostfaktor enthalten, und setzen den Rang der Dokumente herab, die Wörter mit einem negativen Boostfaktor enthalten.

Ein Dokument, das Abfragebegriffe mit hohen Boostfaktoren enthält, wird höher eingestuft als es ohne Anwendung des Boostfaktors der Fall wäre. (Der Boostfaktor ist der einzige Faktor, der zur Bewertung des Dokuments beiträgt.)

Bei der Erstellung des Verzeichnisses können Sie denselben Boostfaktor einer beliebigen Anzahl Wörter zuordnen. Das Verzeichnis kann Begriffe mit einem oder mit mehreren Wörtern enthalten. Begriffe mit mehreren Wörtern werden als ein Ausdruck angesehen.

Wenn ein durch einen Boostwert gewichtetes Wort in einer Abfrage angegeben wird, die den Operator OR verwendet (z. B. dies | das), wird ein gewichtetes Mittel für die Abfragebegriffe berechnet. Die daraus resultierende zusammengefasste Bewertung wird für alle Vorkommen des Abfrageoperanden OR verwendet. Für unterschiedliche OR-Abfrageoperanden werden keine unterschiedlichen Bewertungen berechnet.

Boosting, das auf Verzeichnissen von Boostwörtern basiert, wird für feldspezifische Abfragebegriffe nicht unterstützt. Bei der syntaktischen Analyse der Abfragebegriffe wird nur der Abfragetext und nicht der Feldname zur Berechnung der Dokumentbewertung verwendet. Zur Anwendung von Boostfaktoren auf Abfragebegriffe, die in Feldern vorkommen, können Sie den Boostklassen Feldnamen zuordnen.

Zum Erstellen eines Verzeichnisses von Boostwörtern muss ein Fachmann auf dem Sachgebiet der Objektgruppe eine Liste mit Boostwörtern im XML-Format erstellen oder mit einem Anwendungsentwickler die XML-Datei erstellen. Zum Umwandeln der XML-Datei in eine binäre Datei (.dic ) muss ein Tool für die Unternehmenssuche (**esboosttermdictbuilder**) verwendet werden.

Ein Administrator für die Unternehmenssuche lädt die Binärdatei auf das System hoch und ordnet ihr einen Anzeigenamen zu. Objektgruppenadministratoren können ein Verzeichnis von Boostwörtern für das Durchsuchen von Dokumenten in einer Objektgruppe auswählen, wenn sie die Suchserveroptionen für eine Objektgruppe konfigurieren.

**Einschränkung:** Sie können ein benutzerdefiniertes Verzeichnis von Boostwörtern nicht mehr ändern, wenn Sie es dem System hinzugefügt haben. Gehen Sie wie folgt vor, um die Boostwörter für die Abfrageverarbeitung zu überarbeiten:

1. Aktualisieren Sie die XML-Quellendatei.
2. Wandeln Sie die XML-Quelle in eine neue Verzeichnisdatei um.
3. Entfernen Sie das alte Verzeichnis von Boostwörtern aus allen Objektgruppen, die es verwenden.
4. Löschen Sie das alte Verzeichnis von Boostwörtern vom System.
5. Fügen Sie dem System das neue Verzeichnis von Boostwörtern hinzu.
6. Ordnen Sie das neue Verzeichnis von Boostwörtern den Objektgruppen zu, die es verwenden sollen.

Sie können ein Script schreiben, das diese Schritte einschließt, und das Script anschließend verwenden, um das Wörterverzeichnis erneut in Ihrem System für die Unternehmenssuche zu implementieren.

#### **Zugehörige Konzepte**

„Dokumentrangfolge“ auf Seite 223



Benutzerdefinierte Verzeichnisse von Boostwörtern

## Zugehörige Tasks

 Erstellen einer XML-Datei für Boostwörter

 Erstellen eines Verzeichnisses von Boostwörtern

„Erneutes Implementieren von benutzerdefinierten Wörterverzeichnissen“ auf Seite 216

## Hinzufügen von Verzeichnissen von Boostwörtern zum System

Wenn Sie benutzerdefinierte Verzeichnisse von Boostwörtern erstellen, müssen Sie diese Verzeichnisse dem System für die Unternehmenssuche zuordnen. Sie können später entscheiden, welches Verzeichnis von Boostwörtern Sie zum Durchsuchen einer Objektgruppe verwenden wollen.

### Vorbereitung

Sie müssen ein Administrator für die Unternehmenssuche sein, um dem System benutzerdefinierte Verzeichnisse von Boostwörtern hinzuzufügen.

### Einschränkungen

Die maximale Größe eines Verzeichnisses von Boostwörtern ist 8 MB.

### Vorgehensweise

Gehen Sie wie folgt vor, um dem System für die Unternehmenssuche benutzerdefinierte Boostwörter zuzuordnen:

1. Klicken Sie **System** an, um die Sicht **System** zu öffnen.
2. Klicken Sie  **Bearbeiten** an, um zur Sicht für die Systembearbeitung zu wechseln.
3. Klicken Sie **Verzeichnisse von Boostwörtern konfigurieren** auf der Seite **Suche** an.
4. Klicken Sie **Verzeichnis von Boostwörtern hinzufügen** auf der Seite **Verzeichnisse von Boostwörtern konfigurieren** an.
5. Geben Sie auf der Seite **Verzeichnis von Boostwörtern hinzufügen** einen eindeutigen Anzeigenamen für das Verzeichnis ein. Optional können Sie auch eine Beschreibung eingeben.
6. Geben Sie die Speicherposition der .dic-Datei an. Wenn sich die Datei auf Ihrem lokalen System befindet, können Sie nach der Datei suchen. Wenn sich die Datei auf dem Indexserver befindet, geben Sie den vollständig qualifizierten Pfad ein.
7. Klicken Sie **OK** an. Ihr benutzerdefiniertes Verzeichnis von Boostwörtern wird dem System für die Unternehmenssuche hinzugefügt und ist für das Durchsuchen von Objektgruppen verfügbar.

## Zuordnen eines Verzeichnisses von Boostwörtern zu einer Objektgruppe

Wenn dem System für die Unternehmenssuche Verzeichnisse von Boostwörtern zugeordnet wurden, können Sie eines für die Verwendung beim Suchen in einer Objektgruppe auswählen. Wenn ein Abfragebegriff mit einem Begriff in diesem

Verzeichnis übereinstimmt, wird der Rang der Dokumente mit diesem Begriff erhöht oder herabgesetzt, je nachdem, welcher Boostfaktor diesem Begriff im Verzeichnis zugeordnet ist.

### Vorbereitung

Sie müssen ein Administrator für die Unternehmenssuche oder ein Objektgruppenadministrator der Objektgruppe sein, um ein Verzeichnis von Boostwörtern für eine Objektgruppe auszuwählen.

### Vorgehensweise

Gehen Sie wie folgt vor, um einer Objektgruppe ein Verzeichnis von Boostwörtern zuzuordnen:

1. Bearbeiten Sie eine Objektgruppe, wählen Sie die Seite **Suche** aus, und klicken Sie **Suchserveroptionen konfigurieren** an.
2. Wählen Sie im Feld **Name des Verzeichnisses von Boostwörtern** auf der Seite **Suchserveroptionen** das Verzeichnis von Boostwörtern aus, das verwendet werden soll, wenn Benutzer diese Objektgruppe abfragen.

Die Liste der verfügbaren Verzeichnisse enthält alle Verzeichnisse von Boostwörtern, die dem System für die Unternehmenssuche hinzugefügt wurden.

3. Klicken Sie **OK** an.

---

## Auf URI-Mustern basierende Dokumentrangfolge

Sie können den Rang von Dokumenten erhöhen oder herabsetzen, indem Sie URI-Mustern Boostfaktoren zuordnen.

Allen Dokumenten wird eine statische Standardrangfolgebewertung zugeordnet, wenn sie dem Index hinzugefügt werden. Die Standardbewertung variiert in Abhängigkeit davon, ob eine statische Rangfolge für die Objektgruppe aktiviert wurde, und falls ja, welcher statische Rangfolgetyp (nach Dokumentdatum oder bei Webdokumenten die Anzahl anderer Dokumente, die mit ihnen verknüpft sind) verwendet wird.

Sie können den relativen Rang von Dokumenten beeinflussen, indem Sie URI-Mustern Boostfaktoren zuordnen. Der Boostfaktor wird zusammen mit der statischen Standardrangfolgebewertung und anderen Faktoren verwendet, um die statische Endbewertung eines Dokuments zu ermitteln.

Dabei ist die Reihenfolge der konfigurierten URI-Muster wichtig. Der Indexserver bewertet die URI-Muster in der Reihenfolge, in der Sie aufgelistet sind, wenn die Werte der einzelnen Dokumente in einer Objektgruppe berechnet werden. Für jeden URI gilt Folgendes:

1. Der Indexserver durchsucht die URI-Muster nacheinander.
2. Wenn der Indexserver das erste URI-Muster findet, das mit einem Dokument im Index übereinstimmt, wird der Boostfaktor angewendet, der für dieses URI-Muster für das Dokument konfiguriert ist.
3. Wenn ein Dokument nicht mit einem URI-Muster übereinstimmt, wird die statische Standardrangfolgebewertung verwendet.

Nachdem Sie einen Boostfaktor für ein URI-Muster konfiguriert haben, müssen Sie das URI-Muster in der Reihenfolge positionieren, in der es vom Indexserver überprüft werden soll.

## Beeinflussen der Bewertungen von Dokumenten, die mit URI-Mustern übereinstimmen

Sie können den Rang von Dokumenten erhöhen oder herabsetzen, die mit einem URI-Muster übereinstimmen, indem Sie einen Boostfaktor auf die statische Standardrangfolgebewertung anwenden.

### Vorbereitung

Sie müssen die Rolle eines Administrators für die Unternehmenssuche haben oder ein Objektgruppenadministrator für diese Objektgruppe sein, um den Rang von Dokumenten zu beeinflussen, die mit einem URI-Muster übereinstimmen.

### Informationen zu dieser Task

Der konfigurierte Boostfaktor wird zusammen mit der statischen Standardrangfolgebewertung verwendet, um eine neue statische Bewertung für alle Dokumente zu berechnen, die mit dem angegebenen URI-Muster übereinstimmen.

Die Boostfaktoren beeinflussen nur statische Bewertungen, und diese Faktoren sind nur ein Teil der Berechnung, die den endgültigen Rang eines Dokuments ermittelt. Wenn ein Dokument beispielsweise über eine hohe Anzahl von Links verfügt (was eine hohe Anfangsbewertung zur Folge hat), wird ein Dokument ohne Links immer niedriger eingestuft.

### Vorgehensweise

Gehen Sie wie folgt vor, um die Dokumentbewertungen zu beeinflussen, die mit einem URI-Muster übereinstimmen:

1. Bearbeiten Sie eine Objektgruppe, wählen Sie die Seite **Index** aus, und klicken Sie **Bewertungen durch URI-Mustererkennung beeinflussen** an.
2. Klicken Sie **URI-Muster hinzufügen** auf der Seite **Bewertungen durch URI-Mustererkennung beeinflussen** an.
3. Geben Sie ein URI-Muster für Dokumente an, deren Rang in den Suchergebnissen erhöht oder herabgesetzt werden soll. Beispiel:  
`http://domain.org/hr/*  
db2://*ROI*  
*/afs/*`
4. Geben Sie für den Boostfaktor einen Wert zwischen -10 und 10 ein. Die statische Endbewertung für alle Dokumente, die mit dem URI-Muster übereinstimmen, wird auf Basis dieses Boostfaktors berechnet.
5. Klicken Sie **OK** an.
6. Positionieren Sie das neue URI-Muster auf der Seite **Bewertungen durch URI-Mustererkennung beeinflussen** in der Reihenfolge, in der es vom Indexserver überprüft werden soll.

Der Indexserver berechnet die statischen Rangfolgebewertungen in der Reihenfolge, in der die URIs aufgeführt sind. Führen Sie die genaueren URIs zuerst auf, um bestmögliche Ergebnisse zu erzielen. Im folgenden Beispiel stimmt das Unterverzeichnis `/forms` mit dem URI-Muster `http://www.ibm.com/hr/*` überein. Führen Sie das URI-Muster für das Unterverzeichnis `/forms` zuerst auf, um sicherzustellen, dass die Bewertungen für Dokumente im Unterverzeichnis `/forms` ordnungsgemäß berechnet werden.

```
http://www.ibm.com/hr/forms/* 8
http://www.ibm.com/hr/* -2
```

7. Wenn Sie ein URI-Muster oder den Boostfaktor ändern wollen, wählen Sie das URI-Muster aus, und klicken Sie  **Bearbeiten** an.
8. Wenn Sie ein URI-Muster aus der Liste entfernen wollen, wählen Sie das URI-Muster aus, und klicken Sie  **Entfernen** an.
9. Erstellen Sie den Hauptindex neu, um die Boostfaktoren auf bereits indexierte Dokumente anzuwenden.

---

## Auf Boostklassen basierende Dokumentrangfolge

Durch die Zuordnung von Feldern zu Boostklassen können Sie beeinflussen, wie Dokumente in den Suchergebnissen eingestuft werden.

Bei der syntaktischen Analyse von Dokumenten ordnet der Parser Dokumenttoken *Boostklassen* zu und richtet sich dabei nach den Feldern, zu denen die Token gehören. Diese Boostklassen werden in den Index aufgenommen und im Rahmen der Abfragebewertung zur Berechnung der Bewertungen verwendet, die zur Einstufung der Ergebnisdokumente beitragen.

Sie können numerische Boostfaktoren für die Boostklassen konfigurieren, um Einfluss darauf zu nehmen, wie die Bewertungen berechnet werden. Wenn ein Abfragebegriff mit einem Token in einem Feld übereinstimmt, das einer Boostklasse zugeordnet ist, beeinflusst der Beitrag dieses Tokenvorkommens die Gesamtbewertung des Dokuments. Die Bewertung wird durch Anwendung des in der Boostklasse konfigurierten Boostfaktors berechnet.

Sie möchten zum Beispiel die Bewertungen von Titelfeldern erhöhen. Wenn ein Abfragebegriff im Titel vorkommt, liefert dieses Vorkommen einen hohen Beitrag zur Dokumentbewertung und hilft dem Dokument, in den Suchergebnissen höher eingestuft zu werden.

Verwenden Sie zur Beeinflussung der Dokumentrangfolge die Administrationskonsole für die Unternehmenssuche, um Boostfaktoren für Boostklassen anzugeben und den Boostklassen Felder zuzuordnen. Sechzehn Boostklassen sind für die Unternehmenssuche vorkonfiguriert. Acht Boostklassen sind zur Verwendung mit Inhaltsfeldern vorgesehen, die anderen acht Boostklassen sind für Metadatenfelder vorgesehen. Sie können die Bewertungen bearbeiten, die den Standardboostklassen zugeordnet sind, und außerdem den Boostklassen andere oder zusätzliche Felder zuzuordnen.

Wenn Sie die Feldzuordnungen ändern, müssen Sie die Dokumente erneut mit dem Crawler durchsuchen und syntaktisch analysieren, damit die Änderungen auf bereits indexierte Dokumente angewendet werden können. Wenn Sie die für eine Boostklasse angegebenen Faktoren ändern, überwachen Sie die Suchserver, stoppen Sie die Suchserverprozesse, und starten Sie sie erneut, damit die Änderungen wirksam werden.

### Erkennung von doppelten Dokumenten und Dokumentzusammenfassungen

Bei der Zuordnung eines Felds zu einer Boostklasse müssen Sie angeben, ob das Feld zur Erkennung von doppelten Dokumenten verwendet wird und ob der Inhalt des Felds in Dokumentzusammenfassungen in den Suchergebnissen eingeschlossen werden kann.

- Wenn ein Feld zur Erkennung von doppelten Dokumenten verwendet wird, wird das Feld als Inhaltsfeld betrachtet, und nur die für Inhaltsfelder konzipier-

ten Boostklassen können ausgewählt werden. Der Inhalt dieser Feldtypen kann in dynamischen Dokumentzusammenfassungen in den Suchergebnissen verwendet werden.

- Wenn das Feld nicht zur Erkennung von doppelten Dokumenten verwendet wird, wird das Feld als Metadatenfeld betrachtet und nur die für Metadatenfelder konzipierten Boostklassen können ausgewählt werden. In diesem Fall werden zwei Dokumente, die mit Ausnahme des angegebenen Felds miteinander übereinstimmen, als Duplikate betrachtet, und das Feld wird nicht in dynamischen Dokumentzusammenfassungen verwendet.

## Werte für viele oder wenige Ergebnisse

Bei der Bewertung einer Abfrage nimmt der Suchvorgang eine Schätzung vor, wie viele Ergebnisdokumente zurückgegeben werden. Schwellenwerte legen fest, ob eine Abfrage wenige oder viele Ergebnisse zurückgibt, oder ob die Anzahl zwischen den beiden Schwellenwerten liegt:

### Wenige Ergebnisse

Wenn die geschätzte Anzahl der Ergebnisdokumente unterhalb der unteren Schwelle liegt, handelt es sich um eine Abfrage mit wenigen Ergebnissen.

### Viele Ergebnisse

Wenn die geschätzte Anzahl der Ergebnisdokumente die obere Schwelle übersteigt, handelt es sich um eine Abfrage mit vielen Ergebnissen.

### Ergebnisanzahl zwischen den Schwellenwerten

Wenn die geschätzte Anzahl zwischen den beiden Schwellenwerten liegt, ist der Wert der Abfrage eine Zusammenfassung dieser beiden Schwellenwerte.

Jede Boostklasse gibt Boostfaktoren an, die während der Abfrageverarbeitung Abfragen mit wenigen und vielen Ergebnissen zugeordnet werden. Der niedrige Boostfaktor beeinflusst den relativen Rang von Abfragen mit wenigen Ergebnissen, und der hohe Boostfaktor beeinflusst den relativen Rang von Abfragen mit vielen Ergebnissen. Eine Zusammenfassung der beiden Boostfaktoren beeinflusst den relativen Rang von Abfragen mit einem Wert, der zwischen den beiden Schwellenwerten liegt.

Die Werte der Boostfaktoren steuern den relativen Rang aller Vorkommen eines Abfragebegriffs in einem Dokument. Jedes Vorkommen eines Abfragebegriffs in einem Dokument wird gemäß des entsprechenden Boostfaktors gezählt.

Bei der Konfiguration von Boostklassen für eine Objektgruppe können die Standardboostfaktoren bearbeitet werden. Sie könnten beispielsweise Boostfaktoren angeben, um sicherzustellen, dass die in Titelfeldern vorkommenden Abfragebegriffe fünfmal so viel zählen wie Abfragebegriffe, die in normalem Text vorkommen.

### Zugehörige Konzepte

„Dokumentrangfolge“ auf Seite 223

## Zuordnen von Feldern zu Boostklassen

Sie können den relativen Rang von Feldern beeinflussen, indem Sie Boostklassen Feldnamen zuordnen.

### Vorbereitung

Sie müssen ein Administrator für die Unternehmenssuche oder ein Objektgruppenadministrator der Objektgruppe sein, um Boostklassen Felder zuzuordnen.

### Informationen zu dieser Task

Das System verwendet Boostfaktoren, um die Rangfolge von Dokumenten zu beeinflussen, die Abfragebegriffe in den Feldern enthalten, die Boostklassen zugeordnet werden.

Die Unternehmenssuche reserviert einige Zuordnungen für interne Felder und normalen Text, die bzw. der keine weiteren definierenden Merkmale enthalten bzw. enthält. Sie können den Boostklassen, die die reservierten Felder verwenden, andere Felder zuordnen, aber Sie können die reservierten Felder nicht bearbeiten oder löschen.

### Vorgehensweise

Gehen Sie wie folgt vor, um Boostklassen Felder zuzuordnen:

1. Bearbeiten Sie eine Objektgruppe, wählen Sie die Seite **Syntaxanalyse** aus, und klicken Sie **Felder zu Boostklassen zuordnen** an.
2. Klicken Sie **Feld hinzufügen** auf der Seite **Felder zu Boostklassen zuordnen** an.
3. Geben Sie auf der Seite **Feld einer Boostklasse hinzufügen** den Namen des Felds ein, das Sie einer Boostklasse zuordnen möchten.

Sie können den Namen eines Felds in einer durchsuchten oder in einer externen Quelle angeben, oder Sie können den Namen eines Felds angeben, das von einem XML-Element zugeordnet wird. Sie können auch den Namen eines Felds, das von einem HTML-Metadatenelement zugeordnet wird, oder einen der vordefinierten Feldnamen angeben.

4. Geben Sie an, ob das Feld für die Erkennung von doppelten Dokumenten verwendet wird. Wenn Sie das Markierungsfeld auswählen, enthält die Liste der verfügbaren Boostklassen Klassen, die auf Inhaltsfelder angewendet werden. Wenn ein Dokument mit diesem Feld als Suchergebnis zurückgegeben wird, wird der Inhalt des Felds im Dokumentzusammenfassungsbereich angezeigt. Wenn Sie das Markierungsfeld inaktivieren, enthält die Liste mit verfügbaren Boostklassen Klassen, die auf Metadatenfelder angewendet werden. Der Inhalt des Felds wird nicht im Dokumentzusammenfassungsbereich der Suchergebnisse angezeigt.

5. Wählen Sie eine Boostklasse aus, und klicken Sie **OK** an.

Das hinzugefügte Feld wird auf der Seite **Felder zu Boostklassen zuordnen** angezeigt. Sie können eine Option zum Bearbeiten der Boostklasse auswählen und verschiedene Boostfaktoren konfigurieren, um die Bewertungen der Dokumente zu ermitteln, die dieses Feld enthalten.

6. Klicken Sie  **Bearbeiten** an, um festzulegen, ob ein Feld zur Erkennung von doppelten Dokumenten verwendet werden soll, oder um das Feld einer anderen Boostklasse zuzuordnen. (Für die Unternehmenssuche reservierte Felder können nicht bearbeitet werden.)
7. Klicken Sie  **Entfernen** an, um ein Feld aus einer Boostklasse zu entfernen. (Für die Unternehmenssuche reservierte Felder können nicht entfernt werden.)
8. Zum Anwenden von Änderungen auf bereits indizierte Dokumente müssen diese Dokumente erneut durchsucht und indiziert werden.

### Zugehörige Konzepte

„Dokumentrangfolge“ auf Seite 223

## Konfigurieren von Boostfaktoren für Boostklassen

Die Boostfaktoren, die Sie für Boostklassen konfigurieren, stellen Ihre Einschätzung dar, wie wichtig das Vorhandensein bestimmter Felder in Ergebnisdokumenten für eine Abfrage ist. Boostklassen mit hohen Boostfaktoren können den Rang von Ergebnisdokumenten erhöhen, die Felder enthalten, die den Boostklassen zugeordnet werden.

### Vorbereitung

Sie müssen ein Administrator für die Unternehmenssuche oder ein Objektgruppenadministrator der Objektgruppe sein, um Boostfaktoren für Boostklassen zu konfigurieren.

### Informationen zu dieser Task

Das System verwendet die für eine Boostklasse konfigurierten Boostfaktoren, die statische Rangfolgebewertung und weitere Faktoren zur Berechnung einer neuen Bewertung für Ergebnisdokumente, die Felder enthalten, die der Boostklasse zugeordnet sind.

### Vorgehensweise

Gehen Sie wie folgt vor, um Boostfaktoren für Boostklassen zu konfigurieren:

1. Bearbeiten Sie eine Objektgruppe, wählen Sie die Seite **Syntaxanalyse** aus, und klicken Sie **Felder zu Boostklassen zuordnen** an.
2. Klicken Sie auf der Seite **Felder zu Boostklassen zuordnen** die Option **Boostklassen bearbeiten** an.
3. Suchen Sie auf der Seite **Boostklassen** die Boostklasse, die Sie ändern wollen, und klicken Sie  **Bearbeiten** an.
4. Geben Sie auf der Seite **Boostklasse bearbeiten** neue Werte für die hohen und niedrigen Boostfaktoren an. Sie können für beide Faktoren denselben Wert eingeben.
5. Klicken Sie **OK** an.
6. Damit die Änderungen wirksam werden, überwachen Sie die Suchserver, und wählen Sie die Symbole zum Stoppen und erneuten Starten der Suchvorgänge aus. Wenn Benutzer Abfragen übergeben, wird der relative Rang der Ergebnisdokumente, die Felder enthalten, die dieser Boostklasse zugeordnet sind, durch die neuen Boostfaktoren ermittelt.

### Zugehörige Konzepte

„Dokumentrangfolge“ auf Seite 223

## Standardwerte für Boostklassen

Die Unternehmenssuche stellt 16 Boostklassen zur Verfügung, mit deren Hilfe Sie beeinflussen können, wie Dokumente in den Suchergebnissen eingestuft werden.

Zur Berechnung von Bewertungen für Felder und Text, die bzw. der keine weiteren definierenden Merkmale aufweisen bzw. aufweist, sind die folgenden Felder für die Unternehmenssuche reserviert:

```
es_special_field.regular_text
es_special_field.default_field
es_special_field.default_metadata_field
```

Sie können den Boostklassen, die die reservierten Felder verwenden, andere Felder zuordnen, aber Sie können die reservierten Felder nicht bearbeiten oder löschen.

Für alle anderen Felder können Sie die Boostfaktoren bearbeiten, die das System zur Berechnung von Dokumenträngen verwendet. Ferner können Sie den Boostklassen beliebig viele Felder zuordnen. Dazu gehören auch die Boostklassen, die von den reservierten Feldern verwendet werden.

In der folgenden Tabelle sind die Namen der Boostklassen, die Standardboostfaktoren für Abfragen mit wenigen Ergebnissen, die Standardboostfaktoren für Abfragen mit vielen Ergebnissen und die Namen der vordefinierten Felder aufgelistet, die den Boostklassen in der Standardkonfiguration zugeordnet sind.

Die Standardboostfaktoren variieren in Abhängigkeit von der Methode für die statische Rangfolge, die für die Objektgruppe während ihrer Erstellung ausgewählt wurde. Die Optionen umfassen keine statische Rangfolge, einen Rang, der durch die Anzahl Links auf ein Dokument (für Webquellen) ermittelt wird, und einen Rang, der durch das Dokumentdatum ermittelt wird.

*Tabelle 6. Standardwerte für Boostklassen*

<b>Hohe und niedrige Standardboostfaktoren</b>				
<b>Boostklassenname</b>	<b>Keine statische Rangfolge</b>	<b>Dokumentverknüpfung</b>	<b>Dokumentdatum</b>	<b>Vordefinierte Feldzuordnungen</b>
Inhaltsklasse A	Niedrig: 4 Hoch: 2	Niedrig: 6 Hoch: 1	Niedrig: 4 Hoch: 2	es_special_field.regular_text
Inhaltsklasse B	Niedrig: 5 Hoch: 4	Niedrig: 7 Hoch: 3	Niedrig: 5 Hoch: 4	es_special_field.html_emphasized_text  Enthält die folgenden HTML-Elemente: b, big, caption, dfn, em, h4, h5, h6, strong
Inhaltsklasse C	Niedrig: 7 Hoch: 4	Niedrig: 9 Hoch: 3	Niedrig: 7 Hoch: 4	es_special_field.html_headers  Enthält die folgenden HTML-Elemente: h1, h2, h3
Inhaltsklasse D	Niedrig: 2 Hoch: 5	Niedrig: 1 Hoch: 5	Niedrig: 2 Hoch: 5	Titel
Inhaltsklasse E	Niedrig: 1 Hoch: 1	Niedrig: 5 Hoch: 10	Niedrig: 1 Hoch: 1	es_special_field.anchor
Inhaltsklasse F	Niedrig: 1 Hoch: 1	Niedrig: 1 Hoch: 1	Niedrig: 1 Hoch: 1	es_special_field.anchor_same_dir
Inhaltsklasse G	Niedrig: 1 Hoch: 1	Niedrig: 1 Hoch: 1	Niedrig: 1 Hoch: 1	es_special_field.anchor_same_host
Inhaltsklasse H	Niedrig: 1 Hoch: 1	Niedrig: 1 Hoch: 1	Niedrig: 1 Hoch: 1	es_special_field.default_field
Metadatenklasse A	Niedrig: 1 Hoch: 1	Niedrig: 1 Hoch: 1	Niedrig: 1 Hoch: 1	es_special_field.default_metadata_field
Metadatenklasse B	Niedrig: 1 Hoch: 1	Niedrig: 1 Hoch: 1	Niedrig: 1 Hoch: 1	
Metadatenklasse C	Niedrig: 1 Hoch: 1	Niedrig: 1 Hoch: 1	Niedrig: 1 Hoch: 1	
Metadatenklasse D	Niedrig: 1 Hoch: 1	Niedrig: 1 Hoch: 1	Niedrig: 1 Hoch: 1	
Metadatenklasse E	Niedrig: 1 Hoch: 1	Niedrig: 5 Hoch: 1	Niedrig: 1 Hoch: 1	keywords

Tabelle 6. Standardwerte für Boostklassen (Forts.)

Hohe und niedrige Standardboostfaktoren				
Boostklassenname	Keine statische Rangfolge	Dokumentverknüpfung	Dokumentdatum	Vordefinierte Feldzuordnungen
Metadatenklasse F	Niedrig: 1 Hoch: 1	Niedrig: 3 Hoch: 1	Niedrig: 1 Hoch: 1	es_special_field.urlhost
Metadatenklasse G	Niedrig: 1 Hoch: 1	Niedrig: 1 Hoch: 1	Niedrig: 1 Hoch: 1	es_special_field.urlpath
Metadatenklasse H	Niedrig: 1 Hoch: 1	Niedrig: 1 Hoch: 1	Niedrig: 1 Hoch: 1	Beschreibung

**Zugehörige Konzepte**

„Dokumentrangfolge“ auf Seite 223



---

## Suchanwendungen für die Unternehmenssuche

Mit einer Suchanwendung können Sie Objektgruppen und externe Quellen in Ihrem System für die Unternehmenssuche durchsuchen. Sie können eine beliebige Anzahl Suchanwendungen erstellen, und eine einzelne Suchanwendung kann eine beliebige Anzahl Objektgruppen und externe Quellen durchsuchen.

### Mustersuchanwendung

Die Mustersuchanwendung veranschaulicht viele Such- und Abruffunktionen, die für die Unternehmenssuche verfügbar sind. Diese Musteranwendung ist auch ein funktionsfähiges Beispiel, das veranschaulicht, wie Sie mit IBM SI-API (IBM Search and Index API) interaktive, benutzerdefinierte Suchanwendungen erstellen können, die die Ziele Ihres Unternehmens widerspiegeln.

Mit der Mustersuchanwendung können Sie alle aktiven Objektgruppen und externen Quellen in Ihrem System für die Unternehmenssuche durchsuchen, sofern Sie keine Merkmale in der Standardkonfigurationsdatei ändern. Sie können die Mustersuchanwendung zum Testen von Objektgruppen und externen Quellen verwenden, bevor Sie diese Benutzern zur Verfügung stellen.

Die Mustersuchanwendung wird automatisch allen Objektgruppen und externen Quellen zugeordnet. In einer Produktionsumgebung steuern Administratoren für die Unternehmenssuche, welche Suchanwendungen verschiedene Objektgruppen durchsuchen dürfen.

### Benutzerdefinierte Suchanwendungen

Sie können die Suchanwendungen, die Sie erstellt haben, als eigenständige Webanwendungen in einer IBM WebSphere Application Server-Umgebung ausführen, oder Sie können sie als Portlets in einer IBM WebSphere Portal-Umgebung starten. Wenn Sie die SI-API verwenden, können Sie Suchanwendungen entwerfen, die wie die Mustersuchanwendung in beiden Umgebungen nahtlos arbeiten.

Zur Unterstützung bei der Anpassung der Suchanwendungen können Sie die Anpassungsfunktion für die Suchanwendung verwenden. Diese Anwendung ermöglicht Ihnen, eine Auswahl in einer grafischen Oberfläche vorzunehmen und die Auswirkungen der Änderungen angezeigt zu bekommen, während Sie diese vornehmen. Wenn Sie Ihre Änderungen speichern, aktualisieren Sie die Konfigurationsdatei für die Suchanwendung.

#### Tipp:

Detaillierte Beispiele für die Verwendung der Anpassungsfunktion für die Suchanwendung und die Installation der Suchanwendung als Portlet in WebSphere Portal finden Sie im IBM Redbook *IBM OmniFind Enterprise Edition Configuration and Implementation Scenarios*.

#### Zugehörige Konzepte

„Indexierte Optionen für das Durchsuchen von Dokumenten“ auf Seite 194

„Suchanwendungs-IDs“ auf Seite 283

„Sicherheit auf Dokumentebene“ auf Seite 284

 SI-API-Übersicht (Search and Index API)

 Abfragesyntax

 Festlegen von Abfragemerkmalen

---

## Zuordnen von Suchanwendungen zu Objektgruppen

Bevor Sie eine neue Suchanwendung verwenden können, müssen Sie diese den Objektgruppen zuordnen, die sie durchsuchen soll.

### Vorbereitung

Sie müssen ein Administrator für die Unternehmenssuche sein, damit Sie Suchanwendungen den zu durchsuchenden Objektgruppen zuordnen können.

### Vorgehensweise

Gehen Sie wie folgt vor, um einer Suchanwendung mindestens eine Objektgruppe zuzuordnen:

1. Klicken Sie **Sicherheit** in der Funktionsleiste der Administrationskonsole an.
2. Klicken Sie **Suchanwendungen konfigurieren** auf der Seite **Suchanwendungen** an.
3. Klicken Sie **Suchanwendung hinzufügen** auf der Seite **Suchanwendungen konfigurieren** an.
4. Geben Sie den Namen der Suchanwendung ein.
5. Wählen Sie die Objektgruppen aus, die die Anwendung durchsuchen soll:
  - Klicken Sie **Alle Objektgruppen und externen Quellen** an, wenn die Suchanwendung alle Objektgruppen durchsuchen soll, die Sie dem System hinzufügen.
  - Klicken Sie **Bestimmte Objektgruppen und externe Quellen** an, wenn die Suchanwendung nur die von Ihnen angegebenen Objektgruppen durchsuchen soll.

Wenn Sie diese Option auswählen, wird eine Liste mit Objektgruppennamen und Namen externer Quellen angezeigt. Wählen Sie das Markierungsfeld **Auswählen** für jede Objektgruppe aus, die die Anwendung durchsuchen soll.
6. Klicken Sie **OK** an.

---

## Funktionen der Mustersuchanwendung

Die Mustersuchanwendung für die Unternehmenssuche veranschaulicht die meisten Suchfunktionen, die Sie in Ihre benutzerdefinierten Suchanwendungen integrieren können.

Mit der Mustersuchanwendung können Sie alle Objektgruppen und externen Quellen gleichzeitig durchsuchen. Sie können mit dieser Anwendung alle Objektgruppen und externen Quellen im System für die Unternehmenssuche durchsuchen, sofern die Standardanwendungsmerkmale nicht modifiziert werden.

## Abfragefunktionen

Mit diesen Funktionen können Sie die folgenden Tasks ausführen:

- Angeben von einfachen Freiformatabfragen.
- Angeben von komplexeren Abfragen zur Genauigkeitsverbesserung der Suchergebnisse. Sie können z. B. bestimmte Felder oder XML-Elemente durchsuchen oder mit einer Abfragesyntax nach Dokumenten suchen, die bestimmte Wörter und Ausdrücke einschließen bzw. ausschließen.
- Angeben der Objektgruppen und externen Quellen, die durchsucht werden sollen.
- Durchsuchen bestimmter Quellentypen oder aller Quellentypen.
- Durchsuchen bestimmter Dokumenttypen. Sie können z. B. nur Microsoft Word-Dokumente oder nur Dokumente im PDF-Format durchsuchen.
- Angeben der Sprache für Ihre Abfragebegriffe. Ferner können Sie die Sprachen der Dokumente angeben, die durchsucht werden sollen.
- Durchsuchen bestimmter Untergruppen einer Objektgruppe. Eine Suchanwendung kann z. B. Ihre Anzeige auf einen vordefinierten Dokumentbereich (einen Gültigkeitsbereich) begrenzen. Sie können auch eine Abfrage übergeben, die nur die Dokumente in einer benannten Kategorie durchsucht.
- Erweitern der Abfrage, um Synonyme der Abfragebegriffe aufzunehmen. Wenn der Objektgruppe ein Synonymverzeichnis zugeordnet wurde, werden Dokumente, die Synonyme Ihrer Abfragebegriffe enthalten, als Suchergebnisse zurückgegeben.

## Suchergebnisfunktionen

Mit diesen Funktionen können Sie die folgenden Tasks ausführen:

- Anzeigen der Suchergebnisse, die mit den Abfragebegriffen übereinstimmen.
- Steuern der Anzahl angezeigter Ergebnisdokumente auf jeder Seite und Vorwärts- und Rückwärtsblättern in der Ergebnismenge.
- Aus- und Einblenden von Details zu den Ergebnisdokumenten. Sie können z. B. in jedem Ergebnisdokument Kurzbeschreibungen der Dokumente oder Details wie Feldnamen anzeigen.
- Ausblenden von Dokumenten von derselben Quelle. Wenn z. B. eine Quelle 100 Dokumente zurückgibt, werden die beiden wichtigsten Dokumente in der Ergebnismenge zusammen gruppiert. Sie können die verbleibenden 98 Dokumente anzeigen, indem Sie eine Option zur Anzeige weiterer Dokumente derselben Quelle auswählen.
- Sortieren von Dokumenten nach Relevanz, nach Dokumentdatum oder nach den Werten in einem bestimmten Feld. Beim Sortieren nach Datum oder Feld können Sie angeben, ob die Ergebnisse in aufsteigender oder absteigender Reihenfolge angezeigt werden sollen.
- Beantworten von Korrekturvorschlägen für in der Abfragezeichenfolge festgestellte mögliche Rechtschreibfehler.
- Anzeigen von Informationen zu den Kategorien, zu denen ein Ergebnisdokument gehört (wenn die Objektgruppe Kategorien verwendet) und Durchsuchen nur jener Dokumente, die zu einer bestimmten Kategorie gehören.
- Angeben zusätzlicher Abfragebegriffe für die Suche innerhalb der Suchergebnisse.

## Dokumentabruffunktionen

Mit diesen Funktionen können Sie die folgenden Tasks ausführen:

- Abrufen von Dokumenten durch Anklicken des Dokument-URI und öffnen der Dokumente in einem Web-Browser. Wenn ein Notes-Crawler oder ein Domino Document Manager-Crawler für die Verwendung des DIIOP-Protokolls konfiguriert ist, können Dokumente, die von diesen Crawlern durchsucht werden, von einer Lotus Notes-Clientanzeigeanwendung statt von einem Web-Browser angezeigt werden.

Wenn für einen Crawler die Sicherheit auf Dokumentebene konfiguriert ist, können nur Benutzer mit einer Zugriffsberechtigung für sicheren Inhalt Dokumente abrufen.

- Abrufen von Dokumenten durch Anklicken von Quick Links. Ein Quick Link ordnet URIs Schlüsselwörter zu. Wenn eine Abfrage die angegebenen Schlüsselwörter enthält, werden die zugeordneten URIs (die für diese Schlüsselwörter als relevant vordefiniert wurden) oben in den Suchergebnissen angezeigt.

---

## Suchanwendungsmerkmale

Sie können die Konfigurationsdatei für eine Suchanwendung bearbeiten, um Optionen für Ihre Umgebung anzugeben, das Aussehen der Anwendung zu ändern und die Optionen zu steuern, die dem Benutzer nach dem Start der Suchanwendung zur Verfügung stehen.

Sie können Merkmale auch mit der Anpassungsfunktion für die Suchanwendung bearbeiten. Wenn Sie mit der Anpassungsfunktion eine Auswahl vornehmen, werden die Auswirkungen der Änderung sofort angezeigt. Wenn Sie die gewünschten Einstellungen für die Optionen für das Durchsuchen der Objektgruppen und die Anzeige der Suchergebnisse vorgenommen haben, können Sie die Optionen speichern, um die Konfigurationsdatei für die Suchanwendung zu aktualisieren.

**Wichtig:** Wenn Sie die Suchanwendung als Portlet in WebSphere Portal ausführen, können Sie die Anpassungsfunktion für die Suchanwendung nicht verwenden, um Änderungen an der Suchanwendung im Dialogbetrieb vorzunehmen. Sie müssen die Verwaltungsschnittstelle von WebSphere Portal verwenden, um Merkmale zu bearbeiten und die Portletinstanz zu konfigurieren.

Die Konfigurationsdatei für die Mustersuchanwendung für die Unternehmenssuche ist die Datei `config.properties`. In diesem Abschnitt werden die Merkmale in dieser Datei erläutert und die Standardmerkmale beschrieben. Wenn Sie Konfigurationsdateien für Ihre benutzerdefinierten Suchanwendungen erstellen, weichen die Merkmale in diesen Dateien sowie die für diese Merkmale angegebenen Werte möglicherweise hiervon ab.

## Umgebungsparameter

Sie können Optionen angeben, mit deren Hilfe die Suchanwendung gesteuert wird.

### **applicationName**

Gibt den Namen einer gültigen Suchanwendung an. Der Standardwert ist Default.

Ändern Sie den Standardwert, wenn Sie eine andere Suchanwendung als Standardanwendung verwenden wollen.

**Tipp:** Wenn der Anwendungsname Default ist, können Sie mit der Mustersuchanwendung alle Objektgruppen und externen Quellen mit einer einzigen Abfrage durchsuchen.

**timeout**

Gibt an, wie viele Sekunden auf eine Antwort vom Suchserver gewartet werden soll, bevor eine Suchanforderung das Zeitlimit überschreitet. Sie müssen hier eine ganze Zahl eingeben (z. B. 60, nicht 60,5 oder sechzig). Wenn Sie kein Zeitlimit eingeben, beträgt der Standardwert 30 Sekunden.

**hostname**

Gibt den vollständig qualifizierten Hostnamen des Web-Servers an, der zur Unterstützung Ihrer Instanz von WebSphere Application Server konfiguriert ist. Der Standardwert ist localhost.

Sie können sicherstellen, dass die Suchanwendung ordnungsgemäß funktioniert, indem Sie den Standardwert in den vollständig qualifizierten Hostnamen ändern, für den WebSphere Application Server konfiguriert ist. Wenn z. B. der lokale Computer-Hostname "MeineMaschine" und der Hostname für den Web-Server "www.ibm.com" lautet, geben Sie www.ibm.com an.

**protocol**

Gibt das Protokoll für die Kommunikation mit dem Web-Server an: http oder https. Wenn kein Wert angegeben wird, ist der Standardwert http.

**port**

Gibt die Portnummer des Web-Servers an, der zur Unterstützung Ihrer Instanz von WebSphere Application Server konfiguriert ist. Der Standardwert ist 80. Dies ist typisch für das Protokoll HTTP. Der typische Port für das Protokoll HTTPS ist 443.

**trustStore**

Wenn Sie das Protokoll HTTPS verwenden, müssen Sie den vollständig qualifizierten Pfad für die Schlüsselspeicherdatei (die Datenbankdatei mit den öffentlichen Schlüsseln) angeben. Auch als *Truststore* bezeichnet, ermöglichen diese Informationen die Verwendung des Protokolls SSL (Secure Sockets Layer) für die vertrauenswürdige Kommunikation. Wenn Sie einen Windows-Pfad angeben wollen, müssen Sie dem umgekehrten Schrägstrich einen zweiten umgekehrten Schrägstrich als Escapezeichen hinzufügen. Beispiel: x:\\Application Server\\webserver.key

**trustPassword**

Wenn Sie das Protokoll HTTPS verwenden, müssen Sie ein Kennwort für die angegebene Schlüsselspeicherdatei angeben.

**username**

Die Suchanwendung legt für diesen Wert automatisch den Benutzernamen fest, den der Benutzer bei der Anmeldung an der Suchanwendung angibt. Geben Sie hier nur dann einen Benutzernamen an, wenn Sie das Standardverhalten für die Authentifizierung von Benutzern außer Kraft setzen wollen. Dieses Feld wird nur dann verwendet, wenn Sie die globale Sicherheit in WebSphere Application Server aktiviert haben.

**password**

Die Suchanwendung legt für diesen Wert automatisch das Kennwort fest, das der Benutzer bei der Anmeldung an der Suchanwendung angibt. Geben Sie hier nur dann ein Kennwort an, wenn Sie einen Benutzernamen angegeben haben. Dieses Feld wird nur dann verwendet, wenn Sie die globale Sicherheit in WebSphere Application Server aktiviert haben.

**ssoCookieName**

Gibt den Namen des Cookies an, das die SSO-Tokenzeichenfolge (Single Sign-on) enthält. Der Standardwert ist `LtpaToken`.

**proxyHost**

Gibt den vollständig qualifizierten Hostnamen eines Proxy-Servers an, wenn ein Proxy-Server für den Zugriff auf den Suchserver erforderlich ist.

**proxyPort**

Gibt den Port für den angegebenen Proxy-Server-Host an.

**proxyUser**

Gibt einen Benutzernamen an, der für die Anmeldung am Proxy-Server verwendet werden soll, wenn für den Proxy-Server eine Basisauthentifizierung erforderlich ist.

**proxyPassword**

Gibt das Kennwort für den angegebenen Benutzernamen für den Proxy-Server an.

**filter**

Gibt eine Klasse an, die zum Abrufen von Dokumenten verwendet wird, die in den Suchergebnissen aufgelistet sind. Die Standardklasse ist `com.ibm.es.api.filters.SetDocumentURIFilterFetch`. Ändern Sie diesen Wert nur dann, wenn Sie über eine benutzerdefinierte Klasse verfügen, die Sie stattdessen zum Abrufen von Dokumenten verwenden möchten.

**logging.level**

Gibt die Detailebene für die Protokollierung an:

**OFF** Es werden keine Nachrichten protokolliert.

**SEVERE**

Es werden Nachrichten zu schweren Störungen protokolliert. Dies ist der Standardwert.

**INFO** Es werden Informationsnachrichten protokolliert.

**FINE** Es werden Tracenachrichten mit weniger Details protokolliert. (Diese Option entspricht der Protokollstufe FINE der Java-Klasse `java.util.logging.Level`.)

**ALL** Es werden alle Nachrichten protokolliert.

## Symbole für Quellentypen

Sie können die Bilder anpassen, die den Datenquellentyp darstellen, zu dem ein Ergebnisdokument gehört. Die folgenden Quellentypsymbole, mit denen die bei der Installation von OmniFind Enterprise Edition unterstützten Crawler und externen Quellen angegeben werden, sind in der Datei `config.properties` vordefiniert.

**documentSource.vbr.icon**

Gibt den Pfad und den Namen einer Bilddatei an, die darauf hinweist, dass das Dokument von einem Content Edition-Crawler durchsucht wurde. Das Standardsymbol ist `/images/sourceVBR.gif`.

**documentSource.db2.icon**

Gibt den Pfad und den Namen einer Bilddatei an, die darauf hinweist, dass das Dokument von einem DB2-Crawler durchsucht wurde. Das Standardsymbol ist `/images/sourceDB2.gif`.

**documentSource.cm.icon**

Gibt den Pfad und den Namen einer Bilddatei an, die darauf hinweist, dass das Dokument von einem DB2 Content Manager-Crawler durchsucht wurde. Das Standardsymbol ist /images/sourceCM.gif.

**documentSource.dominodoc.icon**

Gibt den Pfad und den Namen einer Bilddatei an, die darauf hinweist, dass das Dokument von einem Domino Document Manager-Crawler durchsucht wurde. Das Standardsymbol ist /images/sourceDominoDoc.gif.

**documentSource.exchange.icon**

Gibt den Pfad und den Namen einer Bilddatei an, die darauf hinweist, dass das Dokument von einem Exchange Server-Crawler durchsucht wurde. Das Standardsymbol ist /images/sourceExchange.gif.

**documentSource.database.icon**

Gibt den Pfad und den Namen einer Bilddatei an, die darauf hinweist, dass das Dokument von einem JDBC-Datenbankcrawler durchsucht wurde. Das Standardsymbol ist /images/sourceJDBC.gif.

**documentSource.nntp.icon**

Gibt den Pfad und den Namen einer Bilddatei an, die darauf hinweist, dass das Dokument von einem NNTP-Crawler durchsucht wurde. Das Standardsymbol ist /images/sourceNNTP.gif.

**documentSource.notes.icon**

Gibt den Pfad und den Namen einer Bilddatei an, die darauf hinweist, dass das Dokument von einem Notes-Crawler durchsucht wurde. Das Standardsymbol ist /images/sourceNotes.gif.

**documentSource.quickplace.icon**

Gibt den Pfad und den Namen einer Bilddatei an, die darauf hinweist, dass das Dokument von einem QuickPlace-Crawler durchsucht wurde. Das Standardsymbol ist /images/sourceWorkplace.gif.

**documentSource.seedlist.icon**

Gibt den Pfad und den Namen einer Bilddatei an, die darauf hinweist, dass das Dokument von einem Crawler für die Einstiegspunktliste durchsucht wurde. Das Standardsymbol ist /images/sourceSeedlist.gif.

**documentSource.unixfs.icon**

Gibt den Pfad und den Namen einer Bilddatei an, die darauf hinweist, dass das Dokument von einem UNIX-Dateisystem-Crawler durchsucht wurde. Das Standardsymbol ist /images/sourceUnixFS.gif.

**documentSource.web.icon**

Gibt den Pfad und den Namen einer Bilddatei an, die darauf hinweist, dass das Dokument von einem Web-Crawler durchsucht wurde. Das Standardsymbol ist /images/sourceWeb.gif.

**documentSource.wcm.icon**

Gibt den Pfad und den Namen einer Bilddatei an, die darauf hinweist,

dass das Dokument von einem Web Content Management-Crawler durchsucht wurde. Das Standardsymbol ist `/images/sourceWorkplace.gif`.



#### **documentSource.wps.icon**

Gibt den Pfad und den Namen einer Bilddatei an, die darauf hinweist, dass das Dokument von einem WebSphere Portal-Crawler durchsucht wurde. Das Standardsymbol ist `/images/sourceWPS.gif`.



#### **documentSource.winfs.icon**

Gibt den Pfad und den Namen einer Bilddatei an, die darauf hinweist, dass das Dokument von einem Windows-Dateisystem-Crawler durchsucht wurde. Das Standardsymbol ist `/images/sourceWindowsFS.gif`.



#### **documentSource.ldap.icon**

Gibt den Pfad und den Namen einer Bilddatei an, die darauf hinweist, dass das Dokument zu einer externen Quelle gehört, die für einen LDAP-Server erstellt wurde. Das Standardsymbol ist `/images/sourceLDAP.gif`.



#### **documentSource.jdbc.icon**

Gibt den Pfad und den Namen einer Bilddatei an, die darauf hinweist, dass das Dokument zu einer externen Quelle gehört, die für eine JDBC-Datenbanktabelle (Java Database Connectivity) erstellt wurde. Das Standardsymbol ist `/images/sourceJDBC.gif`.

## **Symbole der Clientanzeigefunktion**

Ergebnisdokumente können in einem Web-Browser angezeigt werden. Dokumente, die von Notes-Crawlern oder Domino Document Manager-Crawlern durchsucht wurden, die für die Verwendung des DIIOP-Protokolls konfiguriert sind, können auch von einer Lotus Notes-Clientanzeigeanwendung angezeigt werden.

Wenn Sie ermöglichen wollen, dass Dokumente mit einer Clientanzeigeanwendung angezeigt werden können, müssen Sie sicherstellen, dass das folgende Merkmal auf "true" gesetzt ist:

```
clientViewer.show=true
```

Sie können die Bilder anpassen, die die Clientanzeigeanwendung darstellen. Im folgenden Beispiel weist das Lotus Notes-Symbol darauf hin, dass das Dokument mit der Anzeigeanwendung angezeigt werden kann:

```
client.notes.icon=/images/notes.gif
client.dominodoc.icon=/images/notes.gif
```

In den Suchergebnissen werden das Symbol und der Link für die Clientanzeigeanwendung wie folgt angezeigt:



#### **Clientanzeigefunktion**

## **Dokumentfelder**

Für Datenquellentypen mit Feldern können Sie steuern, welche Felder in den Ergebnisdokumenten angezeigt werden.

**fields.URI prefix=durch\_leerzeichen\_getrennte\_liste\_mit\_feldnamen**

Sie müssen den Doppelpunkt (:) im URI-Präfix codieren, indem Sie einen umgekehrten Schrägstrich (\) als Escapezeichen voranstellen. Wenn Sie

eine Liste mit Feldnamen in einer neuen Zeile fortsetzen wollen, beenden Sie die vorhergehende Zeile mit einem umgekehrten Schrägstrich (\). Beispiel:

```
fields.db2\://=datenbankname tabellenname
fields.domino\://=datenbanktitel dateiname ersteller
fields.dominodoc\://=bibliotheksdatenbanktitel dokumentdatenbanktitel dateiname autor
fields.exchange\://=von ersteller
fields.file\://=verzeichnis dateiname
fields.https\://=dokument-id
fields.http\://=dokument-id
fields.jdbc\://=datenbankname tabellenname
fields.news\://=gruppe von
fields.quickplace\://=bereichstitel raumtitel ersteller
fields.seedlist\://=autor
fields.vbr\://=elementname repositorytyp überarbeitungsbenutzer
fields.wcm\://=autor eigner modifikator
fields.web\://=
fields.wp6\://=
fields.wps\://=
```

## Feldsymbole

Für Datenquellentypen und Dokumente mit Feldern können Sie die Bilder anpassen, die zur Darstellung der Felder verwendet werden. Alle Felder über der Dokumentzusammenfassung enthalten ein Identifizierungsbild. Die folgenden Feldsymbole sind in der Datei `config.properties` vordefiniert.



### **field.icon.databasetitle**

Gibt den Pfad und den Namen einer Bilddatei an, die darauf hinweist, dass das Feld den Dokumenttitel enthält. Das Standardsymbol ist `/images/notesdb.gif`.



### **field.icon.databasename**

Gibt den Pfad und den Namen einer Bilddatei an, die darauf hinweist, dass das Feld den Namen der Datenbank enthält, zu der das Dokument gehört. Das Standardsymbol ist `/images/db2.gif`.



### **field.icon.tablename**

Gibt den Pfad und den Namen einer Bilddatei an, die darauf hinweist, dass das Feld den Namen der Tabelle enthält, zu der das Dokument gehört. Das Standardsymbol ist `/images/table.gif`.



### **field.icon.directory**

Gibt den Pfad und den Namen einer Bilddatei an, die darauf hinweist, dass das Feld den Namen des Verzeichnisses enthält, zu dem das Dokument gehört. Das Standardsymbol ist `/images/closedFolder.gif`.



### **field.icon.filename**

Gibt den Pfad und den Namen einer Bilddatei an, die darauf hinweist, dass das Feld den Dateinamen des Dokuments enthält. Das Standardsymbol ist `/images/document.gif`.

### **field.icon.documentID**

Gibt den Pfad und den Namen einer Bilddatei an, die darauf hinweist, dass das Feld die Dokumentkennung enthält. Dieses leere Bild eignet sich für die Verwendung in Verbindung mit Webdokumenten, beispielsweise, um ein Bild für die URL anzugeben, ohne dass dem Benutzer ein Bild angezeigt wird. Das Standardsymbol ist `/images/dot.gif`.



#### **field.icon.group**

Gibt den Pfad und den Namen einer Bilddatei an, die darauf hinweist, dass das Feld die Dokumentkennung enthält. Dieses leere Bild eignet sich für die Verwendung in Verbindung mit Webdokumenten, beispielsweise, um ein Bild für die URL anzugeben, ohne dass dem Benutzer ein Bild angezeigt wird. Das Standardsymbol ist `/images/document.gif`.



#### **field.icon.from**

Gibt den Pfad und den Namen einer Bilddatei an, die darauf hinweist, dass das Feld jemanden angibt, der das Dokument gesendet hat. Das Standardsymbol ist `/images/author.gif`.



#### **field.icon.creator**

Gibt den Pfad und den Namen einer Bilddatei an, die darauf hinweist, dass das Feld den Ersteller des Dokuments angibt. Das Standardsymbol ist `/images/author.gif`.



#### **field.icon.author**

Gibt den Pfad und den Namen einer Bilddatei an, die darauf hinweist, dass das Feld den Autor des Dokuments angibt. Das Standardsymbol ist `/images/author.gif`.



#### **field.icon.revisionuser**

Gibt den Pfad und den Namen einer Bilddatei an, die darauf hinweist, dass das Feld jemanden angibt, der das Dokument überarbeitet hat. Das Standardsymbol ist `/images/author.gif`.



#### **field.icon.owner**

Gibt den Pfad und den Namen einer Bilddatei an, die darauf hinweist, dass das Feld den Eigner des Dokuments angibt. Das Standardsymbol ist `/images/author.gif`.



#### **field.icon.modifier**

Gibt den Pfad und den Namen einer Bilddatei an, die darauf hinweist, dass das Feld jemanden angibt, der das Dokument modifiziert hat. Das Standardsymbol ist `/images/author.gif`.

## **Standardfeldsymbol**

Sie können ein Bild angeben, das verwendet wird, wenn keine Feldsymbole für Felder konfiguriert sind, die in den Suchergebnissen angezeigt werden. Das folgende Standardfeldsymbol ist in der Datei `config.properties` vordefiniert.



#### **field.defaultIcon**

Gibt den Pfad und den Namen einer Bilddatei an, die das Standardsymbol für Felder in den Suchergebnissen darstellt. Das Standardsymbol ist `/images/database.gif`.

## **Datumsfelder**

Sie können angeben, welche Felder Datumsfelder sind. Die hier angegebenen Feldnamen werden in den Suchergebnissen wie Datumsdaten formatiert. Das Datumsformat stimmt mit den Ländereinstellungen im Web-Browser überein.

**date.fields=durch\_leerzeichen\_getrennte\_liste\_mit\_feldnamen**

Wenn Sie eine Liste mit Feldnamen in einer neuen Zeile fortsetzen wollen, beenden Sie die vorhergehende Zeile mit einem umgekehrten Schrägstrich (\).

Beispiel:

```
date.fields=änderungsdatum erstellungsdatum
```

## Dokumenttitel

Sie können alternative Dokumenttitel angeben, indem Sie den Titeltext durch aussagekräftigere Daten ersetzen (d. h., die Titel können *bereinigt* werden). Wenn z. B. keine Dokumenttitel mit der wenig informativen Bezeichnung "Folie 1" angezeigt werden sollen, können Sie angeben, dass "Folie 1" in den Suchergebnissen unterdrückt wird. Ein aussagekräftigeres Feld, wie z. B. der Dateiname, könnte stattdessen zur Angabe des Ergebnisdokuments verwendet werden.

Ferner haben Sie die Möglichkeit, alternative Titel für Dokumente anzugeben, indem Sie bedeutungslose Wörter aus dem Dokumenttitel entfernen (d. h., die Titel können *abgeschnitten* werden). Wenn z. B. eine Anzahl Ergebnisdokumente mit "Microsoft Word -" beginnt, können Sie die Lesbarkeit der Suchergebnisse verbessern, indem Sie den sich wiederholenden Anfangstext unterdrücken.

**titles.clean=durch\_kommata\_getrennte\_titelliste**

**titles.truncatePrefix=durch\_kommata\_getrennte\_präfixliste**

Die durch Kommata getrennten Listen können Leerzeichen und andere Zeichen mit Ausnahme von Kommata enthalten. Wenn Sie eine Liste in einer neuen Zeile fortsetzen wollen, beenden Sie die vorhergehende Zeile mit einem umgekehrten Schrägstrich (\).

Beispiel:

```
titles.clean=Folie 1, Layout 1, ohne Titel, \
Dokument ohne Titel, PowerPoint-Präsentation, \
(kein Titel für diese Seite)
```

```
titles.truncatePrefix=Microsoft Word -, Microsoft Powerpoint -
```

## Standardwerte für Benutzervorgaben

In der Suchanwendung können Sie Standardwerte für die Vorgabenseite angeben. Wenn ein Benutzer die Vorgaben ändert, stehen die neuen Werte nur für die aktuelle Sitzung des Benutzers zur Verfügung. Die folgenden Vorgaben sind in der Datei `config.properties` vordefiniert.

**preferences.resultsRange=10**

Gibt an, dass auf jeder Seite der Suchergebnisse 10 Ergebnisdokumente aufgelistet werden können.

**preferences.siteCollapsing=Yes**

Gibt an, dass URIs von derselben Quelle in den Suchergebnissen ausgeblendet werden sollen. Das Ausblenden von Sites ist nur dann möglich, wenn die Ergebnisse nach Relevanz sortiert werden. Für Web- und NNTP-Datenquellen werden URIs, die mit dem URI der Root-Site (z. B. `www.ibm.com`) übereinstimmen, automatisch ausgeblendet. Für andere Datenquellentypen und für Websites mit tieferen Pfadebenen (z. B. `www.ibm.com/hr`) müssen Regeln für das Ausblenden von Sites in der Administrationskonsole für die Unternehmenssuche konfiguriert werden.

### **preferences.spellCorrections=Yes**

Gibt an, dass die empfohlenen Korrekturen für Rechtschreibfehler angezeigt werden sollen, wenn ein Benutzer eine Abfrage übergibt, die ein möglicherweise falsch geschriebenes Wort enthält. Beachten Sie, dass Stoppwörter immer entfernt werden, bevor Rechtschreibvorschläge gemacht werden.

### **preferences.extendedHighlighting=No**

Gibt an, dass die Abfragebegriffe nicht zusätzlich zum Dokumentzusammenfassungsfeld in Extrafeldern (wie z. B. dem Dokumenttitel) hervorgehoben werden.

## **Standardobjektgruppen und externe Standardquellen**

Sie können angeben, welche Objektgruppen und externen Quellen auf der Seite **Vorgaben** und der Seite **Erweiterte Suche** vorausgewählt werden. Benutzer können die Standardwerte bearbeiten, damit weniger Objektgruppen und externe Quellen durchsucht werden, als standardmäßig zur Verfügung stehen. Wenn Sie die Anzahl Objektgruppen und externer Quellen hier einschränken, können die Benutzer beliebige für die Suchanwendung zur Verfügung stehende Objektgruppen oder externe Quellen auswählen, wenn sie die Vorgaben oder die Optionen für erweiterte Suche modifizieren.

### **preferences.defaultCollections=\***

**preferences.defaultCollections=durch\_leerzeichen\_getrennte\_liste\_mit\_objektgruppen\_ids** Geben Sie einen Stern (\*) an, damit alle Objektgruppen und externen Quellen durchsucht werden. (Die Objektgruppen und externen Quellen müssen der Suchanwendung in der Administrationskonsole für die Unternehmenssuche zugeordnet werden.) Dies ist die Standardeinstellung in der Datei `config.properties`.

Geben Sie die Objektgruppen-IDs für die Objektgruppen und externen Quellen an, die die Benutzer standardmäßig durchsuchen sollen, um die Suche der Benutzer einzuschränken, wenn diese keine Änderungen an den Vorgaben oder den Optionen für die erweiterte Suche vornehmen.

Beispiel:

```
preferences.defaultCollections=*
preferences.defaultCollections=coll_id1 coll_id2
```

## **Zusätzliche Informationen für die Suchergebnisse**

Sie können festlegen, wie viele Informationen in den Suchergebnissen zurückgegeben werden, und steuern, ob der Benutzer die Suchergebnisse filtern kann. Die folgenden Einstellungen sind die Standardeinstellungen in der Datei `config.properties`.

### **refreshButton.show=false**

Legt fest, ob ein Knopf **Aktualisieren** auf der Basissuchseite angezeigt wird. Der Knopf **Aktualisieren** steht für erweiterte Suchen immer zur Verfügung. Wenn Sie diese Option auf "true" setzen, können die Benutzer die Liste der Objektgruppen und der externen Quellen aktualisieren, die für die Suche zur Verfügung stehen.

Wenn Sie die Anpassungsfunktion für die Suchanwendung verwenden, benötigen Sie keinen Knopf **Aktualisieren**.

Wenn Sie die Anpassungsfunktion für die Suchanwendung nicht verwenden, wollen Sie möglicherweise den Knopf **Aktualisieren** anzeigen, wenn

Sie an der Konfigurationsdatei vorgenommene Änderungen testen. Nach dem Speichern der Änderungen können Sie **Aktualisieren** anklicken, um zu sehen, wie sich die Änderungen auf die Suchanwendung auswirken. Ohne den Knopf **Aktualisieren** müssen Sie die Anwendung "ESSearchServer" in WebSphere Application Server erneut starten, bevor die Änderungen wirksam werden.

Wenn keine Objektgruppen oder externen Quellen für die Suche zur Verfügung stehen (z. B. wenn der falsche Hostname angegeben, die Suchserver nicht gestartet oder die Anwendung "ESSearchServer" in WebSphere Application Server nicht gestartet wurde), wird der Knopf **Aktualisieren** automatisch angezeigt, um Sie bei der Fehlerbehebung zu unterstützen.

**builtQueryString.show=false**

Steuert die Anzeige der vollständig erweiterten Abfragesyntax in einem Bereich, der der Liste mit Ergebnisdokumenten vorangeht. Setzen Sie diese Option auf "true", wenn die tatsächlich verarbeitete Abfrage angezeigt werden soll.

**extraQueryData.show=false**

Steuert die Anzeige zusätzlicher Informationen zu der Abfrage. Setzen Sie diese Option auf "true", wenn Informationen zu Einschränkungen der Zugriffssteuerungsliste, die Namen der durchsuchten Objektgruppen und externen Quellen und die Abfragesprache angezeigt werden sollen.

**refineResults.show=true**

Legt fest, ob der Benutzer die Suchergebnisse durch Angabe zusätzlicher Abfragebegriffe verfeinern kann. Wenn Sie diese Option auf "true" setzen, wird ein Abfragebereich mit der Bezeichnung **In Ergebnissen suchen** unten auf der Seite mit den Suchergebnissen angezeigt.

**sorting.show=true**

Legt fest, ob eine Option für das Sortieren der Suchergebnisse angezeigt wird. Setzen Sie diese Option auf "false", um die Optionen **Sortieren nach** und **Sortierreihenfolge** für das Sortieren von Suchergebnissen zu unterdrücken.

**sourceTypeFilter.show=true**

Legt fest, ob in den Suchergebnissen eine Optionen zum Filtern von Ergebnissen nach Quellentyp angezeigt wird. Setzen Sie diese Option auf "false", wenn Sie den Benutzern das Filtern von Ergebnissen nach Quellentyp nicht ermöglichen wollen.

Wenn Sie verhindern wollen, dass Benutzer Ergebnisse nach Dokumenttyp filtern, müssen Sie einige oder alle Dokumenttypenträge (documentType.bezeichnung=dokumenttypen) in der Konfigurationsdatei löschen.

**filter.showOnTwoLines=true**

Legt fest, ob die Optionen zum Filtern von Ergebnissen nach Quellentyp oder Dateityp in den Suchergebnissen in einer oder in zwei Zeilen angezeigt werden. Beim Anzeigen von Suchergebnissen können Benutzer einen Quellentyp und einen Dateityp auswählen, um nur die Ergebnisdokumente anzuzeigen, die mit den ausgewählten Filtern übereinstimmen.

Setzen Sie dieses Merkmal auf "false", um den für die Anzeige der Suchergebnisse zur Verfügung stehenden Bereich zu maximieren. Zur besseren Lesbarkeit der Filter, insbesondere wenn die verfügbaren Filter über eine Zeile hinaus gehen, können Sie dieses Merkmal auf "true" setzen, damit jeder Filter in einer separaten Zeile angezeigt wird.

**clientViewer.show=true**

Legt fest, ob die Lotus Notes-Clientanwendung für die Anzeige eines Ergebnisdokuments verwendet werden soll. Setzen Sie diese Option auf "false", wenn Sie die Anwendung nicht für die Anzeige von Domino-Dokumenten verwenden wollen.

**showDetails.show=true**

Legt fest, ob die Links **Details anzeigen** und **Details verdecken** in den Suchergebnissen angezeigt werden. Setzen Sie diese Option auf "false", wenn Sie nicht wollen, dass Benutzer zusätzliche Details zu Ergebnisdokumenten anzeigen können.

**showDetailsImage.show=true**

Legt fest, ob Details zu Ergebnisdokumenten in einem Fenster angezeigt werden sollen. Setzen Sie diese Option auf "false", wenn Sie nicht wollen, dass Benutzer zusätzliche Details zu Ergebnisdokumenten anzeigen können, indem sie den Cursor über dem URI eines Dokuments positionieren.

**numberSearchResultsReturned.show=true**

Legt fest, ob die Gesamtzahl der Suchergebnisse angezeigt wird. Setzen Sie diese Option auf "false", wenn Sie nicht wollen, dass Benutzer anzeigen können, wie viele Dokumente in den Suchergebnissen zurückgegeben wurden.

**showMessage.error=true**

Legt fest, ob Fehlermeldungen angezeigt werden. Setzen Sie diese Option auf "false", wenn Sie nicht wollen, dass Fehlermeldungen im oberen Bereich der Suchanwendung angezeigt werden.

**showMessage.warning=true**

Legt fest, ob Warnungen angezeigt werden. Setzen Sie diese Option auf "false", wenn Sie nicht wollen, dass Warnungen im oberen Bereich der Suchanwendung angezeigt werden.

**showMessage.info=true**

Legt fest, ob Informationsnachrichten angezeigt werden. Setzen Sie diese Option auf "false", wenn Sie nicht wollen, dass Informationsnachrichten im oberen Bereich der Suchanwendung angezeigt werden.

**showMessage.success=true**

Legt fest, ob Nachrichten über die erfolgreiche Ausführung angezeigt werden. Setzen Sie diese Option auf "false", wenn Sie nicht wollen, dass Nachrichten über die erfolgreiche Ausführung einer Aktion im oberen Bereich der Suchanwendung angezeigt werden.

## Benutzerdefiniertes Banner und Logo

Sie können die Bilder im Bannerbereich oben in der Suchanwendung anpassen. Sie können z. B. die Standardbilder für OmniFind Enterprise Edition durch Bilder ersetzen, die das Branding Ihres Unternehmens widerspiegeln. Wenn Sie kein Banner anzeigen möchten, setzen Sie mindestens eine der beiden folgenden Zeilen auf Kommentar. Das Merkmal `banner.icon` gibt eine Grafik an, die auf der linken Seite des Bannerbereichs angezeigt wird. Das Merkmal `banner2.icon` gibt eine Grafik an, die auf der rechten Seite des Bannerbereichs angezeigt wird.

```
banner.icon=/images/WS_II_0FEdition.gif
banner2.icon=/images/WS_II_mosaic.gif
```

## Benutzerdefiniertes Hintergrundbild

Sie können die Bilder anpassen, die als Hintergrund der Seiten in der Suchanwendung angezeigt werden. Sie könnten z. B. die Standardbilder für die Unternehmensuche durch Bilder ersetzen, die Marken Ihres Unternehmens widerspiegeln. Wenn kein Hintergrundbild auf einer Seite angezeigt werden soll, setzen Sie mindestens eine der folgenden Zeilen auf Kommentar.

```
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

Die Merkmale im Linkbereich der Datei `config.properties` geben an, dass die Namen der Seiten in der Suchanwendung als Links auf jeder Seite und nicht in der Funktionsleiste oder auf Seiten mit Indexzungen angezeigt werden. Das Anzeigen von Links ist nützlich, wenn Sie die Suchanwendung als Portlet ausführen und den Bereich zum Anzeigen der Suchanwendung auf einer Portalseite minimieren wollen.

Setzen Sie diese Zeilen auf Kommentar, wenn Sie lieber durch Auswahl von Optionen in der Funktionsleiste und auf Seiten mit Indexzungen durch die Suchanwendung navigieren wollen.

## Indexzungen für die Suche

Die Merkmale im Bereich mit den Indexzungen der Datei `config.properties` geben die Namen der JSPs (Java Server Pages) an, die für Seiten mit Indexzungen in der Sicht **Suchen** der Suchanwendung (**Basissuche**, **Erweiterte Suche** und **Kategoriebaum**) verwendet werden. Bearbeiten Sie diese Seiten nur dann, wenn Sie Erfahrung mit Java-Programmierung und JSPs haben.

Beispiele für die Anpassung dieses Bereichs:

- Weiterleiten der Suchanwendung zu benutzerdefinierten JSPs, die ein anderes Aussehen für die Seiten mit Indexzungen zur Verfügung stellen.
- Auf Kommentar setzen der Einträge des Kategoriebaums. Wenn Sie z. B. keine Kategorien für Ihre Objektgruppe konfigurieren, muss die Seite mit dem Kategoriebaum in der Suchanwendung nicht angezeigt werden.
- Kopieren der Einträge für die Seiten mit Indexzungen in den Funktionsleistenbereich der Datei `config.properties` und Setzen dieser Zeilen auf Kommentar. Es wäre z. B. möglich, dass nur die Funktionsleiste und keine Seiten mit Indexzungen angezeigt werden sollen.

## Funktionsleisten

Die Merkmale des Funktionsleistenbereichs der Datei `config.properties` geben die Namen der JSPs an, die für die Funktionsleiste in der Suchanwendung verwendet werden. Bearbeiten Sie diese Seiten nur dann, wenn Sie Erfahrung mit Java-Programmierung und JSPs haben.

Beispiele für die Anpassung dieses Bereichs:

- Weiterleiten der Suchanwendung zu benutzerdefinierten JSPs, die ein anderes Aussehen für die Funktionsleiste zur Verfügung stellen.

- Auf Kommentar setzen der Einträge der Funktionsleiste für Elemente, die nicht angezeigt werden sollen. Es wäre z. B. möglich, dass der Link auf die Seite **Produktinfo** nicht in die Funktionsleiste eingefügt werden soll.
- Versetzen der Funktion zur Anzeige der Seite **Erweiterte Suche** aus dem Bereich der Indexungen in der Datei `config.properties`, damit diese Option nur in der Funktionsleiste zur Verfügung steht.

## Aussagekräftige Bezeichnungen für Dokumenttypen

Sie können die Lesbarkeit des Dokumenttypfilters verbessern, indem Sie den tatsächlichen Namen von Dokumenttypen präzisere und aussagekräftigere Begriffe zuordnen. Die den Suchanwendungen zur Verfügung stehenden Dokumenttypen werden durch die Klasse `AvailableDocumentTypes` der SI-API (Search and Index API) definiert. Die verfügbaren Dokumenttypen sind der Einfachheit halber auch am Ende der Datei `config.properties` aufgelistet.

### **documentType.label=durch\_leerzeichen\_getrennte\_liste\_mit\_dokumenttypen**

Gibt den Namen an, der in der Zeile für den Dokumenttypfilter in den Suchergebnissen angezeigt wird, sowie eine Liste mit den tatsächlichen Dokumenttypen, die bei Auswahl des Filters durch einen Benutzer angezeigt werden sollen.

Sie könnten beispielsweise die Bezeichnung `html` angeben und diesem Namen die Dateierweiterungen und MIME-Typen verschiedener Webdokumente zuordnen. Wenn ein Benutzer zum Filtern der Suchergebnisse **html** anklickt, werden nur Dokumente mit den angegebenen Erweiterungen und MIME-Typen angezeigt.

Die folgenden Dokumenttypzuordnungen sind in der Datei `config.properties` vordefiniert:

```
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
```

Wenn der Wert, den Sie für die Bezeichnung des Dokumenttyps angeben, mit dem Namen eines Merkmals in der Datei `application.properties` übereinstimmt, wird der Wert für das Merkmal aus der Datei `application.properties` angezeigt, und nicht der Wert, den Sie hier angeben. Wenn Sie z. B. `documentType.unixfs` als Bezeichnung für einen Dateitypfilter angeben, wird der Wert für das Merkmal `unixfs` aus der Datei `application.properties` (**UNIX-Dateisystem**) als Name des anklickbaren Dateitypfilters angezeigt.

## Benutzerdefinierte Filter

Sie können benutzerdefinierte Abfragen angeben, um die Anzeige von Ergebnisdokumenten zu filtern.

### **filterCustom.label=abfragebegriffe**

Gibt den Namen an, der in der Zeile für den benutzerdefinierten Filter in den Suchergebnissen angezeigt wird, sowie eine Abfrage, die bei Auswahl des Filters durch einen Benutzer die Suchergebnisse verfeinert. (Beim Anzeigen von Suchergebnissen kann der Benutzer einen benutzer-

definierten Filter auswählen, um nur die Ergebnisdokumente anzuzeigen, die mit der vordefinierten Abfrage übereinstimmen.)

Im folgenden Beispiel werden die Suchergebnisse so gefiltert, dass nur die Dokumente angezeigt werden, die zur Datenbank hr (Human Resources - Personalwesen) gehören.

```
filterCustom.HR_database_only=databasename::hr
```

Wenn ein Benutzer zum Filtern der Suchergebnisse **HR\_database\_only** anklickt, wird die Abfrage `databasename::hr` verarbeitet. Wenn die Suchergebnisse angezeigt werden, werden nur Dokumente aus der Datenbank hr aufgelistet.

Wenn der Wert, den Sie für die Bezeichnung des benutzerdefinierten Filters angeben, mit dem Namen eines Merkmals in der Datei `application.properties` übereinstimmt, wird der Wert für das Merkmal aus der Datei `application.properties` angezeigt, und nicht der Wert, den Sie hier angeben. Wenn Sie z. B. `filterCustom.hostData=` als Bezeichnung für den benutzerdefinierten Filter angeben, wird der Wert des Merkmals `hostData` in der Datei `application.properties` (**Servereinstellungen**) als Name des anklickbaren benutzerdefinierten Filters angezeigt.

Mehrere benutzerdefinierte Filter sind auf Kommentar gesetzt und stehen als Beispiele in der Datei `config.properties` zur Verfügung.

## Erkennen von Kopien

Werden dem Index für die Unternehmenssuche Dokumente hinzugefügt, wird eine Analyse ausgeführt, um Kopien zu entfernen, so dass einem Benutzer in den Suchergebnissen kein Dokument mehrmals angezeigt wird. Sie können die Suchergebnisse weiter filtern, indem Sie eine Option angeben, die beinahe identische Dokumente unterdrückt, sodass diese nicht in den Suchergebnissen angezeigt werden.

### **preferences.nearDuplicateDetection=No**

Gibt an, das Beinaheduplikate eines Dokuments während der Abfrageverarbeitung nicht gefiltert werden.

Wenn Sie **Yes** angeben, werden Dokumente mit ähnlichen Titeln und Zusammenfassungen unterdrückt, wenn ein Benutzer die Suchergebnisse anzeigt. Es wird jedoch eine Nachricht angezeigt, die den Benutzer darüber informiert, dass einige Dokument übergangen wurden, weil sie anderen Dokumenten der Ergebnisliste ähnlich sind. Der Benutzer kann einen Link anklicken, um die Unterdrückung aufzuheben und alle Dokumente in der Ergebnisliste anzuzeigen.

Damit Beinaheduplikate eines Dokuments unterdrückt werden, muss das SI-API-Abfrageobjekt (Search and Index API) Query für die Suchanwendung die Methode `setProperty` angeben, wobei die Zeichenfolge `NearDuplicateDetection` auf **Yes** gesetzt sein muss (Beispiel: `query.setProperty("NearDuplicateDetection", "Yes");`).

## Analyse der häufigsten Ergebnisse (Balkendiagramme für Metadatenfelder)

Sie können Optionen angeben, um Balkendiagramme anzuzeigen, die die Analyse der häufigsten Ergebnisse darstellen. Standardmäßig werden die 500 häufigsten Ergebnisse analysiert. Jedes Diagramm entspricht einem Metadatenfeld, und jeder

Balken eines Diagramms entspricht einem Feldwert. Die Länge des Balkens gibt an, wie oft der Feldwert auftritt. Je länger der Balken ist, desto häufiger ist der Feldwert aufgetreten.

Wenn Sie den Crawlerbereich eines Crawlers konfigurieren, können Sie Optionen für Metadatenfelder angeben. Sie müssen angeben, dass das Feld feldspezifisch durchsucht werden darf und dass es in den Suchergebnissen zurückgegeben werden darf, um ein Diagramm für die besten Ergebnisse für ein Metadatenfeld zu konfigurieren.

Die Merkmale, die Sie für das Balkendiagramm konfigurieren, haben das folgende Format. Dabei ist *zahl* eine Zahl, die das Balkendiagramm in der Suchanwendung eindeutig angibt, *option* ist die Option für das Balkendiagramm, und *wert* ist der Wert der Option:

`topResultsChartszahl.option=wert`

**topResultsChartszahl.titleKey=anwendungsschlüssel**

Gibt einen Titel für das Diagramm an. Dabei ist *zahl* eine Zahl, die das Diagramm in der Suchanwendung eindeutig angibt, und *anwendungsschlüssel* ist eine Bezeichnung für den Diagrammtitel. Diese Bezeichnung kann ein Schlüssel in der Datei `application.properties` für die Suchanwendung oder der Wert sein, den Sie hier angeben. Im folgenden Beispiel wird die Bezeichnung für den Diagrammtitel durch den Wert des Schlüssels `topResults.mostRecentDocuments` in der Datei `application.properties` angegeben:

`topResultsCharts3.titleKey=topResults.mostRecentDocuments.`

In einem anderen Beispiel wird als Diagrammtitel der genaue Wert verwendet, den Sie hier angeben: `topResultsCharts3.titleKey=Organizations`

**topResultsChartszahl.enable=true**

Gibt an, ob das Balkendiagramm angezeigt wird, wenn ein Benutzer die Suchergebnisse anzeigt. Wenn Sie `false` angeben, wird das Diagramm nicht angezeigt.

**topResultsChartszahl.fieldName=feldname**

Gibt den Namen des Metadatenfelds an, dessen Werte für das Diagramm analysiert werden sollen. Beispiel: `databaseTitle`. Sie müssen für jedes Diagramm, das Sie hinzufügen, einen anderen Feldnamen angeben (ein Feld kann nicht in mehreren Diagrammen verwendet werden).

**topResultsChartszahl.maxValues.collapsed=zahl**

Gibt die Anzahl der ausgeblendeten Elemente an, die in diesem Diagramm angezeigt werden sollen. Geben Sie z. B. 5 an, um die fünf Werte anzuzeigen, die in diesem Feld am häufigsten auftreten.

**topResultsChartszahl.maxValues.expanded=zahl**

Gibt die Anzahl der Einträge an, die im Diagramm angezeigt werden sollen, wenn die Diagrammanzeige vollständig erweitert ist. Geben Sie z. B. 10 an, wenn höchstens 10 verschiedene Feldwerte angezeigt werden sollen, wenn das Diagramm erweitert wurde.

**topResultsChartszahl.fieldValueSeparator=zeichen**

Gibt das Zeichen an, das die Werte in dem zu analysierenden Feld voneinander trennt. Wenn ein Feld z. B. mehrere Werte enthält, die durch ein Semikolon voneinander getrennt werden (wie `Sachbearbeiter;Verkäufer;Broker`), können Sie dieses Merkmal verwenden, um das Semikolon (;) als Trennzeichen für die Feldwerte anzugeben, sodass jeder Wert dem Balken-

diagramm als separates Element hinzugefügt wird. Ohne diese Option wird der gesamte Feldwert dem Diagramm als ein Element hinzugefügt.

**topResultsChartszahl.canUserChangeFieldName=true**

Gibt an, ob ein Benutzer während der Anzeige der Suchergebnisse ein anderes Feld auswählen und dessen häufigste Ergebnisse anzeigen kann. Wenn Sie `false` angeben, kann ein Benutzer während der Anzeige der Suchergebnisse kein anderes Feld für die Analyse auswählen.

Wenn Sie `true` angeben, wird eine Liste aller Felder, die in den ersten 500 häufigsten Ergebnissen gefunden wurden, zusammen mit den Ergebnissen des aktuellen Felds angezeigt. Wenn der Benutzer ein Feld aus dieser Liste auswählt, wechselt die Diagrammbezeichnung zu **Diagramm mit dynamischen Feldern**, und die Ergebnisse des Balkendiagramms für das ausgewählte Feld werden solange angezeigt, bis der Benutzer ein anderes Feld auswählt oder den Browser schließt. Wenn der Benutzer die Suchanwendung erneut ausführt, werden wieder die Ergebnisse des ursprünglichen Felds angezeigt.

**topResultsChartszahl.width=zahl**

Gibt die Anzeigebreite des Balkendiagramms in Pixel an. Zum Beispiel 300.

**topResultsChartszahl.barheight=zahl**

Gibt die Höhe der einzelnen Balken des Balkendiagramms in Pixel an. Zum Beispiel 10.

**topResultsChartszahl.color=#farbcode**

Gibt den Hexadezimalcode für die Grundfarbe des Balkens im Balkendiagramm an. Der Standardwert ist Blau (#0309C0).

**topResultsChartszahl.color.gradient=#farbcode**

Gibt die Verlaufsfarbe des Balkens im Balkendiagramm an. Der Standardwert ist Türkis (#00FFFF). Wenn sich die Anzahl Ergebnisse für einen bestimmten Feldwert von selten bis häufig verwendet ändert, verändert sich der Farbton des Balkens von dem in der Farboption `color` angegebenen Wert in den in der Option für die Verlaufsfarbe `color.gradient` angegebenen Wert. Beispiel: 

**topResultsChartszahl.sortKey=frequency**

Gibt an, wie die Elemente im Balkendiagramm sortiert werden sollen. Unterstützte Werte:

**none** Die Elemente des Balkendiagramms sind nicht sortiert.

**label** Die Elemente des Balkendiagramms werden nach dem Feldwert sortiert. Wenn Sie benutzerdefinierte Bezeichnungen mit dem Merkmal `topResultsChartszahl.listOfLabels.prefixKey=feldname` konfigurieren, werden die Elemente im Balkendiagramm auch nach dem Bezeichnungsnamen sortiert.

**frequency**

Die Elemente des Balkendiagramms werden nach der Anzahl der Ergebnisse sortiert, die pro Feldwert zurückgegeben werden.

**topResultsChartszahl.sortOrder=descending**

Gibt die Sortierreihenfolge für die Elemente an, die nach Bezeichnung oder Häufigkeit sortiert werden. Unterstützte Werte:

**ascending**

Elemente, die nach der Bezeichnung sortiert werden, werden in alphabetischer Reihenfolge von A bis Z aufgelistet. Für Elemente,

die nach der Häufigkeit sortiert werden, werden Feldwerte, die häufiger vorkommen, weiter unten in der Liste aufgeführt als Werte, die seltener vorkommen.

### **descending**

Elemente, die nach der Bezeichnung sortiert werden, werden in umgekehrter alphabetischer Reihenfolge von Z bis A aufgelistet. Für Elemente, die nach der Häufigkeit sortiert werden, werden Feldwerte, die häufiger vorkommen, weiter oben in der Liste aufgeführt als Werte, die seltener vorkommen.

### **topResultsChartszahl.listOfLabels.prefixKey=feldname**

Optional. Hier können Sie angeben, ob für bestimmte Feldwerte immer Informationen angezeigt werden sollen, wobei *feldname* das Feld angibt, dessen Wert für das Diagramm analysiert wird. Die Bezeichnungen, die Sie für dieses Merkmal angeben, werden immer angezeigt, auch wenn der Feldwert nicht in den Suchergebnissen vorhanden ist.

Sie können eine beliebige Anzahl Bezeichnungen für ein Feld konfigurieren. Für jede Bezeichnung geben Sie zwei Merkmale an, die das folgende Format aufweisen:

*feldnamezahl.value=wert*

*feldnamezahl.displayValue=anzeigewert*

Dabei gilt Folgendes:

*feldname*

Ist der Name des Felds, für das Sie die Bezeichnungen konfigurieren.

*zahl*

Ist eine Zahl, die die Bezeichnung eindeutig definiert.

*wert*

Ist ein Wert, den Sie in den Analyseergebnissen anzeigen wollen.

*anzeigewert*

Ist die Bezeichnung, die im Balkendiagramm angezeigt wird. Diese Bezeichnung kann ein Schlüssel in der Datei `application.properties` für die Suchanwendung oder der Wert sein, den Sie hier angeben.

Beispiel:

```
topResultsCharts1.listOfLabels.prefixKey=databasetitle
```

```
databasetitle1.value=JK Enterprises Articles & Papers
databasetitle1.displayValue=Articles & papers
```

```
databasetitle2.value=JK Enterprises Blank Forms
databasetitle2.displayValue=Blank forms
```

```
databasetitle3.value=JK Enterprises Bulletins & Guidelines
databasetitle3.displayValue=Bulletins & Guidelines
```

## **Analyse der häufigsten Ergebnisse (benutzerdefinierte HTML)**

Sie können Optionen angeben, um Diagramme der häufigsten Ergebnisse anzuzeigen, indem Sie die Anwendungsprogrammierschnittstelle `com.ibm.es.searchui.charts.servlet.AbstractDynamicChart` mit einer benutzerdefinierten Java-Klasse erweitern. Wenn Sie diese Methode für die Ergebnisanzeige verwenden, können Sie nicht nur die Abfragen, die Metadatenfelder durchsuchen, sondern alle zurückgegebenen Suchergebnisse mit HTML formatieren. Ergebnisse

können für Dokumente in Objektgruppen für die Unternehmenssuche oder aus Suchen in externen Repositorys und auf Websites zurückgegeben werden.

Sie können eine beliebige Anzahl Diagramme für eine Suchanwendung konfigurieren. Verwenden Sie das folgende Merkmal, um jedem Diagramm einen Titel zuzuordnen:

**topResultsChartszahl.titleKey=anwendungsschlüssel**

Gibt einen Titel für das Diagramm an. Dabei ist *zahl* eine Zahl, die das Diagramm in der Suchanwendung eindeutig angibt, und *anwendungsschlüssel* ist eine Bezeichnung für den Diagrammtitel. Diese Bezeichnung kann ein Schlüssel in der Datei `application.properties` für die Suchanwendung oder der Wert sein, den Sie hier angeben. Im folgenden Beispiel wird die Bezeichnung für den Diagrammtitel durch den Wert des Schlüssels `topResults.mostRecentDocuments` in der Datei `application.properties` angegeben:  
`topResultsCharts3.titleKey=topResults.mostRecentDocuments.`

Jedes Element des Diagramms entspricht einem einzelnen Suchergebniswert und umfasst eine Gruppe von Merkmalen, die das folgende Format aufweisen, wobei *zahl* eine Zahl ist, die das Diagramm eindeutig angibt, *option* die Diagrammoption und *wert* der Optionswert ist:

`topResultsChartszahl.option=wert`

**topResultsChartszahl.enable=true**

Gibt an, ob das Diagramm angezeigt werden soll, wenn ein Benutzer Suchergebnisse anzeigt. Wenn Sie `false` angeben, wird das Diagramm nicht angezeigt.

**topResultsChartszahl.maxValues.collapsed=zahl**

Gibt die Anzahl der ausgeblendeten Elemente an, die in diesem Diagramm angezeigt werden sollen. Geben Sie z. B. 5 an, um die fünf am häufigsten auftretenden Ergebnisse anzuzeigen, die die Bedingungen erfüllen.

**topResultsChartszahl.width=zahl**

Gibt die Anzeigebreite des Diagramms in Pixel an. Zum Beispiel 400.

**topResultsChartszahl.dynamicChartClass=benutzerdefinierte\_klasse**

Gibt den Namen Ihrer benutzerdefinierten Java-Klasse an, die die Anwendungsprogrammierschnittstelle `com.ibm.es.searchui.charts.servlet.AbstractDynamicChart` erweitert und definiert, wie die Ausgabe im Diagramm angezeigt werden soll. Beispiel:  
`topResultsCharts.3.dynamicChartClass=com.ibm.es.searchui.charts.servlet.DynamicMostRecentDocuments`

Die folgenden Merkmale, die in der Standarddatei `config.properties` für die Mustersuchanwendung enthalten sind und als Beispiele dienen, werden von der Muster-Java-Klasse `DogearSearchResults` verwendet. In der Dokumentation zur Dogear-Anwendungsprogrammierschnittstelle finden Sie Informationen zu Abfrageparametern, die Sie möglicherweise in Ihre benutzerdefinierte Suchanwendung einschließen möchten.

Sie können diese Merkmale nicht mit der Anpassungsfunktion für die Suchanwendung festlegen:

**topResultsChartszahl.xsl.fileName=style-sheet**

Gibt den Pfad und den Namen eines XSL-Style-Sheets an, das für die Formatierung der Anzeige der häufigsten Ergebnisse im Diagramm verwendet werden soll. Zum Beispiel `/styles/dogear.xsl`.

**topResultsChartszahl.url=url**

Gibt die zu durchsuchende URL an.

**topResultsChartszahl.url.parameters=ps=zahl**

Gibt die Seitengröße an. ps=3 begrenzt die Seitengröße z. B. auf 3.

Hierbei handelt es sich einfach um eine in der Dogear-Anwendungsprogrammierschnittstelle REST definierte Parameterzeichenfolge.

## Analyse der häufigsten Ergebnisse (maximale Ergebnisanzahl)

**topResult.resultSize=zahl**

Der Standard- und Maximalwert für die Anzahl der aus der Analyse der häufigsten Ergebnisse zurückzugebenden Ergebnisse ist 500. Sie können diesen Wert verringern, Sie können ihn jedoch nicht erhöhen. Sie können z. B. eine geringere Anzahl angeben, wenn während der Analyse der häufigsten Ergebnisse Probleme mit der Zeitlimitüberschreitung von Abfragen auftreten. Sie können dieses Merkmal nicht mit der Anpassungsfunktion für die Suchanwendung festlegen.

### Zugehörige Konzepte



Festlegen von Abfragemerkmalen



Java-Klassen zum Anzeigen der häufigsten Ergebnisse

## Bearbeiten der Merkmale der Mustersuchanwendung

Die Mustersuchanwendung für die Unternehmenssuche kann alle aktiven Objektgruppen und externen Quellen in Ihrem System durchsuchen. Sie können eine Konfigurationsdatei bearbeiten, um Optionen für Ihre Web-Server-Umgebung anzugeben, eine andere Suchanwendung als Standardanwendung zu verwenden oder um zu steuern, welche Optionen beim Start der Suchanwendung angezeigt werden.

### Informationen zu dieser Task

Das Installationsprogramm stellt in IBM WebSphere Application Server auf den Suchservern für die Unternehmenssuche eine Mustersuchanwendung für die Unternehmenssuche bereit. Zum Konfigurieren dieser Suchanwendung bearbeiten Sie die mit der Anwendung bereitgestellte Konfigurationsdatei `config.properties`.

Stoppen Sie die Anwendung "ESSearchServer" in WebSphere Application Server, und starten Sie diese erneut, damit die von Ihnen vorgenommenen Änderungen wirksam werden.

### Vorgehensweise

Gehen Sie wie folgt vor, um die Merkmale der Mustersuchanwendung zu bearbeiten:

1. Melden Sie sich am Suchserver als Administrator für die Unternehmenssuche an.
2. Bearbeiten Sie die Datei `config.properties` mit einem Standardtexteditor.

Die Datei `config.properties` wird an der folgenden Speicherposition installiert. Dabei ist `ES_INSTALL_ROOT` das OmniFind Enterprise Edition-Installationsverzeichnis auf dem Suchserver:

```
ES_INSTALL_ROOT/installedApps/ESSearchApplication.ear/
ESSearchApplication.war/WEB-INF/config.properties
```

3. Bearbeiten Sie die Merkmale, und geben Sie Informationen zu Ihrer Web-Server-Umgebung und den Suchvorgaben an. Speichern und schließen Sie dann die Datei. Das Nummernzeichen (#) in der Datei gibt eine Kommentarzeile an.
4. Stoppen Sie die Anwendung "ESSearchServer", und starten Sie sie erneut.

#### **AIX, Linux oder Solaris**

```
./stopServer.sh ESSearchServer
./startServer.sh ESSearchServer
```

#### **Windows**

```
stopServer ESSearchServer
startServer ESSearchServer
```

Diese Scripts befinden sich im Verzeichnis was-installationsverzeichnis/AppServer/bin.

- Bei WebSphere Application Server Version 5 ist der Standardinstallationspfad /usr/WebSphere auf AIX-Systemen, /opt/WebSphere auf Linux- oder Solaris-Systemen oder C:\Programme\WebSphere auf Windows-Systemen.
- Bei WebSphere Application Server Version 6 ist der Standardinstallationspfad /usr/IBM/WebSphere auf AIX-Systemen, /opt/IBM/WebSphere auf Linux- oder Solaris-Systemen oder C:\Programme\IBM\WebSphere auf Windows-Systemen.

#### **Zugehörige Konzepte**

 Festlegen von Abfragemerkmalen

 Java-Klassen zum Anzeigen der häufigsten Ergebnisse

#### **Zugehörige Tasks**

„Konfigurieren der Suchserver für die ausschließliche Annahme von sicheren Suchanforderungen (SSL)“ auf Seite 269

## **Anpassen von Suchanwendungen**

Die Anpassungsfunktion für die Suchanwendung ist eine grafische Oberfläche für die Anpassung von Suchanwendungen für die Unternehmenssuche oder Ihrer benutzerdefinierten Suchanwendungen.

#### **Einschränkungen**

Die Anpassungsfunktion für die Suchanwendung ist als eine eigenständige Anwendung verfügbar. Sie können die Anpassungsfunktion für die Suchanwendung nicht innerhalb von WebSphere Portal starten, um Suchanwendungen anzupassen, die als Portlets ausgeführt werden. Sie müssen die Verwaltungsschnittstelle von WebSphere Portal verwenden, um Merkmale zu bearbeiten und die Portletinstanz zu konfigurieren, um Suchanwendungen anzupassen, die als Portlets ausgeführt werden.

#### **Informationen zu dieser Task**

Die Anpassungsfunktion für die Suchanwendung ermöglicht Ihnen, von Ihnen gewünschte Änderungen darzustellen und eine Suchanwendung zu modifizieren, ohne die Konfigurationsdatei zu bearbeiten. Beispielsweise können Sie die Banner- und Hintergrundbilder ändern, das Layout der Suchschnittstelle anpassen und Optionen für die Arbeit mit Suchergebnissen angeben.

Wenn Sie in der Anpassungsfunktion für die Suchanwendung eine Auswahl vornehmen, werden die Auswirkungen Ihrer Auswahl angezeigt. Wenn Sie die Änderungen speichern, aktualisieren Sie die Konfigurationsdatei für die Suchanwendung.

Stoppen Sie die Anwendung "ESSearchServer" in WebSphere Application Server, und starten Sie diese erneut, damit die von Ihnen vorgenommenen Änderungen wirksam werden.

### **Tipp:**

Detaillierte Beispiele für die Verwendung der Anpassungsfunktion für die Suchanwendung finden Sie im IBM Redbook *IBM OmniFind Enterprise Edition Configuration and Implementation Scenarios*.

### **Vorgehensweise**

Gehen Sie wie folgt vor, um eine Suchanwendung anzupassen:

1. Wenn Sie die Mustersuchanwendung für die Unternehmenssuche anpassen wollen, geben Sie die URL-Adresse für die Anpassungsfunktion für die Suchanwendung in einem Web-Browser ein. Beispiel:

`http://suchserver.com/ESSearchApplication/palette.do`

`suchserver.com` ist der Hostname des Suchservers.

Wenn Ihr Web-Server nicht für Port 80 konfiguriert ist, müssen Sie auch die richtige Portnummer angeben. Beispiel:

`http://suchserver.com:9080/ESSearchApplication/palette.do`

**Tipp:** Wenn Sie Administrator für die Unternehmenssuche sind, können Sie die Anpassungsfunktion für die Suchanwendung auch öffnen, indem Sie die Option **Anpassungsfunktion für die Suche** in der Administrationskonsole für die Unternehmenssuche auswählen.

2. Wenn Sie eine benutzerdefinierte Suchanwendung anpassen wollen, geben Sie die URL-Adresse für die Anpassungsfunktion für die Suchanwendung ein, und hängen Sie den Namen der Konfigurationsdatei für Ihre Suchanwendung an. Beispiel:

`http://suchserver.com/ESSearchApplication/palette.do?configFile=/WEB-INF/meine_konfiguration.properties`

Wenn die von Ihnen angegebene Datei nicht vorhanden ist, werden die Werte in der Datei `config.properties` für die Mustersuchanwendung angezeigt.

**Tipp:** Sie können die Konfigurationsdatei, die mit einer Suchanwendung verwendet werden soll, auch angeben, indem Sie nach dem Start der Anpassungsfunktion für die Suchanwendung **Laden** anklicken und den Namen der Datei angeben.

3. Wenn die globale Sicherheit in WebSphere Application Server aktiviert ist, melden Sie sich mit einer gültigen Benutzer-ID und einem gültigen Kennwort an.
4. Wählen Sie die Optionen aus, die Sie anpassen wollen. Dies können beispielsweise Informationen zum Suchserver, die anzuzeigenden Nachrichtentypen, Abfrage- und Suchergebnisoptionen und die Namen von Bildern sein, mit denen unterschiedliche Dokumenttypen in den Suchergebnissen angegeben werden.

Wenn Sie Hilfe beim Angeben der Optionen benötigen, klicken Sie **Hilfe für Anpassungsfunktion** an. Wenn Sie die Auswirkungen einiger Änderungen

anzeigen wollen, wie z. B. die Darstellung von Suchergebnissen, geben Sie eine Abfrage ein, und klicken Sie **Suchen** an.

5. Wenn Sie alle gewünschten Änderungen vorgenommen haben, klicken Sie **Speichern** an, um die Konfigurationsdatei zu aktualisieren. Wenn Sie **Zurücksetzen** anklicken, werden die in der Anpassungsfunktion für die Suchanwendung angezeigten Optionen auf die in der zuletzt gespeicherten Version der Konfigurationsdatei enthaltenen Werte zurückgesetzt.
6. Melden Sie sich am Suchserver als Administrator für die Unternehmenssuche an, stoppen Sie die Anwendung "ESSearchServer", und starten Sie sie erneut.

#### **AIX, Linux oder Solaris**

```
./stopServer.sh ESSearchServer
./startServer.sh ESSearchServer
```

#### **Windows**

```
stopServer ESSearchServer
startServer ESSearchServer
```

Diese Scripts befinden sich im Verzeichnis was-installationsverzeichnis/AppServer/bin.

- Bei WebSphere Application Server Version 5 ist der Standardinstallationspfad /usr/WebSphere auf AIX-Systemen, /opt/WebSphere auf Linux- oder Solaris-Systemen oder C:\Programme\WebSphere auf Windows-Systemen.
- Bei WebSphere Application Server Version 6 ist der Standardinstallationspfad /usr/IBM/WebSphere auf AIX-Systemen, /opt/IBM/WebSphere auf Linux- oder Solaris-Systemen oder C:\Programme\IBM\WebSphere auf Windows-Systemen.

#### **Zugehörige Tasks**

„Konfigurieren der Suchserver für die ausschließliche Annahme von sicheren Suchanforderungen (SSL)“ auf Seite 269

## **Klonen der Mustersuchanwendung**

Wenn Sie die Mustersuchanwendung als Modell für die Erstellung eigener Suchanwendungen verwenden wollen, können Sie die Datei config.properties kopieren oder die Anpassungsfunktion für die Suchanwendung verwenden.

#### **Informationen zu dieser Task**

Zur Vereinfachung der Erstellung benutzerdefinierter Suchanwendungen können Sie Konfigurationsoptionen klonen, die Sie für die Mustersuchanwendung angeben, und dann die Optionen anpassen, die Sie ändern wollen.

Nachdem Sie die Mustersuchanwendung geklont haben, geben Sie den Namen Ihrer Konfigurationsdatei an, um die neue Suchanwendung zu starten. Außerdem geben Sie den Namen Ihrer Konfigurationsdatei an, um die neue Suchanwendung mit der Anpassungsfunktion für die Suchanwendung anzupassen.

Durch das Klonen der Mustersuchanwendung können Sie schnell Suchanwendungen für bestimmte Zwecke oder Zielgruppen erstellen. Beispielsweise können Sie eine Suchanwendung für Mitarbeiter der Personalabteilung und eine Suchanwendung für Vertriebsmitarbeiter erstellen.

Stoppen Sie die Anwendung "ESSearchServer" in WebSphere Application Server, und starten Sie diese erneut, damit die von Ihnen vorgenommenen Änderungen wirksam werden.

## Vorgehensweise

Gehen Sie wie folgt vor, um die Mustersuchanwendung zu klonen:

1. Gehen Sie wie folgt vor, um eine Konfigurationsdatei zu bearbeiten:

- a. Kopieren Sie die Datei `config.properties` für die Mustersuchanwendung, und benennen Sie sie um.

Die Datei `config.properties` wird an der folgenden Speicherposition installiert. Dabei ist `ES_INSTALL_ROOT` das OmniFind Enterprise Edition-Installationsverzeichnis auf dem Suchserver:

```
ES_INSTALL_ROOT/installedApps/ESSearchApplication.ear/
 ESSearchApplication.war/WEB-INF/config.properties
```

Die von Ihnen erstellte Datei muss sich im Unterverzeichnis `WEB-INF` befinden.

- b. Bearbeiten Sie die Merkmale, die Sie in Ihrer benutzerdefinierten Suchanwendung verwenden wollen, und speichern Sie die Datei. Sie müssen zumindest das Merkmal `applicationName` ändern, um den Namen Ihrer Suchanwendung anzugeben.

2. Gehen Sie wie folgt vor, um die Mustersuchanwendung mit der Anpassungsfunktion für die Suchanwendung zu klonen:

- a. Starten Sie die Anpassungsfunktion für die Suchanwendung, indem Sie den Namen der Konfigurationsdatei anhängen, die Sie erstellen wollen. Im folgenden Beispiel wird eine Datei mit dem Namen `meine_neue_datei.properties` erstellt:

```
http://ESServer.com/ESSearchApplication/palette.do?configFile=/WEB-INF/meine_neue_datei.properties
```

Da die Datei noch nicht vorhanden ist, werden die Werte verwendet, die in der Konfigurationsdatei der Mustersuchanwendung `config.properties` angegeben sind.

**Tipp:** Sie können eine Konfigurationsdatei für eine Suchanwendung auch erstellen, indem Sie nach dem Start der Anpassungsfunktion für die Suchanwendung **Laden** anklicken und den Namen einer Datei angeben. Die Datei wird erstellt, wenn Sie **Speichern** anklicken, um Ihre benutzerdefinierten Optionen zu speichern.

- b. Wenn die globale Sicherheit in WebSphere Application Server aktiviert ist, melden Sie sich mit einer gültigen Benutzer-ID und einem gültigen Kennwort an.
- c. Geben Sie einen Namen für Ihre Suchanwendung an, geben Sie die Optionen an, die Sie für das Durchsuchen von Objektgruppen verwenden wollen, und klicken Sie auf **Speichern**. Die von Ihnen angegebenen Änderungen werden in Ihrer neuen Konfigurationsdatei im Unterverzeichnis `WEB-INF` gespeichert.

Wenn Sie Hilfe beim Angeben der Optionen benötigen, klicken Sie **Hilfe für Anpassungsfunktion** an. Wenn Sie die Auswirkungen einiger Änderungen anzeigen wollen, wie z. B. die Darstellung von Suchergebnissen, geben Sie eine Abfrage ein, und klicken Sie **Suchen** an.

3. Melden Sie sich am Suchserver als Administrator für die Unternehmenssuche an, stoppen Sie die Anwendung "ESSearchServer", und starten Sie sie erneut.

### AIX, Linux oder Solaris

```
./stopServer.sh ESSearchServer
./startServer.sh ESSearchServer
```

## Windows

```
stopServer ESSearchServer
startServer ESSearchServer
```

Diese Scripts befinden sich im Verzeichnis was-installationsverzeichnis/  
AppServer/bin.

- Bei WebSphere Application Server Version 5 ist der Standardinstallationspfad /usr/WebSphere auf AIX-Systemen, /opt/WebSphere auf Linux- oder Solaris-Systemen oder C:\Programme\WebSphere auf Windows-Systemen.
- Bei WebSphere Application Server Version 6 ist der Standardinstallationspfad /usr/IBM/WebSphere auf AIX-Systemen, /opt/IBM/WebSphere auf Linux- oder Solaris-Systemen oder C:\Programme\IBM\WebSphere auf Windows-Systemen.

## Analysieren der häufigsten Ergebnisse

Sie können den Benutzern helfen, eine Suchergebnismenge einzugrenzen, indem Sie Optionen für die Analyse der häufigsten Ergebnisse konfigurieren.

Die Analyse der häufigsten Ergebnisse klassifiziert im Wesentlichen die aktuelle Suchergebnismenge anhand der Häufigkeit, in der die analysierten Werte auftreten. Benutzer können die Suchergebnisse filtern, indem sie einen der analysierten Werte auswählen, wodurch der ausgewählte Wert automatisch den Suchkriterien als neues Schlüsselwort hinzugefügt wird. Die Analyse der häufigsten Ergebnisse hat den Nutzen, dass sie es Benutzern ermöglicht, die Suchergebnisse zu optimieren, ohne eine komplexe oder erweiterte Suchsyntax verwenden zu müssen.

In einer Anwendung für die Unternehmenssuche können Sie Balkendiagramme verwenden, um grafisch darzustellen, welche Metadatenfeldwerte am häufigsten in den Suchergebnissen auftreten. Sie können auch eine benutzerdefinierte Java-Klasse erstellen, um die häufigsten Ergebnisse, einschließlich der Ergebnisse aus unternehmensfremden Suchquellen, in einem beliebigen HTML-Format anzuzeigen.

### Einschränkungen

Stellen Sie sicher, dass die entsprechenden Schriftarten für Ihre Sprache auf dem Computer installiert sind, auf dem Sie WebSphere Application Server und die Suchanwendung ausführen. Dieser Schritt ist erforderlich, um sicherzustellen, dass die Schriftart beim Generieren des Balkendiagramms auf eine Schriftart gesetzt wird, die die Zeichen in der Diagrammbezeichnung unterstützt. Dies ist vor allem bei asiatischen Sprachen ein kritischer Punkt. Wenn Sie die Schriftarten nach der Installation von WebSphere Application Server installieren, müssen Sie WebSphere Application Server erneut starten, damit die Änderungen wirksam werden.

Wenn Sie die Suchanwendung als eigenständige Anwendung ausführen, können Sie die Diagramme für die häufigsten Ergebnisse mit der Anpassungsfunktion für die Suchanwendung konfigurieren, oder Sie bearbeiten die Konfigurationsdatei für die Suchanwendung. Wenn Sie die Suchanwendung als Portlet in WebSphere Portal ausführen, müssen Sie die Verwaltungsschnittstelle von WebSphere Portal verwenden, um die Merkmale zu bearbeiten und die Portletinstanz zu konfigurieren. Sie können nicht die Anpassungsfunktion für die Suchanwendung verwenden, um Optionen für die Analyse der häufigsten Ergebnisse zu konfigurieren.

### Informationen zu dieser Task

Sie können die häufigsten Ergebnisse grafisch darstellen, indem Sie die Metadatenfeldwerte, die am häufigsten vorkommen, in einem Balkendiagramm anzeigen. Sie

können auch eine Java-Klasse verwenden, um die Suchanwendung zu erweitern und die häufigsten Ergebnisse in einem anderen Format anzuzeigen. Sie können beispielsweise HTML verwenden, um die häufigsten Ergebnisse in einer ungeordneten Liste darzustellen.

### **Balkendiagramme für Metadatenfelder**

Sie können Optionen angeben, um Metadatenfelder zu analysieren und die Ergebnisse dieser Analyse in Balkendiagrammen anzuzeigen. Die Diagramme werden zusätzlich zu den Ergebnissen der Benutzerabfrage angezeigt. Jedes Diagramm entspricht einem einzelnen Metadatenfeld (wie z. B. Dokumentgröße, Autor, Datum und so weiter), und jeder Balken eines Diagramms entspricht einem bestimmten Feldwert.

Die Länge des Balkens gibt die Anzahl Dokumente an, die einen bestimmten Feldwert enthalten, im Vergleich zu anderen Dokumenten, die in diesem Feld andere Werte enthalten. Je länger der Balken ist, desto häufiger ist der Feldwert aufgetreten. Benutzer können die Ergebnisse optimieren, indem Sie einen Balken des Balkendiagramms anklicken. Der Feldwert, der durch den ausgewählten Balken dargestellt wird, wird als weiterer Abfragebegriff hinzugefügt, und die neuen Suchergebnisse werden durch die zusätzlichen Suchkriterien eingegrenzt.

Wenn Sie den Crawlerbereich eines Crawlers konfigurieren, können Sie eine Option auswählen, um Suchoptionen für Metadatenfelder anzugeben. Sie können beispielsweise angeben, ob ein Metadatenfeld als freier Text durchsucht wird, ob es nach dem Feldnamen durchsucht wird, ob es in den Suchergebnissen angezeigt wird, ob es als parametrische Daten durchsucht wird und so weiter. Damit Diagramme für Metadatenfelder in den Suchergebnissen angezeigt werden, müssen Sie Metadatenfeldoptionen für den Crawler konfigurieren. Sie müssen mindestens angeben, dass das Feld feldspezifisch durchsucht werden darf und dass es in den Suchergebnissen angezeigt werden darf.

Wenn Sie beim Konfigurieren der Optionen für ein Metadatenfeld die Option **Vollständige Übereinstimmung** auswählen, wird eine Abfrage auf vollständige Übereinstimmung ausgeführt, wenn ein Benutzer einen Balken des Balkendiagramms auswählt. Eine Suche auf vollständige Übereinstimmung gibt an, dass nur Ergebnisse zurückgegeben werden sollen, wenn der Abfragebegriff mit dem gesamten Feldwert übereinstimmt. Wenn das Feld weniger Inhalt oder zusätzlichen Inhalt aufweist, gilt dies nicht als Übereinstimmung.

Wenn Sie beim Konfigurieren der Optionen für ein Metadatenfeld die Option **Vollständige Übereinstimmung** nicht auswählen, wird eine feldspezifische Abfrage ausgeführt, wenn der Benutzer einen Balken des Balkendiagramms auswählt. In diesem Fall muss der zusätzliche Abfragebegriff in dem Feld vorhanden sein, es ist jedoch keine Übereinstimmung mit dem gesamten Feldwert erforderlich.

### **Mit HTML formatierte Ergebnisse**

Sie können eine benutzerdefinierte Java-Klasse verwenden, um die Anwendungsprogrammierschnittstelle `com.ibm.es.searchui.charts.servlet.AbstractDynamicChart` zu erweitern. Wenn Sie diese Methode zum Anzeigen der Ergebnisse verwenden, können Sie die Anzeige beliebiger Suchergebnisse mit HTML formatieren, statt nur Abfragen, die Metadatenfelder durchsuchen. Ergebnisse können für Dokumente in Objektgruppen für die Unternehmenssuche oder aus Suchen in externen Repositories und auf Websites zurückgegeben werden.

Jedes Diagramm entspricht einer einzelnen Java-Klasse, und jede Klasse gibt an, wie die häufigsten Ergebnisse im Diagramm dargestellt werden sollen. Beispiel: Die mit dem Mustercode für die Unternehmenssuche bereitgestellte Musterklasse `com.ibm.es.searchui.charts.servlet.DynamicMostRecentDocuments` stellt die häufigsten Ergebnisse als ungeordnete Liste dar. Die Dokumente sind nach dem Datum sortiert, und es wird nur der Dokumenttitel und das Datum angezeigt.

Eine andere mit der Unternehmenssuche bereitgestellte Musterklasse, `com.ibm.es.searchui.charts.servlet.DogearSearchResults`, zeigt, wie Sie Benutzern eine Liste mit zur Benutzerabfrage gehörenden Lesezeichen von Lotus Connections Dogear bereitstellen können.

Wenn Benutzer ein Suchergebnis in Ihrer benutzerspezifisch formatierten Ausgabe anklicken, wird das Dokument in einem neuen Browserfenster angezeigt.

### Vorgehensweise

Anhand der folgenden Prozedur sehen Sie, wie Sie die Anpassungsfunktion für die Suchanwendung verwenden können, um anzugeben, dass Metadatenfelder analysiert werden sollen. Die am häufigsten auftretenden Metadatenfeldwerte werden grafisch in einem Balkendiagramm dargestellt.

1. Öffnen Sie die Anpassungsfunktion für die Suchanwendung. Wenn sie nicht bereits angezeigt wird, laden Sie die Konfigurationsdatei für die Suchanwendung, die Sie anpassen wollen, und klicken Sie **Anwenden** an.
2. Blättern Sie abwärts zu **Diagramme mit den besten Ergebnissen**, und klicken Sie **Diagramm hinzufügen** an.
3. Wenn die Liste mit den Diagrammoptionen angezeigt wird, lassen Sie das Markierungsfeld **Angepasstes Diagramm** leer (wählen Sie diese Option nur aus, wenn Sie für die Analyse der häufigsten Ergebnisse eine angepasste Java-Klasse erstellt haben), und wählen Sie das Markierungsfeld **Diagramm aktivieren** aus, um sicherzustellen, dass das Balkendiagramm angezeigt wird, wenn ein Benutzer die Suchergebnisse anzeigt.
4. Geben Sie in das Feld **Diagrammtitel** eine beschreibende Bezeichnung für das Diagramm ein. Der Wert, den Sie hier angeben, ersetzt den Platzhaltertext **Neue Diagrammzeile**.
5. Entscheiden Sie, ob Sie das Markierungsfeld **Dynamische Feldauswahl aktivieren** auswählen möchten. Wenn Sie diese Option aktivieren, kann ein Benutzer während der Anzeige der Suchergebnisse andere Felder auswählen und die Analyse der häufigsten Ergebnisse für die betreffenden Felder anzeigen.

Sie können diese Option verwenden, um es Benutzern zu ermöglichen, eine Suchergebnismenge zu optimieren. Ein Benutzer durchsucht möglicherweise eine Datenbank nach Informationen zu weiblichen Mitarbeitern. Nachdem er eine Abfrage eingegeben hat, die ein Feld durchsucht, das das Geschlecht der Mitarbeiter angibt (wie z. B. `sex:F`), werden die Werte für die häufigsten Ergebnisse als Balken im Balkendiagramm dargestellt. Als nächstes wählt der Benutzer `job` aus der Liste der Felder aus, die in der ursprünglichen Ergebnismenge vorhanden sind. Die Balkendiagramme zeigen nun die Ergebnisse für die häufigsten Jobs an, die von Frauen ausgeübt werden. Schließlich wählt der Benutzer `designer` aus der Liste der verfügbaren Felder aus. Nun werden in den Ergebnissen Informationen zu weiblichen Mitarbeitern bereitgestellt, die als Designer arbeiten.

6. Geben Sie in das Feld **Name des Metadatenfelds** den Namen des Metadatenfelds ein, dessen Werte für das Diagramm analysiert werden sollen.
7. Wenn ein Feld mehrere Werte enthält, die durch ein Trennzeichen voneinander getrennt sind, geben Sie dieses Zeichen im Feld **Feldwerttrennzeichen** an. Wenn ein Feld zwei Werte enthält, wie z. B. den Vornamen und den Nachnamen eines Kunden, die durch ein Semikolon voneinander getrennt sind, können Sie das Semikolon hier angeben, damit jeder Wert dem Balkendiagramm einzeln hinzugefügt wird. Wenn Sie das Trennzeichen nicht angeben, wird der gesamte Feldwert als ein einzelnes Element analysiert.
8. Geben Sie Optionen zum Anzeigen des Balkendiagramms an, wie z. B. die Größe und die Farbe der Balken und wie viele Balken angezeigt werden, wenn das Diagramm ausgeblendet oder eingeblendet wird.
9. Geben Sie Optionen zum Sortieren der Analyseergebnisse an. Sie können beispielsweise nach Feldwerten sortieren oder danach, wie häufig die Werte auftreten.
10. Wenn Sie sicherstellen wollen, dass für bestimmte Feldwerte immer Informationen angezeigt werden, selbst wenn der betreffende Feldwert nicht in den Suchergebnissen vorkommt, klicken Sie **Zeile hinzufügen** im Bereich **Angepasste Bezeichnungen** an. Geben Sie den Feldwert an, der im Balkendiagramm angezeigt werden soll, selbst wenn der betreffende Wert nicht in den Suchergebnissen zurückgegeben wird, und geben Sie eine Bezeichnung für diesen Balken im Balkendiagramm an.  
Wenn Sie z. B. immer sehen wollen, ob ein Konkurrent wie z. B. JK Enterprises in den Suchergebnissen vorkommt, selbst wenn er aufgrund mangelnder Häufigkeit nicht als eines der häufigsten Ergebnisse angezeigt würde, geben Sie JK Enterprises als zu analysierenden Quellenwert an, und geben Sie danach eine aussagekräftige Bezeichnung für diesen Balken im Balkendiagramm an.
11. Klicken Sie **Anwenden** neben dem Diagrammtitel an, damit die von Ihnen angegebenen Optionen auf das Diagramm angewendet werden.
12. Werden bereits Suchergebnisse im Bereich der Suchanwendung angezeigt, wird die Analyse der häufigsten Ergebnisse auf die aktuelle Suche angewendet und Ihr neues Diagramm wird angezeigt. Andernfalls müssen Sie eine Abfrage eingeben, um die Anzeige des Balkendiagramms zu testen.
13. Wenn Sie mit dem Diagramm zufrieden sind, klicken Sie **Speichern** an, um die Konfigurationsdatei für die Suchanwendung zu aktualisieren.

#### Zugehörige Konzepte

 Festlegen von Abfragemerkmalen

 Java-Klassen zum Anzeigen der häufigsten Ergebnisse

---

## Zugreifen auf Suchanwendungen

Sie greifen auf eine Suchanwendung zu, indem Sie eine URL-Adresse in einem Web-Browser angeben.

### Vorbereitung

Sie müssen die Suchanwendung für Ihre Web-Server-Umgebung konfigurieren.

### Informationen zu dieser Task

Die Mustersuchanwendung ist auf den Suchservern für die Unternehmenssuche installiert. Sie können diese Anwendung wie bereitgestellt zum Testen von Objektgruppen und externen Quellen verwenden, bevor Sie diese Benutzern zur Verfügung stellen. Sie können die Anwendung auch als Modell zum Erstellen Ihrer eigenen Suchanwendungen verwenden.

### Vorgehensweise

Gehen Sie wie folgt vor, um eine Suchanwendung zu starten:

1. Geben Sie die URL-Adresse für die Suchanwendung in einem Web-Browser ein. Beispiel:

```
http://suchserver.com/ESearchApplication/
```

*suchserver.com* ist der Hostname des Suchservers.

Wenn Ihr Web-Server nicht für Port 80 konfiguriert ist, müssen Sie auch die richtige Portnummer angeben. Beispiel:

```
http://suchserver.com:9080/ESearchApplication/
```

2. Wenn Sie eine benutzerdefinierte Suchanwendung starten wollen, geben Sie die URL-Adresse für die Mustersuchanwendung ein, und hängen Sie den Namen der Konfigurationsdatei für Ihre Suchanwendung an. Beispiel:

```
http://suchserver.com/ESearchApplication/search.do?configFile=/WEB-INF/meine_konfiguration.properties
```

Wenn die von Ihnen angegebene Datei nicht vorhanden ist, wird die Mustersuchanwendung für die Unternehmenssuche angezeigt.

3. Wenn die globale Sicherheit in WebSphere Application Server aktiviert ist, melden Sie sich bei der Anwendung mit einer gültigen Benutzer-ID und einem gültigen Kennwort an.

Wenn für eine der für die Suchanwendung zur Verfügung stehenden Objektgruppen die Sicherheit aktiviert ist und wenn die sicheren Objektgruppen Crawler enthalten, die zur Prüfung der Berechtigungsnachweise von Benutzern während der Abfrageverarbeitung konfiguriert sind, können Sie ein Benutzerprofil konfigurieren. Geben Sie auf der Seite **Mein Profil** Berechtigungsnachweise für den Zugriff auf die sicheren Domänen an. Sie können diese Domänen anschließend durchsuchen, ohne sich bei ihnen anzumelden.

Wenn der Crawler SSO-Sicherheit (Single Sign-on) unterstützt, können Sie gesicherte Domänen durchsuchen, ohne ein Benutzerprofil zu erstellen.

4. Übergeben Sie auf der Seite **Suche** eine Abfrage. Alle Objektgruppen und externen Quellen, die auf der Seite **Vorgaben** für die Suche ausgewählt wurden, werden durchsucht.

---

## Konfigurieren der Suchserver für die ausschließliche Annahme von sicheren Suchanforderungen (SSL)

Sie können die HTTP-Schnittstelle (Hypertext Transfer Protocol) auf den Suchservern inaktivieren und die Server so konfigurieren, dass sie Suchanforderungen nur über SSL (Secure Sockets Layer) und die sichere HTTPS-Schnittstelle (Hypertext Transfer Protocol Secure) akzeptieren.

### Informationen zu dieser Task

Wenn Sie die Suchserver so konfigurieren wollen, dass sie bei der Verarbeitung von Suchanforderungen ausschließlich das SSL-Protokoll verwenden, müssen Sie die HTTP-Schnittstelle inaktivieren. Außerdem müssen Sie sicherstellen, dass auf

beiden Suchservern und auf allen Client-Computern, wie z. B. dem WebSphere Portal-Server, auf dem das Suchportlet für die Unternehmenssuche installiert ist, dieselbe Schlüsselspeicherdatei gespeichert ist. Die Schlüsselspeicherdatei, die auch als *Truststore* bezeichnet wird, enthält öffentliche Schlüssel, die die Verwendung des SSL-Protokolls für die gesicherte Kommunikation ermöglichen.

### Vorgehensweise

Gehen Sie wie folgt vor, um die Suchserver für die ausschließliche Annahme sicherer Suchanforderungen zu konfigurieren:

1. Melden Sie sich als Administrator für die Unternehmenssuche an. Führen Sie bei einer Installation mit mehreren Servern die folgenden Schritte auf dem Indexserver aus:
  - a. Stoppen Sie das System für die Unternehmenssuche:

```
esadmin system stopall
```
  - b. Bearbeiten Sie die Datei `ES_NODE_ROOT/master_config/nodes.ini`.
  - c. Ändern Sie den Wert für `knoten-id.searchserverport` vom HTTP-Port (normalerweise 80) in den HTTPS-Port (normalerweise 443), und speichern Sie die Datei. Aktualisieren Sie bei einer Installation mit mehreren Servern beide Werte für `knoten-id.searchserverport` (für jeden Suchserver).
2. Gehen Sie wie folgt vor, um den Suchserver zu aktualisieren: Führen Sie bei einer Installation mit mehreren Servern die folgenden Schritte auf beiden Suchservern aus:
  - a. Melden Sie sich bei einer Installation mit mehreren Servern als Administrator für die Unternehmenssuche am Suchserver an.
  - b. Bearbeiten Sie die Datei `ES_NODE_ROOT/nodeinfo/es.cfg`.
  - c. Aktualisieren Sie das Merkmal `TrustStore`, um den vollständig qualifizierten Pfad für die SSL-Schlüsselspeicherdatei anzugeben.
  - d. Aktualisieren Sie das Merkmal `HTTPProtocol`, um HTTPS anzugeben, und speichern Sie anschließend die Datei.
  - e. Geben Sie den folgenden Befehl ein, wobei `truststorekennwort` das Kennwort für die Schlüsselspeicherdatei ist. Mit diesem Befehl wird der Kennwortwert verschlüsselt und der Wert `TrustStorePassword` in der Datei `es.cfg` aktualisiert.

#### AIX, Linux oder Solaris

```
eschangetrustpw.sh kennwort_für_truststore
```

#### Windows

```
eschangetrustpw kennwort_für_truststore
```

- f. Stellen Sie sicher, dass die Merkmale `trustStore` und `trustPassword` in der Datei `config.properties` für die Suchanwendung den korrekten, vollständig qualifizierten Pfad und das korrekte Kennwort für die Schlüsselspeicherdatei angeben. Sie können diese Informationen prüfen oder ändern, indem Sie die Datei `config.properties` bearbeiten oder die Anpassungsfunktion für die Suchanwendung verwenden.
    - g. Wenn Sie das Suchportlet für die Unternehmenssuche verwenden, müssen Sie sicherstellen, dass die Portletparameter `trustStore` und `trustPassword` den korrekten, vollständig qualifizierten Pfad und das korrekte Kennwort für die Schlüsselspeicherdatei angeben. Verwenden Sie die Portletverwaltungsoptionen in der WebSphere Portal-Verwaltungsschnittstelle, um diese Informationen zu prüfen oder zu ändern.
3. Starten Sie das System für die Unternehmenssuche erneut:

esadmin system startall

### Zugehörige Tasks

„Bearbeiten der Merkmale der Mustersuchanwendung“ auf Seite 260

„Anpassen von Suchanwendungen“ auf Seite 261

„Konfigurieren der Unternehmenssuche in WebSphere Portal Version 5.1“ auf Seite 375

„Konfigurieren der Unternehmenssuche in WebSphere Portal Version 6“ auf Seite 381

---

## Konfigurieren der Suchserver für die Annahme von Anforderungen über einen Proxy-Server

Sie können den Suchserver für die Annahme von Anforderungen über einen Proxy-Server konfigurieren.

### Vorgehensweise

Gehen Sie wie folgt vor, um Anforderungen an den Suchserver über einen Proxy-Server übergeben zu können:

1. Melden Sie sich als Administrator für die Unternehmenssuche an, und stoppen Sie das System für die Unternehmenssuche. Melden Sie sich bei einer Installation mit mehreren Servern auf dem Indexserver an.  
esadmin system stopall
2. Bearbeiten Sie die Datei `ES_NODE_ROOT/master_config/nodes.ini`.
  - a. Ändern Sie das Merkmal `ProxyServer`, um den vollständig qualifizierten Hostnamen für den Proxy-Server anzugeben.
  - b. Ändern Sie das Merkmal `ProxyServerPort`, um den Port für den Proxy-Server anzugeben.
  - c. Optional: Wenn der Proxy-Server erfordert, dass alle Anforderungen authentifiziert sind, dann müssen Sie das Merkmal `ProxyServerUserName` aktualisieren, um einen gültigen Benutzernamen für den Proxy-Server anzugeben.
  - d. Optional: Wenn der Proxy-Server erfordert, dass alle Anforderungen authentifiziert sind, dann müssen Sie den folgenden Befehl eingeben. Dabei steht *kennwort\_des\_proxy-servers* für das Kennwort für den Benutzernamen des angegebenen Proxy-Servers. Mit diesem Befehl wird der Kennwortwert verschlüsselt und der Wert `ProxyServerUserPassword` in der Datei `es.cfg` aktualisiert.

### AIX, Linux oder Solaris

```
eschangeproxypw.sh kennwort_des_proxy-servers
```

### Windows

```
eschangeproxypw kennwort_des_proxy-servers
```

3. Stellen Sie sicher, dass die Merkmale `proxyHost` und `proxyPort` in der Konfigurationsdatei (Datei mit Endung `.properties`) für die Suchanwendung den korrekten, vollständig qualifizierten Hostnamen und die korrekte Portnummer für den Proxy-Server angeben.

Wenn der Proxy-Server eine Authentifizierung erfordert, müssen Sie sicherstellen, dass die Merkmale `proxyUser` und `proxyPassword` einen gültigen Benutzernamen und ein gültiges Kennwort für den Proxy-Server angeben. Sie können diese Informationen prüfen oder ändern, indem Sie die Konfigurationsdatei bearbeiten oder die Anpassungsfunktion für die Suchanwendung verwenden.

4. Wenn Sie das Suchportlet für die Unternehmenssuche verwenden, müssen Sie sicherstellen, dass die Portletparameter `proxyHost` und `proxyPort` den korrekten, vollständig qualifizierten Hostnamen und die korrekte Portnummer für den Proxy-Server angeben.

Wenn der Proxy-Server eine Authentifizierung erfordert, müssen Sie sicherstellen, dass die Merkmale `proxyUser` und `proxyPassword` einen gültigen Benutzernamen und ein gültiges Kennwort für den Proxy-Server angeben. Verwenden Sie die Portletverwaltungsoptionen in der WebSphere Portal-Verwaltungsschnittstelle, um diese Informationen zu prüfen oder zu ändern.

5. Starten Sie das System für die Unternehmenssuche erneut:  
`esadmin system startall`

---

## Unterstützung für externe Quellen

Eine *externe Quelle* ist eine Datenquelle, die Sie für die Suche mit einer Anwendung für die Unternehmenssuche aktivieren, ohne die Dokumente in der Datenquelle mit einem Crawler durchsuchen, syntaktisch analysieren oder indexieren zu müssen.

Die folgenden Datenquellentypen können als externe Quellen durchsucht werden:

- Datenbanken, die das JDBC-Protokoll (Java Database Connectivity) unterstützen. Nur IBM DB2-, Oracle-, Microsoft SQL Server 2000- und Microsoft SQL Server 2005-Datenbanken werden unterstützt. Für jede Tabelle, für die Sie das Durchsuchen ermöglichen, wird eine separate externe Quelle erstellt.

**Einschränkung:** Die Unterstützung für SQL Server 2005-Datenbanken ist auf Tabellen beschränkt, die keine Daten des Typs 'VARIANT' enthalten. Der JDBC-Treiber für SQL Server 2005 wird auf AIX-Systemen nicht unterstützt.

- LDAP-Server (Lightweight Directory Access Protocol). Für jeden LDAP-Server wird eine externe Quelle erstellt.

Nach der Konfiguration von Informationen zu einer externen Quelle müssen Sie ihr mindestens eine Suchanwendung zuordnen. Benutzer können dann die externe Quelle durchsuchen und gleichzeitig Objektgruppen abfragen, die mittels Crawler-suche, Syntaxanalyse und Indexieren von Daten für die Unternehmenssuche erstellt wurden.

### Zugehörige Konzepte

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

---

## Hinzufügen von externen Quellen zum System

Wenn Sie dem System für die Unternehmenssuche eine externe Quelle hinzufügen, müssen Sie den gewünschten Quellentyp angeben. Mithilfe eines Assistenten können Sie Informationen zur Datenquelle angeben und festlegen, wie diese durchsucht werden soll.

### Vorbereitung

Sie müssen ein Administrator für die Unternehmenssuche sein, um dem System externe Quellen hinzuzufügen.

### Einschränkungen

Zum Durchsuchen einer Oracle-Datenbank als externe Quelle muss das Oracle-Clientprogramm auf den Suchservern für die Unternehmenssuche installiert sein.

Der JDBC-Treiber für Microsoft SQL Server 2005 wird auf AIX-Systemen nicht unterstützt.

### Informationen zu dieser Task

Wenn Sie dem System Informationen zu einer externen Quelle hinzufügen, ermöglichen Sie den Benutzern die Abfrage der Quelle mit einer Anwendung für die

Unternehmenssuche. Sie können festlegen, dass LDAP-Server (Lightweight Directory Access Protocol) und JDBC-Datenbanktabellen (Java Database Connectivity) durchsucht werden.

Bei der Konfiguration eines LDAP-Servers können Sie mithilfe eines Assistenten Informationen angeben, die dem System die Herstellung einer Verbindung zum Server ermöglichen, und Optionen angeben, wie der Server durchsucht werden soll.

Bei der Konfiguration einer JDBC-Datenbank können Sie mithilfe eines Assistenten Informationen angeben, die dem System die Herstellung einer Verbindung zur Datenbank ermöglichen, die Tabellen auswählen, die für die Suche aktiviert werden sollen, und Optionen angeben, wie die Daten in den Tabellen durchsucht werden sollen. Für jede Tabelle, die Sie dem System hinzufügen, wird eine separat durchsuchbare externe Quelle erstellt.

Informationen zu Treibern für SQL Server 2000 finden Sie unter <http://www.microsoft.com/downloads/details.aspx?familyid=07287B11-0502-461A-B138-2AA54BFDC03A&displaylang=en>. Informationen zu Treibern für SQL Server 2005 finden Sie unter <http://www.microsoft.com/downloads/details.aspx?familyid=e22bc83b-32ff-4474-a44a-22b6ae2c4e17&displaylang=en>.

### Vorgehensweise

Gehen Sie wie folgt vor, um dem System eine externe Quelle hinzuzufügen:

1. Führen Sie vor dem Hinzufügen einer externen Quelle die folgenden Schritte aus, um JDBC-Datenbanken in ein System für die Unternehmenssuche einzuschließen. Sie müssen diesen Schritt, der dem System das Lokalisieren der geeigneten JDBC-Treiber ermöglicht, einmal ausführen.
  - a. Melden Sie sich am Crawler-Server als Administrator für die Unternehmenssuche an.
  - b. Bearbeiten Sie die Datei `ES_INSTALL_ROOT/configurations/interfaces/discovery__interface.ini`, und geben Sie an, dass die Umgebungsvariablen `CLASSPATH` und `LD_LIBPATH` den Klassenpfad zu den JDBC-Treibern und den Pfad zu den Bibliotheksdateien enthalten sollen.
  - c. Bearbeiten Sie die Datei `ES_INSTALL_ROOT/configurations/interfaces/customcommunication__interface.ini`, und geben Sie an, dass die Umgebungsvariablen `CLASSPATH` und `LD_LIBPATH` den Klassenpfad zu den JDBC-Treibern und den Pfad zu den Bibliotheksdateien enthalten sollen.
  - d. Optional: Fügen Sie der Umgebungsvariablen `LD_LIBPATH` den Pfad der Oracle-Bibliothek hinzu (z. B. `LD_LIBPATH=../home/oracle/OraHome1/lib32`), und bearbeiten Sie die Datei `escrset.sh` so, dass sie den Bibliothekspfad angibt und die Oracle-Installationsverzeichnisvariable exportiert, damit ein Oracle JDBC-Treiber für lokale oder katalogisierte Datenbanken verwendet wird. Beispiel:

```
ORACLE_HOME=/home/oracle/OraHome1
export ORACLE_HOME
```
  - e. Starten Sie das System für die Unternehmenssuche erneut, einschließlich der allgemeinen Übertragungsschicht (Common Communication Layer, CCL):

#### AIX, Linux oder Solaris

```
esadmin stop
stopccl.sh
startccl.sh
esadmin start
```

### Windows-Eingabeaufforderung

```
esadmin stop
stopccl
startccl
esadmin start
```

### Verwaltungstool für Windows-Services

Gehen Sie wie folgt vor, um die allgemeine Übertragungsschicht im Hintergrund zu starten:

- 1) Geben Sie `esadmin stop` ein.
  - 2) Starten Sie die Windows-Services: **Start** → **Einstellungen** → **Systemsteuerung** → **Verwaltung** → **Dienste**.
  - 3) Klicken Sie **IBM OmniFind Enterprise Edition** mit der rechten Maustaste an, und klicken Sie **Beenden** an. Nachdem der Service gestoppt wurde, klicken Sie **Starten** an.
  - 4) Geben Sie `esadmin start` ein.
2. Klicken Sie **Externe Quellen** an, um die Sicht **Externe Quellen** zu öffnen.
  3. Klicken Sie **Externe Quelle hinzufügen** an.
  4. Wählen Sie den Typ der externen Quelle aus, den Sie hinzufügen wollen (LDAP-Server oder JDBC-Datenbank).
  5. Klicken Sie **Weiter** an, um die Konfiguration der externen Quelle zu starten.

Ein Assistent wird für den Quellentyp geöffnet, den Sie erstellen. Führen Sie die Eingabeaufforderungen des Assistenten aus, um die externe Quelle zu konfigurieren. Sie können auf jeder Seite des Assistenten **Hilfe** anklicken, um mehr über die zur Verfügung stehenden Optionen zu erfahren.

Die folgenden Standardnamen und -speicherpositionen von JDBC-Treibern können Ihnen bei der Konfiguration von Verbindungsinformationen für DB2- und Oracle-Datenbanken behilflich sein:

#### DB2: Traditioneller JDBC-Treiber

Treibername: `COM.ibm.db2.jdbc.app.DB2Driver`  
Beispielspeicherposition: `db2-installationsroot/java/db2java.zip`

#### DB2: Universeller JDBC-Treiber

Treibername: `com.ibm.db2.jcc.DB2Driver`  
Beispielspeicherpositionen:  
`db2-installationsroot/java/db2jcc.jar`  
`db2-installationsroot/java/db2jcc_license_cu.jar`

#### Oracle

Treibername: `oracle.jdbc.driver.OracleDriver`  
Beispielspeicherposition: `oracle_ausgangsverzeichnis/jdbc/lib/ojdbc14.jar`

#### Microsoft SQL Server 2000

Treibername: `com.microsoft.jdbc.sqlserver.SQLServerDriver`  
Beispielspeicherpositionen:  
`mssql-jdbc-ausgangsverzeichnis/lib/mssqlserver.jar`  
`mssql-jdbc-ausgangsverzeichnis/lib/msbase.jar`  
`mssql-jdbc-ausgangsverzeichnis/lib/msutil.jar`

#### Microsoft SQL Server 2005

Treibername: `com.microsoft.sqlserver.jdbc.SQLServerDriver`  
Beispielspeicherposition: `installationsverzeichnis/sqljdbc_1.0/ländereinstellung/sqljdbc.jar`  
Beispiel: `installationsverzeichnis/sqljdbc_1.0/enu/sqljdbc.jar`

6. Klicken Sie nach Angabe der Optionen zum Durchsuchen der externen Quelle **Fertig stellen** an.

Die neue externe Quelle wird in der Sicht **Externe Quellen** zusammen mit anderen externen Quellen angezeigt, die dem System hinzugefügt wurden.

### Zugehörige Konzepte

 SI-API-Föderatoren (Search and Index API)

---

## Zuordnen von Suchanwendungen zu externen Quellen

Bevor Sie eine externe Quelle durchsuchen können, müssen Sie ihr mindestens eine Suchanwendung zuordnen.

### Vorbereitung

Sie müssen ein Administrator für die Unternehmenssuche sein, damit Sie Suchanwendungen den zu durchsuchenden externen Quellen zuordnen können.

### Vorgehensweise

Gehen Sie wie folgt vor, um eine Suchanwendung mindestens einer externen Quelle zuzuordnen:

1. Klicken Sie **Sicherheit** in der Funktionsleiste der Administrationskonsole an.
2. Klicken Sie **Suchanwendungen konfigurieren** auf der Seite **Suchanwendungen** an.
3. Klicken Sie **Suchanwendung hinzufügen** auf der Seite **Suchanwendungen konfigurieren** an.
4. Geben Sie den Namen der Suchanwendung ein.
5. Wählen Sie die externen Quellen aus, die die Anwendung durchsuchen soll:
  - Klicken Sie **Alle Objektgruppen und externen Quellen** an, wenn die Suchanwendung alle externen Quellen durchsuchen soll, die Sie dem System hinzufügen.
  - Klicken Sie **Bestimmte Objektgruppen und externe Quellen** an, wenn die Suchanwendung nur die von Ihnen angegebenen externen Quellen durchsuchen soll.Wenn Sie diese Option auswählen, wird eine Liste mit Objektgruppennamen und Namen externer Quellen angezeigt. Wählen Sie das Markierungsfeld **Auswählen** für jede externe Quelle aus, die die Anwendung durchsuchen soll.

6. Klicken Sie **OK** an.

### Zugehörige Konzepte

 SI-API-Föderatoren (Search and Index API)

---

## Sicherheit für die Unternehmenssuche

Über die Sicherheitsmechanismen für die Unternehmenssuche können Sie Ihre Quellen gegen unberechtigtes Durchsuchen schützen und Verwaltungsfunktionen auf bestimmte Benutzer beschränken.

Mithilfe der Unternehmenssuche kann der Benutzer einen großen Bereich von Datenquellen durchsuchen. Um sicherzustellen, dass nur berechtigte Benutzer auf den betreffenden Inhalt zugreifen können, und um sicherzustellen, dass nur berechtigte Benutzer auf die Administrationskonsole zugreifen können, koordiniert die Unternehmenssuche die Sicherheit auf mehreren Ebenen und setzt sie dort um.

### **Tipp:**

Detaillierte Beispiele für die Konfiguration der Sicherheit für die Unternehmenssuche finden Sie im IBM Redbook *IBM OmniFind Enterprise Edition Configuration and Implementation Scenarios*. Die Szenarien zeigen Ihnen, wie Sie die globale Sicherheit in WebSphere Application Server mit einem LDAP-Repository aktivieren, die Komponente für das Identitätsmanagement einrichten und verschiedene Crawler konfigurieren, um sicherzustellen, dass die Sicherheit auf Dokumentenebene erzwingen wird.

### **Web-Server**

Die erste Sicherheitsebene ist der Web-Server. Wenn Sie die globale Sicherheit in WebSphere Application Server aktivieren, können Sie Benutzern Verwaltungsrollen zuordnen und Benutzer authentifizieren, die das System verwalten. Wenn sich ein Benutzer an der Administrationskonsole anmeldet, sind für diesen Benutzer nur die Funktionen und Objektgruppen verfügbar, die dieser Benutzer verwalten darf.

Suchanwendungen können außerdem Sicherheitsmechanismen von WebSphere Application Server verwenden, um Benutzer zu authentifizieren, die Objektgruppen durchsuchen.

### **Sicherheit auf Objektgruppenebene**

Wenn Sie eine Objektgruppe erstellen, können Sie die Sicherheit auf Objektgruppenebene aktivieren. Sie können diese Einstellung nach der Erstellung der Objektgruppe nicht ändern. Wenn Sie die Sicherheit auf Objektgruppenebene nicht aktivieren, können Sie später keine Sicherheitsmaßnahmen auf Dokumentenebene festlegen.

Bei aktivierter Objektgruppensicherheit gilt Folgendes:

- Die globalen Analyseprozesse der Unternehmenssuche wenden verschiedene Regeln für die Indexierung von doppelten Dokumenten an.
- Sie können Optionen so konfigurieren, dass Sicherheit auf Objektgruppenebene umgesetzt wird. Beispielsweise können Dokumenten beim Durchsuchen Sicherheitstoken zugeordnet werden, für aktuelle Berechtigungsnachweise kann eine Prüfung während der Abfrageverarbeitung erforderlich sein, und es kann angegeben werden, ob Ankertext in Webdokumenten indexiert werden soll.
- Sie können die Sicherheit umsetzen, indem Sie den Objektgruppen und externen Quellen Suchanwendungen (nicht einzelne Benutzer) zuordnen, mit denen sie durchsucht werden können. Sie verwenden dann die Stan-

ardzugriffssteuerungsmaßnahmen, um Benutzern den Zugriff auf Suchanwendungen zu erlauben oder zu verweigern.

Sie müssen zwischen der Aktivierung der Objektgruppensicherheit und der Suchqualität abwägen. Das Aktivieren der Objektgruppensicherheit reduziert die Informationen, die für jedes Dokument indexiert werden. Als Nebeneffekt werden für einige Suchabfragen weniger Ergebnisse gefunden.

### **Sicherheit auf Dokumentebene**

Wenn Sie Crawler für eine Objektgruppe konfigurieren, können Sie die Sicherheit auf Dokumentebene aktivieren. Sie können beispielsweise Optionen angeben, um Daten Sicherheitstoken zuzuordnen, während die Daten von Crawlern gesammelt werden. Ihre Suchanwendungen können diese mit Dokumenten im Index gespeicherten Token verwenden, um Zugriffssteuerungen umzusetzen und sicherzustellen, dass nur Benutzer mit entsprechenden Berechtigungsnachweisen die Möglichkeit haben, die Daten abzufragen und Suchergebnisse anzuzeigen.

Für bestimmte Datenquellentypen können Sie Optionen konfigurieren, um die Anmeldeberechtigungen eines Benutzers mit aktuellen Zugriffssteuerungen während der Abfrageverarbeitung zu prüfen. Diese zusätzliche Sicherheitsebene stellt sicher, dass die Berechtigungen eines Benutzers in Echtzeit mit der nativen Datenquelle geprüft wird. Diese Funktionalität kann vor Situationen schützen, in denen sich die Berechtigungsnachweise eines Benutzers ändern, nachdem ein Dokument und die zugehörigen Sicherheitstoken indexiert wurden.

Die Verarbeitungsphase des Ankertextes innerhalb der globalen Analyse ordnet normalerweise Text, der in einem Dokument erscheint (dem Quelldokument), einem anderen Dokument zu (dem Zieldokument), in dem dieser Text nicht unbedingt erscheinen muss. Wenn Sie einen Web-Crawler konfigurieren, können Sie angeben, ob der Ankertext aus dem Index ausgeschlossen werden soll, wenn der Text mit einem Dokument verknüpft ist, das der Web-Crawler nicht durchsuchen darf.

Die Sicherheit für Ihre Objektgruppe umfasst die Authentifizierung und Zugriffssteuerungsmechanismen, die die Unternehmenssuche verwenden kann, um indexierten Inhalt zu schützen. Es bestehen außerdem Sicherheitsmaßnahmen gegen zerstörerisch tätige und nicht berechtigte Benutzer, die versuchen, auf Daten zuzugreifen, während diese übermittelt werden. Beispiel: Die Suchserver verwenden für die Kommunikation mit dem Indexserver und der Suchanwendung Protokolle wie SSL (Secure Sockets Layer), SSH (Secure Shell) und HTTPS (Secure Hypertext Transfer Protocol).

Zusätzliche Sicherheit wird durch Verschlüsselung geboten. So wird z. B. das Kennwort des Administrators für die Unternehmenssuche, das während der Produktinstallation angegeben wird, in einem verschlüsselten Format gespeichert. Kennwörter, die Benutzer in den Benutzerprofilen angeben, werden auch in einem verschlüsselten Format gespeichert.

Für erhöhte Sicherheit müssen Sie sicherstellen, dass die Serverhardware gesichert und vor unbefugtem Zugriff geschützt wird. Durch die Installation einer Firewall können Sie die Server für die Unternehmenssuche vor unbefugtem Zugriff durch einen anderen Teil Ihres Netzes schützen. Stellen Sie außerdem sicher, dass auf den Servern für die Unternehmenssuche keine offenen Ports vorhanden sind. Konfigurieren Sie das System so, dass es für Anforderungen nur in den Ports empfängsbereit ist, die explizit den Aktivitäten und Anwendungen für die Unternehmenssuche zugewiesen wurden.

---

## Installationssicherheit

Das Installationsprogramm für OmniFind Enterprise Edition richtet eine Umgebung ein, in der die Sicherheit umgesetzt wird, wenn Benutzer Objektgruppen für die Unternehmenssuche verwalten oder durchsuchen.

### Administrator-ID für die Unternehmenssuche

Während des Installationsprozesses wird der Installationsverantwortliche aufgefordert, eine Benutzer-ID und ein Kennwort einzugeben, die für den Administrator für die Unternehmenssuche verwendet werden sollen. Das System verwendet die angegebenen Berechtigungsnachweise, für folgende Zwecke:

- Authentifizieren des Administrators für die Unternehmenssuche, wenn Verwaltungstasks ausgeführt werden.
- Erstellen der internen Datenbank für die Unternehmenssuche.
- Starten aller Sitzungen oder Prozesse der Unternehmenssuche.

Die Benutzer-ID, die während der Installation angegeben wird, muss eine gültige Benutzer-ID des Betriebssystems mit Systemverwaltungsberechtigung sein. Das Installationsprogramm speichert die Berechtigungsnachweise in verschlüsselter Form in einer Merkmaldatei auf jedem Server für die Unternehmenssuche.

### Verschlüsselung

Zum Schutz sensibler Daten wird bei allen Nachrichten, die über das System für die Unternehmenssuche übertragen werden, eine Verschlüsselung verwendet, um den Abschnitt mit den Authentifizierungsdaten zu codieren. Dieser Prozess erfordert wenig Systemaufwand, weil nur die Authentifizierungs-IDs und die Kennwörter verschlüsselt werden. Außerdem werden alle Kennwörter verschlüsselt, die vom System gespeichert werden (in Konfigurationsdateien, in der Datenbank für die Unternehmenssuche und so weiter).

### WebSphere Application Server-Sicherheit

Wenn WebSphere Application Server zuvor nicht auf dem Suchserver installiert war, installiert das Installationsprogramm das Produkt automatisch mit inaktivierter globaler Sicherheit. Wird die globale Sicherheit zu einem späteren Zeitpunkt in WebSphere Application Server aktiviert, ist WebSphere Application Server für die Authentifizierung des Administrators für die Unternehmenssuche zuständig.

Wenn Sie die globale Sicherheit aktivieren, müssen Sie die Administrator-ID und das Administratorkennwort für die Unternehmenssuche, die während der Installation angegeben wurden, der Benutzerregistry von WebSphere Application Server hinzufügen, z. B. einem LDAP-Verzeichnis (LDAP - Lightweight Directory Access Protocol).

Wenn Sie nach der Installation von OmniFind Enterprise Edition die globale Sicherheit aktivieren, müssen Sie Konfigurationswerte aktualisieren und den Befehl **eschangewaspw** ausführen, um die WebSphere Application Server-Berechtigungsnachweise zu verschlüsseln und in einer Merkmaldatei für die Unternehmenssuche zu speichern.

---

## Authentifizierung und Zugriffssteuerung

Die Unternehmenssuche unterstützt die Benutzerauthentifizierung und die Verwendung von Berechtigungen (Zugriffssteuerung), um Ihre Inhalte vor nicht berechtigten Benutzern zu schützen und um den Zugriff auf Verwaltungsfunktionen zu steuern.

### Authentifizierung

Zur Authentifizierung gehören alle Prozesse, mit denen ein System die Identität eines Benutzers prüft, der auf das System zugreifen will. Da die Zugriffssteuerung normalerweise auf der Identität des Benutzers beruht, der den Zugriff auf eine Ressource anfordert, ist die Authentifizierung entscheidend für die effektive Sicherheit.

Die Authentifizierung von Benutzern der Unternehmenssuche wird über Berechtigungsnachweise implementiert, die mindestens aus einer Benutzer-ID und einem Kennwort bestehen.

Die Unternehmenssuche nutzt für die Authentifizierung von Benutzern, die auf die Administrationskonsole zugreifen, die Authentifizierungsunterstützung, die in WebSphere Application Server bereitgestellt wird.

### Berechtigung (Zugriffssteuerung)

Mit Berechtigung ist ein Mechanismus gemeint, mit dem ein System das Recht für den Zugriff auf bestimmte Daten oder das Ausführen einer bestimmten Aktion erteilt oder widerruft. Häufig ist eine bestimmte Art von Authentifizierung erforderlich, mit der sich ein Benutzer an einem System anmelden kann. Zugriffssteuerungsmechanismen bestimmen, welche Operationen der Benutzer ausführen kann und welche nicht. Dazu wird die Identität des Benutzers mit einer Zugriffssteuerungsliste verglichen. Zu den Zugriffssteuerungen gehören folgende Elemente:

- Dateiberechtigungen, wie z. B. das Recht zum Erstellen, Lesen, Bearbeiten oder Löschen einer Datei.
- Programmberechtigungen, wie z. B. das Recht zum Ausführen eines Programms.
- Datenberechtigungen, wie z. B. das Recht zum Abrufen oder Aktualisieren von Informationen in einer Datenbank.

---

## Verwaltungsrollen

Bei der Unternehmenssuche wird das Konzept der Rollen verwendet, um den Zugriff auf verschiedene Funktionen in der Administrationskonsole zu steuern.

Wenn OmniFind Enterprise Edition (OmniFind Enterprise Edition) installiert ist, konfiguriert der Installationsverantwortliche eine Benutzer-ID und ein Kennwort für den Administrator für die Unternehmenssuche. Wenn Sie das erste Mal auf die Administrationskonsole zugreifen, müssen Sie sich mit dieser Benutzer-ID anmelden. Wenn Sie die globale Sicherheit in WebSphere Application Server nicht aktivieren, ist diese Benutzer-ID die einzige Benutzer-ID, mit der Sie auf die Administrationskonsole für die Unternehmenssuche zugreifen können.

Wenn Sie die globale Sicherheit in WebSphere Application Server aktivieren, können Sie zusätzliche Benutzer als Benutzer mit Verwaltungsaufgaben für die Unternehmenssuche registrieren. Durch die Zuordnung von Benutzern zu Rollen können Sie den Zugriff auf bestimmte Objektgruppen beschränken und steuern, welche

Funktionen der Benutzer mit Verwaltungsaufgaben ausführen kann. Die Benutzer-IDs, die Sie Verwaltungsrollen in der Unternehmenssuche zuweisen, müssen in einer Benutzerregistry von WebSphere Application Server hinterlegt sein.

Wenn ein Benutzer mit Verwaltungsaufgaben sich anmeldet, authentifiziert die Unternehmenssuche die Benutzer-ID. Nur die Objektgruppen und Funktionen, die der Benutzer verwalten darf, sind für ihn in der Konsole verfügbar.

Sie können Benutzer für die folgenden Verwaltungsrollen registrieren:

#### **Administrator für die Unternehmenssuche**

Dieser Benutzer erstellt Objektgruppen und hat die Berechtigung, sämtliche Aspekte Ihres Systems für die Unternehmenssuche zu verwalten. Wenn Sie OmniFind Enterprise Edition installieren, geben Sie die Benutzer-ID und das Kennwort für den ersten Benutzer mit Verwaltungsaufgaben für die Unternehmenssuche an. Nach dem erstmaligen Anmelden kann dieser Benutzer der Rolle 'Administrator für die Unternehmenssuche' weitere Benutzer zuordnen.

#### **Objektgruppenadministrator**

Dieser Benutzer kann die Handhabung von Objektgruppen bearbeiten, überwachen und steuern, zu deren Verwaltung er berechtigt ist. Diese Benutzer können keine Objektgruppen erstellen. Objektgruppenadministratoren können Aktivitäten auf Systemebene nur überwachen und ausführen, wenn ihnen diese Berechtigung durch einen Administrator für die Unternehmenssuche erteilt wird.

#### **Bediener**

Dieser Benutzer kann die Handhabung von Objektgruppen überwachen und steuern, zu deren Verwaltung er berechtigt ist. Diese Benutzer können z. B. die Objektgruppenaktivitäten starten und stoppen, jedoch keine Objektgruppen erstellen oder bearbeiten. Ein Bediener kann Aktivitäten auf Systemebene nur überwachen und ausführen, wenn dem Bediener diese Berechtigung durch einen Administrator für die Unternehmenssuche erteilt wird.

#### **Überwachungsbeauftragter**

Dieser Benutzer kann Objektgruppen überwachen, zu deren Verwaltung er berechtigt ist. Diese Benutzer können keine Vorgänge steuern (wie z. B. das Starten und Stoppen von Servern), Objektgruppen erstellen, oder Objektgruppen bearbeiten. Ein Überwachungsbeauftragter kann Aktivitäten auf Systemebene nur beobachten, aber nicht ausführen, wenn dem Überwachungsbeauftragten diese Berechtigung durch einen Administrator für die Unternehmenssuche erteilt wird.

## **Konfigurieren von Benutzern mit Verwaltungsaufgaben**

Indem Sie Verwaltungsrollen konfigurieren, können Sie den Zugriff auf Objektgruppen einschränken und die Funktionen steuern, die einzelne Benutzer mit Verwaltungsaufgaben ausführen können.

### **Vorbereitung**

Bevor Sie einen Benutzer einer Verwaltungsrolle zuordnen, stellen Sie sicher, dass die Sicherheit in WebSphere Application Server aktiviert ist. Stellen Sie weiterhin sicher, dass die Benutzer-ID in einer WebSphere Application Server-Benutzerregistry vorhanden ist.

Sie müssen ein Administrator für die Unternehmenssuche sein, um Benutzer mit Verwaltungsaufgaben zu konfigurieren.

### Vorgehensweise

Gehen Sie wie folgt vor, um Benutzern Verwaltungsrollen zuzuweisen:

1. Klicken Sie **Sicherheit** an, um die Sicht **Sicherheit** zu öffnen.
2. Klicken Sie **Benutzer hinzufügen** auf der Seite **Verwaltungsrollen** an.
3. Geben Sie die Benutzer-ID des Benutzers ein, den Sie registrieren wollen, und wählen Sie die entsprechende Verwaltungsrolle aus.
4. Wenn Sie den Benutzer nicht als Administrator für die Unternehmenssuche registrieren, geben Sie an, ob dieser Benutzer Seiten über die Funktionsleistung **System** abrufen kann.

Beispiel: Sie möchten unter Umständen, dass manche Bediener oder Objektgruppenadministratoren Protokolldateien auf Systemebene überwachen können.

5. Wenn Sie den Benutzer nicht als Administrator für die Unternehmenssuche registrieren, wählen Sie die Objektgruppen und externen Quellen aus, die der Benutzer verwalten soll.

Sie können die Markierungsfelder der einzelnen Objektgruppen und externen Quellen auswählen oder die Verwaltung aller Objektgruppen und externen Quellen für den Benutzer aktivieren.

---

## Sicherheit auf Objektgruppenebene

Für die Bereitstellung von Sicherheit auf Objektgruppenebene konfigurieren Sie Optionen für die Indexierung von Inhalt und Optionen, die es Suchanwendungen erlauben, bestimmte Objektgruppen zu durchsuchen.

Wenn Sie eine Objektgruppe erstellen, können Sie eine Option für die Aktivierung der Objektgruppensicherheit auswählen. Wenn Sie diese Option wählen, können Sie später Sicherheitsmaßnahmen auf Dokumentenebene konfigurieren. Bei aktivierter Objektgruppensicherheit wenden die globalen Analyseprozesse der Unternehmenssuche auch verschiedene Regeln für die Indexierung von doppelten Dokumenten an.

Nachdem Sie eine Suchanwendung erstellt haben, können Sie mithilfe einer Suchanwendungs-ID festlegen, welche Objektgruppen und externen Quellen die Suchanwendung durchsuchen darf und welche Benutzer auf die Suchanwendung zugreifen dürfen.

## Analyse von doppelten Dokumenten und Objektgruppensicherheit

Wenn Sie die Objektgruppensicherheit aktivieren, erkennen die globalen Analyseprozesse keine doppelten Dokumente in der Objektgruppe.

Während der globalen Analyse erkennt der Indexierungsprozess Dokumente, die doppelt sind oder fast die gleichen Werte haben. Dann werden all diese Dokumente einer kanonischen Darstellung des Inhalts zugeordnet. Wenn Sie zulassen, dass doppelte Dokumente erkannt werden, können Sie sicherstellen, dass die Suchergebnisse nicht mehrere Dokumente mit demselben oder fast identischem Inhalt enthalten.

Wenn Sie beim Erstellen einer Objektgruppe die Objektgruppensicherheit aktivieren, werden doppelte Dokumente nicht erkannt und somit nicht einer einheitlichen kanonischen Darstellung zugeordnet. Stattdessen wird jedes Dokument einzeln indexiert. Dadurch wird sichergestellt, dass die Sicherheitsmaßnahmen für jedes Dokument ausgewertet werden, sodass Benutzer nur jene Dokumente durchsuchen können, deren Sicherheitstoken mit ihren Berechtigungsnachweisen übereinstimmen. Zwei Dokumente können nahezu identischen Inhalt haben, aber verschiedene Zugriffssteuerungslisten für die Umsetzung der Sicherheit verwenden.

Angenommen, es gibt zwei doppelte Dokumente, Dokument\_A und Dokument\_B, von denen ein Benutzer nur über die Zugriffsberechtigung für Dokument\_B verfügt. Da Dokument\_B durch die Erkennung doppelter Dokumente eliminiert ist, wird dem Benutzer keines der Dokumente in den Suchergebnissen angezeigt, da er für Dokument\_A keine Zugriffsberechtigung hat.

Die Inaktivierung der Analyse auf doppelte Dokumente kann die Sicherheit von Dokumenten in einer Objektgruppe verbessern, die Suchqualität kann sich jedoch verschlechtern, wenn der Benutzer mehrere Kopien desselben Dokuments als Suchergebnis erhält.

## Suchanwendungs-IDs

Die Funktionalität zum Durchsuchen verschiedener Objektgruppen wird gesteuert, indem Sie den Objektgruppen und externen Quellen Suchanwendungen zuordnen, mit denen sie durchsucht werden können. Eine Anwendung mit dem Namen **Default** ermöglicht, dass die Mustersuchanwendung wie bereitgestellt verwendet werden kann, um alle Objektgruppen und externen Quellen zu durchsuchen.

Alle Suchanwendungen müssen einen gültigen Anwendungsnamen (APPID) an die Anwendungsprogrammierschnittstelle für die Unternehmenssuche übergeben. Nur die Objektgruppen und externen Quellen, die dieser APPID zugeordnet sind, können mit der Suchanwendung durchsucht werden.

Bevor eine Suchanwendung auf eine Objektgruppe oder externe Quelle zugreifen kann, muss ein Administrator für die Unternehmenssuche die Suchanwendung genau den Objektgruppen und Quellen zuordnen, die die Anwendung durchsuchen darf. Eine Suchanwendung kann alle Objektgruppen und externen Quellen in einem System für die Unternehmenssuche oder nur die Objektgruppen und externen Quellen durchsuchen, die Sie angeben.

Die Mustersuchanwendung (ESSearchApplication) verfügt über eine Merkmaldatei, die den zu verwendenden Anwendungsnamen angibt. Die Standardposition für diese Merkmaldatei ist `ES_INSTALL_ROOT\installedApps\ESSearchApplication.ear\ESSearchApplication.war\WEB-INF\config.properties`.

Der Anfangswert für den Anwendungsnamen lautet **Default**. Wenn Sie diesen Wert ändern, ändern Sie die Liste der Objektgruppen und externen Quellen, die die Anwendung ESSearchApplication durchsuchen kann.

Wenn Sie steuern wollen, welche Benutzer welche Objektgruppen durchsuchen können, müssen Sie der Clientanwendung Benutzer (oder Benutzergruppen) zuordnen, indem Sie die Steuerfunktionen für den Standardzugriff von WebSphere Application Server in ähnlicher Weise verwenden, wie Sie sie für die Einschränkung des Zugriffs auf eine URL (Uniform-Resource-Locator) verwenden. Sie können z. B. den Zugriff auf die URL einschränken, über die Ihre Suchanwendung gestartet wird.

Weitere Informationen zu Suchanwendungs-IDs und dazu, wie Sicherheitsmaßnahmen in Ihre benutzerdefinierten Suchanwendungen eingebunden werden können, finden Sie in den Abschnitten über Search and Index API für die Unternehmenssuche.

#### Zugehörige Konzepte

-  SI-API-Übersicht (Search and Index API)
-  SI-API-Sicherheit (Search and Index API)

---

## Sicherheit auf Dokumentebene

Wenn die Sicherheit bei der Erstellung einer Objektgruppe aktiviert ist, können Sie Sicherheitsmaßnahmen auf Dokumentebene konfigurieren. Mit der Zugriffssteuerung auf Dokumentebene wird sichergestellt, dass die Suchergebnisse nur Dokumente enthalten, für deren Anzeige der Benutzer berechtigt ist, der die Suchanforderung übergeben hat.

Ein System für die Unternehmenssuche unterstützt viele Methoden für das Konfigurieren von Sicherheitsmaßnahmen auf Dokumentebene:

- Dokumente können vorgefiltert und Sicherheitstoken zugeordnet werden, bevor sie dem Index hinzugefügt werden.
- Für einige Datentypen können Suchergebnisse nachgefiltert werden, um die Anmeldeberechtigung des Benutzers anhand der aktuellen Zugriffssteuerungsdaten zu prüfen. Die Identitätsmanagementkomponente für die Unternehmenssuche kann die verschiedenen Berechtigungsnachweise verschlüsseln, die Benutzer für den Zugriff auf unterschiedliche Repositories benötigen, und die verschlüsselten Berechtigungsnachweise in Profilen speichern. Wenn die zu durchsuchenden Quellen durch ein Produkt geschützt sind, das SSO-Sicherheit (SSO - Single Sign-on) bereitstellt, kann die Identitätsmanagementkomponente den Zugriff auf Dokumente steuern, ohne dass Benutzer Profile erstellen müssen.
- Für die meisten Crawlertypen kann eine benutzerdefinierte Java-Klasse (Plug-in) verwendet werden, um Dokumenten im Index Sicherheitstoken zuzuordnen.
- Für Dokumente, die von einem Web-Crawler durchsucht werden, kann der Ankertext in Dokumenten, die Links zu unzulässigen Dokumenten enthalten, aus dem Index ausgeschlossen werden.

#### Zugehörige Konzepte

-  Anwendungssicherheit

## Vor- und Nachfiltern von Suchergebnissen

Es gibt zwei verschiedene Methoden zum Filtern von Dokumenten, mit denen sichergestellt werden kann, dass die Suchergebnisse nur Dokumente enthalten, für deren Anzeige der Benutzer berechtigt ist, der die Suchanforderung übergeben hat.

- Bei der ersten Methode werden die nativen Zugriffssteuerungslisten (ACLs - Access Control Lists) des Dokuments während der Crawlersuche in den Index repliziert. Dabei wird die Suchmaschine verwendet, um Benutzerberechtigungen mit den replizierten Dokument-ACLs zu vergleichen. Durch Vorfiltern der Dokumente und durch das Steuern, welche Dokumente dem Index hinzugefügt werden, erzielen Sie die beste Leistung. Es ist jedoch nicht einfach, alle Sicherheitsrichtlinien der verschiedenen Back-End-Quellen im Index zu modellieren und

die Vergleichslogik einheitlich zu implementieren. Diese Methode ist außerdem nicht sehr reaktionsfähig in Bezug auf Änderungen, die in den Quellen-ACLs auftreten können.

- Bei der zweiten Methode werden Dokumente in der Ergebnismenge nachgefiltert, indem aktuelle Sicherheitsdaten von den Back-End-Quellen abgefragt werden. Diese Methode ermöglicht, dass die beteiligten Back-End-Quellen als abschließende Arbiter für die Dokumente verwendet werden können, die an den Benutzer zurückgegeben werden. Sie stellt außerdem sicher, dass die Ergebnismenge dem aktuellen Stand der Zugriffssteuerung entspricht. Diese Methode führt jedoch zu einer verminderten Suchleistung, weil sie es erforderlich macht, dass zu allen Back-End-Quellen eine Verbindung besteht. Wenn eine Quelle nicht zugänglich ist, müssen Links zu Dokumenten zusammen mit Dokumenten, für deren Anzeige der Benutzer nicht berechtigt ist, aus der Ergebnismenge herausgefiltert werden.

**Wichtig:** In einer Konfiguration mit mehreren Servern erfolgt das Nachfiltern für einige Quellentypen auf dem Crawler-Server. Wenn der Crawler-Server für die Wartung heruntergefahren ist, erhalten Benutzer keine Ergebnisse, wenn sie Objektgruppen für die Unternehmenssuche abfragen. Außerdem werden keine Ergebnisse zurückgegeben, wenn nicht auf die Back-End-Server zugegriffen werden kann, die für die Zugriffssteuerung erforderlich sind.

Bei der Unternehmenssuche basiert die Unterstützung für die Umsetzung der Zugriffssteuerung auf einer Kombination dieser zwei Methoden. Der Entwurf stellt eine optimale Leistung bereit und verwaltet dabei die exakten Sicherheitsrichtlinien der Ursprungsdocumentrepositorys. Durch Speichern von Zugriffssteuerungsdaten der höchsten Ebene im Index kann das System eine vorläufige (möglicherweise kleinere) Ergebnismenge bereitstellen, die dann nachgefiltert werden kann, um aktuelle Zugriffssteuerungen zu prüfen. Dabei wird angenommen, dass ein Benutzer, der Zugriff auf das Repository hat, zu dem das Dokument gehört, wahrscheinlich auch auf das Dokument zugreifen kann.

Die Zugriffssteuerungsdaten, die im Index gespeichert werden, variieren je nach Crawlertyp. Der Notes-Crawler beispielsweise kann Zugriffssteuerungen auf Datenbank- und Serverebene speichern, und der QuickPlace-Crawler speichert Zugriffssteuerungen für Server, Bereiche und Räume.

Alle Datenquellentypen in einem System für die Unternehmenssuche unterstützen die Funktionalität zum Indexieren nativer Zugriffssteuerungslisten während der Crawleruche. Einige Datenquellentypen unterstützen außerdem die Funktionalität zum Nachfiltern der Ergebnismenge und zum Prüfen der aktuellen Berechtigungsnachweise des Benutzers. (Diese Art der Unterstützung wird über native Sicherheitsmechanismen oder über die Identitätsmanagementkomponente für die Unternehmenssuche bereitgestellt.)

Dieser Doppelansatz für die Sicherheit umfasst die folgenden Tasks:

- Extrahieren nativer Informationen der Zugriffssteuerungsliste während der Crawleruche.
- Speichern von Server- und Datenbank-ACL-Informationen im Index.
- Erstellen des Sicherheitskontexts des Benutzers, wenn der Benutzer sich anmeldet oder wenn die Sitzung initialisiert wird. Bei dieser Task müssen die verschiedenen Kennungen berücksichtigt werden, die ein einzelner Benutzer verwenden muss, um auf die verschiedenen Back-End-Quellen zuzugreifen.

- Verarbeiten der Suche mit dem Sicherheitskontext des Benutzers und Erstellen einer vorläufigen Ergebnismenge, die nur die Dokumente enthält, auf die der Benutzer auf Repositoryebene zugreifen kann.
- Nachfiltern der vorläufigen Ergebnismenge anhand der aktuellen nativen Informationen der Zugriffssteuerungslisten der Back-End-Quellen, von denen die Dokumente für die Ergebnismenge stammen.

## Prüfung durch gespeicherte Sicherheitstoken

Wenn die Sicherheit für eine Objektgruppe bei ihrer Erstellung aktiviert ist, können Sie Sicherheitsmaßnahmen auf Dokumentenebene konfigurieren, indem Sie im Index Sicherheitsdaten speichern.

Standardmäßig wird jedem Dokument ein allgemein zugängliches Token zugeordnet, wodurch das Dokument für jeden zugänglich wird. Wenn die Sicherheit für die Objektgruppe aktiviert ist, kann das allgemein zugängliche Token durch einen Wert ersetzt werden, der vom Administrator bereitgestellt oder einem Feld im durchsuchten Dokument entnommen wird. Bei der Konfiguration eines Crawlers geben Sie an, dass Sie Sicherheitstoken verwenden möchten, um die Anzahl Benutzer, die auf die vom Crawler durchsuchten Dokumente zugreifen können, einzugrenzen.

Wenn ein Objektgruppenadministrator einen Crawler konfiguriert, kann er Sicherheitsoptionen für einzelne Tabellen, Dateisysteme usw. angeben (d. h., dass für unterschiedliche Datenquellen verschiedene Sicherheitsregeln im Crawlerbereich definiert werden können). Der Administrator kann Folgendes tun:

- Angeben, dass die Dokumente allgemein zugänglich sind (alle Benutzer können die Dokumente durchsuchen)
- Zuordnen von benutzerdefinierten Sicherheitstoken zu jedem Dokument
- Extrahieren von Sicherheitstoken aus einem Feld in den durchsuchten Daten und Zuordnen der extrahierten Token zu jedem Dokument

Sicherheitstoken sind (mit Ausnahme der allgemein zugänglichen Standardtoken) vollständig benutzerdefiniert. Ein Sicherheitstoken kann durch eine Benutzer-ID, eine Gruppen-ID, eine Benutzerrolle oder einen beliebigen anderen Wert dargestellt werden, den sie als gültig für die Datenquelle festlegen.

Ein Administrator kann z. B. angeben, dass das Feld `hrDeptName` verwendet werden soll, um den Zugriff auf Dokumente zu steuern, die von einem Notes-Crawler durchsucht werden. Der Administrator kann außerdem angeben, dass bei Fehlen dieses Felds in einem Dokument oder bei Nichtangabe von Sicherheitsdaten in diesem Feld zwei benutzerdefinierte Token, `hrgroup1` und `hrgroup2`, zur Steuerung des Dokumentenzugriffs verwendet werden sollen.

Die Sicherheitstoken werden dem Crawler über die Konfigurationsdatei des Crawlers verfügbar gemacht. Für jedes Dokument stellt der Crawler den Wert des Sicherheitstokens als Metadaten bereit. Die indexierende Komponente liest das Sicherheitstoken und wendet es auf die Übergabeinformationen für das Dokument im Index an. Wenn der Administrator für die native Datenquelle die Zugriffssteuerungsliste aktualisiert, werden die aktualisierten Sicherheitsmaßnahmen verfügbar, wenn der nächste Haupt- oder Deltaindex erstellt wird.

Sie können benutzerdefinierte Geschäftsregeln anwenden, um den Wert der Sicherheitstoken zu ermitteln, indem Sie die Regeln in einer Java-Klasse codieren. Bei der Konfiguration der Crawlermerkmale geben Sie den Namen des Plug-ins an, das

der Crawler beim Durchsuchen von Dokumenten verwenden soll. Die Sicherheitstoken, die Ihr Plug-in hinzufügt, werden im Index gespeichert und können für die Steuerung des Zugriffs auf Dokumente verwendet werden.

## Verwendung von Sicherheitstoken durch Suchanwendungen

Es ist Aufgabe der Clientsuchanwendung, während des Suchvorgangs die Sicherheitstoken bereitzustellen, damit die Dokumente entsprechend gefiltert werden können. Wird kein Sicherheitstoken bereitgestellt, wird das allgemein zugängliche Standardtoken automatisch während des Suchvorgangs angewendet.

Die Mustersuchanwendung für die Unternehmenssuche veranschaulicht, wie Sie die Sicherheit auf Dokumentenebene implementieren können. Im folgenden Beispiel wird vorausgesetzt, dass der Administrator einer Gruppe von Dokumenten einen Wert für das Sicherheitstoken zugeordnet hat (im Gegensatz zu einem Szenario, in dem das Sicherheitstoken aus einem Feld in einem durchsuchten Dokument extrahiert wird). Die Suchanwendung verwendet die Anmelde-ID des Benutzers, um zu bestimmen, auf welche Dokumente der Benutzer zugreifen kann. Statt der tatsächlichen Benutzer-ID verwendet die Suchanwendung die ID der Gruppe, zu der der Benutzer gehört. Durch Verwendung einer Gruppen-ID als Sicherheitstoken können Benutzer der Gruppe hinzugefügt und aus ihr entfernt werden, ohne dass der Index erneut erstellt werden muss.

Das Sicherheitstoken, das der Administrator einer Gruppe von Dokumenten zugeordnet hat, stellt eine gültige Betriebssystemgruppen-ID dar. Verschiedene Gruppen-IDs werden verschiedenen Dokumenten im Crawlerbereich zugeordnet. Beispiel:

Document1-5: Security token = Group1  
Document6-10: Security token = Group2

## Prüfung aktueller Berechtigungsnachweise während der Abfrageverarbeitung

Wenn die Sicherheit für eine Objektgruppe bei ihrer Erstellung aktiviert wurde, können Sie mit bestimmten Domänentypen die aktuellen Berechtigungsnachweise des Benutzers prüfen, wenn der Benutzer eine Abfrage übergibt.

Bevor die Suchserver auf eine Abfrage antworten, kommunizieren sie mit den nativen Repositories, um die aktuellen Berechtigungen des Benutzers zu prüfen. Dann werden alle Dokumente aus den Suchergebnissen entfernt, für die der Benutzer keine Berechtigung zum Anzeigen hat.

Bei der Konfiguration der folgenden Crawlertypen können Sie eine Option auswählen, mit der Sie die Benutzerberechtigungen prüfen können, indem Sie die Berechtigungsnachweise mit den aktuellen, vom nativen Repository verwalteten Zugriffssteuerungen vergleichen. Nach dem Durchsuchen und Indexieren von Dokumenten werden Benutzer, die sichere Objektgruppen durchsuchen wollen, mithilfe der Identitätsmanagementkomponente für die Unternehmenssuche geprüft.

- Content Edition-Crawler (nur für die Repositorytypen Documentum, FileNet Panagon Content Services, Hummingbird DM, Portal Document Manager und SharePoint)
- DB2 Content Manager-Crawler
- Domino Document Manager-Crawler
- Notes-Crawler

- QuickPlace-Crawler
- Windows-Dateisystem-Crawler

Für die folgenden Crawlertypen können aktuelle Benutzerberechtigungen geprüft werden, wenn Benutzer das Suchportlet in WebSphere Portal verwenden, um Objektgruppen für die Unternehmenssuche zu durchsuchen.

- Web Content Management-Crawler
- WebSphere Portal-Crawler

#### **Zugehörige Konzepte**

„Umsetzung der Sicherheit auf Dokumentenebene für Lotus Domino-Dokumente“ auf Seite 307

„Umsetzung der Sicherheit auf Dokumentenebene für Dokumente des Windows-Dateisystems“ auf Seite 311

#### **Zugehörige Tasks**

„Konfigurieren der Lotus Domino-Trusted Server, um Benutzerberechtigungen zu prüfen“ auf Seite 308

## **Identitätsmanagement für die Unternehmenssuche**

Die Verwaltung mehrerer Benutzerberechtigungen stellt Unternehmen immer wieder vor Probleme. Ein System für die Unternehmenssuche löst diese Probleme durch die Bereitstellung einer optionalen Identitätsmanagementkomponente.

### **Tipp:**

Detaillierte Beispiele für die Konfiguration der Komponente für das Identitätsmanagement für die Unternehmenssuche finden Sie im IBM Redbook *IBM OmniFind Enterprise Edition Configuration and Implementation Scenarios*.

Die in einem Unternehmen vorhandenen Informationen können in verschiedener Form vorliegen. Sie können im gesamten Unternehmen verteilt sein und von der für die Task am besten geeigneten verfügbaren Software verwaltet werden. Beispielsweise verwenden Benutzer des Unternehmens möglicherweise eine SQL-Anwendung, um auf relationale Datenbanken zuzugreifen, oder ein Dokumentverwaltungssystem, um auf Dokumente zuzugreifen, die sie für ihre Arbeit benötigen.

Die Steuerung des Zugriffs auf sensible Informationen in diesen Repositories wird im Allgemeinen von der Verwaltungssoftware umgesetzt. Benutzer identifizieren sich beim Hostsystem mithilfe einer Kombination aus einer Benutzer-ID und einem Kennwort. Nach der Authentifizierung durch das System steuert die Verwaltungssoftware anhand der für den Benutzer definierten Zugriffsberechtigungen, welche Dokumente der Benutzer anzeigen und verwenden darf.

Häufig sind Benutzern für jedes Repository verschiedene Benutzer-IDs und Kennwörter zugeordnet. Ähnlich wie Benutzer aufgefordert werden, sich bei den ursprünglichen unternehmensweiten Repositories zu identifizieren, müssen Benutzer Berechtigungsnachweise angeben, bevor sie Dokumente in einer Objektgruppe für die Unternehmenssuche anzeigen können, für die eine Prüfung der aktuellen Berechtigungsnachweise erforderlich ist. Benutzer mit mehreren Identitäten müssen für jede Identität die entsprechenden Berechtigungsnachweise angeben.

Wenn Sie angeben, dass Sie die Unternehmenssuche für das Identitätsmanagement in der Administrationskonsole verwenden wollen, können die Suchserver die fol-

genden Methoden verwenden, um während der Abfrageverarbeitung die aktuellen Berechtigungsnachweise eines Benutzers zu prüfen:

- Die Suchanwendung kann den Benutzer auffordern, die Berechtigungsnachweise, die er für den Zugriff auf verschiedene Domänen benötigt, in einem Benutzerprofil zu speichern. Das Profil, das verschlüsselt und in einem sicheren Datenspeicher gespeichert ist, ermöglicht dem Benutzer, die sicheren Domänen zu durchsuchen. Wenn für eine Domäne, die eine Prüfung der aktuellen Berechtigungsnachweise erfordert, keine Berechtigungsnachweise angegeben sind, werden Dokumente aus dieser Domäne aus den Suchergebnissen ausgeschlossen.
- Wenn Dokumente in einer Objektgruppe von einem Crawler durchsucht wurden, der SSO-Sicherheit (SSO - Single Sign-on) unterstützt, und Sie angeben, dass SSO-Sicherheit für die Steuerung des Zugriffs auf Dokumente verwendet werden soll, verwendet das System SSO-Sicherheitsmethoden, um Benutzer für die Dauer einer Suchsitzung zu authentifizieren. Der Benutzer braucht kein Profil zu erstellen, das die Berechtigungsnachweise angibt. Er muss auch keine Kombination aus Benutzer-ID und Kennwort angeben, wenn er sichere Domänen durchsucht.

Wenn Benutzer Objektgruppen durchsuchen, für die beim Übergeben einer Abfrage eine Prüfung der aktuellen Berechtigungsnachweise erforderlich ist, kann das System das Profil oder SSO-Sicherheitsmethoden verwenden, um den Zugriff auf Dokumente zu verweigern oder zuzulassen.

## **Abrufen der Gruppeninformationen des Benutzers**

Zum Prüfen der Berechtigungsnachweise eines Benutzers muss das Identitätsmanagement für jede Identität des Benutzers die entsprechenden Gruppeninformationen abrufen und diese Informationen einer USC-Zeichenfolge (User Security Context - Benutzersicherheitskontext) hinzufügen. Mithilfe dieser Gruppeninformationen werden Ergebnisse gefiltert, und zwar in Übereinstimmung mit Zugriffssteuerungsdaten, die im Index für die Unternehmenssuche gespeichert sind, oder in Übereinstimmung mit SSO-Authentifizierungsdaten. Die Identitätsmanagementkomponente führt dies aus, indem sie SSO-Token nutzt oder die Berechtigungsnachweise des Benutzers verwendet, um eine Verbindung zum Back-End-System herzustellen und die Gruppen anzufordern, zu denen der Benutzer gehört.

Wenn Sie Identitätsmanagementoptionen in der Administrationskonsole konfigurieren, können Sie angeben, wie oft diese Gruppeninformationen aktualisiert werden sollen. Sie können bei jeder Anmeldung des Benutzers an der Suchanwendung neue Gruppendaten extrahieren. Sie können die Gruppendaten auch regelmäßig extrahieren, z. B. jeden dritten Tag.

## **Sicherheit ohne Identitätsmanagementkomponente**

Nicht alle Unternehmen wollen die verschiedenen Identitäten ihrer Benutzergruppen mit der Identitätsmanagementkomponente für die Unternehmenssuche verwalten. Wenn Sie die Identitätsmanagementkomponente in der Administrationskonsole für die Unternehmenssuche inaktivieren, müssen Ihre Suchanwendungen die Zeichenfolge für den Benutzersicherheitskontext generieren. Nach der Generierung wird die USC-Zeichenfolge verwendet, um für jede Abfrage den Wert für Einschränkungen der Zugriffssteuerungsliste festzulegen. Beispiel:

```
Query q = factory.createQuery("IBM");
q.setACLConstraints("User's Security Context in XML");
```

**Tipp:** Eine Erweiterung von Search and Index API (SI-API) stellt eine programmorientierte Steuerung für die Identitätsmanagementdatenbank bereit, mit deren Hilfe Sie eine eigene Identitätsmanagementfunktion schreiben können. Diese Anwendungsprogrammierschnittstelle (API) ermöglicht das Generieren des Benutzersicherheitskontexts mit Java-Objekten. Die XML-Zeichenfolge wird dann automatisch erstellt.

Die XML-Abfragezeichenfolge muss das folgende Format aufweisen. Dabei enthält ... die vollständig formatierte XML-Zeichenfolge:

```
@SecurityContext::'...'
```

Das Format der XML-Zeichenfolge ist wie folgt:

```
<identities id="login_UserName">
 <ssoToken>token_value</ssoToken>
 <identity id="security_domain">
 <type>Notes</type>
 <username>domain_UserName</username>
 <password encrypt="no">domain_userPW</password>
 <groups>
 <group id="g1" />
 <group id="g2" />
 </groups>
 <properties>
 <property name="property_name">property_value</property>
 ...
 </properties>
 </identity>
 ...
</identities>
```

#### **identities**

Der Wert des Attributs `id` ist die Benutzer-ID, die der Benutzer angibt, wenn er sich am System anmeldet.

#### **ssoToken**

Optional: Gibt das LTPA-Token (LTPA - Lightweight Third-Party Authentication) an, das für den Benutzer für die Dauer der Browsersitzung erstellt wird. Dieser Parameter wird nur verwendet, wenn die Zieldomäne für SSO aktiviert ist und der Crawler für die Verwendung der SSO-Sicherheit konfiguriert wurde.

#### **identity**

Enthält die Berechtigungsnachweise des Benutzers für eine bestimmte Datenquelle. Der Wert des Attributs `id` ist die Domäne, in der die Berechtigungsinformationen des Benutzers gespeichert sind (im Fall von Domino ist dies der Domino-Domänenname).

#### **type**

Gibt den Datentyp an und entspricht dem Crawlertyp (Notes, DB2, Exchange Server und so weiter).

#### **username**

Gibt den Benutzernamen an, der für das Durchsuchen der Domäne verwendet werden soll.

#### **password**

Gibt das Kennwort für den angegebenen Benutzernamen an. Das Attribut `encrypt` muss auf `no` gesetzt sein. (Die Unternehmenssuche stellt kein Verschlüsselungsverfahren außerhalb der Identitätsmanagementkomponente bereit.)

**groups**

Gibt die Namen der Gruppen an, zu denen der Benutzer gehört. Für jeden Gruppenname wird ein separates Gruppenelement verwendet.

**properties**

Gibt eine Liste von verbindungs-spezifischen Merkmalen an, wie z. B. die Administrator-ID und das verschlüsselte Kennwort, die für die Erstellung des Crawlers verwendet wurden, oder die Angabe, ob SSO für die Quelle aktiviert ist.

**property\_name**

Der Name des Merkmals.

**property\_value**

Der Wert des Merkmals.

**Benutzervalidierung mit Benutzerprofilen**

Suchanwendungen können Benutzer auffordern, die Berechtigungsnachweise, die sie für den Zugriff auf verschiedene Domänen benötigen, in einem Benutzerprofil zu speichern.

Benutzer müssen der Suchanwendung die Berechtigungsnachweise bereitstellen, die sie zur Anmeldung an der Domäne verwenden, um eine Domäne zu durchsuchen, die Benutzerberechtigungen erforderlich macht, die beim Übergeben einer Abfrage geprüft werden sollen. Mit dem Identitätsmanagement für die Unternehmenssuche können Benutzer Berechtigungsnachweise für eine beliebige Anzahl Domänen in einem Benutzerprofil speichern. Die Berechtigungsnachweise werden verschlüsselt und sicher im System für die Unternehmenssuche gespeichert.

Wenn für eine Domäne, die eine Prüfung der aktuellen Berechtigungsnachweise erfordert, keine Berechtigungsnachweise angegeben sind, werden Dokumente aus dieser Domäne aus den Suchergebnissen ausgeschlossen.

Benutzer können ein Benutzerprofil erstellen und ihre Berechtigungsnachweise registrieren, während sie eine Suchanwendung verwenden. In der Mustersuchanwendung für die Unternehmenssuche wird diese Funktion über die Option **Mein Profil** bereitgestellt. In Ihren benutzerdefinierten Suchanwendungen ist diese Funktion möglicherweise anders implementiert.

Objektgruppen können Dokumente aus vielen verschiedenen Quellentypen enthalten. Eine Objektgruppe kann beispielsweise Dokumente enthalten, die bei einer Crawlersuche in einem Windows-Dateisystem und mehreren Lotus Notes-Datenbanken gefunden wurden. Die Identitätsmanagementkomponente unterscheidet zwischen verschiedenen Quellentypen und fordert nur zur Eingabe von Berechtigungsnachweisen auf, die für den Zugriff auf Domänen benötigt werden, für die eine Validierung erforderlich ist.

Standardmäßig ist jeder Berechtigungsnachweis für die Suche aktiviert und macht es somit erforderlich, dass der Benutzer die Benutzer-ID und das Kennwort angeben muss, die den sicheren Domänen jeweils entsprechen. Hat der Benutzer die Benutzer-ID oder das Kennwort für eine bestimmte Domäne vergessen, kann die Domäne durch Inaktivieren des Markierungsfelds für die Suche inaktiviert werden. Durch das Inaktivieren einer Domäne wird verhindert, dass sichere Dokumente in dieser Domäne in einer Ergebnismenge zurückgegeben werden.

Nach der Erstellung eines Profils kann der Benutzer eine Suchanforderung übergeben. Die Identitätsmanagementkomponente hat die erforderlichen Informationen, um die USC-Zeichenfolge (User's Security Context - Benutzersicherheitskontext) zu

erstellen, die in den nachfolgenden Suchanforderungen verwendet werden soll. Wenn Sie die Identitätsmanagementkomponente nicht verwenden, muss die Suchanwendung die USC-Zeichenfolge bereitstellen, wenn Anwender Domänen abfragen, für die die Prüfung aktueller Berechtigungsnachweise erforderlich ist.

Wenn der Benutzer das nächste Mal versucht, Objektgruppen für die Unternehmenssuche zu durchsuchen, wiederholt die Identitätsmanagementkomponente den Prüfprozess für die Berechtigungsnachweise, kann jetzt jedoch das Profil des Benutzers abrufen. Hat es keine Änderungen gegeben, wird der Benutzer automatisch dort positioniert, wo Suchanforderungen übergeben werden können, und wird nicht aufgefordert, ein Profil zu erstellen.

Erkennt die Identitätsmanagementkomponente eine Änderung in einem Berechtigungsnachweis des Benutzers, wird dem Benutzer automatisch die Profilseite angezeigt, wenn auf die Suchanwendung zugegriffen wird. Dies geschieht beispielsweise, wenn bei einer der für die Suche aktivierten Domänen ein Kennwort geändert wurde oder wenn eine Domäne, für die eine Authentifizierung erforderlich ist, einer Objektgruppe hinzugefügt wird.

Benutzer können die Empfehlung zur Aktualisierung des Profils ignorieren. Dies führt jedoch dazu, dass die entsprechenden Dokumente aus den Suchergebnissen ausgeschlossen werden.

In der mit der Unternehmenssuche bereitgestellten Mustersuchanwendung können Benutzer jederzeit Profile aktualisieren, indem Sie **Mein Profil** in der Funktionsleiste auswählen.

### **Benutzervalidierung mit SSO-Sicherheit**

Wenn Dokumente in einer Objektgruppe von einem Crawler durchsucht wurden, der SSO-Sicherheit (SSO - Single Sign-on) unterstützt, können Sie beim Konfigurieren der Identitätsmanagementoptionen angeben, dass SSO-Sicherheit für die Steuerung des Zugriffs auf Dokumente verwendet werden soll.

### **Aktivieren von SSO**

Durch die SSO-Authentifizierung (SSO - Single Sign-on) erhält ein Benutzer die Möglichkeit zu einer einmaligen Authentifizierung und zum Zugriff auf eine Vielzahl von Ressourcen, ohne erneut zur Eingabe von Berechtigungsnachweisen aufgefordert zu werden. In einem System für die Unternehmenssuche verringert die SSO-Authentifizierung den Aufwand für die Verwaltung der vielen Benutzernamen und Kennwörter, die Benutzer angeben müssen, um auf Dokumente in sicheren Objektgruppen zuzugreifen.

IBM WebSphere Application Server und Lotus Domino unterstützen eine Form von Single Sign-on, die als Lightweight Third-Party Authentication (LTPA) bezeichnet wird. Versucht ein Benutzer, auf eines der beiden Produkte zuzugreifen, wird er aufgefordert, sich mit einem Benutzernamen und einem Kennwort zu authentifizieren. Dieser Benutzername und dieses Kennwort werden anhand eines LDAP-Repositorys geprüft, das von beiden Produkten gemeinsam genutzt wird. Nach der Authentifizierung des Benutzers wird ein Sitzungscookie erstellt, das das LTPA-Token enthalten soll. Der Benutzer kann dann auf andere Ressourcen auf jedem Server zugreifen, der dieselbe Authentifizierungskonfiguration aufweist, ohne erneut zur Angabe von Berechtigungsnachweisen aufgefordert zu werden. Dieses Token besteht so lange, wie die Browsersitzung gültig ist.

Gehen Sie wie folgt vor, um die SSO-Unterstützung zur Verwendung mit Objektgruppen für die Unternehmenssuche zu aktivieren:

- Stellen Sie sicher, dass die globale Sicherheit von WebSphere Application Server und eine gültige LDAP-Registry auf den Suchservern für die Unternehmenssuche aktiviert sind. Die LDAP-Registry kann jedes gültige LDAP-Produkt sein, das von WebSphere Application Server unterstützt wird.
- Stellen Sie sicher, dass das WebSphere-Authentifizierungsverfahren so konfiguriert ist, dass es ein aktives Authentifizierungsverfahren von Lightweight Third Party Authentication (LTPA) verwendet. Wenn Sie LTPA konfigurieren, geben Sie einen gültigen, aber flexiblen Domänennamen ein, wie z. B. ihr.server.com.
- Stellen Sie sicher, dass der LTPA-Schlüssel aus WebSphere Application Server exportiert und in andere Produkte in derselben Domäne importiert wurde, für die die Unterstützung für LTPA aktiviert werden soll.

Nachdem Sie mithilfe eines Browsers geprüft haben, dass die oben genannte Sicherheitskonfiguration korrekt arbeitet, können Sie mithilfe der Administrationskonsole für die Unternehmenssuche Crawler konfigurieren, die die SSO-Authentifizierung unterstützen.

## **Single Sign-on (SSO) und Identitätsmanagement**

Wenn Benutzer Objektgruppen durchsuchen, für die eine Prüfung der aktuellen Berechtigungsnachweise erforderlich ist, kann das System die Methoden der SSO-Sicherheit verwenden, um den Zugriff auf Dokumente zu verweigern oder zuzulassen. Benutzer werden nicht zur Eingabe von Berechtigungsnachweisen aufgefordert, wenn sie Quellen durchsuchen, die die SSO-Authentifizierung unterstützen. Die Identitätsmanagementkomponente wird verwendet, wenn alle folgenden Bedingungen zutreffen:

- Single Sign-on (SSO) ist in WebSphere Application Server und in den Ziel-domänen korrekt aktiviert.
- Die Sicherheit ist in mindestens einer der Objektgruppen aktiviert, die die Suchanwendung durchsuchen kann.
- Die Optionen für die Verwendung der Identitätsmanagementkomponente und der SSO-Sicherheit sind in der Administrationskonsole für die Unternehmenssuche aktiviert.
- Die Option für die Verwendung der SSO-Sicherheit und die Optionen für die Umsetzung der Sicherheit auf Dokumentenebene (wie z. B. das Indexieren von Zugriffssteuerungen oder das Prüfen aktueller Berechtigungsnachweise während der Abfrageverarbeitung) wurden ausgewählt, als die folgenden Crawlertypen konfiguriert wurden:
  - Content Edition (nur für Repositorys von Portal Document Manager verfügbar)
  - Domino Document Manager (nur für Crawler verfügbar, die das DIIOP-Protokoll verwenden)
  - Notes (nur für Crawler verfügbar, die das DIIOP-Protokoll verwenden)
  - QuickPlace (nur für Crawler verfügbar, die das DIIOP-Protokoll verwenden)

## **Sicherheit für das Suchportlet**

Wenn Benutzer das Suchportlet für die Unternehmenssuche verwenden, um Objektgruppen über WebSphere Portal zu durchsuchen, wird die Sicherheit außerdem für Dokumente bereitgestellt, die mithilfe des Crawlers für die Einstiegspunktliste, des Web Content Management-Crawlers und des WebSphere Portal-Crawlers durchsucht werden.

Die sichere Suche wird für diese Quellentypen nur unterstützt, wenn Sie das Portlet verwenden, statt einer Suchanwendung, die außerhalb von WebSphere Portal ausgeführt wird. Innerhalb von WebSphere Portal werden die Benutzerberechtigungen über das Suchportlet abgefragt. Nachdem sich ein Benutzer an WebSphere Portal angemeldet hat, enthalten alle Suchanforderungen die Sicherheitsdaten des Benutzers (Benutzername, Gruppenzugehörigkeit usw.). Da diese Informationen immer verfügbar sind, ist keine LTPA-Token-basierte Einzelmeldung erforderlich.

Wenn Sie jedoch ein anderes Produkt verwenden, um Sites und Dokumente auf einem WebSphere Portal-Server zu schützen, müssen Sie SSO-Optionen bei der Konfiguration des Crawlers angeben. Wenn Sie z. B. ein Produkt wie IBM Tivoli Access Manager WebSEAL oder CA SiteMinder SSO Agent for PeopleSoft verwenden, müssen Sie Berechtigungsnachweise angeben, mit denen der Crawler auf Dokumente auf dem Server über Single Sign-on zugreifen kann. In diesem Fall aktivieren Sie SSO für den Crawlerzugriff auf geschützte Inhalte, Sie aktivieren SSO dabei nicht für die sichere Suche.

## Konfigurieren des Identitätsmanagements

Mithilfe der Identitätsmanagementkomponente für die Unternehmenssuche können Sie angeben, wie Benutzerberechtigungen während der Abfrageverarbeitung geprüft werden sollen.

### Vorbereitung

Sie müssen ein Administrator für die Unternehmenssuche sein, um Identitätsmanagementoptionen zu konfigurieren.

### Informationen zu dieser Task

Wenn Benutzer Objektgruppen durchsuchen, für die eine Prüfung der aktuellen Berechtigungsnachweise während der Abfrageverarbeitung erforderlich ist, kann die Identitätsmanagementkomponente für die Unternehmenssuche ein Benutzerprofil oder SSO-Sicherheitsmethoden verwenden, um den Zugriff auf Dokumente zu verweigern oder zuzulassen.

### Vorgehensweise

Gehen Sie wie folgt vor, um das Identitätsmanagement zu konfigurieren:

1. Klicken Sie **Sicherheit** an, um die Sicht **Sicherheit** zu öffnen.
2. Klicken Sie auf der Seite **Suchanwendungen** die Option **Identitätsmanagement konfigurieren** an.
3. Wählen Sie auf der Seite **Identitätsmanagement konfigurieren** das Markierungsfeld für die Verwendung der Identitätsmanagementkomponente für die Unternehmenssuche aus, um die Prüfung von Benutzerberechtigungen während der Abfrageverarbeitung zu steuern. Ist dieses Markierungsfeld nicht ausgewählt, muss die Suchanwendung die USC-Zeichenfolge (USC - User Security Context) bereitstellen, wenn Benutzer Domänen abfragen, für die eine Prüfung der aktuellen Berechtigungsnachweise während der Abfrageverarbeitung erforderlich ist.
4. Geben Sie an, wie oft die Identitätsmanagementkomponente Benutzerberechtigungen aus Gruppendatensätzen in der Benutzerregistry von WebSphere Application Server extrahieren soll. Sie können die Berechtigungsnachweisdaten jedes Mal aktualisieren, wenn der Benutzer auf die Suchanwendung zugreift, oder nach Ablauf einer angegebenen Anzahl Tage.

5. Geben Sie für die SSO-Authentifizierung unterstützenden Crawlertypen an, ob die Identitätsmanagementkomponente SSO-Sicherheitstoken statt Benutzerprofilen verwenden soll, um Benutzer zu prüfen. Sie können das Markierungsfeld auswählen, um die SSO-Authentifizierung für alle Crawlertypen zu verwenden, oder Sie können die Markierungsfelder für einzelne Crawlertypen auswählen.

**Wichtig:** Die Identitätsmanagementkomponente verwendet SSO-Sicherheitsmechanismen nur dann, wenn die SSO-Sicherheit in WebSphere Application Server und in den Zieldomänen korrekt konfiguriert ist.

## Ankertextanalyse

Wenn Sie die Objektgruppensicherheit aktivieren, wenden die globalen Analyseprozesse bestimmte Regeln für die Indexierung des Ankertextes in Dokumenten an, die von Web-Crawlern durchsucht werden. Wenn Sie die Objektgruppensicherheit nicht aktivieren, können Sie bei der Konfiguration der einzelnen Web-Crawler angeben, ob der Ankertext in Links zu verbotenen Dokumenten indiziert werden soll.

Ankertext ist die Information in einer Hypertextverbindung, die die Seite beschreibt, zu der der Link eine Verbindung herstellt. Beispiel: Im folgenden Link ist der Text Abfragesyntax der Ankertext in einem Link, der zu der Seite `syntax.htm` eine Verbindung herstellt:

```
Abfragesyntax
```

Normalerweise folgt der Web-Crawler Links in Dokumenten, um zusätzliche Dokumente zu durchsuchen, und nimmt diese verlinkten Seiten in den Index auf. Während der globalen Analyse ordnen die Indexprozesse den Ankertext nicht nur dem Dokument zu, in das er eingebettet ist (das Quelldokument), sondern auch dem Zieldokument. Im oben genannten Beispiel ist der Ankertext Abfragesyntax der Zielseite `syntax.htm` zugeordnet sowie der Quellenseite, die das Ankerkonstrukt enthält. Durch diese Zuordnung kann das Zieldokument von einer Abfrage abgerufen werden, in der Text angegeben ist, der in dem Quelldokument erscheint. Diese Zuordnung stellt jedoch ein Sicherheitsrisiko dar, falls es dem Benutzer erlaubt ist, das Zieldokument, aber nicht das Quelldokument anzuzeigen.

Wenn Sie beim Erstellen einer Objektgruppe die Objektgruppensicherheit aktivieren, wird die Ankertextverarbeitung inaktiviert. Der Ankertext wird nicht mehr mit einem Dokument indiziert, es sei denn, er tritt im Dokument oder in den Metadaten des Dokuments auf. Durch diese Sicherheitssteuerung wird sichergestellt, dass Benutzern keine Informationen in Dokumenten angezeigt werden, auf die sie nicht zugreifen dürfen. Ein Dokument wird nur als Suchergebnis zurückgegeben, wenn sein Inhalt oder seine Metadaten mit der Abfrage übereinstimmen.

Die Aktivierung der Objektgruppensicherheit kann die Sicherheit von Webdokumenten verbessern, indem Benutzer nur jene Dokumente durchsuchen können, deren Sicherheitstoken mit ihren Berechtigungsnachweisen übereinstimmen. Wenn jedoch der Ankertext nicht verarbeitet wird, enthalten die Suchergebnisse möglicherweise nicht sämtliche Dokumente, die für eine Suchabfrage relevant sein könnten.

Wenn Sie die Objektgruppensicherheit nicht aktivieren, können Sie bei der Konfiguration erweiterter Web-Crawler-Merkmale angeben, ob der Ankertext in Links zu verbotenen Dokumenten indexiert werden soll.

### **Indexieren von Ankertext in Links zu verbotenen Dokumenten**

Wenn ein Dokument Links zu Dokumenten enthält, die der Web-Crawler nicht durchsuchen darf, können Sie bei der Konfiguration eines Web-Crawlers angeben, ob der Ankertext für diese Links im Index beibehalten werden soll.

#### **Vorbereitung**

Sie müssen die Rolle eines Administrators für die Unternehmenssuche haben oder ein Objektgruppenadministrator für den zu konfigurierenden Web-Crawler sein, um Optionen für das Indexieren von Ankertext konfigurieren zu können.

#### **Informationen zu dieser Task**

Steueranweisungen in einer Datei `robots.txt` oder in den Metadaten von Webdokumenten können den Web-Crawler am Zugriff auf Dokumente auf einer Website hindern. Wenn ein Dokument, das der Web-Crawler durchsuchen darf, Links zu verbotenen Dokumenten enthält, können Sie angeben, wie der Ankertext für diese Links behandelt werden soll.

Sie können bei der Konfiguration des Web-Crawlers angeben, ob der Ankertext zu verbotenen Dokumenten indexiert werden soll. Maximale Sicherheit erreichen Sie, wenn Sie angeben, dass der Ankertext in Links zu verbotenen Dokumenten nicht indexiert werden soll. Falls der Ankertext nicht indexiert wird, umfassen die Suchergebnisse möglicherweise jedoch nicht alle für eine Abfrage potenziell relevanten Dokumente.

#### **Vorgehensweise**

Gehen Sie wie folgt vor, um das Indexieren von Ankertext in Links zu verbotenen Dokumenten zu aktivieren bzw. zu inaktivieren:

1. Bearbeiten Sie eine Objektgruppe, suchen Sie dann auf der Seite **Crawler** den Web-Crawler, der konfiguriert werden soll, und klicken Sie  **Crawlermerkmale** an.
2. Klicken Sie **Erweiterte Web-Crawler-Merkmale bearbeiten** an.
3. Wählen Sie das Markierungsfeld **Ankertext in Links zu verbotenen Dokumenten indexieren** aus, um den Ankertext in allen Dokumenten, die dieser Crawler durchsucht, zu indexieren. Benutzer erfahren von Seiten, die der Web-Crawler nicht durchsuchen darf, indem sie nach Text suchen, der im Ankertext der Links enthalten ist, die auf diese Seiten verweisen.

Entfernen Sie die Markierung aus diesem Markierungsfeld, um Ankertext in Links zu verbotenen Dokumenten aus dem Index auszuschließen. Benutzer können dann nicht von Seiten erfahren, die der Web-Crawler nicht durchsuchen darf. Der Ankertext wird zusammen mit den verbotenen Dokumenten aus dem Index ausgeschlossen.

4. Klicken Sie **OK** an, und klicken Sie dann auf der Seite **Web-Crawler-Merkmale** wieder **OK** an.
5. Damit die Änderungen wirksam werden, stoppen Sie den Crawler, und starten Sie ihn erneut.

Für bereits indexierte Dokumente werden die Änderungen erst übernommen, wenn die Dokumente erneut vom Crawler durchsucht wurden, sodass sie erneut

indexiert werden können. Wenn bei einer vorherigen Crawlersuche Informationen zu verbotenen Dokumenten dem Index hinzugefügt wurden, werden diese Informationen anschließend aus dem Index entfernt.

---

## Aktivieren der Sicherheit für die Unternehmenssuche

Wenn Sie die Sicherheit umsetzen wollen, wenn Benutzer ein System für die Unternehmenssuche verwalten oder durchsuchen, müssen Sie die globale Sicherheit in WebSphere Application Server konfigurieren. Außerdem müssen Sie Sicherheitsinformationen in den Konfigurationsdateien für die Unternehmenssuche, in der Administrationskonsole und in den Suchanwendungen konfigurieren.

### Tipp:

Detaillierte Beispiele für die Aktivierung der globalen Sicherheit in WebSphere Application Server mit einem LDAP-Repository, einschließlich Beispielen zur Konfiguration verschiedener Crawler, die gewährleisten, dass die Sicherheit auf Dokumentenebene erzwungen wird, finden Sie im IBM Redbook *IBM OmniFind Enterprise Edition Configuration and Implementation Scenarios*.

### Vorgehensweise

Gehen Sie wie folgt vor, um die Sicherheit für ein System für die Unternehmenssuche zu aktivieren:

1. Wählen Sie, welcher Benutzerregistrytyp für die Authentifizierung von Benutzern verwendet werden soll. Viele WebSphere Application Server-Administratoren entscheiden sich beispielsweise für die Verwendung der LDAP-Benutzerregistry (LDAP - Lightweight Directory Access Protocol).
2. Wählen Sie **Sicherheit** in der Administrationskonsole für die Unternehmenssuche aus, und ordnen Sie mindestens einem Benutzer in der Benutzerregistry von WebSphere Application Server die Verwaltungsrolle **Administrator für die Unternehmenssuche** zu.

Alternativ können Sie der Benutzerregistry von WebSphere Application Server die Administrator-ID für die Unternehmenssuche hinzufügen, die bei der Installation von OmniFind Enterprise Edition angegeben wird.

**Wichtig:** Nach Aktivierung der globalen Sicherheit können nur Benutzer-IDs, die sich in der Benutzerregistry befinden und denen eine Verwaltungsrolle für die Unternehmenssuche zugeordnet wurde, auf die Administrationskonsole zugreifen und die Unternehmenssuche verwalten.

3. Folgen Sie der Vorgehensweise in „Konfigurieren der globalen Sicherheit und einer LDAP-Benutzerregistry in WebSphere Application Server“ auf Seite 298, um die globale Sicherheit zu aktivieren und die Benutzerregistry zu konfigurieren.
4. Wenn Sie nach der Installation von OmniFind Enterprise Edition die globale Sicherheit aktivieren, müssen Sie für das System für die Unternehmenssuche die Benutzer-ID und das Kennwort für WebSphere Application Server bereitstellen. Zum Bereitstellen dieser Informationen verwenden Sie den Befehl **eschangewaspw**:
  - Wenn Sie OmniFind Enterprise Edition auf einem einzelnen Server installiert haben, folgen Sie der Vorgehensweise in „Aktivieren der Sicherheit für ein System für die Unternehmenssuche mit einem Server“ auf Seite 300.

- Wenn Sie OmniFind Enterprise Edition auf mehreren Servern installiert haben, folgen Sie der Vorgehensweise in „Aktivieren der Sicherheit für ein System für die Unternehmenssuche mit mehreren Servern“ auf Seite 301.
5. Führen Sie die Tasks aus, die den Dokumenttypen entsprechen, die Sie durchsuchen wollen. Informationen hierzu finden Sie in „Anforderungen der Crawlerkonfiguration zur Unterstützung der Sicherheit“ auf Seite 302.

## Konfigurieren der globalen Sicherheit und einer LDAP-Benutzerregistry in WebSphere Application Server

Zum Aktivieren der Sicherheit in OmniFind Enterprise Edition müssen Sie zuerst die globale Sicherheit in WebSphere Application Server aktivieren.

### Informationen zu dieser Task

Ein Schritt, den Sie zum Aktivieren der globalen Sicherheit ausführen müssen, ist das Konfigurieren einer Benutzerregistry zum Authentifizieren von Benutzer-IDs. In dieser Task wird erläutert, wie Sie eine LDAP-Benutzerregistry (LDAP - Lightweight Directory Access Protocol) in WebSphere Application Server gleichzeitig konfigurieren, während Sie die globale Sicherheit aktivieren. Obwohl andere Benutzerregistrytypen von WebSphere Application Server unterstützt werden, können Sie nicht die Registry des lokalen Betriebssystems verwenden, um Benutzer der Unternehmenssuche zu authentifizieren. Zur Verwendung der Registry des lokalen Betriebssystems müssen auf den Suchservern für die Unternehmenssuche für jeden Benutzer in Ihrem Unternehmen Betriebssystembenutzerkonten vorhanden sein.

Diese Task basiert auf WebSphere Application Server Version 6. Wenn Sie eine frühere Version von WebSphere Application Server verwenden, unterscheiden sich möglicherweise die Standardpfade und einige Bezeichnungen auf der Benutzeroberfläche. Diese Task verwendet außerdem IBM Tivoli Directory Server für die LDAP-Registry. Wenn Sie einen anderen Registrytyp oder eine benutzerdefinierte Registry verwenden, müssen Sie die für Ihre Registry geeigneten Informationen bereitstellen.

In dieser Task werden die Schritte zusammengefasst, die erforderlich sind, um die globale Sicherheit für die Verwendung mit einem System für die Unternehmenssuche zu konfigurieren. Ausführliche Informationen hierzu finden Sie in der Informationszentrale von *WebSphere Application Server Version 6.0.x* unter der folgenden URL-Adresse: <http://publib.boulder.ibm.com/infocenter/wasinfo/v6r0/index.jsp>

### Vorgehensweise

Gehen Sie wie folgt vor, um die globale Sicherheit in WebSphere Application Server zu aktivieren:

1. Greifen Sie auf dem Suchserver für die Unternehmenssuche auf die folgende URL-Adresse zu, um die Administrationskonsole von WebSphere Application Server zu öffnen, wobei *lokaler\_host* der Wert localhost oder der Servername wie z. B. omnifind.search.xyz.com ist.  
`http://lokaler_host:9060/ibm/console`
2. Klicken Sie **Sicherheit** und dann **Globale Sicherheit** an.
3. Konfigurieren Sie WebSphere für die Verwendung einer LDAP-Registry:
  - a. Klicken Sie **LDAP** unter **Benutzer-Registries** an.

- b. Geben Sie die Serverbenutzer-ID und das Kennwort an, die für das Ausführen des Anwendungsservers verwendet werden.
  - c. Wählen Sie IBM Tivoli Directory Server als Registrytyp aus.
  - d. Geben Sie den Hostnamen des LDAP-Servers an, entweder eine IP-Adresse oder einen DNS-Hostnamen (DNS - Domain Name Service). Die Standardportnummer ist 389.
  - e. Geben Sie den Basis-DN (Distinguished Name - definierter Name) ein, der der Ausgangspunkt für die Suche in der Registry ist, wie z. B. ou=sales,o=ibm,c=us.
  - f. Da einige LDAP-Server keine anonymen Bindungen bei der Suche in der Registry unterstützen, geben Sie den definierten Namen für den Anwendungsserver an, z. B. cn=searchuser,o=ibm,c=us, sowie das Kennwort für den Anwendungsserver. Der Anwendungsserver verwendet diesen definierten Namen (DN) und das Kennwort für die Bindung an die Registry.
  - g. Wählen Sie das Markierungsfeld **SSL aktiviert** aus, um die SSL-Kommunikation (SSL - Secure Sockets Layer) zwischen WebSphere und LDAP (Lightweight Directory Access Protocol) zu verwenden.
  - h. Klicken Sie **Anwenden** und dann **OK** an.
4. Wählen Sie die Markierungsfelder **Globale Sicherheit aktivieren** und **Java-2-Sicherheit erzwingen** unter **Allgemeine Merkmale** aus.
  5. Wählen Sie Simple WebSphere Authentication Mechanism (SWAM) als aktives Authentifizierungsverfahren aus.
  6. Wählen Sie die LDAP-Benutzerregistry (LDAP - Lightweight Directory Access Protocol) als aktive Benutzerregistry aus.
  7. Klicken Sie **OK** an.
  8. Klicken Sie den Link **Speichern** oben auf der Seite an. Wenn Sie aufgefordert werden, Ihre Änderungen zu speichern, klicken Sie den Knopf **Speichern** an.
  9. Klicken Sie **Abmelden** auf der Funktionsleiste an.
  10. Stoppen Sie die Anwendung "ESSearchServer", und starten Sie sie erneut.

#### AIX, Linux oder Solaris

```
./stopServer.sh ESSearchServer
./startServer.sh ESSearchServer
```

#### Windows

```
stopServer ESSearchServer
startServer ESSearchServer
```

Diese Scripts befinden sich im Verzeichnis was-installationsverzeichnis/AppServer/bin.

- Bei WebSphere Application Server Version 5 ist der Standardinstallationspfad /usr/WebSphere auf AIX-Systemen, /opt/WebSphere auf Linux- oder Solaris-Systemen oder C:\Programme\WebSphere auf Windows-Systemen.
  - Bei WebSphere Application Server Version 6 ist der Standardinstallationspfad /usr/IBM/WebSphere auf AIX-Systemen, /opt/IBM/WebSphere auf Linux- oder Solaris-Systemen oder C:\Programme\IBM\WebSphere auf Windows-Systemen.
11. Starten Sie die Administrationskonsole von WebSphere Application Server erneut.

12. Der Server wird jetzt im sicheren Modus gestartet. Geben Sie daher die Serverbenutzer-ID und das Kennwort ein, die Sie angegeben haben, als Sie die LDAP-Benutzerregistry für die Anmeldung an der Konsole konfiguriert haben (siehe Schritt 3b auf Seite 299).

## Aktivieren der Sicherheit für ein System für die Unternehmenssuche mit einem Server

Wenn Sie die globale WebSphere Application Server-Sicherheit aktivieren, nachdem Sie OmniFind Enterprise Edition installiert haben, müssen Sie den Befehl **eschangewaspw** verwenden, um die Konfigurationsdatei für die Unternehmenssuche (`es.cfg`) mit dem Kennwort des WebSphere Application Server-Benutzers zu aktualisieren.

### Vorbereitung

Stellen Sie sicher, dass die Datei `config.properties` für die Anwendung "ESSearchApplication" einen gültigen Benutzernamen und ein gültiges Kennwort für WebSphere Application Server angibt. Die Standardposition dieser Datei ist `ES_INSTALL_ROOT/installedApps/ESSearchApplication.ear/ESSearchApplication.war/WEB-INF`.

### Informationen zu dieser Task

Der Befehl **eschangewaspw** verschlüsselt das Kennwort, bevor er es in der Datei `es.cfg` speichert.

### Vorgehensweise

Gehen Sie wie folgt vor, um ein vorhandenes System für die Unternehmenssuche mit einem einzelnen Server für die Verwendung der globalen Sicherheit zu aktivieren:

1. Melden Sie sich auf dem Server für die Unternehmenssuche als Administrator für die Unternehmenssuche an, und stoppen Sie das System für die Unternehmenssuche:  
`esadmin system stopall`
2. Stellen Sie sicher, dass der Eintrag `WASUser` in der Datei `ES_NODE_ROOT/nodeinfo/es.cfg` einen gültigen WebSphere Application Server-Benutzernamen angibt.
3. Führen Sie das folgende Script aus, wobei *was-kennwort* das Kennwort des WebSphere Application Server-Benutzernamens ist, der in der Datei `ES_NODE_ROOT/nodeinfo/es.cfg` (siehe Schritt 2) angegeben ist.

#### AIX, Linux oder Solaris

```
eschangewaspw.sh was-kennwort
```

#### Windows

```
eschangewaspw was-kennwort
```

4. Wählen Sie unter Windows **Start** → **Einstellungen** → **Systemsteuerung** → **Verwaltung** → **Dienste** aus, und fügen Sie denselben Benutzernamen und dasselbe Kennwort für WebSphere Application Server den WebSphere Application Server- und ESSearchServer-Services hinzu.
5. Starten Sie das System für die Unternehmenssuche erneut:  
`esadmin system startall`

## Aktivieren der Sicherheit für ein System für die Unternehmenssuche mit mehreren Servern

Wenn Sie die globale WebSphere Application Server-Sicherheit aktivieren, nachdem Sie OmniFind Enterprise Edition installiert haben, müssen Sie den Befehl **eschangewaspw** verwenden, um die Konfigurationsdatei für die Unternehmenssuche (`es.cfg`) mit dem Kennwort des WebSphere Application Server-Benutzers zu aktualisieren.

### Vorbereitung

Stellen Sie sicher, dass die Datei `config.properties` für die Anwendung "ESSearchApplication" einen gültigen Benutzernamen und ein gültiges Kennwort für WebSphere Application Server angibt. Die Standardposition dieser Datei auf den Suchservern ist `ES_INSTALL_ROOT/installedApps/ESSearchApplication.ear/ESSearchApplication.war/WEB-INF`.

### Informationen zu dieser Task

Der Befehl **eschangewaspw** verschlüsselt das Kennwort, bevor er es in der Datei `es.cfg` speichert.

### Vorgehensweise

Gehen Sie wie folgt vor, um ein vorhandenes System für die Unternehmenssuche mit mehreren Servern für die Verwendung der globalen Sicherheit zu aktivieren:

1. Führen Sie die folgenden Schritte auf dem Indexserver für die Unternehmenssuche aus:
  - a. Melden Sie sich als Administrator für die Unternehmenssuche an, und stoppen Sie das System für die Unternehmenssuche:  
`esadmin system stopall`
  - b. Stellen Sie sicher, dass der Eintrag `WASUser` in der Datei `ES_NODE_ROOT/nodeinfo/es.cfg` einen gültigen WebSphere Application Server-Benutzernamen angibt.
  - c. Führen Sie das folgende Script aus, wobei *was-kennwort* das Kennwort des WebSphere Application Server-Benutzers ist, der in der Datei `ES_NODE_ROOT/nodeinfo/es.cfg` (siehe Schritt 1b) angegeben ist.

#### AIX, Linux oder Solaris

```
eschangewaspw.sh was-kennwort
```

#### Windows

```
eschangewaspw was-kennwort
```

2. Führen Sie die folgenden Schritte auf dem zweiten Suchserver (bei einer Konfiguration mit zwei Servern) oder auf dem Crawler-Server und beiden Suchservern (bei einer Konfiguration mit vier Servern) aus:
  - a. Melden Sie sich als Administrator für die Unternehmenssuche an.
  - b. Führen Sie das folgende Script aus, wobei *was-kennwort* das Kennwort des WebSphere Application Server-Benutzers ist, der in der Datei `ES_NODE_ROOT/nodeinfo/es.cfg` (siehe Schritt 1b) angegeben ist.

#### AIX, Linux oder Solaris

```
eschangewaspw.sh was-kennwort
```

#### Windows-Eingabeaufforderung

```
eschangewaspw was-kennwort
```

3. Wählen Sie unter Windows **Start** → **Einstellungen** → **Systemsteuerung** → **Verwaltung** → **Dienste** aus, und fügen Sie denselben Benutzernamen und dasselbe Kennwort für WebSphere Application Server den WebSphere Application Server- und ESSearchServer-Services hinzu.
4. Starten Sie das System für die Unternehmenssuche auf dem Indexserver für die Unternehmenssuche erneut:  

```
esadmin system startall
```

---

## Anforderungen der Crawlerkonfiguration zur Unterstützung der Sicherheit

Die Crawler müssen über eine Zugriffsberechtigung für die nativen Sicherheitsdaten verfügen, um Informationen zu sammeln, die die Umsetzung der Sicherheit auf Dokumentenebene ermöglichen. Für einige Datentypen müssen zusätzliche Schritte ausgeführt werden, um eine sichere Umgebung zu konfigurieren.

*Tabelle 7. Anforderungen der Crawlerkonfiguration zur Unterstützung der Sicherheit*

<b>Content Edition-Crawler</b>	
<p>Bevor Sie einen Crawler erstellen, um auf Repositorys im Direktmodus zuzugreifen, konfigurieren Sie das WebSphere Information Integrator Content Edition-System so, dass es im Direktmodus ausgeführt wird und einen Connector für den Crawler-Server konfiguriert.</p> <p>Bevor Sie einen Crawler erstellen, um auf Repositorys im Servermodus zuzugreifen, führen Sie ein Script aus (<code>escrvbr.sh</code> unter AIX, Linux oder Solaris oder <code>escrvbr.vbs</code> unter Windows), um den Crawler-Server zu konfigurieren.</p> <p>Wenn Sie den Crawler konfigurieren, geben Sie eine Benutzer-ID und ein Kennwort an, die dem Crawler ermöglichen, auf alle zu durchsuchenden Repositorys zuzugreifen. Falls erforderlich, können Sie für jedes Repository im Crawlerbereich eine andere Benutzer-ID und ein anderes Kennwort angeben.</p>	<p>Referenzinformationen:</p> <ul style="list-style-type: none"> <li>• „Zugriff auf Content Edition-Repositorys im Direktmodus“ auf Seite 48</li> <li>• „Zugriff auf WebSphere II Content Edition-Repositorys im Servermodus“ auf Seite 49</li> <li>• „Konfigurieren des Crawler-Servers unter UNIX für WebSphere II Content Edition“ auf Seite 50</li> <li>• „Konfigurieren des Crawler-Servers unter Windows für WebSphere II Content Edition“ auf Seite 51</li> </ul>
<b>DB2-Crawler</b>	
<p>Bevor Sie den Crawler erstellen, führen Sie ein Script aus (<code>escrdb2.sh</code> unter AIX, Linux oder Solaris oder <code>escrdb2.vbs</code> unter Windows), um den Crawler-Server zu konfigurieren.</p> <p>Wenn Sie den Crawler für das Durchsuchen ferner, nicht katalogisierter Datenbanken konfigurieren, geben Sie eine Benutzer-ID und ein Kennwort an, die das Durchsuchen aller Datenbanken auf dem Zieldatenbankserver ermöglichen. Falls erforderlich, können Sie für jede Datenbank im Crawlerbereich eine andere Benutzer-ID und ein anderes Kennwort angeben.</p>	<p>Referenzinformationen:</p> <ul style="list-style-type: none"> <li>• „Konfigurieren des Crawler-Servers unter UNIX für DB2-Crawler“ auf Seite 54</li> <li>• „Konfigurieren des Crawler-Servers unter Windows für DB2-Crawler“ auf Seite 55</li> </ul>
<b>DB2 Content Manager-Crawler</b>	

*Tabelle 7. Anforderungen der Crawlerkonfiguration zur Unterstützung der Sicherheit (Forts.)*

---

Bevor Sie den Crawler erstellen, führen Sie ein Script aus (escrcm.sh unter AIX, Linux oder Solaris oder escrcm.vbs unter Windows), um den Crawler-Server zu konfigurieren.

Wenn Sie den Crawler konfigurieren, geben Sie eine Benutzer-ID und ein Kennwort an, die dem Crawler ermöglichen, auf alle zu durchsuchenden Server zuzugreifen. Falls erforderlich, können Sie für jeden Server im Crawlerbereich eine andere Benutzer-ID und ein anderes Kennwort angeben.

Referenzinformationen:

- „Konfigurieren des Crawler-Servers unter UNIX für DB2 Content Manager-Crawler“ auf Seite 63
- „Konfigurieren des Crawler-Servers unter Windows für DB2 Content Manager-Crawler“ auf Seite 65

---

**Domino Document Manager-, Notes- und QuickPlace-Crawler**

---

*Tabelle 7. Anforderungen der Crawlerkonfiguration zur Unterstützung der Sicherheit (Forts.)*

---

Gehen Sie wie folgt vor, um Lotus Domino-Server zu durchsuchen, die das NRPC-Protokoll verwenden (NRPC - Notes Remote Procedure Call):	Referenzinformationen:
----------------------------------------------------------------------------------------------------------------------------------------	------------------------

- Stellen Sie auf einem AIX-System sicher, dass das I/O Completion Port-Modul auf dem Crawler-Server installiert und verfügbar ist.
- Bevor Sie den Crawler erstellen, führen Sie ein Script aus (`escrnote.sh` unter AIX, Linux oder Solaris oder `escrnote.vbs` unter Windows), um den Crawler-Server zu konfigurieren.
- Ein Domino-Server muss auf dem Crawler-Server für die Unternehmenssuche installiert sein, und dieser Domino-Server muss ein Element der zu durchsuchenden Domino-Domäne sein.
- Der zu durchsuchende Domino-Server muss als Lotus Domino Trusted Server konfiguriert sein, damit aktuelle Benutzerberechtigungen geprüft werden können, wenn ein Benutzer eine Suchanforderung übergibt.
- Wenn Sie den Crawler konfigurieren, geben Sie den Pfad für eine Lotus Notes-Benutzer-ID-Datei an, die für den Zugriff auf den Server berechtigt ist, wie z. B. `c:\Programme\lotus\notes\data\name.id` oder `/local/notesdata/name.id`. Geben Sie außerdem das Kennwort für diese ID-Datei an.

- „Konfigurieren des I/O Completion Ports unter AIX für das Durchsuchen von Lotus Domino-Quellen“ auf Seite 88
- „Konfigurieren des Crawler-Servers unter UNIX für das Durchsuchen von Lotus Domino-Quellen“ auf Seite 83
- „Konfigurieren des Crawler-Servers unter Windows für das Durchsuchen von Lotus Domino-Quellen“ auf Seite 85
- „Konfigurieren der Lotus Domino-Trusted Server, um Benutzerberechtigungen zu prüfen“ auf Seite 308
- „Konfigurieren von Servern, die das DIIOP-Protokoll verwenden“ auf Seite 86
- „Konfigurieren des QuickPlace-Servers für die Verwendung der Sicherheit "Local User" (Lokaler Benutzer)" auf Seite 309
- „Konfigurieren von Directory Assistance auf einem QuickPlace-Server“ auf Seite 310

Gehen Sie wie folgt vor, um Lotus Domino-Server zu durchsuchen, die das DIIOP-Protokoll verwenden (DIIOP - Domino Internet Inter-ORB Protocol):

- Stellen Sie auf einem AIX-System sicher, dass das I/O Completion Port-Modul auf dem Crawler-Server installiert und verfügbar ist.
- Konfigurieren Sie den Crawler-Server so, dass er das Protokoll verwenden kann.
- Wenn Sie den Crawler konfigurieren, geben Sie eine vollständig qualifizierte Lotus Notes-Benutzer-ID an, die für den Zugriff auf den Server berechtigt ist, wie z. B. `User Name/Any Town/My Company`. Geben Sie außerdem das Kennwort für diese Benutzer-ID an.

Zum Durchsuchen von QuickPlace-Servern müssen Sie den QuickPlace-Server so konfigurieren, dass er, abhängig vom Typ der zu verwendenden Sicherheit, die Sicherheit für lokale Benutzer oder die Verzeichnisverwaltung unterstützt.

---

## **Exchange Server-Crawler**

---

*Tabelle 7. Anforderungen der Crawlerkonfiguration zur Unterstützung der Sicherheit (Forts.)*

---

Wenn Sie den Crawler konfigurieren, geben Sie eine Benutzer-ID an, die für den Zugriff auf allgemein zugängliche Ordner auf dem zu durchsuchenden Exchange-Server berechtigt ist. Geben Sie außerdem das Kennwort für diese Benutzer-ID an.

Referenzinformationen:

- „Zugriffsprüfung zum Schutz von Exchange Server-Dokumenten“ auf Seite 307

Damit der Crawler beim Durchsuchen von Daten das Exchange Server-Schlüsselmanagement und das SSL-Protokoll (SSL - Secure Sockets Layer) verwenden kann, geben Sie außerdem den vollständig qualifizierten Pfad zur Schlüsselspeicherdatei und ein Kennwort an, das dem Crawler den Zugriff auf diese Datei ermöglicht. Die Schlüsselspeicherdatei muss auf dem Crawler-Server für die Unternehmenssuche vorhanden sein.

---

#### **JDBC-Datenbankcrawler**

Wenn Sie den Crawler konfigurieren, können Sie eine Benutzer-ID und ein Kennwort angeben, die das Durchsuchen von Tabellen in der Zieldatenbank ermöglichen. Falls erforderlich, können Sie für jede Datenbank im Crawlerbereich eine andere Benutzer-ID und ein anderes Kennwort angeben.

---

#### **NNTP-Crawler**

Die zu durchsuchenden NNTP-Server müssen dem Crawler-Server das Lesen von Daten erlauben.

---

#### **UNIX-Dateisystem-Crawler**

Die zu durchsuchenden AIX-, Linux- und Solaris-Unterverzeichnisse müssen dem Crawler-Server das Lesen von Daten erlauben.

---

#### **Web-Crawler**

---

*Tabelle 7. Anforderungen der Crawlerkonfiguration zur Unterstützung der Sicherheit (Forts.)*

---

<p>Der Web-Crawler hält das Protokoll zum Sperren von Websitebereichen für Robots ein. Wenn ein Web-Server eine Datei robots.txt auf der obersten Ebene des Serververzeichnis enthält, analysiert der Crawler die Datei und durchsucht nur dann Websites auf diesem Server, wenn er dazu berechtigt ist. Informationen zu diesem Protokoll finden Sie unter <a href="http://www.robotstxt.org/wc/exclusion.html">http://www.robotstxt.org/wc/exclusion.html</a>.</p>	<p>Referenzinformationen:</p> <ul style="list-style-type: none"><li>• „Durch HTTP-Basisauthentifizierung geschützte Websites“ auf Seite 109</li><li>• „Durch formularbasierte Authentifizierung geschützte Websites“ auf Seite 110</li></ul>
----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Beim Konfigurieren des Web-Crawlers:

- Sie müssen einen Benutzeragentennamen für den Crawler angeben. Regeln in den Dateien robots.txt der zu durchsuchenden Server können angeben, ob dieser Name zu einer Zugriffsberechtigung oder Zugriffsverweigerung führt.
- Optional: Wenn ein Web-Server eine HTTP-Basisauthentifizierung verwendet, um den Zugriff auf Websites einzuschränken, können Sie die Authentifizierungsnachweise angeben, mit denen der Web-Crawler auf kennwortgeschützte Seiten zugreifen kann.
- Optional: Wenn ein Web-Server HTML-Formulare verwendet, um den Zugriff auf Websites einzuschränken, können Sie die Authentifizierungsnachweise angeben, mit denen der Web-Crawler auf kennwortgeschützte Seiten zugreifen kann.

---

#### **Einstiegspunktliste-, Web Content Management- und WebSphere Portal-Crawler**

---

<p>Bevor Sie einen Crawler erstellen, müssen Sie ein Setup-Script ausführen, um die Unternehmenssuche in einen WebSphere Portal-Server zu integrieren. Für verschiedene Versionen von WebSphere Portal werden verschiedene Scripts bereitgestellt.</p>	<p>Referenzinformationen:</p> <ul style="list-style-type: none"><li>• „Setup-Scripts zum Integrieren der Unternehmenssuche in WebSphere Portal“ auf Seite 374</li></ul>
--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Wenn Sie den Crawler konfigurieren, geben Sie einen vollständig qualifizierten definierten Namen (DN) an, der dem Crawler ermöglicht, Seiten vom zu durchsuchenden Server abzurufen, z. B. `uid=admin,cn=RegularEmployees,ou=SoftwareGroup,o=IBM,c=US`. Geben Sie außerdem das Kennwort für diesen definierten Namen an. Der definierte Name muss mit einem definierten Namen übereinstimmen, der in WebSphere Portal konfiguriert ist.

Stellen Sie sicher, dass die Berechtigungen für den angegebenen Benutzer-DN in der PAC-Komponente (Portal Access Control) von WebSphere Portal definiert sind. Der Crawler verwendet die PAC-Komponente, um Zugriffssteuerungsdaten für die Dokumente abzurufen, die er durchsucht.

---

#### **Windows-Dateisystem-Crawler**

---

Tabelle 7. Anforderungen der Crawlerkonfiguration zur Unterstützung der Sicherheit (Forts.)

Die zu durchsuchenden Unterverzeichnisse müssen dem Crawler-Server das Lesen von Daten erlauben. Wenn Sie den Crawler für das Durchsuchen ferner Dateisysteme konfigurieren, geben Sie eine Benutzer-ID an, die dem Crawler den Zugriff auf ferne Daten ermöglicht. Geben Sie außerdem ein Kennwort für diese Benutzer-ID an.

Referenzinformationen:

- „Umsetzung der Sicherheit auf Dokumentenebene für Dokumente des Windows-Dateisystems“ auf Seite 311
- „Sicheres Suchen in vertrauenswürdigen Windows-Domänen“ auf Seite 313

Stellen Sie sicher, dass die Domänenkonten korrekt konfiguriert sind, damit aktuelle Benutzerberechtigungen geprüft werden können, wenn ein Benutzer eine Suchanforderung übergibt. Die Anforderungen für das Einrichten von Domänenkonten für Dateien, die auf dem lokalen Computer durchsucht wurden, unterscheiden sich von den Anforderungen für Dateien, die auf einem fernen Windows-Server durchsucht wurden.

## Zugriffsprüfung zum Schutz von Exchange Server-Dokumenten

Wenn ein Exchange Server-Crawler zum Durchsuchen von Dokumenten verwendet werden soll, die durch eine Firewall geschützt sind, müssen Sie prüfen, ob der Crawler-Server auf den Server mit allgemein zugänglichen Ordnern von Microsoft Exchange Server zugreifen kann.

### Informationen zu dieser Task

Wenn der Crawler-Server nicht auf einen sicheren Exchange Server-Server zugreifen kann, gibt der Server den HTTP-Code 501 (Nicht implementiert) zurück. Möglicherweise werden auch Nachrichten angezeigt, die auf das Auftreten einer unerwarteten HTTP-Antwort hinweisen.

### Vorgehensweise

Gehen Sie wie folgt vor, um sicherzustellen, dass der Crawler-Server auf Dokumente hinter der Firewall zugreifen kann:

1. Starten Sie einen Web-Browser auf dem Crawler-Server.
2. Rufen Sie die URL-Adresse für den zu durchsuchenden Exchange Server-Server mit allgemein zugänglichen Ordnern auf. Beispiel: `http://exchange.Ihr_Unternehmen.com/public/`
3. Prüfen Sie, ob Sie die Exchange Server-Seite öffnen können.

Wenn Sie nicht auf den Exchange Server-Server zugreifen können, wenden Sie sich an Ihren unternehmensinternen Serveradministrator.

## Umsetzung der Sicherheit auf Dokumentenebene für Lotus Domino-Dokumente

Wenn der zu durchsuchende Domino-Server das NRPC-Protokoll (Notes Remote Procedure Call) verwendet, müssen Sie den Crawler-Server entsprechend konfigurieren, damit die Zugriffssteuerungen auf Dokumentenebene umgesetzt werden können.

Sie müssen auf dem Crawler-Server einen Domino-Server installieren, um die Sicherheit auf Dokumentenebene für Dokumente auf einem Domino-Server umzusetzen, der das NRPC-Protokoll verwendet. Dieser Domino-Server muss Teil Ihrer Domino-Domäne sein. Führen Sie die Anweisungen zur Installation und Konfiguration des Domino-Servers in der Lotus Domino-Dokumentation aus.

Sie müssen auch die folgenden Tasks ausführen, damit die Suchserver prüfen können, ob ein Benutzer, der eine sichere Objektgruppe durchsucht, über die Berechtigung zum Anzeigen von Dokumenten verfügt, die mit den Suchkriterien übereinstimmen:

- „Konfigurieren der Lotus Domino-Trusted Server, um Benutzerberechtigungen zu prüfen“.
- „Konfigurieren der globalen Sicherheit und einer LDAP-Benutzerregistry in WebSphere Application Server“ auf Seite 298.

#### **Zugehörige Konzepte**

„Prüfung aktueller Berechtigungsnachweise während der Abfrageverarbeitung“ auf Seite 287

„Notes-Crawler“ auf Seite 79

### **Konfigurieren der Lotus Domino-Trusted Server, um Benutzerberechtigungen zu prüfen**

Die zu durchsuchenden Domino-Server müssen als Lotus Domino-Trusted Server konfiguriert werden, um die Sicherheit für Dokumente umzusetzen, die von einem Notes-Crawler durchsucht wurden, der das NRPC-Protokoll (Notes Remote Procedure Call) verwendet.

#### **Vorbereitung**

Diese Vorgehensweise ist erforderlich, wenn Sie beim Durchsuchen ferner Datenbanken die Sicherheit auf Dokumentenebene umsetzen wollen. Diese Vorgehensweise ist nicht erforderlich, um Datenbanken zu durchsuchen, die für den Crawler-Server lokal sind.

Ein Domino-Server muss auf dem Crawler installiert sein, um Trusted Server konfigurieren zu können. Dieser Domino-Server muss Teil Ihrer Domino-Domäne sein.

#### **Informationen zu dieser Task**

Wenn Sie für einen Notes-Crawler Optionen für die Sicherheit auf Dokumentenebene konfigurieren, geben Sie an, ob Sie Zugriffssteuerungen umsetzen wollen, indem Sie die aktuellen Berechtigungsnachweise des Benutzers prüfen, wenn der Benutzer eine Abfrage übergibt. Bei den Domino-Servern, die durchsucht werden sollen, muss es sich um Lotus Domino-Trusted Server handeln, damit dieser Sicherheitstyp umgesetzt werden kann.

Wenn Benutzer eine Domäne durchsuchen, die eine Prüfung ihrer aktuellen Berechtigungsnachweise erforderlich macht, ermöglicht es der Trusted Server der Domino-Server-ID, den Kontext auf die aktuelle Benutzer-ID zu wechseln. Die Domino-Datenbank wird geöffnet, als ob der aktuelle Benutzer sie geöffnet hätte, und alle Informationen zu den Zugriffssteuerungslisten für die Datenbank werden für diesen Benutzer umgesetzt.

Die Möglichkeit, Kontexte auf diese Weise zu wechseln, steht im Allgemeinen nur für Datenbanken zur Verfügung, die im Datenverzeichnis des lokalen Domino-Servers gespeichert sind. Ab Lotus Domino Version 6.5.1 wird diese Funktion über

den Trusted Server bereitgestellt. Zum Konfigurieren des Trusted Servers gibt ein Domino-Administrator an, welche Domino-Server für die Ausführung sensibler Operationen als vertrauenswürdig eingestuft werden sollen, wie zum Beispiel das Agieren als anderer Benutzer, wenn von einem fernen Computer auf eine Datenbank zugegriffen wird.

### Vorgehensweise

Führen Sie die folgenden Schritte auf allen Domino-Servern aus, die von einem Notes-Crawler durchsucht wurden, um einen Trusted Server zu konfigurieren:

1. Verwenden Sie auf einem Domino-Server die ID-Datei des Domino-Domänenadministrators, um den Lotus Domino Administrator-Client zu öffnen.
2. Klicken Sie **Datei** an, und wählen Sie dann **Server öffnen** (Open Server) aus.
3. Geben Sie den Namen des Domino-Servers ein, für den Sie die Trusted Server-Funktionalität aktivieren wollen.
4. Wählen Sie die Indexzunge **Konfiguration** (Configuration) aus.
5. Erweitern Sie das Objekt **Server**, wählen Sie das Dokument **Aktueller Server** (Current Server) aus, und klicken Sie **Server bearbeiten** (Edit Server) an.
6. Wählen Sie die Indexzunge **Sicherheit** (Security) aus, blättern Sie im Dokument nach unten, suchen Sie nach dem Eintrag **Trusted Server**, und klicken Sie den Abwärtspfeil an.
7. Geben Sie eine der folgenden Optionen an:

#### **LocalDomainServers**

Wählen Sie diese Option aus, wenn alle Server in der Domino-Domäne als Trusted Server angesehen werden sollen.

#### *servername*

Geben Sie den Namen eines Domino-Servers an, auf dem eine Crawler-suche als Trusted Server durchgeführt werden soll.

Wenn sich der Domino-Server, der durchsucht werden soll, in einer anderen Domino-Domäne befindet, müssen Sie den Servernamen angeben oder die Gruppe **OtherDomainServers** auswählen. Sie müssen auch die Domino-spezifischen Schritte für die gegenseitige Zertifizierung der Domino-Server-ID-Datei für die Unternehmenssuche mit der anderen Domino-Domäne ausführen. Informationen zu diesen Schritten finden Sie in der Dokumentation für den Domino-Server.

8. Klicken Sie **Speichern und Schließen** (Save and Close) an, um Ihre Änderungen zu speichern.
9. Stoppen Sie die fernen Domino-Server, die Sie aktiviert haben, damit sie als Trusted Server agieren, und starten Sie sie erneut.

#### **Zugehörige Konzepte**

„Prüfung aktueller Berechtigungsnachweise während der Abfrageverarbeitung“ auf Seite 287

„Notes-Crawler“ auf Seite 79

### **Konfigurieren des QuickPlace-Servers für die Verwendung der Sicherheit "Local User" (Lokaler Benutzer)**

Wenn Sie für einen QuickPlace-Crawler die Nutzung der Option **Local User** (Lokaler Benutzer) zur Implementierung von Sicherheit konfigurieren möchten, müssen Sie Domino Directory auf dem Lotus QuickPlace-Server konfigurieren, bevor Sie den Crawler erstellen.

### **Informationen zu dieser Task**

Bei der Konfiguration eines QuickPlace-Crawlers wählen Sie für die Umsetzung der Sicherheit auf Dokumentenebene einen Sicherheitsmodus für den zu verwenden den Crawler aus. Wenn Sie den Modus **Local User** (Lokaler Benutzer) auswählen, müssen Sie sicherstellen, dass alle lokalen Benutzer-IDs und lokalen Gruppen in Domino Directory registriert sind (die Domino Directory-Hierarchie muss mit der QuickPlace-Hierarchie identisch sein).

Weiterhin müssen Sie sicherstellen, dass die Benutzer-ID und das Kennwort, die Sie für den zu verwendenden Crawler angeben, in Domino Directory registriert und zum Lesen der zu durchsuchenden Datenbank berechtigt ist.

Für die Nutzung von QuickPlace ist lediglich der Benutzername erforderlich. Zum Durchsuchen von QuickPlace-Quellen ist jedoch die vollständig erweiterte Benutzer-ID erforderlich. Die erweiterte Benutzer-ID hat das folgende Format:

*benutzername/berichsname/QP/domänenname*

Führen Sie die folgende Prozedur aus, um die vollständig erweiterte Version der Benutzer-ID zu ermitteln, um sicherzustellen, dass diese Benutzer-ID zum Lesen der QuickPlace-Datenbank berechtigt ist, und um Domino Directory die Benutzer-ID hinzuzufügen. In Domino Directory müssen die Benutzer-ID, die zum Durchsuchen von QuickPlace-Datenbanken verwendet werden soll, sowie alle lokalen Benutzer und lokalen Gruppen von QuickPlace vorhanden sein (die Domino Directory-Hierarchie muss mit der QuickPlace-Hierarchie identisch sein).

### Vorgehensweise

Gehen Sie wie folgt vor, um den QuickPlace-Server für die Verwendung der Sicherheit **Local User** (Lokaler Benutzer) zu konfigurieren:

1. Überprüfen Sie die Benutzer-ID-Berechtigungen:
  - a. Öffnen Sie das Serverdokument auf dem QuickPlace-Server.
  - b. Öffnen Sie die Seite **Files** (Dateien), und öffnen Sie anschließend die Zugriffssteuerungsliste (ACL - Access Control List) für die zu durchsuchende Datenbank.
  - c. Vergewissern Sie sich, dass die lokale Benutzer-ID, für die der Crawler konfiguriert wird, in der Zugriffssteuerungsliste vorhanden ist und dass diese Benutzer-ID Leseberechtigung für die Datenbank hat.  
In Schritt 2 müssen Sie die vollständig erweiterte Form dieser Benutzer-ID angeben.
2. Fügen Sie Domino Directory den Benutzer zu:
  - a. Öffnen Sie das Serverdokument auf dem QuickPlace-Server.
  - b. Fügen Sie auf der Seite **People and Groups** (Personen und Gruppen) im Baumstrukturelement für Personen die vollständig erweiterte Benutzer-ID ein, die Sie in Schritt 1 überprüft haben.
  - c. Geben Sie das Kennwort für diese Benutzer-ID im Feld **Internet password** (Internetkennwort) an.

### Konfigurieren von Directory Assistance auf einem QuickPlace-Server

Wenn Sie einen QuickPlace-Crawler so konfigurieren möchten, dass dieser ein LDAP-Verzeichnis zur Implementierung der Sicherheit verwendet, müssen Sie auf dem Lotus QuickPlace-Server eine Directory Assistance-Datenbank erstellen, bevor Sie den Crawler konfigurieren.

### Einschränkungen

Der QuickPlace-Server, den Sie durchsuchen wollen, muss die DIIOP- und HTTP-Tasks ausführen.

### Vorgehensweise

Gehen Sie wie folgt vor, um LDAP-Directory Assistance auf einem QuickPlace-Server zu konfigurieren:

1. Erstellen Sie eine Directory Assistance-Datenbank:
  - a. Öffnen Sie das Serverdokument auf dem QuickPlace-Server.
  - b. Erstellen Sie mithilfe der Schablone **Directory Assistance(6)** (Verzeichnisverwaltung 6) eine Datenbank. Diese Schablone befindet sich auf dem Server.
  - c. Klicken Sie **Add Directory Assistance** (Verzeichnisverwaltung hinzufügen) an, um in der Datenbank ein Dokument zu erstellen.
  - d. Öffnen Sie die Indexzunge **Basic** (Grundeinstellungen), und wählen Sie **LDAP** im Feld **DomainType** (Domänentyp) aus.
  - e. Öffnen Sie die Indexzunge **Naming Contexts** (Namenskontext), und stellen Sie sicher, dass das Markierungsfeld **Trusted for credentials** (Berechtigungs-nachweise vertrauenswürdig) ausgewählt ist.
  - f. Öffnen Sie die Indexzunge **LDAP**, und geben Sie Informationen zum LDAP-Server an.
  - g. Speichern und schließen Sie das Serverdokument.
2. Konfigurieren Sie den QuickPlace-Server für die Verwendung der Directory Assistance-Datenbank:
  - a. Öffnen Sie das Serverdokument auf dem QuickPlace-Server.
  - b. Öffnen Sie die Indexzunge **Basic** (Grundeinstellungen), und geben Sie im Feld **Directory assistance database name** (Name der Verzeichnis-verwaltungsdatenbank) den Namen der Datenbank an, die Sie in Schritt 1 erstellt haben.
  - c. Speichern und schließen Sie das Serverdokument.

Der QuickPlace-Server ist nun in der Lage, den LDAP-Server als sekundäres Domino-Verzeichnis zu verwenden.

## Umsetzung der Sicherheit auf Dokumentenebene für Dokumente des Windows-Dateisystems

Damit aktuelle Berechtigungsnachweise geprüft werden, wenn ein Benutzer Dokumente sucht, die von einem Crawler für Windows-Dateisysteme durchsucht wurden, müssen Sie auf dem Crawler-Server und auf dem Microsoft Windows-Server Domänenkontoinformationen konfigurieren.

Beim Konfigurieren eines Crawlers für das Windows-Dateisystem geben Sie an, ob Unterverzeichnisse auf dem lokalen Computer oder Unterverzeichnisse auf einem fernen Computer durchsucht werden sollen. Wenn die Sicherheit für die Objektgruppe aktiviert ist, können Sie auch Optionen für die Steuerung des Zugriffs auf Dokumente in den durchsuchten Unterverzeichnissen angeben.

Wenn Sie das Umsetzen von Zugriffssteuerungen auswählen, indem Sie beim Übergeben einer Abfrage die aktuellen Berechtigungsnachweise des Benutzers prüfen, müssen Sie sicherstellen, dass die Domänenkonten korrekt konfiguriert sind. Die Anforderungen für das Einrichten von Domänenkonten für Dateien, die auf dem lokalen Computer durchsucht wurden, unterscheiden sich von den Anforderungen für Dateien, die auf einem fernen Windows-Server durchsucht wurden.

**Wichtig:** Benutzerberechtigungen können nicht während der Abfrageverarbeitung geprüft werden, wenn die folgenden zwei Bedingungen gleichzeitig zutreffen:

- Der zu durchsuchende Windows-Server ist kein Element einer Domäne.
- Das zu durchsuchende Verzeichnis ist ein fernes Verzeichnis, wie z. B. \\servername\hostname.

## Prüfung mit lokalen Zugriffssteuerungsdaten

Das System verwendet lokale Benutzerkontoinformationen und Domänenkontoinformationen (falls der Computer zu einer Windows-Domäne gehört), um aktuelle Benutzerberechtigungen zu prüfen. Beide Benutzernamen müssen in den Sicherheitsinformationen zu den zu durchsuchenden Dokumenten aufgelistet sein, damit Berechtigungsnachweise während der Abfrageverarbeitung geprüft werden können.

### Lokale Konten

Bei einem lokalen Konto hat der Benutzername das folgende Format:

*COMPUTERNAME\BENUTZERNAME*

Zum Anmelden geben Benutzer nur den Benutzernamen an, aber für die korrekt angegebene Zuordnung der Windows-Benutzerberechtigungen wird der vollständige Name verwendet. Beispiel: Wenn der Benutzername für das lokale Konto abcuser lautet, wäre z. B. WINSERVER1\abcuser der vollständige Kontoname.

Wenn die Benutzer eine Suchanwendung verwenden und ein Profil für das Durchsuchen von sicheren Dokumenten auf einem lokalen System konfigurieren, müssen sie den Benutzernamen angeben, den sie zum Anmelden unter Windows verwenden (zum Beispiel abcuser).

### Domänenkonten

Bei einem Domänenkonto hat der Benutzername das folgende Format:

*DOMÄNENNAME\BENUTZERNAME*

Zum Anmelden geben die Benutzer diese Informationen im folgenden Format an:

*BENUTZERNAME@DOMÄNENNAME*

Beispiel: Falls Sie die Zuordnung von Benutzerberechtigungen für eine Datei konfigurieren und die Domäne WIN1\abcuser auswählen, wird das Konto anschließend wie folgt angezeigt: abcuser@win1.company.com.

Wenn die Benutzer eine Suchanwendung verwenden und ein Profil konfigurieren, mit dem Sie Dokumente in einer sicheren Domäne durchsuchen können, müssen sie den Benutzernamen angeben, den sie zum Anmelden unter Windows verwenden (zum Beispiel abcuser@win1.company.com).

Die vom Crawler-Server verwendeten Benutzerkonten müssen die folgenden Windows-Benutzerberechtigungen haben, damit die Prüfung der aktuellen Berechtigungsnachweise auf lokalen Computern umgesetzt werden kann. Für das Zuordnen von Benutzerberechtigungen verwenden Sie die Windows-Verwaltungstools: **Verwaltung** → **Sicherheitsrichtlinie für Domänen**.

- Die Benutzer-ID, mit der der Crawler-Server ausgeführt wird, muss über die Berechtigung **Als Teil des Betriebssystems handeln** verfügen. Diese Berechtigung wird für den Benutzer mit Verwaltungsaufgaben für die Unternehmenssuche auf dem Crawler-Server konfiguriert, wenn OmniFind Enterprise Edition installiert wird.

- Die Benutzer müssen über die Benutzerberechtigung **Lokal anmelden** verfügen.

## Prüfung mit Zugriffssteuerungsdaten für die ferne Domäne

Auf dem Windows-Betriebssystem wird jedes Verzeichnis, das mit `\\servername` beginnt, als fernes Verzeichnis angesehen. Beispiel:

```
\\software\utilities\IBM
```

Die Benutzer geben ihre Benutzernamen im folgenden Format an, um auf ein fernes Verzeichnis zuzugreifen.

```
BENUTZERNAME@DOMÄNENNAME
```

Wenn die Benutzer eine Suchanwendung verwenden und ein Profil konfigurieren, mit dem Sie Dokumente auf einem fernen System durchsuchen können, müssen sie den Benutzernamen angeben, den sie für den Zugriff auf das ferne Windows-System verwenden (zum Beispiel `abcuser@win1.company.com`).

Die Benutzerkonten müssen über die folgenden Windows-Benutzerberechtigungen verfügen, damit die Prüfung der aktuellen Berechtigungsnachweise auf fernen Computern umgesetzt werden kann. Für das Zuordnen von Benutzerberechtigungen verwenden Sie die Windows-Verwaltungstools: **Verwaltung** → **Sicherheitsrichtlinie für Domänen**.

- Der Crawler-Server und der Windows-Server, die durchsucht werden sollen, müssen Mitglieder derselben Domäne sein.
- Die Benutzer-ID, mit der der Crawler-Server ausgeführt wird, muss über die Berechtigung **Als Teil des Betriebssystems handeln** verfügen. Diese Berechtigung wird für den Benutzer mit Verwaltungsaufgaben für die Unternehmenssuche auf dem Crawler-Server konfiguriert, wenn OmniFind Enterprise Edition installiert wird.
- Die Benutzer müssen über die Benutzerberechtigung **Anmelden als Stapelverarbeitungsauftrag** verfügen.

### Zugehörige Konzepte

„Prüfung aktueller Berechtigungsnachweise während der Abfrageverarbeitung“ auf Seite 287

„Crawler für das Windows-Dateisystem“ auf Seite 123

„Sicheres Suchen in vertrauenswürdigen Windows-Domänen“

## Sicheres Suchen in vertrauenswürdigen Windows-Domänen

Damit die Sicherheit auf Dokumentenebene für ferne Windows-Dateisysteme erzwungen werden kann, unterstützt das System für die Unternehmenssuche die Überprüfung von Zugriffssteuerungslisten über vertrauenswürdige Domänen.

## Konfigurieren des Crawlers

Sie müssen Optionen in einer neuen Konfigurationsdatei angeben, um den Windows-Dateisystem-Crawler für die Unterstützung vertrauenswürdiger Domänen zu konfigurieren. Die Administrationskonsole für die Unternehmenssuche bietet keine Unterstützung für die Konfiguration dieser Funktion.

1. Bearbeiten Sie die folgende Datei, um die Sicherheit auf Dokumentenebene über vertrauenswürdige Windows-Domänen zu unterstützen:

```
ES_NODE_ROOT/master_config/sitzungs-id/winfscrawler_ext.xml
```

**Tipp:** Sie können die Sitzungs-ID für den Windows-Dateisystem-Crawler ermitteln, den Sie konfigurieren wollen, indem Sie den Crawler in der

Administrationskonsole für die Unternehmenssuche überwachen oder den Befehl **esadmin report collections** verwenden.

2. Geben Sie den Namen der Windows-Domäne und den NetBIOS-Namen von Active Directory an. Beispiel:

```
<ExtendedProperties>
 <SetAttribute XPath="/Crawler/DataSources/Server/Target"
 Name="Domain">jk.enterprises.com
 </SetAttribute>
 <SetAttribute XPath="/Crawler/DataSources/Server/Target"
 Name="NetBIOSDomain">JKE1
 </SetAttribute>
</ExtendedProperties>
```

3. Stoppen Sie den Crawler, und starten Sie ihn erneut, damit die Änderungen wirksam werden.

### Einschränkungen

- Dokumente dürfen keine Zugriffssteuerungslisten mehrerer Windows-Domänen enthalten. Domänenbenutzer und -gruppen müssen pro Objektgruppe zu einer einzigen Windows-Domäne gehören.
- Damit die Zugriffsüberprüfung über ferne Dateisysteme unterstützt wird, müssen die Windows-Server in einer Windows-Domäne oder in vertrauenswürdigen Windows-Domänen ausgeführt werden.
- Der Windows-Dateisystem-Crawler liest den NetBIOS-Namen von Active Directory, der dem zu durchsuchenden Windows-Server zugeordnet ist, und verwendet den NetBIOS-Namen, um die Dateizugriffssteuerungsliste zu filtern. Das vom Crawler-Server verknüpfte Active Directory vertraut dem anderen Active Directory, das Benutzer- und Gruppenkonten definiert.
- Das von Ihnen angegebene Benutzerkonto, das der Crawler für den Zugriff auf einen fernen Windows-Server verwenden soll, muss zu der Windows-Domäne gehören, in der Sie die Zugriffssteuerung erzwingen und prüfen wollen.
- Das Windows-Betriebssystem erlaubt es nur einem Konto, eine Verbindung zu Netzordnern auf einem Dateiserver herzustellen. Andere Konten können nicht gleichzeitig eine Verbindung zu demselben Dateiserver herstellen. Aus diesem Grund können Sie nicht verschiedene Konten für verschiedene Crawler konfigurieren, die denselben Windows-Server durchsuchen sollen, selbst wenn sich die Crawler in verschiedenen Objektgruppen befinden.

#### Zugehörige Konzepte

„Umsetzung der Sicherheit auf Dokumentenebene für Dokumente des Windows-Dateisystems“ auf Seite 311

---

## Inaktivieren der Sicherheit für die Unternehmenssuche

Sie können die Sicherheit für eine Anwendung für die Unternehmenssuche in WebSphere Application Server inaktivieren. Wenn Sie zuvor Sicherheitsmaßnahmen auf Dokumentenebene konfiguriert haben, können Sie angeben, dass die Maßnahmen ignoriert werden sollen. Sicherheitseinstellungen wirken sich außerdem darauf aus, wie ausgeblendete Ergebnisse in den Suchergebnissen angezeigt werden.

### Inaktivieren der Sicherheit für eine Unternehmensanwendung in WebSphere Application Server

Sie können die globale Sicherheit für einzelne Unternehmensanwendungen in WebSphere Application Server inaktivieren, um zu steuern, für welche Aktivitäten für die Unternehmenssuche eine Benutzerauthentifizierung erforderlich ist.

## Informationen zu dieser Task

Das Installationsprogramm von OmniFind Enterprise Edition implementiert drei Unternehmensanwendungen in WebSphere Application Server:

- Die Anwendung "ESAdmin" enthält die Schnittstelle für die Administrationskonsole für die Unternehmenssuche.
- Die Anwendung "ESSearchApplication" enthält die Schnittstelle für die Mustersuchanwendung.
- Die Anwendung "ESSearchServer" stellt die gesamte ferne Kommunikation für die SI-API-Implementierung bereit und aktiviert die SI-API-Schnittstellen für die Kommunikation mit den Suchservern.

Standardmäßig unterstützen alle drei Unternehmensanwendungen die globale Sicherheit von WebSphere Application Server. Wenn diese Anwendungen entdecken, dass die globale Sicherheit aktiviert ist, authentifizieren sie von diesem Zeitpunkt an alle Anforderungen, die sie erhalten.

Einige Unternehmen möchten die Sicherheit möglicherweise für bestimmte Unternehmensanwendungen aktivieren bzw. inaktivieren. Beispiel: Sie möchten alle Benutzer authentifizieren, die auf die Administrationskonsole für die Unternehmenssuche zugreifen, aber nicht die Benutzer, die die SI-API-Schnittstellen oder die Mustersuchanwendung verwenden.

## Vorgehensweise

Gehen Sie wie folgt vor, um die Sicherheit für eine Unternehmensanwendung zu inaktivieren:

1. Starten Sie auf dem Suchserver die Administrationskonsole von WebSphere Application Server.  
Sie können die Administrationskonsole auf eine der folgenden Weisen öffnen:
  - Wählen Sie das Programm über das Windows-Menü **Start** aus.
  - Für WebSphere Application Server Version 5 öffnen Sie einen Web-Browser, und gehen Sie zur folgenden Adresse: `http://hostname:port/admin`. Dabei ist *hostname* der Hostname des Suchservers und *port* die Portnummer für die Administrationskonsole von WebSphere Application Server. Normalerweise ist 9090 die Nummer des Ports der Administrationskonsole.
  - Für WebSphere Application Server Version 6 öffnen Sie einen Web-Browser, und gehen Sie zur folgenden Adresse: `http://hostname:port/ibm/console`. Dabei ist *hostname* der Hostname des Suchservers und *port* die Portnummer für die Administrationskonsole von WebSphere Application Server. Normalerweise ist 9060 die Nummer des Ports der Administrationskonsole.
2. Wenn Sie zur Eingabe von Benutzer-ID und Kennwort aufgefordert werden, geben Sie die Administrator-ID und das Kennwort ein, die bei der Aktivierung der globalen Sicherheit in WebSphere Application Server angegeben wurden.
3. Nachdem Sie sich an der Administrationskonsole angemeldet haben, klicken Sie **Anwendungen** und anschließend **Enterprise-Anwendungen** an.
4. Wählen Sie das Markierungsfeld neben dem Namen der Unternehmensanwendung aus, für die Sie die Sicherheit inaktivieren wollen.
5. Blättern Sie nach unten, und klicken Sie den Link **Benutzern/Gruppen Sicherheitsaufgabenbereiche zuordnen** an.
6. Suchen Sie die Rolle **Alle Authentifizierten** (ALLAUTHENTICATED), und wählen Sie das Markierungsfeld unter der Spalte **Jeder?** aus.

7. Klicken Sie **OK** an.
8. Klicken Sie den Link **Speichern** an, um Ihre Änderungen zu speichern.
9. Wenn Sie WebSphere Network Deployment verwenden, wählen Sie das Markierungsfeld **Änderungen mit Knoten synchronisieren** aus.
10. Klicken Sie **Speichern** an.
11. Stoppen Sie die Anwendung "ESSearchServer", und starten Sie sie erneut.

#### **AIX, Linux oder Solaris**

```
./stopServer.sh ESSearchServer
./startServer.sh ESSearchServer
```

#### **Windows**

```
stopServer ESSearchServer
startServer ESSearchServer
```

Diese Scripts befinden sich im Verzeichnis was-installationsverzeichnis/AppServer/bin.

- Bei WebSphere Application Server Version 5 ist der Standardinstallationspfad /usr/WebSphere auf AIX-Systemen, /opt/WebSphere auf Linux- oder Solaris-Systemen oder C:\Programme\WebSphere auf Windows-Systemen.
- Bei WebSphere Application Server Version 6 ist der Standardinstallationspfad /usr/IBM/WebSphere auf AIX-Systemen, /opt/IBM/WebSphere auf Linux- oder Solaris-Systemen oder C:\Programme\IBM\WebSphere auf Windows-Systemen.

## **Inaktivieren der Sicherheit auf Dokumentenebene**

Unabhängig davon, ob für die Dokumente im Index Zugriffssteuerungen vergeben wurden, können Sie es Benutzern erlauben, eine Objektgruppe zu durchsuchen. Für Crawler, die eine Prüfung der aktuellen Berechtigungsnachweise unterstützen, können Sie auch das Durchsuchen einer Objektgruppe aktivieren, ohne dass aktuelle Zugriffssteuerungen während der Abfrageverarbeitung geprüft werden.

### **Vorbereitung**

Sie müssen ein Administrator für die Unternehmenssuche oder ein Objektgruppenadministrator der Objektgruppe sein, um die Sicherheit auf Dokumentenebene für alle Dokumente in einer Objektgruppe aktivieren bzw. inaktivieren zu können.

### **Einschränkungen**

Sie können Optionen für die Sicherheit auf Dokumentenebene nur angeben, wenn bei der Erstellung der Objektgruppe die Sicherheit aktiviert wurde.

### **Informationen zu dieser Task**

Sie können Crawler so konfigurieren, dass Dokumenten beim Durchsuchen Sicherheitstoken zugeordnet werden. Ihre Suchanwendungen können diese im Index gespeicherten Token verwenden, um Zugriffssteuerungen umzusetzen, wenn Benutzer die Objektgruppe durchsuchen. Für manche Crawler können Sie auch angeben, dass Sie beim Übergeben von Abfragen aktuelle Zugriffssteuerungen prüfen wollen, die den Dokumenten in ihren nativen Repositories zugeordnet sind.

Wenn Sie diese Sicherheitseinschränkungen entfernen wollen, können Sie angeben, dass die Suchserver alle mit einer Abfrage weitergeleiteten Sicherheitstoken igno-

rieren sollen. Sie können es den Benutzern auch ermöglichen, Dokumente abzufragen, ohne dass ihre Berechtigungsnachweise mit den aktuellen Zugriffssteuerungen verglichen werden.

Unter Umständen ist es sinnvoll, die Sicherheit auf Dokumentenebene zeitweise zu inaktivieren, zum Beispiel, wenn Sie eine neue Objektgruppe testen oder wenn Sie einen Fehler an einer Suchanwendung beheben wollen.

### Vorgehensweise

Gehen Sie wie folgt vor, um die Zugriffssteuerung auf Dokumentenebene zu inaktivieren:

1. Bearbeiten Sie eine Objektgruppe, wählen Sie die Seite **Allgemein** aus, und klicken Sie **Sicherheit auf Dokumentenebene aktivieren oder inaktivieren** an.
2. Wählen Sie auf der Seite **Sicherheit auf Dokumentenebene für alle Dokumente** das Markierungsfeld **Zugriffssteuerungen auf Dokumentenebene im Index ignorieren** aus, wenn die den Dokumenten von den Crawlern zugeordneten Sicherheitstoken nicht verwendet werden sollen, wenn die Objektgruppe von Benutzern abgefragt wird.

Die Crawler fügen den Dokumenten zwar weiterhin Sicherheitstoken hinzu, diese werden jedoch von den Suchservern ignoriert, sodass Benutzer die vormals geschützten Dokumente durchsuchen können.

3. Wählen Sie das Markierungsfeld **Aktuelle Berechtigungsnachweise während der Abfrageverarbeitung nicht prüfen** aus, wenn Sie beim Übergeben von Abfragen die aktuellen Berechtigungsnachweise nicht prüfen wollen, die den Dokumenten in ihren nativen Repositorys zugeordnet sind. Dieses Markierungsfeld ist nur für Dokumente verfügbar, die von Crawlern durchsucht wurden, die diese Funktion unterstützen.

Wenn Sie dieses Markierungsfeld auswählen, bleiben andere Optionen für die Sicherheit auf Dokumentenebene wirksam. Wenn Sie z. B. beim Konfigurieren des Crawlers Optionen zum Speichern von Zugriffssteuerungen im Index angeben haben, gelten jene Sicherheitsmaßnahmen weiterhin, solange Sie nicht auch das Markierungsfeld **Zugriffssteuerungen auf Dokumentenebene im Index ignorieren** ausgewählt haben.

## Inaktivieren der Sicherheit für ausgeblendete Suchergebnisse

Wenn die Objektgruppensicherheit aktiviert ist, können Suchergebnisse derselben Site nicht in den Suchergebnissen ausgeblendet werden, es sei denn, Sie geben an, dass während der Abfrageverarbeitung keine Benutzerberechtigungen geprüft werden sollen.

### Vorbereitung

Sie müssen ein Administrator für die Unternehmenssuche oder ein Objektgruppenadministrator der Objektgruppe sein, um die Prüfung der aktuellen Berechtigungsnachweise für alle Dokumente in einer Objektgruppe aktivieren bzw. inaktivieren zu können.

### Einschränkungen

Sie können Optionen für die Sicherheit auf Dokumentenebene nur angeben, wenn bei der Erstellung der Objektgruppe die Sicherheit aktiviert wurde.

### Informationen zu dieser Task

Wenn aktuelle Berechtigungsnachweise geprüft werden, wird für jedes Dokument die Quelle überprüft und gegebenenfalls zur Prüfung weitergeleitet. Sie müssen die Prüfung der aktuellen Berechtigungsnachweise inaktivieren, wenn Dokumente, die dasselbe URI-Präfix aufweisen oder zu einer zuvor konfigurierten ausgeblendeten URI-Gruppe gehören, in den Suchergebnissen ausgeblendet werden sollen.

### Vorgehensweise

Gehen Sie wie folgt vor, um die Prüfung der aktuellen Berechtigungsnachweise zu inaktivieren, damit Dokumente in den Suchergebnissen ausgeblendet werden können:

1. Bearbeiten Sie eine Objektgruppe, wählen Sie die Seite **Allgemein** aus, und klicken Sie **Sicherheit auf Dokumentenebene aktivieren oder inaktivieren** an.
2. Wählen Sie auf der Seite **Sicherheit auf Dokumentenebene für alle Dokumente** das Markierungsfeld **Aktuelle Berechtigungsnachweise während der Abfrageverarbeitung nicht prüfen** aus.
3. Überwachen Sie die Objektgruppe, wählen Sie die Seite **Suche** aus, stoppen Sie die Prozesse des Suchservers, und starten Sie sie erneut.

Wenn Benutzer die Objektgruppe abfragen, werden Dokumente, die dasselbe URI-Präfix aufweisen oder zu Sites gehören, die für das Ausblenden konfiguriert sind, in den Suchergebnissen ausgeblendet. In der Mustersuchanwendung können Benutzer die ausgeblendeten Ergebnisse anzeigen, indem Sie den Link **Weitere Ergebnisse aus derselben Quelle** anklicken.

---

## Starten und Stoppen eines Systems für die Unternehmenssuche

Nach dem Erstellen einer Objektgruppe müssen Sie die Server für Crawlersuche, Syntaxanalyse und Indexierung von Daten starten (die Suchserver werden automatisch gestartet). Stoppen Sie die Server, und starten Sie sie erneut, nachdem Sie Änderungen an der Objektgruppe vorgenommen haben.

Die meisten Server für die Unternehmenssuche können kontinuierlich oder nach den angegebenen Zeitplänen ausgeführt werden. Sie können z. B. Zeitpläne für die Haupt- und Deltaindexerstellung angeben. Nachdem Sie das System für die Unternehmenssuche gestartet haben, müssen Sie die Serverprozesse in der Regel nur dann stoppen und erneut starten, wenn Sie die Konfigurationseinstellungen ändern (wie z. B. das Aktualisieren der Kategorien oder das Vergrößern des Suchcaches).

Wenn Sie den Inhalt einer Objektgruppe oder die Regeln für das Sammeln von Daten aus den Quellen in Ihrem Unternehmen durch Crawler ändern, müssen Sie die Crawler stoppen und erneut starten, damit die Änderungen wirksam werden. Wenn Sie die Crawlersuchregeln nicht ändern, werden der Web-Crawler kontinuierlich und andere Crawler nach den angegebenen Zeitplänen ausgeführt.

Wenn Sie die Verfügbarkeit der Suchserver erweitern wollen, wenn der Indexserver und die Administrationskonsole nicht verfügbar sind, können Sie Befehle zum Starten der Suchserver für eine Objektgruppe im Standalone-Modus angeben.

---

## Starten eines Systems für die Unternehmenssuche

Damit Benutzer eine Objektgruppe durchsuchen können, müssen Sie die Systemprozesse und anschließend die Server für die Crawlersuche, die Syntaxanalyse und die Indexierung der Objektgruppe starten (die Suchserver werden automatisch gestartet).

### Vorbereitung

Konfigurieren Sie die Datenquellen, die Sie durchsuchen wollen, und geben Sie Optionen an, wie die Daten syntaktisch analysiert, indiziert und gesucht werden sollen. Wenn Benutzer z. B. Kategoriedetails in den Suchergebnissen anzeigen können sollen, müssen Sie Kategorien konfigurieren, bevor Sie den Parser starten.

Sie müssen ein Administrator für die Unternehmenssuche, der Objektgruppenadministrator einer Objektgruppe oder ein Bediener mit der Berechtigung, die Objektgruppe zu verwalten sein, um die Server für die Unternehmenssuche zu starten.

Die Server für die Unternehmenssuche müssen in der richtigen Reihenfolge für eine Objektgruppe gestartet werden. Sie müssen z. B. einen Crawler starten und die Daten durchsuchen, bevor Sie durchsuchte Daten indexieren können.

### Einschränkungen

Zum Starten eines Systems für die Unternehmenssuche müssen Sie ein Benutzerkonto verwenden, das mit lokaler Authentifizierung authentifiziert werden kann. Wenn Sie versuchen, das System mit einem AFS-Konto (Andrew File System) zu starten, treten Fehler auf.

### Vorgehensweise

Gehen Sie wie folgt vor, um ein System für die Unternehmenssuche zu starten:

1. Wenn Sie die Unternehmenssuche in einer Konfiguration mit zwei oder vier Servern verwenden, müssen Sie sich als Administrator für die Unternehmenssuche anmelden und die allgemeine Übertragungsschicht (CCL = Common Communication Layer) auf jedem Server starten:

#### AIX, Linux oder Solaris

```
startccl.sh -bg
```

#### Windows-Eingabeaufforderung

```
startccl
```

#### Verwaltungstool für Windows-Services

Gehen Sie wie folgt vor, um die allgemeine Übertragungsschicht im Hintergrund zu starten:

- a. Starten Sie die Windows-Services: **Start** → **Einstellungen** → **Systemsteuerung** → **Verwaltung** → **Dienste**.
  - b. Klicken Sie **IBM OmniFind Enterprise Edition** mit der rechten Maustaste an, und klicken Sie **Start** an.
2. Starten Sie die Komponenten des Systems für die Unternehmenssuche:
    - a. Melden Sie sich als Administrator für die Unternehmenssuche auf einem beliebigen Server für die Unternehmenssuche an.
    - b. Starten Sie alle Systemkomponenten: `esadmin system startall`  
Dieser Befehl startet die folgenden Prozesse und Anwendungen:
      - Den Web-Server (in einer Konfiguration mit mehreren Servern wird der Web-Server auf beiden Suchservern gestartet)
      - Die Anwendungen "ESSearchServer" und "ESAdmin" in WebSphere Application Server (in einer Konfiguration mit mehreren Servern werden die Anwendungen auf beiden Suchservern gestartet)
      - Die Sitzung "ESAdmin" auf dem Indexserver
      - Die allgemeine Übertragungsschicht auf dem Computer, auf dem Sie den Befehl ausführen, falls die allgemeine Übertragungsschicht nicht bereits ausgeführt wird
      - Den Datenbanknetzserver für die Unternehmenssuche
      - Die Informationszentrale für die Unternehmenssuche (in einer Konfiguration mit mehreren Servern wird die Informationszentrale für die Unternehmenssuche auf beiden Suchservern gestartet)
  3. Starten Sie die Administrationskonsole für die Unternehmenssuche, und melden Sie sich als Administrator für die Unternehmenssuche an. Wenn Sie Verwaltungsrollen verwenden, können Sie sich als Objektgruppenadministrator oder Bediener mit Berechtigung für die Objektgruppe anmelden, die Sie starten wollen.
  4. Suchen Sie in der Sicht **Objektgruppen** die Objektgruppe, die Sie verwalten wollen, und klicken Sie  **Überwachen** an.
  5. Klicken Sie auf der Seite **Crawler** für jeden Crawler, den Sie starten wollen,  **Starten** an.

- Wenn Sie einen Web-Crawler starten, beginnt der Crawler sofort, Daten zu durchsuchen. Diese Crawlertypen führen die Crawlersuche kontinuierlich aus, wobei sie Webdokumente durchsuchen und erneut durchsuchen.
- Wenn Sie einen anderen Crawlertyp starten, wird die Crawlersitzung gestartet. Der Crawler beginnt zum terminierten Datum und zur terminierten Zeit mit der Crawlersuche. Wenn Sie den Crawler nicht terminiert haben oder wenn Sie den Crawler früher starten wollen, überwachen Sie den Crawler, und klicken Sie das Startsymbol für jede Datenquelle an, die durchsucht werden soll.

Sie können den Crawler nach seinem Start kontinuierlich ausführen. Wenn Sie den Crawler terminiert haben, wird er am terminierten Datum zur terminierten Zeit erneut ausgeführt.

6. Nachdem Daten durchsucht wurden, öffnen Sie die Seite **Syntaxanalyse**, und klicken Sie  **Starten** an, um den Parser zu starten.

Sie können den Parser kontinuierlich ausführen. Normalerweise ist es erst erforderlich, den Parser zu stoppen, wenn Sie Änderungen an der Syntaxanalyse der Daten vornehmen (wie z. B. die Aktualisierung von Kategorien oder XML-Feldzuordnungen).

7. Optional: Wenn Sie den Start der Indexierungsprozesse erzwingen wollen, statt auf den Start zum terminierten Zeitpunkt zu warten, öffnen Sie die Seite **Index**, und klicken Sie  **Starten** im Bereich **Reorganisation** an.

Sie können die Indexierungsprozesse kontinuierlich ausführen. Der Index wird zu den terminierten Zeitpunkten erstellt.

**Tipp:** Die Suchserver werden automatisch gestartet, und Sie können sie kontinuierlich ausführen. Normalerweise ist es erst erforderlich, die Suchserver zu stoppen, wenn Sie Änderungen an den Einstellungen des Suchcaches oder der Dokumentzusammenfassung vornehmen. Zum erneuten Starten der Suchserver öffnen Sie die Seite **Suche**, klicken Sie  **Stoppen** und anschließend  **Starten** an.

Wenn Sie die Verfügbarkeit der Suchserver erweitern wollen, wenn die Administrationskonsole nicht verfügbar ist, können Sie Befehle zum Starten der Suchserver für eine Objektgruppe im Standalone-Modus angeben. Wenn der Indexserver nicht aktiv ist, ist die Administrationskonsole nicht verfügbar.

#### **Zugehörige Konzepte**

„Verwaltungsrollen“ auf Seite 280

#### **Zugehörige Tasks**

„Anmelden an der Administrationskonsole“ auf Seite 19

„Steuern, welche Komponenten gestartet oder gestoppt werden sollen“ auf Seite 323

„Verwalten der Suchserver im Standalone-Modus“ auf Seite 324

---

## **Stoppen eines Systems für die Unternehmenssuche**

Manchmal ist es erforderlich, einen Server für die Unternehmenssuche zu stoppen und erneut zu starten, beispielsweise, um Änderungen an der Konfiguration vorzunehmen oder Fehler zu beheben.

### **Vorbereitung**

Sie müssen ein Administrator für die Unternehmenssuche, der Objektgruppen-administrator einer Objektgruppe oder ein Bediener mit der Berechtigung sein, die Objektgruppe zu verwalten, um Server für die Unternehmenssuche zu stoppen.

### Informationen zu dieser Task

Sie können die Server für die Unternehmenssuche unabhängig von einander stoppen. Wenn Sie z. B. einen Crawler stoppen und erneut starten, damit Änderungen übernommen werden, die Sie an seiner Konfiguration vorgenommen haben, ist es nicht erforderlich, auch die Suchserver zu stoppen und erneut zu starten.

### Vorgehensweise

Gehen Sie wie folgt vor, um Server für die Unternehmenssuche zu stoppen:

1. Suchen Sie in der Sicht **Objektgruppen** die Objektgruppe, die Sie verwalten wollen, und klicken Sie  **Überwachen** an.
2. Suchen Sie den Crawler, den Sie verwalten wollen, auf der Seite **Crawler**, und stoppen Sie ihn bzw. halten Sie ihn an.

Wenn Sie den Crawlerbereich oder Crawlermerkmale ändern, müssen Sie den Crawler stoppen und erneut starten, damit die Änderungen wirksam werden. Wenn Sie den Crawlerbereich ändern und diese Änderungen auch auf bereits indizierte Dokumente anwenden wollen, müssen Sie diese Dokumente außerdem erneut durchsuchen.

**Tipp:** Es wird möglicherweise eine Nachricht angezeigt, dass die angeforderte Operation das Zeitlimit überschritten hat, selbst wenn der Prozess noch im Hintergrund aktiv ist. Klicken Sie **Aktualisieren** in der Administrationskonsole an (klicken Sie nicht **Aktualisieren** im Web-Browser an), um zu ermitteln, ob die Task beendet wurde. Der Prozess ist beendet, sobald das Statussymbol für den Crawler anzeigt, dass er gestoppt ist.

3. Klicken Sie  **Stoppen** auf der Seite **Syntaxanalyse** an, um den Parser zu stoppen.

Wenn Sie die Regeln für die Syntaxanalyse der Daten ändern, müssen Sie den Parser stoppen und erneut starten, damit die Änderungen wirksam werden. Die Änderungen werden nur für Dokumente angewendet, die nach der Änderung durchsucht werden. Zum Anwenden der Änderungen auf Dokumente, die sich bereits im Index befinden, müssen Sie eine vollständige Crawler suche starten, damit alle Dokumente erneut durchsucht werden. Danach können sie erneut syntaktisch analysiert und indiziert werden.

4. Klicken Sie  **Stoppen** auf der Seite **Index** an, um einen Index zu stoppen, der gerade erstellt wird.

Sie können auch eine Indexerstellung stoppen, während Sie die Indexierungswarteschlange überwachen. Wählen Sie hierzu **System** in der Funktionsleiste aus, öffnen Sie die Seite **Index**, und klicken Sie anschließend  **Stoppen** für den Index an, dessen Erstellung Sie stoppen wollen.

5. Klicken Sie  **Stoppen** auf der Seite **Suche** an, um die Suchserver zu stoppen. Normalerweise müssen Sie die Suchserver nur stoppen und erneut starten, wenn Sie die Einstellungen des Suchcaches oder der Dokumentzusammenfassung ändern.

6. Gehen Sie wie folgt vor, um das System für die Unternehmenssuche und nicht einzelne Server zu stoppen:

- a. Melden Sie sich als Administrator für die Unternehmenssuche auf einem beliebigen Server für die Unternehmenssuche an.

b. Stoppen Sie alle Systemkomponenten: `esadmin system stopall`

Dieser Befehl stoppt die folgenden Prozesse und Anwendungen:

- Den Web-Server (in einer Konfiguration mit mehreren Servern wird der Web-Server auf beiden Suchservern gestoppt)
- Die Anwendungen "ESSearchServer" und "ESAdmin" in WebSphere Application Server (in einer Konfiguration mit mehreren Servern werden die Anwendungen auf beiden Suchservern gestoppt)
- Die Sitzung "ESAdmin" auf dem Indexserver
- Die allgemeine Übertragungsschicht (Common Communications Layer, CCL) für die Unternehmenssuche auf dem Computer, auf dem Sie den Befehl ausführen
- Den Datenbanknetzserver für die Unternehmenssuche
- Die Informationszentrale für die Unternehmenssuche (in einer Konfiguration mit mehreren Servern wird die Informationszentrale für die Unternehmenssuche auf beiden Suchservern gestoppt)

#### Zugehörige Tasks

„Anmelden an der Administrationskonsole“ auf Seite 19

„Steuern, welche Komponenten gestartet oder gestoppt werden sollen“

„Verwalten der Suchserver im Standalone-Modus“ auf Seite 324

---

## Steuern, welche Komponenten gestartet oder gestoppt werden sollen

Mit den Befehlen `esadmin system startall` und `esadmin system stopall` können Sie steuern, welche Komponenten gestartet oder gestoppt werden.

### Informationen zu dieser Task

Die Datei `ES_INSTALL_ROOT/default_config/AutoRunComponents.properties` enthält eine Liste der Komponenten für die Unternehmenssuche, die mit den Befehlen `esadmin system startall` und `esadmin system stopall` gestartet bzw. gestoppt werden können. Alle hier aufgelisteten Komponenten werden standardmäßig mit diesen Befehlen gestartet und gestoppt.

Wenn Sie verhindern wollen, dass bestimmte Komponenten gestartet oder gestoppt werden können, müssen Sie die Merkmaldatei bearbeiten.

### Vorgehensweise

Gehen Sie wie folgt vor, um die Komponenten anzugeben, die beim Starten oder Stoppen des Systems für die Unternehmenssuche gestartet bzw. gestoppt werden sollen:

1. Melden Sie sich als Administrator für die Unternehmenssuche an dem Server an, auf dem Sie den Befehl `esadmin system startall` oder `esadmin system stopall` ausführen wollen.
2. Bearbeiten Sie die Datei `ES_INSTALL_ROOT/default_config/AutoRunComponents.properties`.
3. Soll das Starten einer Komponente verhindert werden, fügen Sie für die Komponente ein Feld mit dem Namen `Component.startable.komponenten-id=false` hinzu. Dabei ist `komponenten-id` die Komponente, die nicht gestartet werden soll.

4. Soll das Stoppen einer Komponente verhindert werden, fügen Sie für die Komponente ein Feld mit dem Namen `Component.stopable.komponenten-id=false` hinzu. Dabei ist `komponenten-id` die Komponente, die nicht gestoppt werden soll.
5. Speichern und schließen Sie die Datei.  
 Beim nächsten Ausführen des Befehls **esadmin system startall** oder **esadmin system stopall** wird die von Ihnen modifizierte Komponente entsprechend den Änderungen, die Sie in der Merkmaldatei vorgenommen haben, nicht gestartet bzw. nicht gestoppt.

**Beispiel:** Im folgenden Beispiel wird der HTTP-Server mit dem Befehl **esadmin system startall** gestartet (Standardeinstellung), aber nicht mit dem Befehl **esadmin system stopall** gestoppt (dies wird über die in diesem Beispiel hervorgehobene Zeile gesteuert):

```
#####
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
```

#### Zugehörige Tasks

„Starten eines Systems für die Unternehmenssuche“ auf Seite 319

„Stoppen eines Systems für die Unternehmenssuche“ auf Seite 321

---

## Verwalten der Suchserver im Standalone-Modus

Wollen Sie die Hochverfügbarkeit der Suchserver sicherstellen, können Sie die Suchserver für einzelne Objektgruppen starten, auch wenn der Indexserver nicht aktiv ist.

### Einschränkungen

Sie können die Suchserver nur als Administrator für die Unternehmenssuche stoppen und starten.

Das Starten und Stoppen von Suchservern im Standalone-Modus ist über die Administrationskonsole für die Unternehmenssuche nicht möglich. Wenn der Indexserver nicht aktiv ist, kann auf die Administrationskonsole nicht zugegriffen werden.

Bevor Sie die Suchserver im Standalone-Modus starten und stoppen können, müssen sie sicherstellen, dass die Crawler-, Parser-, Index- und Suchserver vorher zumindest einmal gestartet worden sind. Dies ist für die Synchronisierung der Dateien auf den Suchservern erforderlich.

Wenn die Sicherheit auf Dokumentenebene für die Objektgruppe aktiviert ist, stellen Sie sicher, dass der Crawler-Server ebenfalls gestartet ist. Dies ist erforderlich, damit die Maßnahmen für die Sicherheit auf Dokumentenebene umgesetzt werden können. Wenn der Crawler-Server nicht aktiv ist, werden in den Suchergebnissen nur Dokumente zurückgegeben, für die keine Authentifizierung erforderlich ist.

### Informationen zu dieser Task

Wenn der Indexserver nicht verfügbar ist, können Sie durch Starten der Suchserver im Standalone-Modus sicherstellen, dass die Benutzer das System weiterhin durchsuchen können. Bei einer Installation mit mehreren Servern können Sie die Befehle zum Starten und Stoppen der Suchserver über jeden Server für die Unternehmenssuche ausführen. Die Befehle versuchen, die Suchserver für die angegebene Objektgruppe auf allen verfügbaren Suchservern zu starten bzw. zu stoppen.

### Vorgehensweise

Gehen Sie wie folgt vor, um die Suchserver im Standalone-Modus zu starten oder zu stoppen:

1. Gehen Sie wie folgt vor, um die Suchserver für eine Objektgruppe zu starten, wenn der Indexserver nicht aktiv ist:
  - a. Melden Sie sich am Crawler-Server und an den Suchservern als Administrator für die Unternehmenssuche an, und starten Sie dann den Service der allgemeinen Übertragungsschicht (CCL):

#### AIX, Linux oder Solaris

```
startccl.sh -bg
```

#### Windows-Eingabeaufforderung

```
startccl
```

#### Verwaltungstool für Windows-Services

Gehen Sie wie folgt vor, um die allgemeine Übertragungsschicht im Hintergrund zu starten:

- 1) Starten Sie die Windows-Services: **Start** → **Einstellungen** → **Systemsteuerung** → **Verwaltung** → **Dienste**.
- 2) Klicken Sie **IBM OmniFind Enterprise Edition** mit der rechten Maustaste an, und klicken Sie **Start** an.

- b. Führen Sie auf den Suchservern das Script **startServer** aus, das sich im Verzeichnis *was-installationsstammverzeichnis/AppServer/bin* befindet, um die Anwendung **ESSearchServer** in WebSphere Application Server zu starten:

#### AIX, Linux oder Solaris

```
./startServer.sh ESSearchServer
```

#### Windows

```
startServer ESSearchServer
```

- c. Geben Sie den folgenden Befehl ein. Dabei gibt *objektgruppen-id* die Objektgruppe an, zu der die zu startenden Suchserver gehören:

```
esadmin startSearch -cid objektgruppen-id
```
2. Gehen Sie wie folgt vor, um die Suchserver für eine Objektgruppe zu stoppen, wenn der Indexserver nicht aktiv ist:
  - a. Melden Sie sich als Administrator für die Unternehmenssuche auf einem beliebigen Server für die Unternehmenssuche an.
  - b. Geben Sie den folgenden Befehl ein. Dabei gibt *objektgruppen-id* die Objektgruppe an, zu der die zu stoppenden Suchserver gehören:

```
esadmin stopSearch -cid objektgruppen-id
```

### Zugehörige Tasks

„Starten eines Systems für die Unternehmenssuche“ auf Seite 319

„Stoppen eines Systems für die Unternehmenssuche“ auf Seite 321



---

## Überwachen der Aktivitäten der Unternehmenssuche

Wenn Sie System- und Objektgruppenaktivitäten überwachen, können Sie den Status verschiedener Prozesse anzeigen, potenzielle Probleme aufspüren oder Konfigurationseinstellungen anpassen, um die Leistung zu verbessern.

Über die Administrationskonsole für die Unternehmenssuche können Sie das System überwachen und nach Bedarf Optionen anpassen. Sie können für jede bedeutende Aktivität (Crawlersuche, Syntaxanalyse, Indexieren und Suche) detaillierte Statistiken anzeigen. Statistikdaten sind unter anderem durchschnittliche Antwortzeiten und Fortschrittsinformation, wie z. B. die Anzahl der bei einer Sitzung durchsuchten bzw. indexierten Dokumente.

Sie können die meisten Aktivitäten durch das Anklicken der entsprechenden Symbole stoppen und starten. Mit diesen Operationen können Sie eine Aktivität anhalten, ihre Konfiguration ändern oder einen Fehler beheben und die Verarbeitung zum gewünschten Zeitpunkt erneut starten.

### **Zugehörige Tasks**

„Starten eines Systems für die Unternehmenssuche“ auf Seite 319

„Stoppen eines Systems für die Unternehmenssuche“ auf Seite 321

---

## Schätzen der Anzahl Dokumente in einer Objektgruppe

Wenn Sie eine Objektgruppe für die Unternehmenssuche erstellen oder bearbeiten, geben Sie eine Schätzung über die zu erwartende Anzahl Dokumente in der Objektgruppe ab. Das System schätzt mithilfe dieser Zahl den Speicherplatz und die Datenträgerressourcen, die für diese Objektgruppe erforderlich sind, setzt damit aber keine Begrenzung der Objektgruppengröße um.

### **Vorbereitung**

Sie müssen die Rolle eines Administrators für die Unternehmenssuche haben oder ein Objektgruppenadministrator für diese Objektgruppe sein, um die geschätzte Größe einer Objektgruppe zu ändern.

### **Informationen zu dieser Task**

Wenn die Objektgruppe die von Ihnen geschätzte Größe erreicht, fügt das System dem Index trotzdem weiterhin Dokumente hinzu. Falls Sie Alerts für die Objektgruppe konfigurieren und die Option auswählen, dass Sie bei Überschreitung des Grenzwerts für die Dokumentanzahl im Index benachrichtigt werden, stimmt dieser Standardgrenzwert mit dem Wert überein, den Sie als Schätzung für die in der Objektgruppe enthaltene Anzahl Dokumente angegeben haben. Das System überwacht diese Schätzung sowie den von Ihnen angegebenen Prozentsatz für den Alert-Grenzwert und sendet eine E-Mail, wenn die für diese Objektgruppe konfigurierte maximale Dokumentanzahl bald erreicht wird.

### **Vorgehensweise**

Gehen Sie wie folgt vor, um eine Schätzung für die mögliche Größe einer Objektgruppe abzugeben:

1. Bearbeiten Sie eine Objektgruppe, wählen Sie die Seite **Allgemein** aus, und klicken Sie **Allgemeine Optionen konfigurieren** an.
2. Geben Sie in das Feld **Geschätzte Anzahl Dokumente** eine Zahl für die von Ihnen erwartete Größe der Objektgruppe ein. Der Standardwert ist 1 000 000 Dokumente.

---

## Überwachen einer Objektgruppe

Sie können allgemeine Informationen zum Status jeder Komponente in einer Objektgruppe anzeigen oder Optionen auswählen, um detaillierte Informationen zu einzelnen Komponenten und URIs anzuzeigen.

### Vorbereitung

Alle Benutzer mit Verwaltungsaufgaben der Unternehmenssuche können Objektgruppen überwachen. Sie müssen ein Administrator für die Unternehmenssuche, ein Objektgruppenadministrator der Objektgruppe oder ein Bediener der Objektgruppe sein, um Komponenten zu starten oder zu stoppen bzw. um Zeitpläne zu aktivieren oder zu inaktivieren.

### Vorgehensweise

Gehen Sie wie folgt vor, um eine Objektgruppe zu überwachen:

1. Suchen Sie in der Sicht **Objektgruppen** die Objektgruppe, die Sie überwachen wollen, und klicken Sie  **Überwachen** an. Informationen zum aktuellen Status der einzelnen Objektgruppenkomponenten werden angezeigt.

**Tipp:** Wenn Sie eine Objektgruppe bearbeiten und sich bereits auf der Seite **Allgemein** befinden, können Sie  **Überwachen** anklicken, um zur Sicht für die Überwachung der Objektgruppe zu wechseln.

2. Wenn Sie detaillierte Informationen zu einem URI anzeigen wollen, klicken Sie  **URI-Details** an.

Sie wollen z. B. prüfen, ob sich ein bestimmter URI im Index befindet oder ob der Index, in dem sich der URI befindet, auf die Suchserver kopiert wurde.

3. Wenn Sie eine einzelne Komponente überwachen und detaillierte Angaben zur Aktivität der betreffenden Komponente anzeigen wollen, klicken Sie das Symbol **Status** an.

---

## Anzeigen von Details zu einem URI

Sie können detaillierte Informationen zu einem URI anzeigen. Sie können aktuelle und protokollierte Informationen dazu anzeigen, wie das Dokument, das durch diesen URI dargestellt wird, durchsucht, indexiert und gesucht wird.

### Vorbereitung

Bevor Sie eine Anforderung zum Anzeigen eines URI-Berichts oder Senden eines Berichts an eine E-Mail-Adresse übergeben, müssen Sie sicherstellen, dass die Komponente aktiv ist, zu der Sie Informationen anfordern. Wenn Sie z. B. Details dazu anzeigen wollen, wie ein Dokument durchsucht, indexiert und gesucht wird, müssen Sie sicherstellen, dass der Web-Crawler, der Indexserver und die Suchserver aktiv sind. Zum Verfolgen eines gelöschten Dokuments stellen Sie sicher, dass die Protokollierungsoptionen für die Dokumentüberwachung konfiguriert sind.

### Informationen zu dieser Task

Das Erfassen von Informationen zu einem URI ist ein zeitintensiver Prozess. Sie können eine Option zum Anzeigen der gewünschten Informationen auswählen und warten, bis diese angezeigt werden. Es ist allerdings effizienter, den Bericht an eine von Ihnen angegebene E-Mail-Adresse zu senden.

Bevor Sie einen Bericht empfangen können, müssen Sie sicherstellen, dass Informationen zu Ihrem E-Mail-Server für die Unternehmenssuche konfiguriert worden sind. Sie geben diese Informationen an, während Sie die E-Mail-Optionen auf der Seite **Protokoll** der Sicht **System** konfigurieren.

Der Indexserver und die Suchserver können Informationen zu allen URIs bereitstellen (beispielsweise ob sich ein URI im Index befindet und ob er auf die Suchserver kopiert wurde). Wenn Sie Informationen dazu anzeigen wollen, wie ein Dokument durchsucht wurde, müssen Sie den URI eines von einem Web-Crawler durchsuchten Dokuments angeben.

### Vorgehensweise

Gehen Sie wie folgt vor, um Details zu einem URI anzuzeigen:

1. Suchen Sie in der Sicht **Objektgruppen** die Objektgruppe, die Sie überwachen wollen, und klicken Sie  **Überwachen** an.

**Tipp:** Wenn Sie eine Objektgruppe bearbeiten und sich bereits auf der Seite **Allgemein** befinden, können Sie  **Überwachen** anklicken, um zur Sicht für die Überwachung der Objektgruppe zu wechseln.

2. Klicken Sie  **URI-Details** an.
3. Geben Sie auf der Seite **URI-Details** den URI ein, zu dem Sie Informationen anzeigen wollen.
4. Wählen Sie die Markierungsfelder der Informationen aus, die Sie anzeigen wollen:

#### **Crawlerdetails (nur für Web-Crawler verfügbar)**

Wählen Sie dieses Markierungsfeld aus, wenn Sie Informationen dazu, wie ein Dokument von einem Web-Crawler durchsucht wurde, und Informationen zu seinem aktuellen Status im Crawlerbereich anzeigen wollen.

#### **Indexdetails**

Wählen Sie dieses Markierungsfeld aus, wenn Sie anzeigen wollen, ob ein Dokument indexiert und auf die Suchserver kopiert wurde.

#### **Suchdetails**

Wählen Sie dieses Markierungsfeld aus, um Informationen dazu anzuzeigen, wie das Dokument gesucht werden kann, und ob das Dokument für die Suche zur Verfügung steht.

#### **Durch den Parser gelöschte Dokumente**

Wählen Sie dieses Markierungsfeld aus, um zu prüfen, ob das Dokument aus dem System für die Unternehmenssuche gelöscht wurde, während es syntaktisch analysiert wurde, und, falls dies zutrifft, warum es gelöscht wurde.

#### **Aus dem Index gelöschte Dokumente**

Wählen Sie dieses Markierungsfeld aus, um zu prüfen, ob ein Dokument aus dem System für die Unternehmenssuche gelöscht wurde, während es indexiert oder analysiert wurde, und, falls dies zutrifft, warum es gelöscht wurde.

5. Wenn Sie warten wollen, bis der Bericht angezeigt wird, klicken Sie **Bericht anzeigen** an.
6. Wenn Sie den Bericht an eine E-Mail-Adresse senden und später ansehen wollen, klicken Sie **Bericht senden** an.
  - a. Geben Sie auf der Seite **Detaillierten URI-Bericht senden** eine E-Mail-Adresse für den Empfang des Berichts in das Feld **E-Mail-Adresse für Benachrichtigung** ein.
  - b. Klicken Sie **Bericht senden** an.

#### Zugehörige Konzepte

„Dokumentüberwachung“ auf Seite 346

#### Zugehörige Tasks

„Anzeigen von Berichten über gelöschte Dokumente“ auf Seite 348

#### Zugehörige Verweise

„URI-Formate in einem Index für die Unternehmenssuche“ auf Seite 130

---

## Überwachen von Crawlern

Sie können allgemeine Informationen zum Status der einzelnen Crawler in der Objektgruppe anzeigen oder Optionen auswählen, um detaillierte Informationen zur Crawleraktivität anzuzeigen.

### Vorbereitung

Wenn sich Ihre Verwaltungsrolle auf das Überwachen von Objektgruppen beschränkt, können Sie die Crawlerstatistik zwar anzeigen, nicht jedoch das Verhalten eines Crawlers ändern (wie z. B. einen Crawler starten oder stoppen).

### Vorgehensweise

Gehen Sie wie folgt vor, um einen Crawler zu überwachen:

1. Suchen Sie in der Sicht **Objektgruppen** die Objektgruppe, die Sie überwachen wollen, und klicken Sie  **Überwachen** an.
2. Öffnen Sie die Seite **Crawler**.

**Tipp:** Wenn Sie eine Objektgruppe bearbeiten und sich bereits auf der Seite **Crawler** befinden, können Sie  **Überwachen** anklicken, um zur Sicht für die Crawlerüberwachung zu wechseln.

3. Wenn der Crawler aktiv ist oder angehalten wurde und Sie detaillierte Statusinformationen zum Crawler anzeigen wollen, klicken Sie  **Details** an. Je nach Crawlertyp wird eine andere Statistik angezeigt.

Wenn es Ihre Verwaltungsrolle für eine Objektgruppe erlaubt, können Sie einen Crawler starten, stoppen und anhalten, während Sie Details zur Crawleraktivität anzeigen. Wenn der Crawler terminiert werden kann, können Sie auch den Zeitplan für die Crawlersuche aktivieren oder inaktivieren.

4. Wenn ein Crawler gestoppt ist oder angehalten wurde, und Sie eine Crawler-sitzung starten wollen, klicken Sie  **Starten** oder **Fortsetzen** an.

#### Für Web-Crawler:

Wurde der Crawler gestoppt, beginnt er erneut mit der Crawlersuche und durchsucht den gesamten Crawlerbereich. Wurde der Crawler angehalten, nimmt er die Crawlersuche am Anfang des Ziels wieder auf, bei dem er angehalten wurde.

Klicken Sie das Symbol **Details** und anschließend das Symbol **Vollständige erneute Crawlersuche starten** an, wenn Sie den Crawler zwingen wollen, unverzüglich eine vollständige Crawlersuche zu starten. Der Crawler beginnt mit dem Durchsuchen des gesamten Crawlerbereichs, einschließlich der Seiten, die sich nicht geändert haben, seitdem sie das letzte Mal durchsucht wurden. Sie wollen möglicherweise alle Dokumente erneut durchsuchen, wenn Sie beispielsweise die Regeln für die Syntaxanalyse von Dokumenten geändert haben und diese Regeln auf zuvor indexierte Dokumente angewendet werden sollen.

#### **Für alle übrigen Crawlertypen:**

Wurde der Crawler gestoppt, beginnt er zum terminierten Datum und zur terminierten Zeit mit der Crawlersuche. Wenn der Crawler zum ersten Mal eine Datenquelle durchsucht, führt er eine vollständige Crawlersuche durch. Wird eine terminierte Crawlersuche wiederholt, durchsucht der Crawler entweder alle Aktualisierungen der Datenquelle (hinzugefügte Dokumente, Löschen und Änderungen) oder nur hinzugefügte Dokumente und Änderungen. Sie können den Typ der Crawlersuche im Crawlerzeitplan konfigurieren.

Wenn Sie den Crawler nicht terminiert haben, oder wenn Sie den Crawler früher starten wollen, klicken Sie das Symbol **Details** an. Klicken Sie anschließend im Bereich mit den Crawlerbereichsdetails das Symbol für den Typ der Crawlersuche an, die gestartet werden soll: Vollständige Crawlersuche, alle Aktualisierungen oder nur neue bzw. modifizierte Dokumente. Sie müssen das entsprechende Startsymbol für jede Datenquelle anklicken, die durchsucht werden soll (wie zum Beispiel ein Server, eine Datenbank oder ein Unterordner).

5. Wenn der Crawler aktiv ist und Sie ihn stoppen wollen, klicken Sie  **Stoppen** oder  **Anhalten** an. Der Crawler stoppt das Durchsuchen der Daten, bis sie den Crawler erneut starten oder die Suche wieder aufnehmen.

Wird der angehaltene Crawler fortgesetzt, nimmt er die Crawlersuche am Anfang des Ziels wieder auf, auf dem er angehalten wurde. So nimmt der DB2-Crawler beispielsweise die Crawlersuche in der ersten Zeile der Tabelle wieder auf, die er gerade durchsucht hat, als er angehalten wurde.

---

## **Anzeigen von Details zur Web-Crawler-Aktivität**

Wenn Sie Details zur Web-Crawler-Aktivität anzeigen, können Sie die Gesamtleistung besser einschätzen und die Web-Crawler-Merkmale sowie die Crawlerbereichsdefinitionen nach Bedarf anpassen.

### **Vorbereitung**

Alle Benutzer mit Verwaltungsaufgaben für die Unternehmenssuche können die Crawleraktivitäten überwachen. Sie müssen Mitglied der Rolle eines Administrators für die Unternehmenssuche, ein Objektgruppenadministrator für die Objektgruppe oder ein Bediener für die Objektgruppe sein, um einen Crawler zu starten oder zu stoppen.

### **Vorgehensweise**

Gehen Sie wie folgt vor, um Details für die Aktivität eines Web-Crawlers anzuzeigen:

1. Suchen Sie in der Sicht **Objektgruppen** die Objektgruppe, die Eigner des zu überwachenden Web-Crawlers ist, und klicken Sie  **Überwachen** an.

2. Öffnen Sie die Seite **Crawler**.

**Tipp:** Wenn Sie eine Objektgruppe bearbeiten und sich bereits auf der Seite **Crawler** befinden, können Sie  **Überwachen** anklicken, um zur Sicht für die Crawlerüberwachung zu wechseln.

3. Wenn der zu überwachende Web-Crawler aktiv oder angehalten ist, klicken Sie  **Details** an.

4. Wählen Sie auf der Seite mit den Details für den Web-Crawler die folgenden Optionen aus oder zeigen Sie sie an, um ausführliche Statistiken über die aktuelle und die vergangene Crawleraktivität anzuzeigen.

- Klicken Sie **Threddetails** an, um anzuzeigen, wie viele Threads aktiv die Websites durchsuchen, und wie viele sich im inaktiven Status befinden.
- Klicken Sie **Aktive Sites** an, um Informationen zu den Websites anzuzeigen, die der Crawler aktiv durchsucht.
- Klicken Sie **Zuletzt durchsuchte URL-Adressen** an. Diese Informationen zeigen an, was der Crawler zuletzt durchsucht hat. Falls die Elemente in der Liste sich nicht ändern, wenn Sie die Sicht aktualisiert anzeigen, findet keine Crawlersuche statt.
- Klicken Sie **Crawlerverlauf** an, um Berichte über vergangene Crawleraktivitäten anzuzeigen.
- Geben Sie im Bereich **URL-Status** eine URL-Adresse ein, für die Sie Informationen anzeigen möchten.
  - a. Klicken Sie **URL-Details** an, um die Statusinformationen für die URL-Adresse anzuzeigen. Sie können URL-Details nur für URL-Adressen anfordern, die zuvor mit einem Crawler durchsucht wurden.
  - b. Klicken Sie **Sitedetails** an, um Informationen anzugeben, die in einem Bericht über die Website enthalten sein sollen, zu der die URL-Adresse gehört. Sie können Sitedetails für eine zuvor mit einem Crawler durchsuchte Website oder für eine noch nicht durchsuchte Website anfordern.

Verwenden Sie diese Option z. B., wenn Sie sehen möchten, ob eine URL sich im Crawlerbereich befindet, ob sie durchsucht oder lediglich aufgespürt wurde, wann sie erneut durchsucht werden soll, und um Informationen zu erhalten, wann der letzte Versuch stattgefunden hat, die Website zu durchsuchen. Sie können auch den Inhalt der Datei `robots.txt` für die Website anzeigen, was Ihnen möglicherweise hilft zu ermitteln, warum die Site nicht durchsucht wird.

## Details von Web-Crawler-Threads

Sie können den Web-Crawler überwachen, um anzuzeigen, wie viele Threads aktiv die Websites durchsuchen, und wie viele sich im inaktiven Status befinden.

Wenn Sie Details zu einem Web-Crawler anzeigen, während Sie eine Objektgruppe überwachen, können Sie den Status des Crawler-Threads anzeigen. Zu den Status, die Sie höchstwahrscheinlich sehen werden, gehören die Folgenden:

### Warten

Gibt an, dass für den Thread keine URL-Adresse zum Durchsuchen verfügbar ist. Dieser Zustand kann auftreten, wenn ein Thread einen Suchvorgang beendet und der Crawler keine weiteren URLs finden kann, die schnell genug durchsucht werden können. Wenn z. B. der Wert für das Crawlermerkmal zu hoch ist, das steuert, wie lange der Crawler warten

muss, bevor eine weitere Seite von derselben Site abgerufen werden kann, kann dieses Merkmal verhindern, dass die URLs schnell genug zur Verfügung gestellt werden.

**Abruf** Gibt an, dass der Thread gerade eine Seite von einer Website herunterlädt.

**Beendet**

Gibt an, dass der Thread gerade die Seiten, die er durchsucht hat, an den Rest des Crawlers sendet, aber noch nicht bereit ist, eine weitere URL zu durchsuchen.

**Ausgesetzt**

Gibt an, dass der Crawler angehalten wurde.

Idealerweise rufen alle Threads ununterbrochen Seiten ab. Wenn sich ein Thread häufig im Status "Beendet" befinden, hat die Datenbank möglicherweise Durchsatzprobleme.

Wenn Threads sich häufig im Wartestatus befinden, sollten Sie den Wert überprüfen, den Sie im Feld **Maximale Anzahl aktiver Hosts** in den Crawlermerkmalen angegeben haben. Wenn der Wert niedrig ist, befinden sich möglicherweise nicht genug Sites im Crawlerbereich, damit die Threads beschäftigt sind, oder es sind nicht genug auswählbare URLs zum Durchsuchen vorhanden. Zu den Bedingungen, die eine niedrige Aktivität verursachen können, gehören DNS-Suchfunktionsfehler und Robot-Suchfunktionsfehler.

## Aktive Web-Crawler-Sites

Sie können den Web-Crawler überwachen, um Informationen zu Websites anzuzeigen, die der Crawler gerade durchsucht.

Wenn Sie Details zu einem Web-Crawler anzeigen, während Sie eine Objektgruppe überwachen, können Sie Statistikdaten über die aktiven Sites anzeigen. Die Statistikdaten zeigen Folgendes:

- Wie viele URLs der Crawler aus seiner internen Datenbank in den Speicher für das Durchsuchen zu diesem Zeitpunkt gestellt hat.
- Wie viele URLs der Crawler bisher versucht hat, zu durchsuchen.
- Wie viel Zeit für diese Iteration des Crawlers verbleibt, bevor eine Site inaktiviert und aus dem Speicher entfernt wird.
- Wie lange eine Site bisher im Speicher vorhanden war.

Diese Informationen ändern sich von Moment zu Moment, in dem der Crawler durch die Crawlersuchregeln fortschreitet, die für ihn konfiguriert wurden. Im Idealfall ist die Anzahl aktivierter URLs dicht an dem Wert, der für das Feld **Maximale Anzahl aktiver Hosts** in den Crawler-Speichermerkmalen konfiguriert ist.

Wenn die Anzahl aktivierter URLs nahe null ist, findet der Crawler keine auswählbaren URLs. Zu den Bedingungen, die eine solch niedrige Aktivität verursachen können, gehören DNS-Suchfunktionsfehler, Netzkonnektivitätsprobleme, Datenbankfehler und Probleme bei der Crawlerbereichsdefinition. Beispiel:

- Wenn sich viele Sites bereits lange im Speicher befinden und wenige URLs bisher durchsucht wurden, sollten Sie nach Netzkonnektivitätsproblemen suchen.
- Wenn nicht genug Sites in der Liste sind, suchen Sie nach Problemen bei der Crawlerbereichsdefinition oder DNS-Suchfunktionsproblemen.
- Wenn Sites in einer angemessenen Geschwindigkeit durchsucht werden, den Speicher jedoch verlassen, bevor viele URLs durchsucht werden konnten, bear-

beiten Sie die Crawler-Speichermerkmale und passen die Zeitlimitwerte im Feld **Zeitdauer, die jeder Host aktiv bleiben kann** an, um die Sites länger im Speicher zu behalten.

## Suchrate des Web-Crawlers

Sie können den Web-Crawler überwachen, um Informationen darüber anzuzeigen, wie schnell der Crawler Seiten aus Websites herunterlädt.

Wenn Sie Details zu einem Web-Crawler anzeigen, während Sie eine Objektgruppe überwachen, können Sie Statistiken anzeigen, wie schnell der Crawler Daten durchsucht (die Suchrate). Sie können auch Statistiken darüber anzeigen, wie viele URLs der Crawler seit dem Beginn der aktuellen Sitzung durchsucht hat.

Die Suchrate ist die Anzahl Seiten, die pro Sekunde durchsucht werden. Diese Anzahl korreliert mit mehreren Merkmalen, die Sie für den Web-Crawler konfigurieren können:

- Die Anzahl Crawler-Threads
- Die Anzahl aktiver Sites
- Die Zeit, die der Crawler warten muss, bis er eine weitere Seite desselben Web-Servers abrufen kann

Wenn der Crawler eine aktive Site pro Crawler-Thread hat und der Crawler zwei Sekunden warten muss, bevor er eine weitere Seite vom selben Web-Server abrufen kann, kann der Crawler nicht schneller als eine Seite pro Thread in zwei Sekunden durchsuchen. Wenn z. B. der Crawler die Standardanzahl von Threads verwendet (200), kann der Crawler 100 Seiten pro Sekunde für 200 Threads durchsuchen.

Wenn doppelt so viele aktive Sites wie Crawler-Threads vorhanden sind und der Crawler zwei Sekunden warten muss, bevor er eine weitere Seite vom selben Web-Server abrufen kann, kann der Crawler eine Seite pro Thread pro Sekunde erreichen. Jedoch würden dann die Übertragungsgeschwindigkeit im Netz für Downloads und der Datenbankdurchsatz zu begrenzenden Faktoren werden. Ein Hinweis auf eine gute Crawlerleistung liegt dann vor, wenn die Suchrate mit der Anzahl Crawler-Threads, der aktiven Sites und der Crawlerwartezeit abgestimmt ist.

Ein weiterer Faktor, den Sie bei der Überwachung der Web-Crawler-Leistung prüfen müssen, ist die Anzahl URLs, die der Crawler seit dem Beginn der aktuellen Crawler Sitzung durchsucht hat. Teilen Sie diese Anzahl durch die gesamte Zeit, die der Crawler aktiv war, um so den durchschnittlichen langfristigen Durchsatz zu errechnen. Falls diese Zahl nicht ansteigt, hat der Crawler die Operation entweder beendet, oder er kann nicht fortfahren. So können z. B. Netzkonnektivitätsprobleme, Datenbankfehler und DNS-Suchfunktionsfehler den Verarbeitungsfortschritt des Crawlers blockieren.

## Erstellen von Web-Crawler-Berichten

Wenn Sie Berichte über vergangene Web-Crawler-Aktivitäten prüfen, können Sie die Gesamtleistung besser einschätzen und die Web-Crawler-Merkmale sowie die Crawlerbereichsdefinitionen nach Bedarf anpassen.

### Vorbereitung

Wenn Ihre Verwaltungsrolle auf die Überwachung von Objektgruppen beschränkt ist, können Sie Crawlerstatistikdaten anzeigen und Berichte über die Crawleraktivität erstellen, aber Sie können das Verhalten des Crawlers nicht beeinflussen (z. B. den Crawler starten oder stoppen).

### Informationen zu dieser Task

Die unterschiedlichen Berichtstypen stellen Ihnen Informationen zur Web-Crawler-Aktivität zur Verfügung. Für bestimmte Berichtstypen werden die Informationen so schnell zurückgegeben, wie sie aus der internen Datenbank des Crawlers abgerufen werden können. Die Erstellung des Siteberichts und des HTTP-Statuscodeberichts benötigt eine gewisse Zeit. Wenn Sie diese Berichtstypen erstellen, können Sie eine E-Mail-Adresse für den Empfang des Berichts angeben, statt darauf zu warten, dass die Ergebnisse an die Administrationskonsole für die Unternehmenssuche zurückgegeben werden.

Wenn Sie Informationen zum Interpretieren der Statistikdaten in den Berichten benötigen, klicken Sie **Hilfe** an, während Sie den Web-Crawler überwachen und die Berichte erstellen.

### Vorgehensweise

Gehen Sie wie folgt vor, um Web-Crawler-Berichte zu erstellen:

1. Suchen Sie in der Sicht **Objektgruppen** die Objektgruppe, die Eigner des zu überwachenden Web-Crawlers ist, und klicken Sie  **Überwachen** an.
2. Öffnen Sie die Seite **Crawler**.

**Tipp:** Wenn Sie eine Objektgruppe bearbeiten und sich bereits auf der Seite **Crawler** befinden, können Sie  **Überwachen** anklicken, um zur Sicht für die Crawlerüberwachung zu wechseln.

3. Falls der Web-Crawler, für den Sie Berichte erstellen möchten, aktiv oder angehalten ist, klicken Sie  **Details** an.
4. Wählen Sie auf der Seite mit den Details für den Web-Crawler eine Option für den Berichtstyp aus, den Sie erstellen möchten:
  - Klicken Sie im Bereich **Crawlerstatusübersicht** die Option **Crawlerverlauf** an, um Berichte über den Crawler und sämtliche Sites zu erstellen, die er aufspürt oder durchsucht.
  - Geben Sie im Bereich **URL-Status** die URL-Adresse der Site an, für die Sie einen Bericht erstellen möchten, und klicken Sie anschließend **Sitedetails** an.
5. Sie können für Crawlerverlaufs- und Siteberichte die Markierungsfelder aller Statistikdaten auswählen, die Sie in einem Bericht sehen möchten, und anschließend **Bericht anzeigen** anklicken.

Für diese Statistiktypen gibt der Crawler einen Bericht an die Administrationskonsole so schnell zurück, wie er die Informationen aus seiner internen Datenbank abrufen kann.

6. Wenn Sie einen Crawlerverlaufsbericht erstellen, können Sie Optionen für das Erstellen eines Siteberichts angeben und anschließend **Bericht ausführen** anklicken.

Dieser Bericht wird auf der Basis der Statistiken erstellt, die Sie für die Aufnahme in den Bericht ausgewählt und in einer von Ihnen angegebenen Datei gespeichert haben (der Dateiname muss absolut sein). Sie können angeben, dass Sie nach dem Erstellen des Berichts eine E-Mail erhalten möchten.

7. Wenn Sie einen Crawlerverlaufsbericht erstellen, können Sie Optionen für das Erstellen eines HTTP-Statuscodeberichts angeben und anschließend **Bericht ausführen** anklicken.

Dieser Bericht enthält Informationen über die Anzahl pro Site ausgegebener HTTP-Statuscodes. Der Bericht wird in einer von Ihnen angegebenen Datei gespeichert (der Dateiname muss absolut sein). Sie können angeben, dass Sie nach dem Erstellen des Berichts eine E-Mail erhalten möchten.

Verwenden Sie diesen Bericht, um festzustellen, welche Sites eine große Anzahl Statuscodes 4xx (die auf nicht gefundene Seiten hinweisen), Statuscodes 5xx (die auf Serverprobleme hinweisen), Statuscodes 6xx (die auf Konnektivitätsprobleme hinweisen) usw. zurückgeben.

Dieser Bericht ist besonders nützlich, wenn der Crawler einige Zeit aktiv ist (z. B. ein Crawler, der seit Wochen aktiv ist). Durch ihn können Sie verschwundene Sites erkennen, neu angekommene Sites, Sites mit einer großen Anzahl an URLs (die möglicherweise auf redundantes Durchsuchen einer Lotus Notes-Datenbank hinweisen) sowie Sites mit einem rekursiven Dateisystem, die von einem HTTP-Server bedient werden. Wenn die Sites mit einer großen Anzahl HTTP-Statuscodes keine Daten zum Index beitragen, können Sie die Leistung des Crawlers verbessern, indem Sie diese Sites aus dem Crawlerbereich entfernen.

## An den Web-Crawler zurückgegebene HTTP-Statuscodes

Wenn Sie einen Web-Crawler überwachen, können Sie Informationen zu den HTTP-Statuscodes anzeigen, die der Crawler von den Seiten empfängt, die er zu durchsuchen versucht.

### Tabellenzusammenfassung

Wenn Sie den Web-Crawler-Verlauf oder den Status einer bestimmten URL-Adresse überwachen, können Sie die Informationen zu den HTTP-Statuscodes anzeigen, die an den Crawler zurückgegeben wurden. Sie können diese Informationen verwenden, um den Crawlerbereich zu verwalten und die Crawlerleistung zu optimieren. Wenn der Crawler z. B. eine große Anzahl HTTP-Statuscodes für eine URL empfängt und der Statuscode angibt, dass die Seiten unter dieser Adresse nicht durchsucht werden können, können Sie die Leistung verbessern, indem Sie die betreffende URL aus dem Crawlerbereich entfernen.

In der folgenden Tabelle werden die HTTP-Statuscodes und ihre Interpretation durch den Web-Crawler aufgelistet. Die Werte von 100 bis 505 sind Standard-HTTP-Statuscodes (weitere Informationen siehe <http://www.w3.org/Protocols/rfc2616/rfc2616.html>). Die übrigen HTTP-Statuscodes gehören zur Unternehmenssuche und zum Web-Crawler.

Tabelle 8. HTTP-Statuscodes vom Web-Crawler

Code	Beschreibung	Code	Beschreibung	Code	Beschreibung	Code	Beschreibung
NULL	Nicht durchsucht	400	Fehlerhafte Anforderung	500	Interner Serverfehler	693	Auswahl fehlgeschlagen (URLFetcher)
100	Weiter	401	Nicht berechtigt	501	Nicht implementiert	694	Schreibfehler (URLFetcher)
101	Protokolle wechseln	402	Zahlung erforderlich	502	Fehlerhaftes Gateway	695	Unvollständiger Blockheader (URLFetcher)

Tabelle 8. HTTP-Statuscodes vom Web-Crawler (Forts.)

Code	Beschreibung	Code	Beschreibung	Code	Beschreibung	Code	Beschreibung
200	Erfolgreich	403	Nicht erlaubt	503	Service nicht verfügbar	699	Unerwarteter Fehler (URLFetcher)
201	Erstellt	404	Nicht gefunden	504	Gateway-Zeitlimitüberschreitung	700	Parserfehler (Kein Headerende)
202	Akzeptiert	405	Methode nicht zulässig	505	HTTP-Version nicht unterstützt	710	Parserfehler (Header)
203	Nicht berechnete Informationen	406	Nicht zulässig	611	Lesefehler	720	Parserfehler (kein HTTP-Code)
204	Kein Inhalt	407	Proxy-Authentifizierung erforderlich	612	Verbindungsfehler	730	Parserfehler (Hauptteil)
205	Inhalt zurücksetzen	408	Zeitlimitüberschreitung der Anforderung	613	Lesezeitüberschreitung	740 oder 4044	Ausgeschlossen durch die Datei robots.txt
206	Teilinhalte	409	Konflikt	614	SSL-Handshake fehlgeschlagen	741	Robots vorübergehend nicht verfügbar
300	Multiple Choice	410	Fehlend	615	Anderer Lesefehler	760	Ausgeschlossen durch Crawlerbereichsdefinition
301	Die Site ist umgezogen	411	Länge erforderlich	616	FBA-Unregelmäßigkeit	761	Unzulässig für lokalen Crawlerbereich; zulässig für globalen Crawlerbereich
302	Gefunden	412	Bedingung fehlgeschlagen	617	Verschlüsselungsfehler	770	Fehlerhaftes Protokoll oder vom Standard abweichender Systemport
303	Siehe Sonstiges	413	Anforderungseinheit zu groß	618	Umleitung ohne Umleitungs-URL	780	Ausgeschlossen durch Dateitypausschlüsse
304	Nicht geändert	414	Anforderungs-URI zu groß	680	Störung der DNS-Suchfunktion	786	Ungültige URL
305	Proxy verwenden	415	Datenträgertyp wird nicht unterstützt	690	Fehlerhafte URL	2004	Kein Index-META-Tag
306	(Wird nicht verwendet)	416	Ungültiger Anfragebereich	691	Getrennte Verbindung (URLFetcher)	3020	Weiche Umleitung
307	Temporäre Umleitung	417	Erwartung ist fehlgeschlagen	692	Schreibzeitlimit (URLFetcher)		

## Anmerkungen zur Tabelle

### Statuscodes 4xx

Sie werden nur selten den Code 400 (fehlerhafte Anforderung) erhalten. Gemäß den HTTP-Statuscodestandards sollen die Codes 4xx angeben, dass der Client (= der Crawler) fehlgeschlagen ist. Das Problem liegt jedoch häufig beim Server oder einer URL, die der Crawler als Link empfangen hat. Einige Web-Server tolerieren z. B. keine URLs, die versuchen, über ein Stammverzeichnis der Site hoch zu navigieren (wie z. B. <http://xyz.ibm.com/../../sales>). Für andere Web-Server stellt diese Aufwärtsnavigation kein Problem dar, sie ignorieren den Operator des übergeordneten Verzeichnisses (..) bereits, wenn der Crawler im Stammverzeichnis ist.

Einige Server behandeln eine Anforderung an das Stammverzeichnis als einen Fehler, und einige obsoletere Verknüpfungen fordern möglicherweise Operationen an, die nicht mehr erkannt oder implementiert werden. Wenn vom Anwendungsserver eine Seite angefordert wird, die er nicht mehr bedient, löst der Anwendungsserver eine Ausnahmebedingung aus, wodurch der Web-Server den HTTP-Statuscode 400 zurückgibt, da die Anforderung als nicht mehr gültig betrachtet wird.

- 615** Gibt an, dass der Crawler-Server, der Daten von Websites herunterlädt, auf eine unerwartete Ausnahmebedingung gestoßen ist. Eine große Anzahl von Statuscodes dieses Typs kann darauf hindeuten, dass der Crawler einen Fehler hat.

### Statuscodes 61x

Mit Ausnahme des Codes 615 weisen Statuscodes 61x auf zu erwartende Probleme bei der Crawlersuche hin, wie z. B. Zeitlimitüberschreitungen. Für die folgenden Statuscodes sind möglicherweise Korrekturmaßnahmen erforderlich:

#### 611, 612 und 613

Diese Probleme werden möglicherweise durch langsame Sites oder eine schlechte Netzleistung verursacht.

- 611** Gibt an, dass ein Fehler aufgetreten ist, als der Crawler ein Dokument abgerufen hat.
- 612** Gibt an, dass ein Fehler aufgetreten ist, als der Crawler versuchte, eine Verbindung zu einem Web-Server herzustellen.
- 613** Gibt an, dass eine Zeitlimitüberschreitung aufgetreten ist, während der Crawler ein Dokument abrief.
- 614** Gibt an, dass der Crawler nicht in der Lage ist, sichere Sites (HTTPS-Sites) zu durchsuchen. Wenn Sie jedoch der Meinung sind, dass diese Sites zugänglich sind, überprüfen Sie, ob die Zertifikate auf dem Crawler-Server und dem Ziel-Web-Server korrekt eingerichtet sind. Wenn eine Site z. B. von einer anerkannten Zertifizierungsstelle zertifiziert wurde, können Sie dem Truststore, der vom Crawler verwendet wird, neue Zertifizierungsstellen hinzufügen.

Prüfen Sie auch, wie die selbst unterzeichneten Zertifikate der Sites konfiguriert sind, die Sie gerade durchsuchen möchten. Der Crawler ist so konfiguriert, dass selbst unterzeichnete Zertifikate akzeptiert werden. Einige Sites erstellen ein selbst unterzeichnetes Zertifikat für eine Stamm-URL (wie z. B. <http://sales.ibm.com/>) und versuchen dann, dieses Zertifikat für untergeordnete Domänen zu verwenden (wie z. B. <http://internal.sales.ibm.com/>). Der Crawler

kann in dieser Weise verwendete Zertifikate nicht akzeptieren. Er akzeptiert selbst unterzeichnete Zertifikate nur dann, wenn der Domänenname des Subjekts (sales.ibm.com) und der Unterzeichner des Zertifikats mit dem Domännennamen der Seite übereinstimmen, die angefordert wird.

- 616** Gibt an, dass das Anmeldeformular für die formularbasierte Authentifizierung (FBA) auch nach erneuter Authentifizierung immer noch im Download angezeigt wird.

Wenn der Crawler mit den in der FBA-Konfigurationsdatei bereitgestellten Informationen (Anmeldeformular und Authentifizierungsdaten wie z. B. Benutzername, Kennwort usw.) nicht authentifiziert werden kann, wird der Statuscode 616 allen Seiten zugeordnet, die von der formularbasierten Authentifizierung abhängig sind. Der Administrator muss untersuchen, warum die FBA-Konfiguration nicht funktioniert.

- 617** Gibt an, dass aus dem Byteinhalt des Dokuments keine Zeichenfolge erstellt werden kann, weil die Codierungszeichenfolge (Zeichensatz) ungültig ist oder das Dokument ungültige Byte enthält.

- 618** Gibt an, dass die Umleitungs-URL ungültig ist, wenn der Crawler die folgenden HTTP-Statuscodes empfängt. Möglicherweise ist die Position des HTTP-Antwortheaders ungültig.

301 Die Site ist umgezogen  
302 Gefunden

- 680** Gibt an, dass der Crawler die IP-Adressen für Hosts im Crawlerbereich nicht beschaffen konnte, möglicherweise aufgrund eines Problems beim Netzzugriff. Diese Art von Fehler bedeutet, dass der Crawler nicht in der Lage ist, ganze Sites zu durchsuchen und nicht etwa, dass er nur einzelne URLs nicht durchsuchen konnte. Eine große Anzahl Statuscodes dieses Typs verringert den Durchsatz erheblich.

#### **Statuscodes 69x**

Die Statuscodes 690 bis 699 werden nie in der persistenten Datenbank des Crawlers aufgezeichnet. Diese Codes stellen Ergebnisse dar, die nicht das tatsächliche Ergebnis eines Downloads von einem fernen Host wiedergeben, sondern vielmehr einen temporären Zustand im Crawler (z. B., dass eine Komponente heruntergefahren wird, während eine andere auf ein Ergebnis wartet oder ein Ergebnis sendet). Diese Statuscodes werden in manche Protokolle, aber nicht in den persistenten Datensatz aufgenommen und sollten daher nicht als Auswahlgruppenwerte verwendet werden.

#### **Statuscodes 7xx**

Die Statuscodes 7xx entstehen hauptsächlich aufgrund von Regeln im Crawlerbereich:

##### **710 - 730**

Geben an, dass der Crawler aufgrund von Problemen keinen vollständigen Download ausführen konnte oder dass der Crawler in einer Site auf ungültige HTML-Daten gestoßen ist. Wenn Sie viele Statuscodes dieses Typs erhalten, wenden Sie sich an den zuständigen Ansprechpartner der Unterstützung für die Unternehmenssuche.

##### **740 oder 4044**

Geben an, dass der Inhalt einer Datei nicht indiziert werden kann,

da das Dokument durch entsprechende Einschränkungen in der Datei robots.txt der Site ausgeschlossen wird.

**740** Gibt an, dass Ankerverknüpfungen, die auf das ausgeschlossene Dokument zeigen, in den Index aufgenommen werden können.

**4044** Gibt an, dass die Ankerverknüpfungen in Dokumenten, die auf das ausgeschlossene Dokument zeigen, ebenfalls vom Index ausgeschlossen sind.

**741** Gibt an, dass eine Site eine Datei robots.txt enthält, die das Durchsuchen zulässt, der Download ist jedoch fehlgeschlagen. Wenn es auch nach mehreren Wiederholungen nicht möglich ist, die URL zu durchsuchen, wird die URL aus dem Crawlerbereich entfernt. Bei einer großen Anzahl Statuscodes dieses Typs sollten Sie prüfen, ob die Zielsite vorübergehend oder dauerhaft nicht verfügbar ist. Wenn die Zielsite nicht mehr verfügbar ist, entfernen Sie sie aus dem Crawlerbereich.

Die anderen Statuscodes 7xx treten hauptsächlich auf, wenn Sie Änderungen am Crawlerbereich durchführen, nachdem der Crawler bereits einige Zeit aktiv war. Solche Statuscodes deuten normalerweise nicht auf Probleme hin, die Sie beheben müssten.

**3020** Gibt an, dass ein Dokument mit dem Statuscode 200 einen Positionsheder enthält, der den Benutzeragenten auf eine andere URL verweist.

---

## Überwachen des Parsers

Überwachen Sie den Parser, wenn es erforderlich ist, Informationen zu den vom Parser syntaktisch analysierten Dokumenten anzuzeigen, bevor sie dem Index für die Unternehmenssuche hinzugefügt werden. Optionen ermöglichen es Ihnen, Statistikdaten zu prüfen und die Parseraktivität zu verwalten.

### Vorbereitung

Wenn Ihre Verwaltungsrolle Sie auf das Überwachen von Objektgruppen beschränkt, können Sie den Parserstatus zwar anzeigen, den Parser jedoch nicht starten oder stoppen.

### Informationen zu dieser Task

Beim Überwachen der Parserdetails wird eine Momentaufnahme der Parseraktivität angezeigt, die Statistikdaten zu den Aktivitäten der Syntaxanalyse zu einem bestimmten Zeitpunkt bereitstellt. Die Statistik zeigt die Anzahl durchsuchter Dokumente an, die syntaktisch analysiert werden oder deren Syntaxanalyse noch ansteht. Weiterhin wird die Anzahl syntaktisch analysierter Dokumente angezeigt, die noch im Index gespeichert werden müssen.

Wenn der Parser aktiv ist, übermitteln Ihnen Nachrichten weitere Informationen zum Status des Parsers. Beispiel:

- Der Parser kann aktiv Dokumente syntaktisch analysieren.
- Der Parser kann inaktiv sein. Der Parser ist so lange inaktiviert, bis weitere Dokumente für die Syntaxanalyse verfügbar sind. Wenn Fehler auftreten, wartet der Parser darauf, dass er erneut gestartet wird. Der Parser startet sich selbst erneut, falls keine Parser-Services verfügbar sind (beispielsweise erfolgt ein auto-

matischer Neustart, wenn die Verbindung zum Parser-Service nicht hergestellt werden kann oder wenn alle Java Virtual Machines des Parsers mit anderen Objektgruppen beschäftigt sind).

- Der Parser kann angehalten sein (beispielsweise kann der Parser angehalten sein, bis eine Indexerstellung abgeschlossen ist).

### Vorgehensweise

Gehen Sie wie folgt vor, um den Parser für eine Objektgruppe zu überwachen:

1. Suchen Sie in der Sicht **Objektgruppen** die Objektgruppe, die Sie überwachen wollen, und klicken Sie  **Überwachen** an.
2. Öffnen Sie die Seite **Syntaxanalyse**.

**Tipp:** Wenn Sie eine Objektgruppe bearbeiten und sich bereits auf der Seite **Syntaxanalyse** befinden, können Sie  **Überwachen** anklicken, um zur Sicht für die Überwachung der Objektgruppe zu wechseln.

3. Wenn der Parser aktiv ist und Sie ausführliche Statusinformationen zur Parsingaktivität anzeigen wollen, klicken Sie  **Details** an.

Wenn es Ihnen Ihre Verwaltungsrolle ermöglicht, Prozesse für eine Objektgruppe zu verwalten, können Sie den Parser starten und stoppen, während Sie die Details zur Parsingaktivität anzeigen.

4. Wenn der Parser gestoppt ist und Sie ihn starten wollen, klicken Sie  **Starten** an.

Wenn Sie zum ersten Mal eine Objektgruppe erstellen, starten Sie den Parser erst, nachdem der Crawler mit der Crawlersuche nach Daten begonnen hat. Dadurch wird sichergestellt, dass dem Parser Daten für die Analyse und die Kategorisierung zur Verfügung stehen. Solange Sie keine Änderungen an den Parserregeln vornehmen, können Sie den Parser kontinuierlich ausführen lassen.

5. Wenn der Parser aktiv ist und Sie ihn stoppen wollen, klicken Sie  **Stoppen** an.

Wenn Sie Änderungen an den Parserregeln vornehmen, müssen Sie den Parser stoppen und erneut starten. Wenn Sie z. B. die Parserkonfiguration geändert haben, müssen Sie den Parser stoppen und erneut starten, bevor Ihre Änderungen wirksam werden.

---

## Überwachen der Indexaktivität für eine Objektgruppe

Überwachen Sie den Index einer Objektgruppe, wenn Sie den Fortschritt einer Indexerstellung verfolgen, den Indexzeitplan aktivieren oder inaktivieren bzw. die Indexaktivität starten oder stoppen wollen.

### Vorbereitung

Jeder Benutzer mit Verwaltungsaufgaben der Unternehmenssuche kann Indexaktivitäten überwachen. Sie müssen ein Administrator für die Unternehmenssuche, der Objektgruppenadministrator einer Objektgruppe oder ein Bediener der Objektgruppe sein, um eine Indexerstellung zu starten oder stoppen.

### Vorgehensweise

Gehen Sie wie folgt vor, um den Index einer Objektgruppe zu überwachen:

1. Suchen Sie in der Sicht **Objektgruppen** die Objektgruppe, die Sie überwachen wollen, und klicken Sie  **Überwachen** an.
2. Öffnen Sie die Seite **Index**.

**Tipp:** Wenn Sie eine Objektgruppe bearbeiten und sich bereits auf der Seite **Index** befinden, können Sie  **Überwachen** anklicken, um zur Sicht für die Überwachung der Objektgruppe zu wechseln.

3. Wenn ein Index terminiert ist und Sie seine Erstellung nicht zum terminierten Datum und Zeitpunkt wünschen, klicken Sie  **Zeitplan inaktivieren** an. Der Index wird erst erstellt, wenn Sie den Zeitplan aktivieren oder den Indexerstellungsprozess starten.
4. Wenn ein Index terminiert ist, aber der Zeitplan für seine Erstellung inaktiviert ist, klicken Sie  **Zeitplan aktivieren** an. Der Index wird für die Erstellung an dem Datum und zu der Zeit eingereiht, die Sie im Indexzeitplan angegeben haben.
5. Wenn ein Index gestoppt ist und Sie ihn starten wollen, klicken Sie  **Starten** an.

In der Regel tritt das Indexieren auf regelmäßig terminierter Basis auf. Wenn Sie einen Index während seiner Erstellung stoppen oder wenn Sie den Zeitplan für einen Index inaktivieren, können Sie **Starten** anklicken, um den Beginn der Indexerstellung zu erzwingen.

6. Wenn eine Indexerstellung aktiv ist und Sie diese stoppen wollen, klicken Sie  **Stoppen** an.

Unter Umständen ist es erforderlich, eine Deltaindexerstellung zu stoppen, z. B. um die Erstellung des Hauptindex zu erzwingen, nachdem Sie den in der Objektgruppe verwendeten Kategorisierungstyp geändert haben.

7. Wenn während einer Indexerstellung Fehler aufgetreten sind, klicken Sie  **Fehler** an.

Die Seite **Inhalt der Protokolldatei** wird angezeigt, damit Sie weitere Informationen zu Indexierungsfehlern anzeigen können. Sie können auf dieser Seite einzelne Fehlernachrichten auswählen, um Details zu diesem Problem anzeigen zu lassen.

---

## Überwachen der Indexierungswarteschlange für die Unternehmenssuche

Sie können den Status aller Indexerstellungen in der Indexierungswarteschlange anzeigen, eine Indexerstellung stoppen oder einen Index aus der Warteschlange löschen.

### Vorbereitung

Sie müssen ein Administrator für die Unternehmenssuche sein, um die Indexierungswarteschlange zu verwalten.

### Informationen zu dieser Task

Es können gleichzeitig mehrere Indizes erstellt werden, es darf sich jedoch jeweils nur ein Index für eine Objektgruppe in der Warteschlange befinden. Beim Konfigurieren der Indexoptionen für das System geben Sie an, wie viele Indizes die Warteschlange und die Ressourcen für die Indexerstellung gleichzeitig gemeinsam nutzen können.

## Vorgehensweise

Gehen Sie wie folgt vor, um die Indexierungwarteschlange zu überwachen:

1. Klicken Sie **System** an, um die Sicht **System** zu öffnen.
2. Wählen Sie die Seite **Index** aus.

Es wird eine Liste von Objektgruppen angezeigt, die Indizes haben, die sich in der Indexierungwarteschlange befinden. Für jeden Index wird der zu erstellende Indextyp (Delta oder vollständig), der Zeitpunkt, zu dem der Index in die Indexierungwarteschlange aufgenommen wurde, und die Startzeit der Indexerstellung (wenn die Erstellung bereits begonnen hat) angezeigt.

3. Wenn Sie einen einzelnen Index verwalten wollen, klicken Sie das Symbol **Status** an.

Sie wollen zum Beispiel anzeigen, wie lange es noch dauert, bis ein Index vollständig erstellt ist, wie viele Dokumente der Index enthält oder den Indexzeitplan inaktivieren.

4. Wenn Sie einen Index stoppen wollen, der gerade erstellt wird, klicken Sie  **Stoppen** an.

Wenn Sie z. B. Kategorisierungsregeln geändert haben, wollen Sie die Delta-indexerstellung stoppen, damit Sie stattdessen den Start einer Hauptindexerstellung erzwingen können.

Wenn Sie eine Indexerstellung starten wollen, nachdem Sie sie gestoppt haben, können Sie warten, bis der Index zu seiner nächsten terminierten Startzeit erneut in die Indexierungwarteschlange aufgenommen wird, oder Sie klicken das Symbol **Status** an, um den Index zu überwachen. Klicken Sie anschließend  **Starten** an, um eine Indexerstellung zu starten.

5. Wenn Sie einen Index aus der Indexierungwarteschlange entfernen wollen, klicken Sie  **Entfernen** an.

---

## Überwachen der Suchserver

Sie können detaillierte Statusinformationen zur Aktivität des Suchservers für eine bestimmte Objektgruppe oder detaillierte Statusinformationen zu den Suchservern für Ihr gesamtes System für die Unternehmenssuche anzeigen.

### Vorbereitung

Jeder Benutzer mit Verwaltungsaufgaben für die Unternehmenssuche kann Suchserver für die Objektgruppen überwachen, für deren Verwaltung er berechtigt ist. Sie müssen ein Administrator für die Unternehmenssuche sein, um alle Suchserver Ihres Systems für die Unternehmenssuche zu überwachen.

Sie müssen ein Administrator für die Unternehmenssuche, der Objektgruppenadministrator einer Objektgruppe oder ein Bediener der Objektgruppe sein, um einen Suchserver zu starten oder zu stoppen.

### Vorgehensweise

1. Gehen Sie wie folgt vor, um die Suchserver für eine einzelne Objektgruppe zu überwachen:
  - a. Suchen Sie in der Sicht **Objektgruppen** die Objektgruppe, die Sie überwachen wollen, und klicken Sie  **Überwachen** an.
  - b. Öffnen Sie die Seite **Suche**.

**Tipp:** Wenn Sie eine Objektgruppe bearbeiten und sich bereits auf der Seite **Suche** befinden, können Sie  **Überwachen** anklicken, um zur Sicht für die Überwachung der Objektgruppe zu wechseln.

2. Gehen Sie wie folgt vor, um alle Suchserver Ihres Systems für die Unternehmenssuche zu überwachen:
  - a. Klicken Sie **System** an, um die Sicht **System** zu öffnen.
  - b. Wählen Sie die Seite **Suche** aus.
3. Wenn ein Suchserver gestoppt ist und Sie ihn starten wollen, klicken Sie  **Starten** an.
4. Wenn ein Suchserver aktiv ist und Sie ihn stoppen wollen, klicken Sie  **Stoppen** an.

Wenn Sie den Suchcache aktivieren oder inaktivieren, Änderungen an der Suchcachegröße vornehmen oder Quick Links ändern, müssen Sie die Suchserver stoppen und erneut starten, damit Ihre Änderungen wirksam werden.
5. Wenn Sie eine Zusammenfassung über die Zeit anzeigen wollen, die ein Suchserver für die Verarbeitung von Suchanforderungen aufwendet, klicken Sie **Antwortzeit (Verlauf)** an.

Der Bericht zeigt die durchschnittliche Dauer in Millisekunden an, die der Suchserver benötigt hat, um an einem bestimmten Datum auf Suchanforderungen zu antworten.

Die durchschnittliche Antwortzeit dient als Bezugswert für die Systemleistung und entspricht der Servicequalität. Eine Erhöhung der Antwortzeit ist ein möglicher Hinweis darauf, dass das System einer Lastspitze ausgesetzt ist. Die Anzahl durchsuchter Objektgruppen und die Größe der Objektgruppen können z. B. die Leistung des Systems überschreiten.
6. Wenn Sie die Liste der am häufigsten übergebenen Abfragen anzeigen wollen, klicken Sie **Häufige Abfragen** an.

Der Bericht zeigt Ihnen die Schlüsselwörter der 50 am häufigsten übergebenen Abfragen an und wie oft eine bestimmte Abfrage übergeben wurde.

Durch das Prüfen der häufigsten Abfragen, können Sie feststellen, welche Quick Links sinnvoll sind. Durch Erstellen von Quick Links können Sie die Suchqualität für viele Benutzer verbessern. Sie können sicherstellen, dass suchrelevante Dokumente immer in den Suchergebnissen zurückgegeben werden.

Möglicherweise ist es auch sinnvoll, im Unternehmensportal Links zu den Ressourcen zu erstellen, die diese Abfragen beantworten. Wenn Benutzer z. B. häufig nach Informationen zu Kostenkonten suchen, fügen Sie einen Link zu der Seite hinzu, auf der die Vorgehensweise für Kostenkonten auf Ihrer Intranet-Homepage erläutert wird.
7. Wenn Sie die Liste der zuletzt übergebenen Abfragen anzeigen wollen, klicken Sie **Zuletzt durchgeführte Abfragen** an.

Der Bericht zeigt Ihnen die Schlüsselwörter der 50 zuletzt übergebenen Abfragen an.

Durch das Überprüfen der zuletzt übergebenen Abfragen können Sie aktuelle Trends verfolgen und dringende Situationen in Ihrem Unternehmen erkennen. Sie können z. B. eine Bedarfsspitze für ein bestimmtes Thema feststellen. Die Bedarfsspitze zeigt möglicherweise an, dass ein Quick Link zu diesem Thema sinnvoll ist oder dass Sie den Benutzern dieses Thema auf andere Weise zugänglich machen sollten (beispielsweise durch Bereitstellen eines Links im Unternehmensportal).

## Ändern der Berechnung von Abfragestatistiken

Sie können ändern, wie das System die Anzahl der häufigen Abfragen und der letzten Abfragen berechnet.

### Informationen zu dieser Task

Wenn Sie die Suchserver überwachen, können Sie Optionen auswählen, um eine Liste der 50 häufigsten Abfragen und eine Liste der 50 letzten verarbeiteten Abfragen anzuzeigen. In der Standardkonfiguration des Suchservers werden Abfragen, die eine funktional entsprechende Abfragezeichenfolge und unterschiedliche Bereichseinstellungen für die Ergebnisse haben, als unabhängige Abfragen gezählt. So wird beispielsweise der Abfragezähler erhöht, wenn ein Benutzer eine Option anklickt, um die nächste Ergebnisseite anzuzeigen.

Sie können die Berechnung der Abfragestatistik durch das System ändern, indem Sie die Datei `runtime-generic.properties` für den Suchserver bearbeiten. Wenn Sie den Parameter **distinctRecentQueryCheck** auf `true` setzen, zählt das System nur Abfragen, die die erste Seite mit Ergebnissen zurückgeben, als unabhängige Abfragen.

Sie müssen sich als Administrator für die Unternehmenssuche anmelden, um die Merkmale des Suchservers zu bearbeiten. Sie müssen ein Administrator für die Unternehmenssuche, der Objektgruppenadministrator einer Objektgruppe oder ein Bediener der Objektgruppe sein, um einen Suchserver zu starten oder zu stoppen.

### Vorgehensweise

Gehen Sie wie folgt vor, um die Berechnung der Abfragestatistik durch das System zu ändern:

1. Melden Sie sich als Administrator für die Unternehmenssuche an. Diese Benutzer-ID wurde bei der Installation von OmniFind Enterprise Edition angegeben. Melden Sie sich in einer Konfiguration mit mehreren Servern auf dem Suchserver an.
2. Bearbeiten Sie die folgende Datei, wobei *objektgruppen-id* die ID der Objektgruppe, die Sie ändern möchten, und *knoten-id* den Suchserver angibt, auf dem Sie die Änderungen vornehmen:  
`ES_NODE_ROOT/master_config/objektgruppen-id.runtime.knoten-id/runtime-generic.properties`
3. Setzen Sie den Wert des Parameters **distinctRecentQueryCheck** auf `true`, und speichern und schließen Sie die Datei.
4. Wiederholen Sie bei einer Konfiguration mit mehreren Servern die obigen Schritte auf dem zweiten Suchserver.
5. Wenn Sie das Verhalten der Abfragestatistik für weitere Objektgruppen ändern wollen, müssen Sie die obigen Schritte für jede Objektgruppe wiederholen, die Sie ändern wollen.
6. Damit die Änderungen wirksam werden, melden Sie sich an der Administrationskonsole für die Unternehmenssuche an, und starten Sie die Suchserver erneut:
  - a. Klicken Sie **System** an, um die Sicht **System** zu öffnen.
  - b. Wählen Sie die Seite **Suche** aus.
  - c. Klicken Sie für jeden geänderten Suchserver  **Stoppen** an.
  - d. Klicken Sie für jeden geänderten Suchserver  **Starten** an.

---

## Überwachen der Datenempfangsfunktion

Überwachen Sie die Datenempfangsfunktion, um ihren Status und Details zur Aktivität der Datenempfangsfunktions-Clientanwendung anzuzeigen.

### Vorbereitung

Sie müssen ein Administrator für die Unternehmenssuche sein, um die Datenempfangsfunktion zu überwachen.

**Wichtig:** Die Datenempfangsfunktion wird in zukünftigen Releases nicht unterstützt. Verwenden Sie SI-API (Search and Index API) anstelle der Anwendungsschnittstellen der Datenempfangsfunktion, um Clientanwendungen für die Unternehmenssuche zu entwickeln. Die folgenden Informationen werden für Benutzer bereitgestellt, die vorher Anwendungen der Datenempfangsfunktion erstellt haben.

### Vorgehensweise

Gehen Sie wie folgt vor, um die Datenempfangsfunktion zu überwachen:

1. Klicken Sie **System** an, um die Sicht **System** zu öffnen.
2. Zeigen Sie auf der Seite **Datenempfangsfunktion** die Statussymbole an, um zu prüfen, ob die Datenempfangsfunktion aktiv ist oder ob sie gestoppt wurde.
3. Wenn die Datenempfangsfunktion aktiv ist und Sie ausführliche Statusinformationen zur Aktivität der Clientanwendung anzeigen wollen, klicken Sie  **Details** an.

Statussymbole auf der Seite **Details der Datenempfangsfunktion** geben an, ob die Datenempfangsfunktion aktiv ist oder ob sie gestoppt wurde. Die Statistik zeigt an, wie viele Anforderungen noch zu verarbeiten sind, sie zeigt den aktuellen Status der Threads an, die Anforderungen von Clientanwendungen bearbeiten und wie viele Threads eines bestimmten Threadstatus aktiv sind.

4. Wenn Sie die Portnummer der Datenempfangsfunktion oder die maximale zulässige Anzahl Dokumente im temporären Speicher ändern, klicken Sie  **Erneut starten** an.

Die Datenempfangsfunktion wird gestartet, wenn das System für die Unternehmenssuche gestartet wird. Sie müssen die Datenempfangsfunktion erst erneut starten, wenn Sie mindestens eine dieser Konfigurationsoptionen ändern.

---

## Dokumentüberwachung

Dokumente können in verschiedenen Verarbeitungsschritten aus dem System gelöscht werden. Sie können Optionen angeben, um zu erfahren, wann ein Dokument gelöscht wurde und welche Probleme zum Löschen geführt haben.

Wenn der Parser einen Fehler feststellt, der die Syntaxanalyse eines Dokuments verhindert, wird eine Nachricht mit einem Ursachencode für das gelöschte Dokument protokolliert. (Dieser Fehlertyp hat nicht zur Folge, dass ältere Versionen des Dokuments aus dem Index entfernt werden.)

Dokumente können während der Indexierungsschritte gelöscht werden, und diese Informationen werden auch protokolliert. URIs und URI-Muster z. B. können explizit gelöscht werden. Ein Dokument kann unter Umständen von einem Crawler durchsucht worden sein, der später gelöscht wurde. Das Quelldokument

kann möglicherweise nicht länger vorhanden sein (ein negativer HTTP-Code wird dem Dokument zugeordnet) oder der dem Dokument zugeordnete HTTP-Code kann evtl. unbekannt sein. Dokumente können auch gelöscht werden, wenn die Rangfolgeinformationen für ein Dokument fehlen, das eine globale Analyse erfordert.

Wenn Sie wissen, dass ein Dokument durchsucht wurde, aber nicht im Index angezeigt wird, können Sie die Administrationskonsole für die Unternehmenssuche verwenden, um den Verarbeitungsablauf für das Dokument im System zu verfolgen. Detaillierte Berichte können zeigen, wann, wo und warum das Dokument gelöscht wurde. Beispiel: Im Bericht kann angegeben sein, dass das Dokument während der globalen Analyse unerwartet gelöscht wurde oder dass ein Administrator den URI aus dem Index gelöscht hat.

#### **Zugehörige Tasks**

„Anzeigen von Details zu einem URI“ auf Seite 328

## **Konfigurieren von Protokolldateien für die Dokumentüberwachung**

Sie können Protokolldateien zum Protokollieren von Informationen zu gelöschten Dokumenten konfigurieren, um zu ermitteln, wann, wo und warum ein Dokument aus dem System gelöscht wurde.

#### **Vorbereitung**

Sie müssen ein Administrator für die Unternehmenssuche oder ein Objektgruppenadministrator der Objektgruppe sein, um Optionen für das Protokollieren gelöschter Dokumente zu konfigurieren.

#### **Informationen zu dieser Task**

Damit verhindert wird, dass Protokolldateien zu viel Plattenspeicherplatz belegen, wechselt das System turnusmäßig die Protokolldateien und startet immer eine neue Protokolldatei, wenn sich das aktuelle Datum ändert. Wenn eine Protokolldatei ihre maximal zulässige Größe erreicht und das Datum sich nicht geändert hat, erstellt das System eine neue Protokolldatei. Wenn die maximale Anzahl Protokolldateien erreicht ist, wird die älteste Protokolldatei gelöscht, sodass eine neue erstellt werden kann.

#### **Vorgehensweise**

Gehen Sie wie folgt vor, um Protokolldateien für die Dokumentüberwachung zu konfigurieren:

1. Bearbeiten Sie eine Objektgruppe, wählen Sie die Seite **Protokoll** aus, und klicken Sie **Dokumentüberwachung konfigurieren** an.
2. Stellen Sie sicher, dass das Markierungsfeld für die Überwachung von Dokumenten auf der Seite **Dokumentüberwachung** ausgewählt ist.
3. Geben Sie die Anzahl Protokolldateien an, die zum Protokollieren von Informationen zu Dokumenten verwendet werden sollen, die aus dem System gelöscht wurden. Diese Protokolldateien werden von allen Sitzungen gemeinsam genutzt, in denen Dokumente gelöscht werden können.

## Anzeigen von Berichten über gelöschte Dokumente

Sie können detaillierte Informationen zu Dokumenten anzeigen, die aus einem System für die Unternehmenssuche gelöscht wurden. Diese Informationen sind nur verfügbar, wenn die Dokumentüberwachung für die Objektgruppe aktiviert ist.

### Vorbereitung

Bevor Sie eine Anforderung zum Anzeigen eines Berichts über gelöschte Dokumente oder zum Senden eines Berichts an eine E-Mail-Adresse übergeben, müssen Sie sicherstellen, dass die Sitzungen aktiv sind, zu denen Sie Informationen anfordern. Beispiel: Stellen Sie sicher, dass die Parser- und Indexsitzungen für die Objektgruppe gestartet werden, um etwas über Dokumente zu erfahren, die während der Syntaxanalyse oder der Indexierung gelöscht wurden.

Bevor Sie einen Bericht empfangen können, stellen Sie sicher, dass Informationen zu Ihrem E-Mail-Server für die Unternehmenssuche konfiguriert worden sind. Sie geben diese Informationen an, während Sie die E-Mail-Optionen auf der Seite **Protokoll** der Sicht **System** konfigurieren.

### Informationen zu dieser Task

Das Erfassen von Informationen zu gelöschten Dokumenten ist ein zeitintensiver Prozess. Sie können eine Option zum Anzeigen der Informationen auswählen und warten, bis diese angezeigt werden. Es ist allerdings effizienter, den Bericht an eine von Ihnen angegebene E-Mail-Adresse zu senden.

Wenn ein Dokument gelöscht wurde, zeigt der Bericht das Datum und die Uhrzeit an, an dem bzw. zu der das Dokument gelöscht wurde, die Bewertungsstufe des Fehlers, die Komponente und Sitzung, bei der das Problem auftrat, und die Fehlermeldung.

### Vorgehensweise

Gehen Sie wie folgt vor, um Details zu gelöschten Dokumenten anzuzeigen:

1. Suchen Sie in der Sicht **Objektgruppen** die Objektgruppe, die Sie überwachen wollen, und klicken Sie  **Überwachen** an.

**Tipp:** Wenn Sie eine Objektgruppe bearbeiten und sich bereits auf der Seite **Allgemein** befinden, können Sie  **Überwachen** anklicken, um zur Sicht für die Überwachung der Objektgruppe zu wechseln.

2. Klicken Sie  **URI-Details** an.
3. Geben Sie auf der Seite **URI-Details** den URI ein, zu dem Sie Informationen anzeigen wollen.
4. Wählen Sie die Markierungsfelder der Informationen aus, die Sie anzeigen wollen:

#### Durch den Parser gelöschte Dokumente

Wählen Sie dieses Markierungsfeld aus, um zu prüfen, ob das Dokument gelöscht wurde, während es syntaktisch analysiert wurde, und, falls dies zutrifft, warum es gelöscht wurde.

#### Aus dem Index gelöschte Dokumente

Wählen Sie dieses Markierungsfeld aus, um zu prüfen, ob ein Dokument gelöscht wurde, während es indiziert oder analysiert wurde, und, falls dies zutrifft, warum es gelöscht wurde.

5. Geben Sie an, wie der Bericht angezeigt werden soll:
  - Wenn Sie warten wollen, bis der Bericht angezeigt wird, klicken Sie **Bericht anzeigen** an.
  - Wenn Sie den Bericht an eine E-Mail-Adresse senden und später ansehen wollen, klicken Sie **Bericht senden** an.

Geben Sie auf der Seite **Detaillierten URI-Bericht senden** eine E-Mail-Adresse für den Empfang des Berichts in das Feld **E-Mail-Adresse für Benachrichtigung** ein, und klicken Sie anschließend **Bericht senden** an.

#### Zugehörige Tasks

„Anzeigen von Details zu einem URI“ auf Seite 328

#### Zugehörige Verweise

„URI-Formate in einem Index für die Unternehmenssuche“ auf Seite 130

## Anzeigen von Protokolldateien zu gelöschten Dokumenten

Sie können protokollierte Nachrichten zu Dokumenten anzeigen, die aus einem System für die Unternehmenssuche gelöscht wurden. Diese Informationen sind nur verfügbar, wenn die Dokumentüberwachung für die Objektgruppe aktiviert ist.

### Informationen zu dieser Task

Sie müssen den URI des Dokuments kennen, um einen Bericht über ein gelöschttes Dokument anzeigen zu können. Wenn Sie die Protokolldateien der gelöschten Dokumente anzeigen, sehen Sie das Datum und die Uhrzeit, an dem bzw. zu der die Dokumente gelöscht wurden, die Bewertungsstufe des Fehlers, die Komponente und die Sitzung, bei der das Problem auftrat, und die ausführliche Fehlermeldung.

### Vorgehensweise

Gehen Sie wie folgt vor, um Protokolldateien zu gelöschten Dokumenten anzuzeigen:

1. Suchen Sie in der Sicht **Objektgruppen** die Objektgruppe, die Sie überwachen wollen, und klicken Sie  **Überwachen** an.

**Tipp:** Wenn Sie eine Objektgruppe bearbeiten und sich bereits auf der Seite **Allgemein** befinden, können Sie  **Überwachen** anklicken, um zur Sicht für die Überwachung der Objektgruppe zu wechseln.

2. Klicken Sie  **Protokolldateien zu gelöschten Dokumenten** an.
3. Wählen Sie auf der Seite **Protokolldateien zu gelöschten Dokumenten** die Protokolldatei aus, die Sie anzeigen wollen. Der Name jeder Protokolldatei gibt an, ob das Dokument vom Parser (pd) oder während einer Indexerstellung (in) gelöscht wurde, und schließt das Datum ein, an dem die Datei erstellt wurde. Wenn mehrere Protokolldateien desselben Typs an demselben Datum erstellt werden, gibt ein numerisches Suffix die Reihenfolge an, in der die Dateien an dem Datum erstellt wurden. Beispiel:

```
dropped_doc_in_20060525.log
dropped_doc_pd_20060524.log (enthält die neuesten Einträge an diesem Datum)
dropped_doc_pd_20060524.log.1
dropped_doc_pd_20060524.log.2 (enthält die ältesten Einträge an diesem Datum)
```

4. Klicken Sie **Protokoll anzeigen** an.

Für jede Nachricht auf der Seite **Inhalt der Protokolldatei** werden die folgenden Informationen angezeigt: das Datum und die Zeit, zu der die Nachricht

erstellt wurde, die Bewertungsstufe der Nachricht, der Name der Sitzung, die die Nachricht abgesetzt hat, die Nachrichten-ID und der Fehlertext.

Sie können die entsprechenden Knöpfe anklicken, um zur ersten, letzten, vorherigen oder nächsten Seite der Protokolldatei zu wechseln. Sie können auch eine Seitenzahl angeben und direkt zu dieser Seite wechseln.

5. Klicken Sie  **Details** an, damit Ihnen detailliertere Informationen zu einer Nachricht angezeigt werden.

Auf der Seite **Details der Protokollnachricht** werden die folgenden Informationen angezeigt: der Hostname des Servers für die Unternehmenssuche, auf dem die Nachricht aufgetreten ist, der Name der Datei, die den Fehler verursacht hat, der Funktionsname und die Zeilennummer, in der der Fehler aufgetreten ist, die Prozess-ID und die Thread-ID.

Sie können Knöpfe anklicken, um zur nächsten bzw. vorherigen Nachricht in der Protokolldatei zu blättern.

---

## Protokolldateien und Alerts

Sie können die Nachrichtentypen auswählen, die Sie für eine Objektgruppe und für das System protokollieren möchten, sowie Optionen für das Erstellen und Anzeigen von Protokolldateien, für das Empfangen von Alerts und für das Empfangen von E-Mails beim Auftreten von Nachrichten angeben.

Bei normalen Operationen schreiben die Komponenten für die Unternehmenssuche die Protokollnachrichten in eine allgemeine Protokolldatei. Diese Protokolldatei befindet sich im Verzeichnis `ES_NODE_ROOT/logs` auf dem Indexserver. Mit der Administrationskonsole können Sie diese allgemeinen Protokolldaten anzeigen.

Wenn ein Problem wie z. B. ein Netzübertragungsfehler auftritt, schreiben die Komponenten Protokollnachrichten in ein Verzeichnis `logs` auf dem Server, auf dem die Komponenten installiert sind. Verwenden Sie zum Anzeigen dieser lokalen Protokolldateien eine Dateianzeigefunktion auf diesem Computer, wie z. B. das Dienstprogramm `tail` auf einem UNIX-System. Über die Administrationskonsole können Sie diese Protokolldateitypen nicht anzeigen.

Wenn Sie Protokolldateien konfigurieren, können Sie die Nachrichtentypen auswählen, die Sie protokollieren möchten (wie z. B. Fehler- oder Warnmeldungen); Sie können angeben, wie oft alte Protokolldateien gelöscht werden sollen, um Platz für neue Protokolldateien zu schaffen, die maximale Größe für die Protokolldateien angeben und die Sprache für die Nachrichten wählen. Sie können auch Optionen zum Empfang von E-Mails angeben, wenn bestimmte Ereignisse auftreten oder bestimmte Nachrichten oder Nachrichtentypen protokolliert werden.

Wenn Sie Protokolldateien überwachen, können Sie wählen, welche Protokolldatei Sie öffnen möchten. Sie können den Inhalt der Protokolldatei filtern, sodass Sie nur die Nachrichten einer bestimmten Wertigkeit anzeigen (z. B. nur Fehlernachrichten), oder Nachrichten, die von einer bestimmten Sitzung der Unternehmenssuche produziert wurden. Wenn Sie eine Protokolldatei anzeigen, können Sie Details zu einzelnen Nachrichten anzeigen. Beispiel: Sie möchten den Namen der Funktion anzeigen, die die Nachricht produziert hat, und sonstige Informationen, mit deren Hilfe Sie Maßnahmen zur Fehlerbehebung ergreifen können, falls notwendig.

### Zugehörige Konzepte



Nachrichten für die Unternehmenssuche

---

## Alerts

Sie können Ihre Unternehmenssuche so konfigurieren, dass sie bei Auftreten bestimmter Ereignisse Nachrichten in die Protokolldatei schreibt.

Nachrichten, die durch Ereignisse ausgelöst werden (so genannte Alerts), können Sie über Bedingungen informieren, die einen Eingriff Ihrerseits erfordern, wie z. B. dass für eine Ressource der Speicherbereich knapp wird. Beim Konfigurieren von Alerts für die Unternehmenssuche geben Sie die Bedingungen an, die vom System überwacht werden sollen. Wenn die Bedingung auftritt, schreibt das System automatisch eine Nachricht in die Protokolldatei.

Zur direkten Benachrichtigung über eine Bedingung können Sie Optionen zum Empfang einer E-Mail angeben, wenn eine der überwachten Nachrichten protokolliert wird.

Sie können Alerts für Ereignisse auf Objektgruppenebene und auf Systemebene konfigurieren. Auf der Objektgruppenebene kann das System folgende Funktionen übernehmen:

- Überwachen der Anzahl durchsuchter Dokumente durch die jeweiligen Crawler und Absetzen eines Alerts bei Erreichen der maximalen Anzahl zulässiger Dokumente
- Überwachen der Anzahl Dokumente, die dem Index für Ihre Objektgruppen hinzugefügt werden, und Absetzen eines Alerts bei Erreichen der maximalen Anzahl zulässiger Dokumente
- Benachrichtigen, wenn das angegebene Zeitlimit für Antworten auf Suchanforderungen überschritten wird

Auf der Systemebene kann das System den Plattenspeicherplatz auf jedem Server für die Unternehmenssuche überwachen und einen Alert absetzen, wenn der Wert für den freien Speicherbereich niedrig ist.

## Konfigurieren von Alerts auf Objektgruppenebene

Durch das Konfigurieren von Alerts können Sie sicherstellen, dass Nachrichten in die Protokolldatei geschrieben werden, wenn bestimmte Ereignisse auf Objektgruppenebene auftreten. Sie können auch eine E-Mail empfangen, wenn Nachrichten über diese Ereignisse protokolliert werden.

### Vorbereitung

Sie müssen ein Administrator für die Unternehmenssuche oder ein Objektgruppenadministrator der Objektgruppe sein, um Alerts für eine Objektgruppe zu konfigurieren.

### Vorgehensweise

Gehen Sie wie folgt vor, um Alerts auf Objektgruppenebene zu konfigurieren:

1. Bearbeiten Sie eine Objektgruppe, wählen Sie die Seite **Protokoll** aus, und klicken Sie **Alerts konfigurieren** an.
2. Wenn das System die Anzahl von den einzelnen Crawlern durchsuchter Dokumente überwachen soll, gehen Sie wie folgt vor:
  - a. Wählen Sie das Markierungsfeld **Wenn die Anzahl der durch einen Crawler durchsuchten Dokumente einen Prozentsatz des zulässigen Maximums erreicht** aus.
  - b. Geben Sie im Feld **Prozentsatz** an, wann eine Nachricht protokolliert werden soll. Geben Sie diese Anzahl als Prozentsatz der maximalen Anzahl Dokumente an, die die Crawler durchsuchen können. (Sie geben die Option **Maximale Anzahl zu durchsuchender Dokumente** beim Konfigurieren der Crawlermerkmale an.) Der Standardwert ist 90 Prozent.

Da Sie unterschiedliche Grenzwerte für verschiedene Crawler konfigurieren können, werden für jeden Crawler separate Nachrichten protokolliert. Beispiel: Wenn Sie den Standardschwellenwert für einen Alert verwenden und einen DB2-Crawler 2 000 000 Dokumente bzw. einen Notes-Crawler 1 000 000 Dokumente durchsuchen lassen, wird eine Nachricht protokol-

liert, wenn der DB2-Crawler 1 800 000 Dokumente durchsucht hat, und eine weitere Nachricht wird protokolliert, wenn der Notes-Crawler 900 000 Dokumente durchsucht hat.

3. Wenn das System die Anzahl dem Index hinzugefügter Dokumente überwachen soll, gehen Sie wie folgt vor:
  - a. Wählen Sie das Markierungsfeld **Wenn die Anzahl Dokumente in der Objektgruppe einen Prozentsatz der geschätzten Größe erreicht** aus.
  - b. Geben Sie im Feld **Prozentsatz** an, wann eine Nachricht protokolliert werden soll. Geben Sie diese Anzahl als Prozentsatz der geschätzten Anzahl Dokumente an, die die Objektgruppe enthalten kann. Der Standardwert ist 85 Prozent.

Das Feld **Grenzwert** enthält die aktuelle geschätzte Größe der Objektgruppe. Wenn Sie diesen Wert ändern wollen, öffnen Sie die Seite **Allgemein** der Objektgruppe, wählen Sie die Option zum Konfigurieren der allgemeinen Optionen aus, und geben Sie im Feld **Geschätzte Anzahl Dokumente** einen neuen Wert an.

**Achtung:** Dieser Grenzwert und die geschätzte Anzahl Dokumente, die Sie für eine Objektgruppe konfigurieren, werden nur für die Überwachung des Wachstums der Objektgruppe verwendet. Sie setzen keinen absoluten Grenzwert für die Größe um, auf die der Index anwachsen kann.

4. Wenn das System Ihnen mitteilen soll, wenn die für das Antworten auf Suchanforderungen erforderliche Zeit einen Grenzwert überschreitet, gehen Sie wie folgt vor:
  - a. Wählen Sie das Markierungsfeld **Wenn die Suchantwortzeit einen Grenzwert überschreitet** aus.
  - b. Geben Sie in das Feld **Grenzwert** die Anzahl Sekunden ein, die Sie als maximale Suchantwortzeit akzeptieren.

Wenn diese Anzahl überschritten wird, gibt das System eine Protokollnachricht zum Ereignis aus. Beispiel: Wenn Sie den Standardwert beibehalten, erstellt das System eine Protokollnachricht, wenn ein Suchserver durchschnittlich fünf Sekunden oder länger benötigt, um auf Suchanforderungen zu antworten.

Die normalen Antwortzeiten liegen unter einer halben Sekunde. Liegt der durchschnittliche Wert über einer Sekunde, weist dies unter Umständen darauf hin, dass Ihr Betriebssystem optimiert werden muss, um eine bessere Leistung zu erzielen, oder dass ein Problem in den Konfigurationseinstellungen des Suchservers vorhanden ist. Sie könnten z. B. die Speicherplatzgröße erhöhen, die Sie dem Suchcache zuordnen.

5. Klicken Sie **OK** an.

Wenn Sie per E-Mail benachrichtigt werden wollen, wenn das System Nachrichten zu diesen Ereignissen protokolliert, öffnen Sie die Seite **Protokoll**, und klicken Sie **E-Mail-Optionen für Nachrichten konfigurieren** an, um eine E-Mail-Adresse anzugeben. Die Nachrichten-IDs der Alerts, die Sie aktiviert haben, werden automatisch der Liste der Nachrichten-IDs hinzugefügt, für die eine E-Mail gesendet werden soll.

Bevor Sie E-Mails empfangen können, müssen Sie allerdings sicherstellen, dass die Angaben zu Ihrem Mail-Server konfiguriert sind. Hierzu muss ein Administrator für die Unternehmenssuche in der Funktionsleiste **System** auswählen, die Seite **Protokoll** öffnen und anschließend **E-Mail-Optionen für Nachrichten konfigurieren** anklicken.

#### Zugehörige Tasks

## Konfigurieren von Alerts auf Systemebene

Durch das Konfigurieren von Alerts können Sie sicherstellen, dass Nachrichten in die Protokolldatei geschrieben werden, wenn bestimmte Ereignisse auf Systemebene auftreten. Sie können auch eine E-Mail empfangen, wenn Nachrichten über diese Ereignisse protokolliert werden.

### Vorbereitung

Sie müssen ein Administrator für die Unternehmenssuche sein, um Alerts auf Systemebene zu konfigurieren.

### Vorgehensweise

Gehen Sie wie folgt vor, um Alerts auf Systemebene zu konfigurieren:

1. Klicken Sie **System** an, um die Sicht **System** zu öffnen.
2. Klicken Sie  **Bearbeiten** an, um zur Sicht für die Systembearbeitung zu wechseln.
3. Klicken Sie **Alerts konfigurieren** auf der Seite **Protokoll** an.
4. Wenn das System den freien Speicherbereich überwachen soll, der auf den Servern für die Unternehmenssuche verfügbar ist, wählen Sie das Markierungsfeld **Wenn der verfügbare Speicherplatz für das Dateisystem einen Prozentsatz des Gesamtspeicherplatzes erreicht** aus.
5. Geben Sie im Feld **Prozentsatz** an, wann Sie vom System benachrichtigt werden wollen, dass der freie Speicherbereich eines Servers einen niedrigen Wert erreicht hat. Geben Sie diesen Wert als Prozentsatz des gesamten Dateisystemspeicherbereichs an. Der Standardwert ist 80 Prozent.  
Ist Ihr System für die Unternehmenssuche auf mehreren Servern installiert, wird für jeden Server eine separate Protokollnachricht erstellt. So werden Sie z. B. mit einer Nachricht informiert, wenn der Speicherbereich auf dem Crawler-Server zu klein wird. Andere Nachrichten informieren Sie über Speicherplatzeinschränkungen auf den Index- und Suchservern.
6. Klicken Sie **OK** an.

Wenn Sie per E-Mail benachrichtigt werden wollen, wenn das System eine Nachricht zu diesem Ereignis protokolliert, öffnen Sie die Seite **Protokoll**, und klicken Sie **E-Mail-Optionen für Nachrichten konfigurieren** an, um eine E-Mail-Adresse und Informationen zu Ihrem Mail-Server anzugeben.

### Zugehörige Tasks

„Empfangen von E-Mail zu protokollierten Nachrichten“ auf Seite 357

---

## Konfigurieren von Protokolldateien

Sie können angeben, welche Nachrichtentypen Sie protokollieren wollen, sowie Optionen zum Erstellen von Protokolldateien.

### Vorbereitung

Sie müssen ein Administrator für die Unternehmenssuche oder ein Objektgruppenadministrator der Objektgruppe sein, um Protokolldateien auf Objektgruppenebene zu konfigurieren. Sie müssen ein Administrator für die Unternehmenssuche sein, um Protokolldateien auf Systemebene zu konfigurieren.

## Informationen zu dieser Task

Damit verhindert wird, dass Protokolldateien zu viel Plattenspeicherplatz belegen, wechselt das System turnusmäßig die Protokolldateien und startet immer eine neue Protokolldatei, wenn sich das aktuelle Datum ändert. Wenn eine Protokolldatei ihre maximal zulässige Größe erreicht und das Datum sich nicht geändert hat, erstellt das System eine neue Protokolldatei. Wenn die maximale Anzahl Protokolldateien erreicht ist, wird die älteste Protokolldatei gelöscht, sodass eine neue erstellt werden kann.

Damit Sie E-Mail zu protokollierten Nachrichten empfangen können, müssen Sie erst Informationen angeben, wie die E-Mail zugestellt werden soll. Anschließend geben Sie an, zu welchen Nachrichten Sie E-Mail empfangen wollen.

## Vorgehensweise

Gehen Sie wie folgt vor, um Protokolldateien für die Unternehmenssuche zu konfigurieren:

1. Wenn Sie Optionen zum Erstellen und Rollieren von Protokolldateien auf Systemebene konfigurieren wollen:
  - a. Klicken Sie **System** an, um die Sicht **System** zu öffnen.
  - b. Klicken Sie  **Bearbeiten** an, um zur Sicht für die Systembearbeitung zu wechseln.
  - c. Klicken Sie **Protokolldateioptionen konfigurieren** auf der Seite **Protokoll** an. Die Seite **Protokolldateioptionen auf Systemebene** wird angezeigt.
2. Wenn Sie Optionen zum Erstellen und Rollieren von Protokolldateien auf Objektgruppenebene konfigurieren wollen:
  - a. Suchen Sie in der Sicht **Objektgruppen** die Objektgruppe, für die Sie Optionen angeben wollen, und klicken Sie  **Bearbeiten** an.
  - b. Klicken Sie **Protokolldateioptionen konfigurieren** auf der Seite **Protokoll** an. Die Seite **Protokolldateioptionen auf Objektgruppenebene** wird angezeigt.
3. Wählen Sie im Feld **Zu protokollierende Informationstypen** die Nachrichtentypen aus, die Sie protokollieren wollen:

### Nur Fehlernachrichten

Fehlernachrichten geben an, dass eine unerwünschte Situation oder ein nicht erwartetes Verhalten aufgetreten ist und dass der Prozess nicht fortgesetzt werden kann. Sie müssen Maßnahmen ergreifen, um den Fehler zu beheben.

### Fehlernachrichten und Warnungen

Warnungen geben einen möglichen Konflikt oder Inkonsistenzen an, sie stoppen aber keinen Prozess. Diese Option ist die Standardeinstellung.

### Alle Nachrichten

Informationsnachrichten bieten allgemeine Informationen zum System oder zur aktuellen Task und erfordern keine Korrekturmaßnahme.

**Wichtig:** Die Auswahl dieser Option kann sich negativ auf die Systemleistung auswirken. Protokollieren Sie nur alle Nachrichten, wenn Sie Fehler beheben müssen oder wenn Sie von der IBM Unterstützungsfunktion dazu aufgefordert werden.

4. Geben Sie in das Feld **Maximale Größe jeder Protokolldatei** die maximale Anzahl Megabyte für jede Protokolldatei ein. Der Standardwert ist 10 MB.

Wenn die Protokolldatei diese Größe erreicht, wird eine neue Protokolldatei erstellt, bis die maximale Anzahl zulässiger Protokolldateien erreicht ist. Indem Sie die Protokolldateien relativ klein halten, ist ihre Anzeige effizienter.

5. Geben Sie in das Feld **Maximale Anzahl Protokolldateien** die maximale Anzahl Protokolldateien ein, die Sie erstellen wollen. Der Standardwert ist 16. Wenn Sie sicherstellen wollen, dass ältere Protokollnachrichten zur Überprüfung verfügbar sind, erhöhen Sie diesen Wert. Wenn Sie mehr an aktuellen Nachrichten interessiert sind und kein langes Aktivitätsprotokoll verwalten müssen, setzen Sie diesen Wert herab.
6. Wählen Sie im Feld **Standardländereinstellung** die Sprache aus, die Sie zum Protokollieren der Nachrichten verwenden wollen. Der Standardwert ist Englisch.
7. Klicken Sie **OK** an.
8. Damit Ihre Änderungen wirksam werden, geben Sie den folgenden Befehl ein, um das System für die Unternehmenssuche zu stoppen und erneut zu starten.

```
esdmin system stopall
esdmin system startall
```

## Konfigurieren von SMTP-Serverinformationen

Bevor Sie E-Mail zu den Aktivitäten der Unternehmenssuche empfangen können, müssen Sie Angaben zu Ihrem SMTP-Server (Simple Mail Transfer Protocol) konfigurieren.

### Vorbereitung

Sie müssen ein Administrator für die Unternehmenssuche sein, um Informationen zu Ihrem SMTP-Server zu konfigurieren.

### Informationen zu dieser Task

Es stehen Ihnen mehrere Verwaltungsfunktionen der Unternehmenssuche für den Empfang von E-Mail zur Verfügung. Bevor Sie E-Mail von diesen Funktionen empfangen können, müssen Sie Informationen zu Ihrem SMTP-Server angeben:

- Wenn Sie Alerts auf Objektgruppenebene oder Systemebene konfigurieren, können Sie E-Mail empfangen, wenn diese Nachrichten protokolliert werden. Sie können auch E-Mail empfangen, wenn andere Nachrichten protokolliert werden, und nicht nur bei Nachrichten, die von überwachten Ereignissen ausgelöst werden.
- Wenn Sie detaillierte Informationen zu einem URI im Index oder einem aus dem System für die Unternehmenssuche gelöschten Dokument anzeigen wollen, können Sie diesen Bericht als E-Mail empfangen.
- Wenn Sie einen Web-Crawler überwachen, und angeben, dass Verlaufsprotokolle für den Web-Crawler erstellt werden sollen, können Sie sich per E-Mail benachrichtigen lassen, wenn ein Bericht erstellt wurde.

### Vorgehensweise

Gehen Sie wie folgt vor, um Ihre SMTP-Serverinformationen zu konfigurieren:

1. Klicken Sie **System** an, um die Sicht **System** zu öffnen.
2. Klicken Sie  **Bearbeiten** an, um zur Sicht für die Systembearbeitung zu wechseln.

3. Klicken Sie **E-Mail-Optionen für Nachrichten konfigurieren** auf der Seite **Protokoll** an.
4. Geben Sie auf der Seite **E-Mail-Optionen für Systemnachrichten** in das Feld **SMTP-E-Mail-Server für die Zustellung von E-Mail** den vollständig qualifizierten Hostnamen oder die IP-Adresse des SMTP-Servers ein, den Sie verwenden wollen.  
Das System verwendet diesen Server, um E-Mail an die von Ihnen angegebenen Adressen zu senden.
5. Geben Sie in das Feld **Häufigkeit der Überprüfung auf neue E-Mail** ein, wie oft das System in Frage kommende Nachrichten abfragen und E-Mail-Benachrichtigungen senden soll.  
Das System fasst alle Nachrichten für eine bestimmte E-Mail-Adresse in einer Nachricht zusammen und sendet diese Nachricht in der von Ihnen angegebenen Häufigkeit.
6. Klicken Sie **OK** an.

## Empfangen von E-Mail zu protokollierten Nachrichten

Sie können Optionen angeben, um E-Mail zu empfangen, wenn bestimmte Nachrichten oder Nachrichtentypen protokolliert werden.

### Vorbereitung

Sie müssen ein Administrator für die Unternehmenssuche sein, um E-Mail-Optionen zu Nachrichten auf Systemebene zu konfigurieren. Sie müssen ein Administrator für die Unternehmenssuche oder ein Objektgruppenadministrator der Objektgruppe sein, um E-Mail-Optionen zu Nachrichten auf Objektgruppenebene zu konfigurieren.

Bevor Sie E-Mail empfangen können, müssen Sie Angaben zu Ihrem SMTP-Server (Simple Mail Transfer Protocol) konfigurieren, damit die E-Mails zugestellt werden können.

### Informationen zu dieser Task

Beim Konfigurieren von Alerts können Sie eine Option auswählen, um Nachrichten zu protokollieren, wenn bestimmte Ereignisse auftreten. Nachdem Sie diese Optionen aktiviert haben, können Sie Optionen konfigurieren, damit Sie automatisch E-Mails empfangen, wenn diese Nachrichten protokolliert werden. Sie können auch Optionen angeben, um E-Mail zu empfangen, wenn andere Nachrichten als durch Ereignisse ausgelöste Nachrichten protokolliert werden.

### Vorgehensweise

Gehen Sie wie folgt vor, um E-Mail-Optionen zu Nachrichten zu konfigurieren:

1. Wenn Sie E-Mail zu Systemnachrichten empfangen wollen:
  - a. Klicken Sie **System** an, um die Sicht **System** zu öffnen.
  - b. Klicken Sie  **Bearbeiten** an, um zur Sicht für die Systembearbeitung zu wechseln.
  - c. Klicken Sie **E-Mail-Optionen für Nachrichten konfigurieren** auf der Seite **Protokoll** an.
  - d. Wählen Sie das Markierungsfeld **E-Mail zu Nachrichten auf Systemebene senden** auf der Seite **E-Mail-Optionen für Systemnachrichten** aus.

- e. Geben Sie in das Feld **E-Mail-Adresse für den Empfang von E-Mail** mindestens eine E-Mail-Adresse ein. In der Regel sollte ein Administrator für die Unternehmenssuche Informationen zu Systemnachrichten empfangen. Trennen Sie die Adressen mit einem Komma. Beispiel:  
steinbeck@us.ibm.com, yeats@ireland.ibm.com, dante@it.ibm.com.
  - f. Wenn Sie E-Mail zu allen protokollierten Fehlernachrichten empfangen wollen, wählen Sie das Markierungsfeld **E-Mail zu allen Fehlernachrichten senden** aus.
  - g. Wenn Sie nur E-Mail erhalten wollen, wenn bestimmte Nachrichten auf Systemebene protokolliert werden, geben Sie die Nachrichten-IDs dieser Nachrichten in den Bereich **E-Mail zu bestimmten Nachrichten senden** ein. Geben Sie eine Nachrichten-ID pro Zeile ein. Beispiel:  
FFQC4819E  
FFQ00005E  
Einige Nachrichten-IDs werden standardmäßig aufgelistet. (Klicken Sie **Hilfe** an, um eine Beschreibung dieser Nachrichten anzuzeigen.)
  - h. Klicken Sie **OK** an.
2. Wenn Sie E-Mail zu Objektgruppennachrichten empfangen wollen:
- a. Klicken Sie **Objektgruppen** an, um die Sicht **Objektgruppen** zu öffnen.
  - b. Suchen Sie in der Liste der Objektgruppen die zu konfigurierende Objektgruppe, und klicken Sie  **Bearbeiten** an.
  - c. Klicken Sie **E-Mail-Optionen für Nachrichten konfigurieren** auf der Seite **Protokoll** an.
  - d. Wählen Sie das Markierungsfeld **E-Mail zu Nachrichten auf Objektgruppenebene senden** auf der Seite **E-Mail-Optionen für Objektgruppennachrichten** aus.
  - e. Geben Sie in das Feld **E-Mail-Adresse für den Empfang von E-Mail** mindestens eine E-Mail-Adresse ein. In der Regel sollte ein Objektgruppenadministrator Informationen zu Nachrichten auf Objektgruppenebene empfangen. Trennen Sie die Adressen mit einem Komma. Beispiel:  
steinbeck@us.ibm.com, yeats@ireland.ibm.com, dante@it.ibm.com.
  - f. Wenn Sie E-Mail zu allen protokollierten Fehlernachrichten empfangen wollen, wählen Sie das Markierungsfeld **E-Mail zu allen Fehlernachrichten senden** aus.
  - g. Wenn Sie nur E-Mail erhalten wollen, wenn bestimmte Nachrichten auf Objektgruppenebene protokolliert werden, geben Sie die Nachrichten-IDs dieser Nachrichten in den Bereich **E-Mail zu bestimmten Nachrichten senden** ein. Geben Sie eine Nachrichten-ID pro Zeile ein. Beispiel:  
FFQC4819E  
FFQ00005E  
Einige Nachrichten-IDs werden standardmäßig aufgelistet. (Klicken Sie **Hilfe** an, um eine Beschreibung dieser Nachrichten anzuzeigen.)
  - h. Klicken Sie **OK** an.

### Zugehörige Konzepte

 Nachrichten für die Unternehmenssuche

### Zugehörige Tasks

„Konfigurieren von Alerts auf Objektgruppenebene“ auf Seite 352

„Konfigurieren von Alerts auf Systemebene“ auf Seite 354

## Ändern der Größe des Abfrageprotokolls

Sie können die Größe der Protokolldateien, die für die Abfrageverarbeitung erstellt werden, vergrößern oder verkleinern, indem Sie eine Konfigurationsdatei bearbeiten. Die Administrationskonsole für die Unternehmenssuche bietet keine Unterstützung für diese Task.

### Informationen zu dieser Task

Während der Abfrageverarbeitung werden die Protokolldaten in die Datei *objektgruppen-id\_OmniFindQueryLog\_datum.log* geschrieben, wobei *objektgruppen-id* die zu konfigurierende Objektgruppe und *datum* das Datum angibt, an dem die Protokolldatei erstellt wird. Sie können die Größe dieser Protokolldatei herauf- oder heruntersetzen, je nach dem wie viele Daten Sie protokollieren möchten, bevor eine neue Protokolldatei erstellt wird.

### Vorgehensweise

Gehen Sie wie folgt vor, um die Größe der Protokolldatei für die Abfrageverarbeitung zu ändern:

1. Melden Sie sich als Administrator für die Unternehmenssuche an. Melden Sie sich in einem System für die Unternehmenssuche mit mehreren Servern am Indexserver an.
2. Öffnen Sie die Datei `ES_ROOT_NODE/master_config/objektgruppen-id.runtime.node1/runtime-generic.properties`.
3. Suchen Sie das Merkmal **MaxFileSize**. Erhöhen oder verringern Sie den Wert dieses Merkmals, um die Größe der Protokolldateien zu erhöhen bzw. zu verringern, und speichern Sie Ihre Änderungen.
4. Gehen Sie in einem System für die Unternehmenssuche mit einem Server wie folgt vor:
  - a. Überwachen Sie über die Administrationskonsole für die Unternehmenssuche die Objektgruppe, die Sie geändert haben, und stoppen Sie die Suchserver.
  - b. Starten Sie die Anwendung `ESSearchServer` in `WebSphere Application Server` erneut.
  - c. Starten Sie die gestoppten Suchserver über die Administrationskonsole erneut.
  - d. Öffnen Sie die Suchanwendung in einem neuen Browser.
5. Gehen Sie in einem System für die Unternehmenssuche mit mehreren Servern wie folgt vor:
  - a. Melden Sie sich als Administrator für die Unternehmenssuche am Indexserver an.
  - b. Geben Sie die folgenden Befehle ein, um das System für die Unternehmenssuche erneut zu starten:

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

---

## Anzeigen von Protokolldateien

Sie können die Protokollnachrichten anzeigen, die von System- und Objektgruppenkomponenten in eine gemeinsame Protokolldatei geschrieben werden. Sie können auch Filter angeben, um Nachrichten einer bestimmten Bewertungsstufe und Nachrichten von bestimmten Sitzungen für die Unternehmenssuche anzuzeigen.

### Vorbereitung

Jeder Benutzer mit Verwaltungsaufgaben für die Unternehmenssuche kann die Protokolldateien der Objektgruppen anzeigen, für deren Verwaltung er berechtigt ist. Sie müssen ein Administrator für die Unternehmenssuche sein oder die Berechtigung für den Zugriff auf die Funktionsleiste **System** haben, um Protokolldateien auf Systemebene anzuzeigen.

### Vorgehensweise

1. Gehen Sie wie folgt vor, um die Protokolldateien einer einzelnen Objektgruppe anzuzeigen:
  - a. Klicken Sie **Objektgruppen** an, um die Sicht **Objektgruppen** zu öffnen.
  - b. Suchen Sie die Objektgruppenliste in der Objektgruppe, die Sie anzeigen wollen, klicken Sie  **Überwachen** an, und öffnen Sie die Seite **Protokoll**.

**Tipp:** Wenn Sie eine Objektgruppe bearbeiten und sich bereits auf der Seite **Protokoll** befinden, können Sie  **Überwachen** anklicken, um zur Sicht für die Überwachung der Objektgruppe zu wechseln.

2. Gehen Sie wie folgt vor, um Protokolldateien auf Systemebene anzuzeigen:
  - a. Klicken Sie **System** an, um die Sicht **System** zu öffnen.
  - b. Wählen Sie die Seite **Protokoll** aus.
3. Wählen Sie im Feld **Protokolldatei** die Protokolldatei aus, die Sie anzeigen wollen.

Der Name jeder Protokolldatei schließt den Protokolldateityp (wie zum Beispiel einen System- oder Objektgruppennamen) und das Datum der Dateierstellung ein. Wenn mehrere Protokolldateien desselben Typs an demselben Datum erstellt werden, gibt ein numerisches Suffix die Reihenfolge an, in der die Dateien erstellt wurden. Beispiel:

```
protokolldateityp_20060526.log (enthält die neuesten Einträge an diesem Datum)
protokolldateityp_20060526.log.1
protokolldateityp_20060526.log.2 (enthält die ältesten Einträge an diesem Datum)
protokolldateityp_20060525.log (enthält die neuesten Einträge an diesem Datum)
protokolldateityp_20060525.log.1
protokolldateityp_20060525.log.2
protokolldateityp_20060525.log.3 (enthält die ältesten Einträge an diesem Datum)
```

4. Wählen Sie die entsprechenden Markierungsfelder im Feld **Fehlerkategorie** aus, um nur Nachrichten mit bestimmten Bewertungsstufen anzuzeigen.
5. Wählen Sie die entsprechenden Markierungsfelder im Feld **Sitzung** aus, um nur die Nachrichten bestimmter Sitzungen anzuzeigen.
6. Klicken Sie **Protokoll anzeigen** an.

Für jede Nachricht auf der Seite **Inhalt der Protokolldatei** werden die folgenden Informationen angezeigt: das Datum und die Zeit, zu der die Nachricht erstellt wurde, die Bewertungsstufe der Nachricht, der Name der Sitzung, die die Nachricht abgesetzt hat, die Nachrichten-ID und der Fehlertext.

Sie können die entsprechenden Knöpfe anklicken, um zur ersten, letzten, vorherigen oder nächsten Seite der Protokolldatei zu wechseln. Sie können auch eine Seitenzahl angeben und direkt zu dieser Seite wechseln.

7. Klicken Sie  **Details** an, damit Ihnen detailliertere Informationen zu einer Nachricht angezeigt werden.

Auf der Seite **Details der Protokollnachricht** werden die folgenden Informationen angezeigt: der Hostname des Servers für die Unternehmenssuche, auf dem die Nachricht aufgetreten ist, der Name der Datei, die den Fehler verursacht hat, der Funktionsname und die Zeilennummer, in der der Fehler aufgetreten ist, die Prozess-ID und die Thread-ID.

Sie können Knöpfe anklicken, um zur nächsten bzw. vorherigen Nachricht in der Protokolldatei zu blättern.



---

# Durchführen von Backup und Restore für ein System für die Unternehmenssuche

Mit Backup- und Restore-Scripts können Sie das System für die Unternehmenssuche sichern und wiederherstellen.

## Was wird von den Scripts gesichert?

Die Scripts können die folgenden Dateien sichern und wiederherstellen:

- Konfigurationsdateien im Verzeichnis `ES_NODE_ROOT/master_config`
- Datenbankdateien für die Crawler einschließlich aller Crawlermetadaten wie z. B. die Angabe, wann die Datenquellen zuletzt durchsucht wurden
- Alle Dateien im Verzeichnis `ES_NODE_ROOT/data`
- Indexierte Dateien für Objektgruppen, die mit nicht standardmäßigen Datenverzeichnissen konfiguriert sind

## Backup-Verzeichnisstruktur

Das Backup-Script erstellt die folgenden Unterverzeichnisse unter einem Verzeichnis, das Sie bei der Ausführung des Scripts angeben. Die ID des Administrators für die Unternehmenssuche muss über eine Schreibberechtigung für das Verzeichnis verfügen, das Sie angeben.

### **master\_config**

Enthält die Konfigurationsdateien aus dem Verzeichnis `ES_NODE_ROOT/master_config`

### **database**

Enthält die Datenbankdateien des Crawler-Servers

### **data**

Enthält die Indexdateien des Indexservers

## Verwendungsrichtlinien

- Sie können die Daten auf einem Computer sichern und auf einem andern Computer wiederherstellen. Beachten Sie aber Folgendes:
  - Dateien, die von einer Version von OmniFind Enterprise Edition gesichert wurden, können nicht auf einem System wiederhergestellt werden, auf dem eine andere Version von OmniFind Enterprise Edition aktiv ist.
  - Sie müssen Daten auf einem System wiederherstellen, das mindestens die gleiche Anzahl Server für die Unternehmenssuche verwendet. Wenn Sie beispielsweise ein System für die Unternehmenssuche sichern, das auf einem einzelnen Server ausgeführt wird, können Sie die Daten auf einem System wiederherstellen, das zwei oder vier Server für die Unternehmenssuche verwendet. Daten, die von einem System mit vier Servern gesichert wurden, können nicht auf einem System wiederhergestellt werden, das zwei oder einen Server verwendet.
  - Dateien, die von einem Betriebssystem gesichert wurden, können nicht auf einem System wiederhergestellt werden, auf dem ein anderes Betriebssystem aktiv ist. Wenn Sie z. B. Ihre Unternehmenssuche auf einem AIX-System installiert haben und Sie diese nun unter Linux ausführen wollen, müssen Sie ein neues System für die Unternehmenssuche auf Ihren Linux-Servern installieren.

- Erstellen Sie den Hauptindex, bevor Sie mit der Sicherung beginnen, damit die neuesten indexierten Daten gesichert werden.
- Alle Einstellungen für das Installationsverzeichnis (ES\_INSTALL\_ROOT), das Datenverzeichnis (ES\_NODE\_ROOT) sowie die ID und das Kennwort für den Administrator für die Unternehmenssuche müssen auf dem gesicherten System und auf dem System, auf dem die Daten wiederhergestellt werden sollen, identisch sein.
- Führen Sie bei einer Konfiguration mit mehreren Servern den Backup und den Restore des Systems über den Indexserver für die Unternehmenssuche aus. Da sich alle Crawlerdaten in Datenbanken auf dem Crawler-Server befinden, führen die Scripts ferne Befehle aus, um die Crawlerdaten zu sichern und wiederherzustellen.
- Sie müssen über ausreichend Plattenspeicherplatz verfügen, um die Systemdateien für die Unternehmenssuche in einem anderen Verzeichnis zu sichern. Die Backup- und Restore-Scripts überprüfen diese Dateien nicht.
- Alle Systemsitzungen werden gestoppt, während die Backup- und Restore-Scripts ausgeführt werden. Damit keine fehlerhaften oder inkonsistenten Systeminformationen angezeigt werden, verwenden Sie die Administrationskonsole für die Unternehmenssuche nicht, während die Scripts ausgeführt werden.
- Wenn das System aufgrund eines nicht behebbaren Fehlers fehlschlägt, müssen Sie OmniFind Enterprise Edition erneut installieren und anschließend das Restore-Script ausführen.

---

## Sichern des Systems für die Unternehmenssuche

Sie können das System für die Unternehmenssuche sichern, indem Sie unter AIX, Linux oder Solaris das Script `esbackup.sh` oder unter Microsoft Windows das Script `esbackup.bat` verwenden.

### Einschränkungen

Die ID des Administrators für die Unternehmenssuche muss über eine Schreibberechtigung für das Verzeichnis verfügen, das Sie angeben, wenn Sie das Backup-Script ausführen.

Alle Systemsitzungen werden gestoppt, während die Backup- und Restore-Scripts ausgeführt werden. Damit keine fehlerhaften oder inkonsistenten Systeminformationen angezeigt werden, verwenden Sie die Administrationskonsole für die Unternehmenssuche nicht, während die Scripts ausgeführt werden.

**Achtung:** Wenn Sie Strg+C drücken, um das Backup-Script zu unterbrechen, gerät das System in einen inkonsistenten Status. Sie müssen den folgenden Befehl eingeben, um alle Servicesitzungen und alle aktiven Sitzungen, die während der Sicherungsverarbeitung gestoppt wurden, zu starten:

```
esadmin system startall
```

### Vorgehensweise

Gehen Sie wie folgt vor, um das System für die Unternehmenssuche zu sichern:

1. Melden Sie sich am Indexserver als Administrator für die Unternehmenssuche an. Diese Benutzer-ID wurde bei der Installation von OmniFind Enterprise Edition angegeben.

2. Starten Sie die allgemeine Übertragungsschicht (Common Communication Layer, CCL) für die Unternehmenssuche, wenn Sie nicht ausgeführt wird:

**AIX, Linux oder Solaris**

```
startccl.sh -bg
```

**Windows-Eingabeaufforderung**

```
startccl
```

**Verwaltungstool für Windows-Services**

Gehen Sie wie folgt vor, um die allgemeine Übertragungsschicht im Hintergrund zu starten:

- a. Starten Sie die Windows-Services: **Start** → **Einstellungen** → **Systemsteuerung** → **Verwaltung** → **Dienste**.
  - b. Klicken Sie **IBM OmniFind Enterprise Edition** mit der rechten Maustaste an, und klicken Sie **Start** an.
3. Stellen Sie bei einer Konfiguration mit mehreren Servern sicher, dass die allgemeine Übertragungsschicht auf allen Servern gestartet ist. Wiederholen Sie die vorangegangenen Schritte, falls erforderlich.
  4. Starten Sie den Backup, wobei *backup-verzeichnis* ein Verzeichnis für die gesicherten Daten ist:

**AIX, Linux oder Solaris**

```
esbackup.sh backup-verzeichnis
```

**Windows-Eingabeaufforderung**

```
esbackup.bat backup-verzeichnis
```

**Zugehörige Verweise**

„Befehle, Rückkehrcodes und Sitzungs-IDs der Unternehmenssuche“ auf Seite 401

---

## Wiederherstellung des Systems für die Unternehmenssuche

Nachdem Sie OmniFind Enterprise Edition neu installiert haben, können Sie mit dem Script `esrestore.sh` für AIX, Linux und Solaris oder dem Script `esrestore.bat` für Microsoft Windows ein System für die Unternehmenssuche wiederherstellen.

### Einschränkungen

Alle Systemsitzungen werden gestoppt, während die Backup- und Restore-Scripts ausgeführt werden. Damit keine fehlerhaften oder inkonsistenten Systeminformationen angezeigt werden, verwenden Sie die Administrationskonsole für die Unternehmenssuche nicht, während die Scripts ausgeführt werden.

Dateien, die von einer Version von OmniFind Enterprise Edition gesichert wurden, können nicht auf einem System wiederhergestellt werden, auf dem eine andere Version von OmniFind Enterprise Edition aktiv ist. Darüber hinaus muss das System, auf dem Sie die Daten wiederherstellen, mindestens gleich viele Server für die Unternehmenssuche wie das System verwenden, von dem die Daten gesichert wurden.

### Vorgehensweise

Gehen Sie wie folgt vor, um das System für die Unternehmenssuche wiederherzustellen:

1. Melden Sie sich am Indexserver als Administrator für die Unternehmenssuche an. Diese Benutzer-ID wurde bei der Installation von OmniFind Enterprise Edition angegeben.
2. Starten Sie die allgemeine Übertragungsschicht (Common Communication Layer, CCL) für die Unternehmenssuche, wenn Sie nicht ausgeführt wird:

**AIX, Linux oder Solaris**

```
startccl.sh -bg
```

**Windows-Eingabeaufforderung**

```
startccl
```

**Verwaltungstool für Windows-Services**

Gehen Sie wie folgt vor, um die allgemeine Übertragungsschicht im Hintergrund zu starten:

- a. Starten Sie die Windows-Services: **Start** → **Einstellungen** → **Systemsteuerung** → **Verwaltung** → **Dienste**.
  - b. Klicken Sie **IBM OmniFind Enterprise Edition** mit der rechten Maustaste an, und klicken Sie **Start** an.
3. Stellen Sie bei einer Konfiguration mit mehreren Servern sicher, dass die allgemeine Übertragungsschicht auf allen Servern gestartet ist. Wiederholen Sie die vorangegangenen Schritte, falls erforderlich.
  4. Stoppen Sie den Controller:

```
esadmin stop
```

5. Stellen Sie Daten der Unternehmenssuche wieder her, wobei *backup-verzeichnis* das Verzeichnis ist, in dem Sie die Dateien gesichert haben:

**AIX, Linux oder Solaris**

```
esrestore.sh backup-verzeichnis
```

**Windows-Eingabeaufforderung**

```
esrestore.bat backup-verzeichnis
```

**Zugehörige Verweise**

„Befehle, Rückkehrcodes und Sitzungs-IDs der Unternehmenssuche“ auf Seite 401

---

## Exportieren und Importieren von Objektgruppenkonfigurationen

Sie können einzelne Objektgruppenkonfigurationen exportieren und importieren. Dabei werden nur die Objektgruppenkonfigurationsdateien exportiert und importiert, nicht die Objektgruppendaten.

Sie können eine Objektgruppe aus einem System für die Unternehmenssuche exportieren und die Objektgruppe anschließend in ein anderes System für die Unternehmenssuche importieren. Sie können die Objektgruppen jedoch nur zwischen Systemen exportieren und importieren, auf denen dieselbe Version der Unternehmenssuche aktiv ist. So können Sie z. B. eine Objektgruppe aus einem System der Version 8.4 exportieren und anschließend in ein System der Version 8.4 oder der Version 8.4.0.150 importieren. Sie können jedoch nicht eine aus einem System der Version 8.3 exportierte Objektgruppe in ein System der Version 8.4 importieren.

Wenn Sie eine Objektgruppe exportieren, können Sie diese nur in ein System für die Unternehmenssuche importieren, das dasselbe Betriebssystem verwendet. So ist es z. B. nicht möglich, eine Objektgruppe, die Sie aus einem unter Linux installier-

ten System für die Unternehmenssuche exportiert haben, in ein System für die Unternehmenssuche zu importieren, das unter Windows installiert ist.

## Exportieren von Objektgruppen

Zum Exportieren einer Objektgruppe verwenden Sie den Befehl **esadmin export**, um jeweils eine Objektgruppe zu exportieren. Die Administrationskonsole für die Unternehmenssuche bietet keine Unterstützung für den Export von Objektgruppen.

1. Melden Sie sich als Administrator für die Unternehmenssuche an. In einer Konfiguration mit mehreren Servern müssen Sie sich auf dem Indexserver anmelden. (Der Befehl schlägt fehl, wenn Sie versuchen, ihn auf einem beliebigen anderen Server für die Unternehmenssuche auszuführen.)
2. Geben Sie die folgenden Befehle ein:

```
esadmin export -help
esadmin export -cid objektgruppen-id [-fname name_der_exportdatei] [-verbose]
```

Dabei gilt Folgendes:

### **-help**

Stellt Hilfetext für den Befehl bereit.

### **-cid *objektgruppen-id***

Gibt die Objektgruppen-ID der zu exportierenden Objektgruppe an.

**Tipp:** Sie können die Objektgruppen-ID der Objektgruppe, die Sie exportieren wollen, mithilfe der Administrationskonsole für die Unternehmenssuche ermitteln oder indem Sie den Befehl **esadmin report collections** verwenden.

### **-fname *name\_der\_exportdatei***

Gibt den Pfad zur Exportdatei an. Wenn es sich nicht um einen absoluten Dateinamen handelt, wird das Verzeichnis `ES_NODE_ROOT/dump` vorausgesetzt. Wenn Sie diese Option auslassen, wird im Verzeichnis `ES_NODE_ROOT/dump` eine Datei erstellt, die die folgende Namenskonvention verwendet:

```
_export_jjjjmmmt_hhmmssz.zip
```

Dabei gilt Folgendes:

*jjjmmmt*

Ist das aktuelle Datum (Jahr, Monat und Tag), an dem der Exportbefehl ausgeführt wird.

*hhmmss*

Ist die aktuelle Zeit (Stunde, Minute und Sekunde), zu der der Exportbefehl ausgeführt wird.

*z* Ist die Zeitzonendifferenz zur westeuropäischen Zeit, wann der Exportbefehl ausgeführt wird. Die Pazifik Standardzeit wird beispielsweise als Westeuropäische Zeit -0800 angegeben.

### **-verbose**

Zeigt Informationen an, die bei der Fehlersuche in der Exportoperation verwendet werden können.

## Importieren von Objektgruppen

Zum Importieren einer Objektgruppe verwenden Sie den Befehl **esadmin import**, um jeweils eine Objektgruppe zu importieren. Die Administrationskonsole für die Unternehmenssuche bietet keine Unterstützung für den Import von Objektgruppen.

1. Melden Sie sich als Administrator für die Unternehmenssuche an. In einer Konfiguration mit mehreren Servern müssen Sie sich auf dem Indexserver anmelden. (Der Befehl schlägt fehl, wenn Sie versuchen, ihn auf einem beliebigen anderen Server für die Unternehmenssuche auszuführen.)
2. Geben Sie die folgenden Befehle ein:

```
esadmin import -help
esadmin import -fname name_der_importdatei
 [-cid neue_objektgruppen-id]
 [-name neuer_objektgruppenname]
 [-colDataDir neues_datenverzeichnis_der_objektgruppe]
 [-force]
 [-verbose]
```

Dabei gilt Folgendes:

### **-help**

Stellt Hilfetext für den Befehl bereit.

### **-fname** *name\_der\_importdatei*

Gibt den Pfad zur Importdatei an. Wenn es sich nicht um einen absoluten Dateinamen handelt, wird das Verzeichnis `ES_NODE_ROOT/dump` vorausgesetzt.

### **-cid** *neue\_objektgruppen-id*

Gibt eine neue Objektgruppen-ID an, wenn die Objektgruppe mit einer anderen ID importiert werden muss.

**Tipp:** Sie können die Objektgruppen-ID der Objektgruppe, die Sie importieren wollen, mithilfe der Administrationskonsole für die Unternehmenssuche ermitteln oder indem Sie den Befehl **esadmin report collections** verwenden.

### **-name** *neuer\_objektgruppenname*

Gibt einen neuen Objektgruppennamen an, wenn die Objektgruppe unter einem anderen Namen importiert werden muss.

### **-colDataDir** *neues\_datenverzeichnis\_der\_objektgruppe*

Gibt das Datenverzeichnis für die Objektgruppe an. Wenn Sie diese Angabe auslassen, wird ein Standardverzeichnis zugeordnet.

### **-force**

Erzwingt den Import einer Objektgruppe, die bereits im Zielsystem vorhanden ist. Das System verwendet die Objektgruppen-ID der importierten Objektgruppe um zu ermitteln, welche Objektgruppe im Zielsystem ersetzt wird.

### **Wichtig:**

- Nachdem Sie eine Objektgruppe importiert haben, können Sie keine weitere Objektgruppe mit derselben Objektgruppen-ID importieren, es sei denn Sie verwenden diese Option, um den Import der Objektgruppe zu erzwingen.

- Eine Objektgruppe, deren Import erzwungen wird, behält die Textanalysesteuerkomponenten, Wörterverzeichnisse und Crawler-Plug-ins nicht bei. Diese Angaben werden von den in der importierten Objektgruppe enthaltenen Informationen überschrieben.
- Alle der Objektgruppe zugeordneten Crawler werden entfernt und durch die Crawler ersetzt, die in der Importdatei angegeben sind. Da die Crawler entfernt werden, werden auch alle Crawlermetadaten und Dokumente entfernt, die noch nicht syntaktisch analysiert wurden. Nach dem Import der Objektgruppe müssen Sie für alle Dokumente eine erneute Crawler-suche ausführen, um diese Daten für die importierten Crawler neu zu generieren.

**-verbose**

Zeigt Informationen an, die bei der Fehlersuche in der Importoperation verwendet werden können.

## Verwendungsrichtlinien

### Textanalysesteuerkomponenten und Wörterverzeichnisse

Alle Textanalysesteuerkomponenten und Wörterverzeichnisse (wie z. B. Verzeichnisse von Synonymen, Stoppwörtern und Boostwörtern), die einer Objektgruppe im Exportsystem zugeordnet sind, werden nicht mit den Objektgruppenkonfigurationsdaten exportiert. Es werden lediglich die Namenszuordnungen zu Textanalysesteuerkomponenten und Wörterverzeichnissen exportiert.

Wenn beim Import eine Textanalysesteuerkomponente oder ein Wörterverzeichnis mit demselben Namen im Zielsystem vorhanden ist, werden diese der importierten Objektgruppe zugeordnet. Andernfalls wird eine Warnung angezeigt, die angibt, dass die Zuordnungen nicht ausgeführt werden konnten. Aus diesem Grund ist es sinnvoll, die Textanalysesteuerkomponenten und Wörterverzeichnisse in das Zielsystem zu laden und dieselben Namen zu verwenden, die auch in der importierten Objektgruppe verwendet werden.

Wenn keine Textanalysesteuerkomponente oder kein Wörterverzeichnis mit demselben Namen im Zielsystem vorhanden ist, wird die Zuordnung aufgehoben. Nach der Aufhebung der Zuordnung, kann die Objektgruppe ordnungsgemäß verwendet werden, die entsprechenden Textanalysesteuerkomponenten oder Wörterverzeichnisse werden jedoch nicht von der Objektgruppe verwendet.

### Speichermodelle

Das für das Exportsystem konfigurierte Speichermodell wird nicht beibehalten. Das für das Zielsystem konfigurierte Speichermodell bleibt so, wie es konfiguriert wurde. Wenn Sie versuchen, eine Objektgruppe in ein System zu importieren, das für ein kleineres Speichermodell konfiguriert ist, wird eine Warnung angezeigt. Es ist möglich, dass die Objektgruppe sich nicht wie erwartet verhält und unter Umständen ernsthafte Auswirkungen auf die Ressourcennutzung des kleineren Systems hat. Sie sollten eine Objektgruppe in ein System importieren, das dasselbe Modell oder ein Modell mit einem größerem Speicher als das Exportsystem verwendet, um diese Probleme zu vermeiden.

### Crawler

Sie müssen sicherstellen, dass die Voraussetzungen für Crawler im Zielsystem installiert und konfiguriert sind.

Importierte Crawler funktionieren nur, wenn die Datenquellen zugänglich sind, die von den Crawlern durchsucht werden. Wenn die Objektgruppe z. B. einen Dateisystemcrawler enthält, der ein lokales Dateisystem durchsucht, kann der Crawler erst mit der Crawlersuche beginnen, wenn dieselbe Verzeichnisstruktur im Zielsystem vorhanden ist.

Crawler-Plug-ins werden nicht exportiert. Beim Import einer Objektgruppe wird eine Warnung angezeigt, und anschließend wird der Import fertig gestellt. Nach dem Import der Objektgruppe müssen Sie die Crawlermerkmale aktualisieren und die Crawler-Plug-ins zuordnen. Sie müssen sicherstellen, dass die Crawler-Plug-ins auf dem Zielsystem implementiert sind, damit eine korrekte Crawlersuche durch den Crawler gewährleistet ist.

### **Zugehörige Verweise**

„Befehle, Rückkehrcodes und Sitzungs-IDs der Unternehmenssuche“ auf Seite 401

---

## Integration in Lotus Notes Version 8

Sie können die Suchfunktionalitäten von IBM Lotus Notes-Clientimplementierungen der Version 8 erweitern, indem Sie das Lotus Notes-Such-Plug-in von Omni-Find Enterprise Edition implementieren. Mit diesem Plug-in können Benutzer Objektgruppen für die Unternehmenssuche über die Suchsymbolleiste des Lotus Notes-Clients durchsuchen.

Damit Sie die Funktionalität der Unternehmenssuche in Lotus Notes integrieren können, müssen Sie die Update-Site für das Plug-in erstellen. Danach können die Benutzer das Plug-in auf ihren lokalen Lotus Notes-Clientensystemen der Version 8 erstellen.

---

### Erstellen der Update-Site für das Plug-in für die Unternehmenssuche

Damit Sie die Unternehmenssuche in Lotus Notes Version 8 integrieren können, müssen Sie die Eclipse-Update-Site für die Unternehmenssuche auf einem Web-Server Ihres Unternehmens implementieren. Mit der Update-Site können Benutzer das Plug-in für die Unternehmenssuche in ihren lokalen Lotus Notes-Clientinstallationen implementieren.

#### Informationen zu dieser Task

Die Dateien, die für die Erstellung der Update-Site für die Unternehmenssuche erforderlich sind, werden in der Datei `com.ibm.es.notes.search.plugin_8.4.0.150.zip` bereitgestellt.

#### Vorgehensweise

Gehen Sie wie folgt vor, um die Eclipse-Update-Site für die Unternehmenssuche zu erstellen:

1. Kopieren Sie die Datei `com.ibm.es.notes.search.plugin_8.4.0.150.zip` auf einen Web-Server Ihres Unternehmens.
2. Entpacken Sie den Inhalt der komprimierten Datei in einem Unterverzeichnis eines gemeinsam genutzten Verzeichnisses Ihres Web-Server-Installationspfads.  
Beispiel: Wenn Sie die Update-Site auf dem Web-Server implementieren wollen, den Sie für die Unternehmenssuche verwenden, entpacken Sie den Inhalt der komprimierten Datei im Verzeichnis `C:\Programme\IBM\HTTP Server\htdocs\en_US\omnifind`.

---

### Installieren des Plug-ins für die Unternehmenssuche in einem Lotus Notes-Client der Version 8

Wenn Sie die Funktionen der Unternehmenssuche für die Abfrageverarbeitung verwenden wollen, können Sie das Plug-in für die Unternehmenssuche der Lotus Notes-Suchleiste hinzufügen.

#### Informationen zu dieser Task

Die rechte obere Ecke des Lotus Notes-Clients enthält eine Suchleiste. Die Suchleiste enthält mehrere Such-Plug-ins, die mit der Basisinstallation des Lotus Notes-

Clients bereitgestellt werden. Nachdem Sie das Plug-in für die Unternehmenssuche installiert haben, können Sie Objektgruppen für die Unternehmenssuche über die Suchleiste durchsuchen.

### Vorgehensweise

Gehen Sie wie folgt vor, um das Plug-in für die Unternehmenssuche im Lotus Notes-Client zu installieren:

1. Öffnen Sie die Datei NOTES\_INSTALL\_ROOT\notes.ini, wobei NOTES\_INSTALL\_ROOT in der Regel c:\lotus\notes ist.
2. Fügen Sie am Ende der Datei ein Merkmal namens OMNIFIND\_ENTERPRISE\_EDITION\_SEARCH\_SERVER\_URL ein, und geben Sie `http://hostname:port/ESearchApplication/search.do?q=` als Wert ein, wobei *hostname* der Hostname des Suchservers für die Unternehmenssuche und *port* der Port des Web-Servers ist.
3. Speichern und schließen Sie die Datei.
4. Öffnen Sie die Lotus Notes-Clientanwendung.
5. Wählen Sie **File** → **Application** → **Install** aus.
6. Wählen Sie den Radioknopf **Search for new features to install** aus.
7. Klicken Sie **Add Remote Location** an.
8. Geben Sie einen Namen für die neue Speicherposition an, beispielsweise Unternehmenssuche. Geben Sie in das URL-Feld die Stamm-URL-Adresse der Update-Site für die Unternehmenssuche ein. Wenn Sie die Datei `com.ibm.es.notes.search.plugin_8.4.0.150.zip` z. B. im Unterverzeichnis `omnifind` des englischen Stammverzeichnis von IBM HTTP Server entpackt haben, geben Sie die folgende URL-Adresse an: `http://hostname:port/omnifind/`.
9. Klicken Sie **Finish** an, um den neuen fernen Standort zu speichern.
10. Wählen Sie das Markierungsfeld neben dem Namen des Standorts aus, und klicken Sie **Next** an.
11. Lesen Sie die Lizenzvereinbarung, wählen Sie **I accept the terms in the license agreement** aus, und klicken Sie **Next** an.
12. Klicken Sie **Finish** an, um das Plug-in zu installieren. Wenn eine Warnung angezeigt wird, dass Sie ein nicht signiertes Plug-in installieren, wählen Sie den Radioknopf **Install this plug-in** aus, und klicken Sie **OK** an.
13. Klicken Sie **Yes** an, um den Lotus Notes-Client erneut zu starten, wenn Sie dazu aufgefordert werden.

Sie können nun das Menü neben der Lotus Notes-Suchleiste anklicken, und **Omni-Find Enterprise Edition** als Suchoption auswählen. Wenn Sie einen Abfragebegriff eingeben und den Knopf **Search** anklicken, wird die Anforderung an den Server für die Unternehmenssuche übertragen, und die Ergebnisse werden im Fenster des Lotus Notes-Web-Browsers angezeigt.

---

## Integration in WebSphere Portal

Sie können die Suchfunktionalität von IBM WebSphere Portal erweitern, indem Sie das Suchportlet für Portlets für die Unternehmenssuche in WebSphere Portal implementieren und WebSphere Portal für die Verwendung des Suchportlets als Standardsuchmaschine konfigurieren.

### Integrationspunkte

Das OmniFind Enterprise Edition-Installationsprogramm stellt Setup-Skripts für die Integration der Unternehmenssuche in WebSphere Portal bereit. Nachdem Sie diese Skripts ausgeführt haben, kann Ihr System für die Unternehmenssuche auf verschiedene Weisen in WebSphere Portal integriert werden:

#### Suchportlet für die Unternehmenssuche

WebSphere Portal bietet dem Benutzer einen zentralen Zugang für die Interaktion mit Anwendungen, Inhalt, Prozessen und Personen. Das WebSphere Portal-Framework ermöglicht die Integration und die Implementierung neuer Anwendungen, so genannter Portlets, ohne dass andere Anwendungen im Portal beeinträchtigt werden.

Wenn Sie das Portlet für die Unternehmenssuche in WebSphere Portal implementieren, können Sie über die Schnittstelle von WebSphere Portal Objektgruppen für die Unternehmenssuche durchsuchen und mit den Suchergebnissen arbeiten. Anhand der Konfigurationseinstellungen von WebSphere Portal können Sie sicherstellen, dass das Portlet für die Unternehmenssuche dieselbe Darstellung und Funktionsweise (Look-and-Feel) aufweist, wie andere Portlets in Ihrer WebSphere Portal-Umgebung.

#### WebSphere Portal-Suchzentrale

Die WebSphere Portal-Suchzentrale bietet einen zentralen Ausgangspunkt zum Durchsuchen aller Quellen, die für das Durchsuchen über WebSphere Portal zur Verfügung stehen. Die Suchzentrale und das universale Suchportlet ermöglichen Ihnen, WebSphere Portal-Inhalt sowie andere in der Suchzentrale registrierte Objektgruppen zu durchsuchen.

Wenn Sie die Setup-Skripts für die Integration der Unternehmenssuche in WebSphere Portal Version 5.1 ausführen, wird eine Seite für die Unternehmenssuche einer Seite in der Schnittstelle der Suchzentrale hinzugefügt. Sie können diese Seite auswählen, um lediglich Objektgruppen für die Unternehmenssuche zu durchsuchen, oder Sie können eine Suchabfrage eingeben, die sowohl die Objektgruppen für die Unternehmenssuche als auch andere Objektgruppen durchsucht, die in der Suchzentrale verfügbar sind.

Wenn Sie die Setup-Skripts für die Integration der Unternehmenssuche in WebSphere Portal Version 6 ausführen, wird die Funktionalität für die Unternehmenssuche als föderierter Service integriert, mit dem Sie Objektgruppen für die Unternehmenssuche und andere Objektgruppen durchsuchen können, die in der Suchzentrale verfügbar sind.

#### WebSphere Portal-Suchleiste

Die rechte obere Ecke aller WebSphere Portal-Schnittstellenthemen enthält eine Suchleiste. Das Standardverhalten dieser Leiste ist es, alle Suchanforderungen an die Suchmaschine der Suchzentrale weiterzuleiten. Wenn Sie die leistungsfähigeren Funktionen der Unternehmenssuche für die Abfrage-

verarbeitung verwenden wollen, können Sie das Standardverhalten ändern, damit alle Suchanforderungen stattdessen an das Suchportlet der Unternehmenssuche umgeleitet werden.

### **WebSphere Portal- und Web Content Management-Crawler**

Sie können mit der Administrationskonsole für die Unternehmenssuche WebSphere Portal- und Web Content Management-Crawler konfigurieren, um WebSphere Portal-Sites und IBM Workplace Web Content Management-Sites einem Index für die Unternehmenssuche hinzuzufügen. Danach können Sie mit dem Portlet für die Unternehmenssuche oder mit einer Suchanwendung den indexierten Inhalt durchsuchen.

Der WebSphere Portal-Crawler kann WebSphere Portal-Sites der Versionen 5.1 und 6 durchsuchen. Der Web Content Management-Crawler kann Sites auf einem WebSphere Portal-Server der Version 6 durchsuchen.

### **IBM Lotus Quickr-Dokumente**

Wenn Sie IBM Lotus Quickr-Dokumente in einen Index für die Unternehmenssuche einschließen wollen, können Sie die Administrationskonsole für die Unternehmenssuche verwenden, um einen Crawler für die Einstiegspunktliste zu konfigurieren. Danach können Sie das Portlet für die Unternehmenssuche in WebSphere Portal oder eine eigenständige Suchanwendung verwenden, um den indexierten Inhalt zu durchsuchen.

Der Crawler für die Einstiegspunktliste kann Lotus Quickr-Inhaltsbibliotheken bzw. Dokumentbibliotheken auf einem WebSphere Portal-Server der Version 6 durchsuchen.

## **Vorteile der Integration**

Die Unternehmenssuche stellt eine funktionale Erweiterung der WebSphere Portal-Suchumgebung dar, indem Unterstützung für das Durchsuchen einer größeren Bandbreite von Datenquellentypen geboten wird. Mit dem Portlet für die Unternehmenssuche können Sie Websites und alle anderen Datenquellentypen durchsuchen, die von einem System für die Unternehmenssuche unterstützt werden.

Die Unternehmenssuche bietet auch im Hinblick auf die Skalierbarkeit Vorteile. Portal Search Engine ist nützlich für kleine oder mittelständische Unternehmen, in denen ein einziger Server ausreichend ist, um die Such- und Abrufauslastung zu unterstützen. Zur Unterstützung der Kapazität eines Unternehmens kann die Auslastung der Unternehmenssuche auf mehrere Server verteilt werden, wobei zwei Server die Unterstützung für die Verarbeitung von Suchen und Abfragen bereitstellen.

---

## **Setup-Scripts zum Integrieren der Unternehmenssuche in WebSphere Portal**

Sie können mit dem OmniFind Enterprise Edition-Installationsprogramm bereitgestellte Setup-Scripts ausführen, um die Unternehmenssuche in IBM WebSphere Portal zu integrieren.

Sie müssen die JAR-Datei, die die Setup-Scripts für Ihre Version von WebSphere Portal enthält, vom Server für die Unternehmenssuche auf den Server kopieren, auf dem WebSphere Portal installiert ist. Mit den Setup-Scripts wird Folgendes erreicht:

- Implementieren von EAR-Dateien, mit denen Sie die Unternehmenssuche in WebSphere Portal verwenden können, und Erstellen von Crawlern, mit denen

Inhalt von WebSphere Portal und IBM Workplace Web Content Management Objektgruppen für die Unternehmenssuche hinzugefügt werden kann.

- Implementieren von WAR-Dateien, die für das Portlet für die Unternehmenssuche erforderlich sind.
- Erstellen von Seiten in WebSphere Portal und Zuordnen der Dateien des Portlets für die Unternehmenssuche zu diesen Seiten.
- Kopieren aller erforderlichen JAR-Dateien in das Installationsverzeichnis von WebSphere Portal. (JAR-Dateien, die sich bereits in den Installationsverzeichnissen befinden, werden gesichert, bevor die JAR-Dateien für die Unternehmenssuche kopiert werden.)
- Bereitstellen eines Integrationspunkts für WebSphere Information Integrator Content Edition, damit Portal Document Manager-Dokumente durchsucht werden können.

Nach Ausführung der Scripts müssen Sie die Verwaltungsschnittstelle von WebSphere Portal verwenden, um Merkmale des Suchportlets zu aktualisieren und Informationen zum Suchserver für die Unternehmenssuche anzugeben.

### Verwendungsrichtlinien

- Die Scripts konfigurieren alle Integrationspunkte zwischen der Unternehmenssuche und WebSphere Portal. Sie können beispielsweise nicht selektiv das Portlet installieren, ohne EAR-Dateien zu installieren, die die WebSphere Portal- und Web Content Management-Crawler unterstützen.
- Wenn Sie WebSphere Information Integrator Content Edition nicht konfigurieren und sich zu einem späteren Zeitpunkt für die Verwendung eines Portlets zum Durchsuchen von Portal Document Manager-Dokumenten entscheiden, müssen Sie ein Script ausführen, um die Unternehmenssuche aus WebSphere Portal zu entfernen. Sie können das Setup-Script dann erneut ausführen und den WebSphere Information Integrator Content Edition-Installationspfad angeben.
- Die Scripts stoppen WebSphere Portal und starten es erneut. Sie können das Script nach der normalen Arbeitszeit ausführen, um sicherzustellen, dass Ihre Benutzergruppe nicht von der Nichtverfügbarkeit von Portalservices betroffen ist.
- Tritt ein Fehler auf, während die Setup-Scripts ausgeführt werden, führen Sie das Setup-Script erneut aus. Tasks, die beim ersten Versuch erfolgreich abgeschlossen wurden, melden möglicherweise Fehler. Der Konfigurationsprozess wird aber fortgesetzt und führt die übrigen Tasks aus.
- Wenn Sie nach Ausführung des Setup-Scripts erstmals auf die Seite des Portlets für die Unternehmenssuche zugreifen, wird die Seite möglicherweise nur langsam aufgebaut, weil das System JavaServer Pages (JSP-Dateien) für das Portlet kompilieren muss.

---

## Konfigurieren der Unternehmenssuche in WebSphere Portal Version 5.1

Zur Integration eines Systems für die Unternehmenssuche in WebSphere Portal Version 5.1.0 oder höher verwenden Sie das Script `wp5_install`.

### Informationen zu dieser Task

Die für die Integration der Unternehmenssuche in WebSphere Portal erforderlichen Dateien werden in der Datei `es.wp5.install.jar` bereitgestellt. Wenn Sie diese Datei entpacken, werden die folgenden Dateien extrahiert:

- ESSearchPortlet.war
- ESSearchAdapterPortlet.war
- ESSearchAdapter.ear
- ESPACServer.ear
- esapi.jar
- siapi.jar
- es.security.jar
- Scriptdateien, Stapeldateien sowie XML- und JACL-Dateien, die für die Installation erforderlich sind

### Vorgehensweise

Gehen Sie wie folgt vor, um die Unternehmenssuche in ein WebSphere Portal-System der Version 5.1 zu integrieren:

1. Kopieren Sie die Datei `es.wp5.install.jar` vom Server für die Unternehmenssuche auf den WebSphere Portal-Server, und entpacken Sie dann die Datei mit dem Java-Befehl **JAR** (oder dem Befehl **TAR**).
2. Optional: Wenn Sie die Integration mit WebSphere Portal Document Manager (PDM) unterstützen wollen, führen Sie einen der folgenden Schritte aus:
  - Führen Sie das WebSphere Information Integrator Content Edition-Installationsprogramm aus, wählen Sie die Option für die Installation eines einzelnen Connectors aus, und installieren Sie den PDM-Connector auf dem WebSphere Portal-Server.
  - Erstellen Sie die WebSphere Information Integrator Content Edition-Verzeichnisstruktur auf dem WebSphere Portal-Server, und kopieren Sie die Dateien einer vorhandenen WebSphere Information Integrator Content Edition-Installation auf den WebSphere Portal-Server, wobei `CE_ROOT` das Installationsstammverzeichnis von WebSphere Information Integrator Content Edition angibt:

```
CE_ROOT/lib/vbr.jar
CE_ROOT/ejb/vbr_pdm.jar
CE_ROOT/war/services.war
CE_ROOT/vbr_services.properties
```

3. Führen Sie an einer Eingabeaufforderung den Befehl **wp5\_install.bat** (unter Windows) bzw. den Befehl **wp5\_install.sh** (unter AIX, Linux oder Solaris) aus. Das folgende Beispiel zeigt Parameter aus Gründen der Lesbarkeit in separaten Zeilen. Sie müssen die Parameter mit dem Befehl angeben:

```
wp5_install.bat
-WASDir "C:\\Programme\\WebSphere\\AppServer"
-WASUser wpsbind -WASPassword wpsbind
-WPSDir "C:\\Programme\\WebSphere\\PortalServer"
-WPSUser wpsadmin -WSPassword wpsadmin
-WPSHost "portalserver.ibm.com:9081"
-IICEDir "C:\\IICE"
```

#### **WASDir**

Der vollständig qualifizierte Pfad für das WebSphere Application Server-Installationsverzeichnis.

#### **WASUser**

Der Benutzername für den WebSphere Application Server-Benutzer mit Verwaltungsaufgaben. Er ist nur erforderlich, wenn die globale Sicherheit in WebSphere Application Server aktiviert ist.

**WASPassword**

Das Kennwort für den WebSphere Application Server-Benutzer mit Verwaltungsaufgaben, falls angegeben.

**WPSDir**

Der vollständig qualifizierte Pfad für das WebSphere Portal-Installationsverzeichnis.

**WPSUser**

Der Benutzername für den WebSphere Portal-Benutzer mit Verwaltungsaufgaben.

**WSPassword**

Das Kennwort für den angegebenen WebSphere Portal-Benutzer mit Verwaltungsaufgaben.

**WPSHost**

Der Hostname und die Portnummer des WebSphere Portal-Servers.

**IICEDir**

Der vollständig qualifizierte Pfad für das Installationsverzeichnis von WebSphere Information Integrator Content Edition. Er ist nur erforderlich, wenn Sie den Portal Document Manager-Connector bereits auf dem WebSphere Portal-Server konfiguriert haben.

4. Nachdem Sie das Script (WebSphere Portal wird gestoppt und erneut gestartet) ausgeführt haben, aktualisieren Sie das Portlet für die Unternehmenssuche, um den Suchserver anzugeben:
  - a. Melden Sie sich mit der Portal-Administrator-ID und dem Portal-Kennwort an WebSphere Portal an.
  - b. Klicken Sie **Verwaltung** in der rechten oberen Ecke an.
  - c. Klicken Sie **Portlet-Verwaltung** im Navigationsbereich auf der linken Seite an, und klicken Sie anschließend **Portlets** an.
  - d. Ändern Sie die Option **Suchen nach** in **Der Titel enthält**.
  - e. Geben Sie Unternehmenssuche in das Feld **Suchen** ein, und klicken Sie den Knopf **Suchen** an.
  - f. Wenn neue Symbole auf der rechten Seite angezeigt werden, klicken Sie das Schraubenschlüsselsymbol an, um das Suchportlet für die Unternehmenssuche zu konfigurieren.
  - g. Modifizieren Sie in der Liste mit den Portletparametern die folgenden Parameter:

**hostname**

Geben Sie den vollständig qualifizierten Hostnamen eines Suchservers für die Unternehmenssuche an.

**port**

Geben Sie die Portnummer an, die WebSphere Application Server auf dem Suchserver für die Unternehmenssuche verwendet. Der Standardwert ist 80. (Der Standardwert für die SSL-Kommunikation ist 443.)

**benutzername**

Geben Sie einen Benutzernamen an, der in einer WebSphere Application Server-Benutzerregistry gültig ist, wenn die globale Sicherheit auf dem Suchserver in WebSphere Application Server aktiviert ist.

**kennwort**

Wenn Sie einen WebSphere Application Server-Benutzernamen angegeben haben, geben Sie das entsprechende Kennwort an.

**protokoll**

Geben Sie das Protokoll an, das für die Kommunikation zwischen WebSphere Portal und dem Suchserver verwendet wird. Der Standardwert ist HTTP. Geben Sie HTTPS an, wenn Sie SSL verwenden.

**truststore**

Geben Sie den vollständig qualifizierten Pfad (mit dem Dateinamen) für den SSL-Zertifikatsspeicher an, wenn Sie SSL verwenden.

**trustkennwort**

Geben Sie das Kennwort für die angegebene Truststore-Datei an, wenn Sie SSL (Secure Sockets Layer) verwenden.

**ssoCookieName**

Geben Sie den Namen des Cookies an, das die SSO-Tokenzeichenfolge (Single Sign-on) enthält. Der Standardwert ist LtpaToken.

**proxyHost**

Geben Sie den vollständig qualifizierten Hostnamen eines Proxy-Servers an, wenn ein Proxy-Server erforderlich ist, um auf den Suchserver für die Unternehmenssuche zuzugreifen.

**proxyPort**

Geben Sie die Portnummer für den Proxy-Server an, wenn Sie einen Proxy-Server angegeben haben.

**proxyUser**

Wenn der Proxy-Server eine Basisauthentifizierung erfordert, geben Sie einen Benutzernamen an, der für die Anmeldung am Proxy-Server verwendet werden soll.

**proxyPassword**

Wenn Sie einen Benutzernamen für den Proxy-Server angegeben haben, geben Sie das entsprechende Kennwort an.

- h. Klicken Sie **OK** an, um Ihre Änderungen zu speichern.

## Konfigurieren der Suchleiste von WebSphere Portal Version 5.1 Search für die Verwendung der Unternehmenssuche

Sie können WebSphere Portal Version 5.1.0 oder höher so konfigurieren, dass es die Unternehmenssuche statt der standardmäßigen WebSphere Portal-Suchmaschine verwendet, wenn Benutzer in der Suchleiste Abfragen übergeben.

### Vorbereitung

Bevor Sie Suchanforderungen an die Unternehmenssuche umleiten können, müssen Sie das Setup-Skript **wp5\_install** ausführen, um die Unternehmenssuche in WebSphere Portal zu integrieren. Sie müssen außerdem die Portletparameter für die Unternehmenssuche aktualisieren, um den Hostnamen, den Port und andere Informationen zum Suchserver für die Unternehmenssuche anzugeben.

### Informationen zu dieser Task

Die rechte obere Ecke aller WebSphere Portal-Schnittstellenthemen enthält eine Suchleiste. Das Standardverhalten dieser Leiste ist es, alle Suchanforderungen an das Portlet der Suchzentrale weiterzuleiten. Wenn Sie die leistungsfähigeren Funktionen der Unternehmenssuche für die Abfrageverarbeitung verwenden wollen, können Sie das Standardverhalten ändern, damit alle Suchanforderungen stattdessen an das Suchportlet der Unternehmenssuche umgeleitet werden.



## Entfernen der Unternehmenssuche aus WebSphere Portal Version 5.1

Zum Entfernen der Unternehmenssuche aus einem WebSphere Portal-System der Version 5.1.0 oder höher verwenden Sie das Script `wp5_uninstall`.

### Informationen zu dieser Task

Wenn Sie die Unternehmenssuche aus WebSphere Portal entfernen, werden die Portletparameter nicht gespeichert, die Sie als Teil des Installationsprozesses für das Portlet der Unternehmenssuche angegeben haben.

Wenn Sie das Script starten, stoppt das Script den WebSphere Portal-Server. Nachdem die Software der Unternehmenssuche entfernt wurde, startet das Script den WebSphere Portal-Server erneut.

### Vorgehensweise

Gehen Sie wie folgt vor, um die Unternehmenssuche von einem WebSphere Portal-System der Version 5.1 zu entfernen:

Führen Sie an einer Eingabeaufforderung den Befehl **wp5\_uninstall.bat** (unter Windows) bzw. den Befehl **wp5\_uninstall.sh** (unter AIX, Linux oder Solaris) aus. Das folgende Beispiel zeigt Parameter aus Gründen der Lesbarkeit in separaten Zeilen. Sie müssen die Parameter mit dem Befehl angeben:

```
wp5_uninstall.bat
-WASDir "C:\\Programme\\WebSphere\\AppServer"
-WASUser wpsbind -WASPassword wpsbind
-WPSDir "C:\\Programme\\WebSphere\\PortalServer"
-WPSUser wpsadmin -WSPassword wpsadmin
-WPSHost "portalserver.ibm.com:9081"
```

#### **WASDir**

Der vollständig qualifizierte Pfad für das WebSphere Application Server-Installationsverzeichnis.

#### **WASUser**

Der Benutzername für den WebSphere Application Server-Benutzer mit Verwaltungsaufgaben. Er ist nur erforderlich, wenn die globale Sicherheit in WebSphere Application Server aktiviert ist.

#### **WASPassword**

Das Kennwort für den WebSphere Application Server-Benutzer mit Verwaltungsaufgaben, falls angegeben.

#### **WPSDir**

Der vollständig qualifizierte Pfad für das WebSphere Portal-Installationsverzeichnis.

#### **WPSUser**

Der Benutzername für den WebSphere Portal-Benutzer mit Verwaltungsaufgaben.

#### **WSPassword**

Das Kennwort für den angegebenen WebSphere Portal-Benutzer mit Verwaltungsaufgaben.

#### **WPSHost**

Der Hostname und die Portnummer des WebSphere Portal-Servers.

---

## Konfigurieren der Unternehmenssuche in WebSphere Portal Version 6

Zur Integration eines Systems für die Unternehmenssuche in WebSphere Portal Version 6 verwenden Sie das Script `wp6_install`.

### Informationen zu dieser Task

Die für die Integration der Unternehmenssuche in WebSphere Portal erforderlichen Dateien werden in der Datei `es.wp6.install.jar` bereitgestellt. Wenn Sie diese Datei entpacken, werden die folgenden Dateien extrahiert:

- `ESSearchPortlet.war`
- `ESPACServer.ear`
- `esapi.jar`
- `es.search.provider.jar`
- `es.security.jar`
- Symbole für die Suchanwendungsquellentypen, die auf der Ergebnisseite des Suchanbieters verwendet werden
- Scriptdateien, Stapeldateien sowie XML- und JACL-Dateien, die für die Installation erforderlich sind

### Vorgehensweise

Gehen Sie wie folgt vor, um die Unternehmenssuche in ein WebSphere Portal-System der Version 6 zu integrieren:

1. Kopieren Sie die Datei `es.wp6.install.jar` vom Server für die Unternehmenssuche auf den WebSphere Portal-Server, und entpacken Sie dann die Datei mit dem Java-Befehl **JAR** (oder dem Befehl **TAR**).
2. Optional: Wenn Sie die Integration mit WebSphere Portal Document Manager (PDM) unterstützen wollen, führen Sie einen der folgenden Schritte aus:
  - Führen Sie das WebSphere Information Integrator Content Edition-Installationsprogramm aus, wählen Sie die Option für die Installation eines einzelnen Connectors aus, und installieren Sie den PDM-Connector auf dem WebSphere Portal-Server.
  - Erstellen Sie die WebSphere Information Integrator Content Edition-Verzeichnisstruktur auf dem WebSphere Portal-Server, und kopieren Sie die Dateien einer vorhandenen WebSphere Information Integrator Content Edition-Installation auf den WebSphere Portal-Server, wobei `CE_ROOT` das Installationsstammverzeichnis von WebSphere Information Integrator Content Edition angibt:

```
CE_ROOT/lib/vbr.jar
CE_ROOT/ejb/vbr_pdm.jar
CE_ROOT/war/services.war
CE_ROOT/vbr_services.properties
```

3. Führen Sie an einer Eingabeaufforderung den Befehl **wp6\_install.bat** (unter Windows) bzw. den Befehl **wp6\_install.sh** (unter AIX, Linux oder Solaris) aus. Das folgende Beispiel zeigt Optionen aus Gründen der Lesbarkeit in separaten Zeilen. Sie müssen die Optionen mit dem Befehl angeben:

```
wp6_install.bat
-WPSProfileDir "C:\\Programme\\IBM\\WebSphere\\profiles\\wp_profile"
-WASDir "C:\\Programme\\IBM\\WebSphere\\AppServer"
-WASUser wpsbind -WASPassword wpsbind
```

```
-WPSDir "C:\\Programme\\IBM\\WebSphere\\PortalServer"
-WPSUser wpsadmin -WSPassword wpsadmin
-WPSHost "portalserver.ibm.com:9081"
-IICEDir "C:\\IICE"
```

**WSPProfileDir**

Der vollständig qualifizierte Pfad für das WebSphere Portal-Profilverzeichnis. Auf AIX-Systemen ist der Standardpfad /usr/IBM/WebSphere/AppServer/profiles/wp\_profile, auf Linux- und Solaris-Systemen /opt/IBM/WebSphere/AppServer/profiles/wp\_profile und auf Windows-Systemen C:\\Programme\\IBM\\WebSphere\\profiles\\wp\_profile.

**WASDir**

Der vollständig qualifizierte Pfad für das WebSphere Application Server-Stammverzeichnis. Er ist nur auf AIX-, Linux- und Solaris-Systemen erforderlich. Auf AIX-Systemen ist das Standardstammverzeichnis /usr/IBM/WebSphere/AppServer, auf Linux- und Solaris-Systemen /opt/IBM/WebSphere/AppServer und auf Windows-Systemen C:\\Programme\\IBM\\WebSphere\\AppServer.

**WASUser**

Der Benutzername für den WebSphere Application Server-Benutzer mit Verwaltungsaufgaben. Er ist nur erforderlich, wenn die globale Sicherheit in WebSphere Application Server aktiviert ist.

**WASPassword**

Das Kennwort für den WebSphere Application Server-Benutzer mit Verwaltungsaufgaben, falls angegeben.

**WPSDir**

Der vollständig qualifizierte Pfad für das WebSphere Portal-Installationsverzeichnis.

**WPSUser**

Der Benutzername für den WebSphere Portal-Benutzer mit Verwaltungsaufgaben.

**WSPPassword**

Das Kennwort für den angegebenen WebSphere Portal-Benutzer mit Verwaltungsaufgaben.

**WPSHost**

Der Hostname und die Portnummer des WebSphere Portal-Servers.

**IICEDir**

Der vollständig qualifizierte Pfad für das Installationsverzeichnis von WebSphere Information Integrator Content Edition. Er ist nur erforderlich, wenn Sie den Portal Document Manager-Connector bereits auf dem WebSphere Portal-Server konfiguriert haben.

4. Nachdem Sie das Script ausgeführt und WebSphere Portal gestoppt und erneut gestartet haben, aktualisieren Sie das Portlet für die Unternehmenssuche, um den Suchserver anzugeben:
  - a. Melden Sie sich mit der Portal-Administrator-ID und dem Portal-Kennwort an WebSphere Portal an.
  - b. Klicken Sie **Verwaltung** in der linken unteren Ecke an.
  - c. Klicken Sie **Portlet-Verwaltung** im Navigationsbereich auf der linken Seite an, und klicken Sie anschließend **Portlets** an.
  - d. Ändern Sie die Option **Suchen nach** in **Der Titel enthält**.
  - e. Geben Sie Unternehmenssuche in das Feld **Suchen** ein, und klicken Sie den Knopf **Suchen** an.

- f. Wenn neue Symbole auf der rechten Seite angezeigt werden, klicken Sie das Schraubenschlüsselsymbol an, um das Suchportlet für die Unternehmenssuche zu konfigurieren.
- g. Modifizieren Sie in der Liste mit den Portletparametern die folgenden Parameter:

**hostname**

Geben Sie den vollständig qualifizierten Hostnamen eines Suchservers für die Unternehmenssuche an.

**port**

Geben Sie die Portnummer an, die WebSphere Application Server auf dem Suchserver für die Unternehmenssuche verwendet. Der Standardwert ist 80. (Der Standardwert für die SSL-Kommunikation ist 443.)

**benutzername**

Geben Sie einen Benutzernamen an, der in einer WebSphere Application Server-Benutzerregistry gültig ist, wenn die globale Sicherheit auf dem Suchserver in WebSphere Application Server aktiviert ist.

**kennwort**

Wenn Sie einen WebSphere Application Server-Benutzernamen angegeben haben, geben Sie das entsprechende Kennwort an.

**protokoll**

Geben Sie das Protokoll an, das für die Kommunikation zwischen WebSphere Portal und dem Suchserver verwendet wird. Der Standardwert ist HTTP. Geben Sie HTTPS an, wenn Sie SSL verwenden.

**truststore**

Geben Sie den vollständig qualifizierten Pfad (mit dem Dateinamen) für den SSL-Zertifikatsspeicher an, wenn Sie SSL verwenden.

**trustkennwort**

Geben Sie das Kennwort für die angegebene Truststore-Datei an, wenn Sie SSL (Secure Sockets Layer) verwenden.

**ssoCookieName**

Geben Sie den Namen des Cookies an, das die SSO-Tokenzeichenfolge (Single Sign-on) enthält. Der Standardwert ist LtpaToken.

**proxyHost**

Geben Sie den vollständig qualifizierten Hostnamen eines Proxy-Servers an, wenn ein Proxy-Server erforderlich ist, um auf den Suchserver für die Unternehmenssuche zuzugreifen.

**proxyPort**

Geben Sie die Portnummer für den Proxy-Server an, wenn Sie einen Proxy-Server angegeben haben.

**proxyUser**

Wenn der Proxy-Server eine Basisauthentifizierung erfordert, geben Sie einen Benutzernamen an, der für die Anmeldung am Proxy-Server verwendet werden soll.

**proxyPassword**

Wenn Sie einen Benutzernamen für den Proxy-Server angegeben haben, geben Sie das entsprechende Kennwort an.

- h. Klicken Sie **OK** an, um Ihre Änderungen zu speichern.

## Konfigurieren der Suchzentrale von WebSphere Portal Version 6 für die Unternehmenssuche

Sie können WebSphere Portal Version 6 so konfigurieren, dass es Objektgruppen für die Unternehmenssuche durchsucht, wenn Benutzer in der WebSphere Portal-Suchzentrale Abfragen übergeben.

### Einschränkungen

Wenn die zu durchsuchenden Objektgruppen für die Unternehmenssuche geschützt sind, müssen die Benutzer das Suchportlet für die Unternehmenssuche ausführen und ein Benutzerprofil konfigurieren. Das Profil wird verschlüsselt und in einem geschützten Speicher für die Unternehmenssuche gespeichert. Das Profil muss vorhanden sein, damit Benutzer Abfragen zum Durchsuchen sicherer Objektgruppen über die WebSphere Portal-Suchzentrale übergeben können.

### Informationen zu dieser Task

Die Suchzentrale in WebSphere Portal Version 6 unterstützt Funktionen für die föderierte Suche über mehrere Objektgruppen. Die Objektgruppen können verschiedene Arten von Inhalt enthalten, wie z. B. Portaldokumentbibliotheken und Portalinhalte (Seiten und Portlets). Wenn Sie die Setup-Scripts für die Integration der Unternehmenssuche in WebSphere Portal ausgeführt haben, können Sie die Suchzentrale so konfigurieren, dass sie auch Objektgruppen für die Unternehmenssuche durchsucht.

### Vorgehensweise

Gehen Sie wie folgt vor, um die Suchzentrale zum Durchsuchen von Objektgruppen für die Unternehmenssuche zu konfigurieren:

1. Melden Sie sich mit der Portal-Administrator-ID und dem Portal-Kennwort an WebSphere Portal an.
2. Klicken Sie **Verwaltung** in der linken unteren Ecke an.
3. Klicken Sie **Verwaltung der Suchfunktion** im Navigationsbereich auf der linken Seite an, und klicken Sie anschließend **Suchläufe verwalten** an.
4. Klicken Sie **Suchservices** an, und klicken Sie anschließend **Neuer Suchservice** an.
5. Wählen Sie den Suchservice der Unternehmenssuche im Feld **Implementierung des Suchservice** aus, und geben Sie dann in das Textfeld **ServiceName** den Namen ein, den Sie für den Service verwenden wollen.
6. Modifizieren Sie in der Liste mit den Parametern die folgenden Parameter:

#### **hostname**

Geben Sie den vollständig qualifizierten Hostnamen eines Suchservers für die Unternehmenssuche an.

#### **port**

Geben Sie die Portnummer an, die WebSphere Application Server auf dem Suchserver für die Unternehmenssuche verwendet. Der Standardwert ist 80. (Der Standardwert für die SSL-Kommunikation ist 443.)

#### **benutzername**

Geben Sie einen Benutzernamen an, der in einer WebSphere Application Server-Benutzerregistry gültig ist, wenn die globale Sicherheit auf dem Suchserver in WebSphere Application Server aktiviert ist.

**kennwort**

Wenn Sie einen WebSphere Application Server-Benutzernamen angegeben haben, geben Sie das entsprechende Kennwort an.

**protokoll**

Geben Sie das Protokoll an, das für die Kommunikation zwischen WebSphere Portal und dem Suchserver verwendet wird. Der Standardwert ist HTTP. Geben Sie HTTPS an, wenn Sie SSL verwenden.

**truststore**

Geben Sie den vollständig qualifizierten Pfad (mit dem Dateinamen) für den SSL-Zertifikatsspeicher an, wenn Sie SSL verwenden.

**trustkennwort**

Geben Sie das Kennwort für die angegebene Truststore-Datei an, wenn Sie SSL (Secure Sockets Layer) verwenden.

**ssoCookieName**

Geben Sie den Namen des Cookies an, das die SSO-Tokenzeichenfolge (Single Sign-on) enthält. Der Standardwert ist LtpaToken.

**proxyHost**

Geben Sie den vollständig qualifizierten Hostnamen eines Proxy-Servers an, wenn ein Proxy-Server erforderlich ist, um auf den Suchserver für die Unternehmenssuche zuzugreifen.

**proxyPort**

Geben Sie die Portnummer für den Proxy-Server an, wenn Sie einen Proxy-Server angegeben haben.

**proxyUser**

Wenn der Proxy-Server eine Basisauthentifizierung erfordert, geben Sie einen Benutzernamen an, der für die Anmeldung am Proxy-Server verwendet werden soll.

**proxyPassword**

Wenn Sie einen Benutzernamen für den Proxy-Server angegeben haben, geben Sie das entsprechende Kennwort an.

7. Klicken Sie **OK** an, um Ihre Änderungen zu speichern.

## Konfigurieren der Suchleiste von WebSphere Portal Version 6 Search für die Verwendung der Unternehmenssuche

Sie können WebSphere Portal Version 6 so konfigurieren, dass es die Unternehmenssuche statt der standardmäßigen WebSphere Portal-Suchmaschine verwendet, wenn Benutzer in der Suchleiste Abfragen übergeben.

### Vorbereitung

Bevor Sie Suchanforderungen an die Unternehmenssuche umleiten können, müssen Sie das Setup-Skript **wp6\_install** ausführen, um die Unternehmenssuche in WebSphere Portal zu integrieren. Sie müssen außerdem die Portletparameter für die Unternehmenssuche aktualisieren, um den Hostnamen, den Port und andere Informationen zum Suchserver für die Unternehmenssuche anzugeben.

### Informationen zu dieser Task

Die rechte obere Ecke aller WebSphere Portal-Schnittstellenthemen enthält eine Suchleiste. Das Standardverhalten dieser Leiste ist es, alle Suchanforderungen an das Portlet der Suchzentrale weiterzuleiten. Wenn Sie die leistungsfähigeren Funk-

tionen der Unternehmenssuche für die Abfrageverarbeitung verwenden wollen, können Sie das Standardverhalten ändern, damit alle Suchanforderungen stattdessen an das Suchportlet der Unternehmenssuche umgeleitet werden.

Wenn Sie die Suchleiste umleiten, wirkt sich die Änderung auf Seiten aus, die dasselbe WebSphere Portal-Thema wie das Suchportlet für die Unternehmenssuche verwenden. Diese Seiten müssen die Datei `banner_searchControl.jspf` aufrufen. Seiten, die ein anderes Thema verwenden oder die nicht die Datei `banner_searchControl.jspf` aufrufen, verwenden weiterhin das Standardportlet der Suchzentrale.

Nachdem Sie diese Task abgeschlossen haben, können Sie die Suchzentrale nicht mehr verwenden, bis Sie die Änderungen rückgängig machen. (Sie können beispielsweise die ursprüngliche Datei `banner_searchControl.jspf` wiederherstellen.)

### Vorgehensweise

Gehen Sie wie folgt vor, um das Portlet der Unternehmenssuche zu verwenden, wenn Benutzer Abfragen über die WebSphere Portal-Suchleiste übergeben:

1. Stoppen Sie die WebSphere Portal-Anwendungsserverinstanz.
2. Wechseln Sie auf dem WebSphere Portal-Server in das Verzeichnis `WPS_PROFILE_ROOT/installedApps/knotenname/wps.ear/wps.war/themes/html/aktuelles_thema`, wobei `knotenname` der Knotenname für den WebSphere Portal-Server und `aktuelles_thema` das derzeit für Ihren WebSphere Portal-Server angewendete Thema ist. Der Standardthemenname für einen WebSphere Portal-Server ist IBM.
3. Erstellen Sie ein Backup der Datei `banner_searchControl.jspf`, indem Sie sie kopieren und umbenennen (z. B. in `banner_searchControl.jspf.BACKUP`).
4. Bearbeiten Sie die Datei `banner_searchControl.jspf`, und ersetzen Sie den Inhalt durch den folgenden Text. Ersetzen Sie im Attribut `action=` des Maskenelements `localhost:10038` durch den Hostnamen und die Portnummer Ihres WebSphere Portal-Servers.

```
<%@ taglib uri="/WEB-INF/tld/SearchMenuControl.tld" prefix="searchmenu" %>
<%String ic = (bidiImageRTL == null) ? "icons/scope_search_submit.gif" :
"icons/scope_search_submit"+bidiImageRTL+".gif";%>

<searchmenu:adminlinkinfo name="SEARCH_CENTER">
<div class="searchControl">
 <form name="SearchForm" style="margin: 0px;" method="GET"
 action="http://localhost:10038/wps/omnifind/portalSearchBar.jsp">
 <table border="0" cellpadding="0" cellspacing="0">
 <tr>
 <td>
 <portal-fmt:text key="search.theme.control.label" bundle="nls.engine"/></td>
 <td valign="middle" style="padding: 0px 4px 0px 4px;">
 <input type="text" name="q"></input></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=">" />" type="image"></input></td>
 </tr>
 </table>
 </form>
 </div>
</searchmenu:adminlinkinfo>
```

5. Öffnen Sie die Datei `banner.jspf`, und speichern Sie die Datei. Dieser Schritt, durch den das Änderungsdatum der Datei aktualisiert wird, um sicherzustellen, dass die Datei erneut kompiliert wird, ist optional, wenn Sie Ihr eigenes Thema anstelle des WebSphere Portal-Standardthemas verwenden.
6. Öffnen Sie die Datei `Default.jsp`, und speichern Sie die Datei.

7. Starten Sie die WebSphere Portal-Anwendungsserverinstanz erneut.

## Konfigurieren des Portlets für die Unternehmenssuche für Lotus Quickr

Sie können das Portlet für die Unternehmenssuche in WebSphere Portal Version 6 so konfigurieren, dass Lotus Quickr-Quellen durchsucht werden.

### Vorbereitung

Führen Sie den Befehl **wp6\_install.bat** (unter Windows) bzw. den Befehl **wp6\_install.sh** (unter AIX, Linux oder Solaris) aus, und befolgen Sie die Anweisungen zum Einrichten der Unternehmenssuche in WebSphere Portal Version 6.

### Vorgehensweise

Gehen Sie wie folgt vor, um das Portlet für die Unternehmenssuche in WebSphere Portal Version 6 zum Durchsuchen von Lotus Quickr-Quellen zu konfigurieren:

1. Aktualisieren Sie die Portletparameter in der WebSphere Portal-Konfiguration:
  - a. Melden Sie sich mit der Lotus Quickr-Administrator-ID und dem entsprechenden Kennwort an WebSphere Portal an.
  - b. Klicken Sie **Site-Verwaltung** und anschließend **Erweiterte Verwaltung** an.
  - c. Klicken Sie **Portlet-Verwaltung** im Navigationsbereich auf der linken Seite an, und klicken Sie anschließend **Portlets** an.
  - d. Ändern Sie die Option **Suchen nach** in **Der Titel enthält**.
  - e. Geben Sie Unternehmenssuche in das Feld **Suchen** ein, und klicken Sie den Knopf **Suchen** an.
  - f. Wenn neue Symbole auf der rechten Seite angezeigt werden, klicken Sie das Schraubenschlüsselsymbol an, um das Suchportlet für die Unternehmenssuche zu konfigurieren.
  - g. Modifizieren Sie in der Liste mit den Portletparametern die folgenden Parameter:

#### hostname

Geben Sie den vollständig qualifizierten Hostnamen eines Suchservers für die Unternehmenssuche an.

#### port

Geben Sie die Portnummer an, die WebSphere Application Server auf dem Suchserver für die Unternehmenssuche verwendet. Der Standardwert ist 80. (Der Standardwert für die SSL-Kommunikation ist 443.)

#### benutzername

Geben Sie einen Benutzernamen an, der in einer WebSphere Application Server-Benutzerregistry gültig ist, wenn die globale Sicherheit auf dem Suchserver in WebSphere Application Server aktiviert ist.

#### kennwort

Wenn Sie einen WebSphere Application Server-Benutzernamen angegeben haben, geben Sie das entsprechende Kennwort an.

#### protokoll

Geben Sie das Protokoll an, das für die Kommunikation zwischen WebSphere Portal und dem Suchserver verwendet wird. Der Standardwert ist HTTP. Geben Sie HTTPS an, wenn Sie SSL verwenden.

**truststore**

Geben Sie den vollständig qualifizierten Pfad (mit dem Dateinamen) für den SSL-Zertifikatsspeicher an, wenn Sie SSL verwenden.

**trustkennwort**

Geben Sie das Kennwort für die angegebene Truststore-Datei an, wenn Sie SSL (Secure Sockets Layer) verwenden.

**ssoCookieName**

Geben Sie den Namen des Cookies an, das die SSO-Tokenzeichenfolge (Single Sign-on) enthält. Der Standardwert ist LtpaToken.

**proxyHost**

Geben Sie den vollständig qualifizierten Hostnamen eines Proxy-Servers an, wenn ein Proxy-Server erforderlich ist, um auf den Suchserver für die Unternehmenssuche zuzugreifen.

**proxyPort**

Geben Sie die Portnummer für den Proxy-Server an, wenn Sie einen Proxy-Server angegeben haben.

**proxyUser**

Wenn der Proxy-Server eine Basisauthentifizierung erfordert, geben Sie einen Benutzernamen an, der für die Anmeldung am Proxy-Server verwendet werden soll.

**proxyPassword**

Wenn Sie einen Benutzernamen für den Proxy-Server angegeben haben, geben Sie das entsprechende Kennwort an.

- h. Klicken Sie **OK** an, um Ihre Änderungen zu speichern.
2. Gehen Sie wie folgt vor, um nach der Konfiguration auf das Portlet zuzugreifen:
  - a. Melden Sie sich am Lotus Quickr-Server an.
  - b. Ändern Sie die URL-Adresse im Browserfenster folgendermaßen:  
`http://hostname:port/lotus/myquickr/ESearchPortlet`

## Entfernen der Unternehmenssuche aus WebSphere Portal Version 6

Zum Entfernen der Unternehmenssuche aus einem WebSphere Portal-System der Version 6 verwenden Sie das Script `wp6_uninstall`.

### Informationen zu dieser Task

Wenn Sie die Unternehmenssuche aus WebSphere Portal entfernen, werden die Portletparameter nicht gespeichert, die Sie als Teil des Installationsprozesses für das Portlet der Unternehmenssuche angegeben haben.

Wenn Sie das Script starten, stoppt das Script den WebSphere Portal-Server. Nachdem die Software der Unternehmenssuche entfernt wurde, startet das Script den WebSphere Portal-Server erneut.

### Vorgehensweise

Gehen Sie wie folgt vor, um die Unternehmenssuche von einem WebSphere Portal-System der Version 6 zu entfernen:

Führen Sie an einer Eingabeaufforderung den Befehl **wp6\_uninstall.bat** (unter Windows) bzw. den Befehl **wp6\_uninstall.sh** (unter AIX, Linux oder Solaris) aus. Das folgende Beispiel zeigt Parameter aus Gründen der Lesbarkeit in separaten Zeilen. Sie müssen die Parameter mit dem Befehl angeben:

```
wp6_uninstall.bat
-WPSProfileDir "C:\\Programme\\IBM\\WebSphere\\AppServer\\profiles\\wp_profile"
-WASDir "C:\\Programme\\IBM\\WebSphere\\AppServer"
-WASUser wpsbind -WASPassword wpsbind
-WPSDir "C:\\Programme\\IBM\\WebSphere\\PortalServer"
-WPSUser wpsadmin -WSPassword wpsadmin
-WPSHost "portalserver.ibm.com:9081"
```

**WPSProfileDir**

Der vollständig qualifizierte Pfad für das WebSphere Portal-Profilverzeichnis.

**WASDir**

Der vollständig qualifizierte Pfad für das WebSphere Application Server-Stammverzeichnis. Er ist nur auf AIX-, Linux- und Solaris-Systemen erforderlich.

**WASUser**

Der Benutzername für den WebSphere Application Server-Benutzer mit Verwaltungsaufgaben. Er ist nur erforderlich, wenn die globale Sicherheit in WebSphere Application Server aktiviert ist.

**WASPassword**

Das Kennwort für den WebSphere Application Server-Benutzer mit Verwaltungsaufgaben, falls angegeben.

**WPSDir**

Der vollständig qualifizierte Pfad für das WebSphere Portal-Installationsverzeichnis.

**WPSUser**

Der Benutzername für den WebSphere Portal-Benutzer mit Verwaltungsaufgaben.

**WSPassword**

Das Kennwort für den angegebenen WebSphere Portal-Benutzer mit Verwaltungsaufgaben.

**WPSHost**

Der Hostname und die Portnummer des WebSphere Portal-Servers.

---

## Integration der Unternehmenssuche in WebSphere Portal-Clustere

Sie können das Portlet für die Unternehmenssuche so einrichten, dass es in einem WebSphere Portal-Clustereystem der Version 6 ausgeführt wird.

### Verwendungsrichtlinien

- Bevor Sie die Setup-Skripte für die Unternehmenssuche ausführen, stellen Sie sicher, dass WebSphere Application Server Network Deployment Manager aktiv ist und dass alle Knoten des Clusters aktiv sind.
- Die Skripte konfigurieren alle Integrationspunkte zwischen der Unternehmenssuche und WebSphere Portal. Sie können beispielsweise nicht selektiv das Portlet installieren, ohne EAR-Dateien zu installieren, die die WebSphere Portal- und Web Content Management-Crawler unterstützen.

- Die Scripts stoppen und starten alle Instanzen des WebSphere Portal-Servers im Cluster erneut. Sie können das Script nach der normalen Arbeitszeit ausführen, um sicherzustellen, dass Ihre Benutzergruppe nicht von der Nichtverfügbarkeit von Portalservices betroffen ist.
- Tritt ein Fehler auf, während die Setup-Scripts ausgeführt werden, führen Sie das Setup-Script erneut aus. Tasks, die beim ersten Versuch erfolgreich abgeschlossen wurden, melden möglicherweise Fehler. Der Konfigurationsprozess wird aber fortgesetzt und führt die übrigen Tasks aus.
- Wenn Sie nach Ausführung des Setup-Scripts erstmals auf die Seite des Portlets für die Unternehmenssuche zugreifen, wird die Seite möglicherweise nur langsam aufgebaut, weil das System JavaServer Pages (JSP-Dateien) für das Portlet kompilieren muss.

## Konfigurieren der Unternehmenssuche in einem WebSphere Portal-Clustersystem

Zur Integration eines Systems für die Unternehmenssuche in einem WebSphere Portal-Clustersystem der Version 6 verwenden Sie das Script `wp6_cluster_install`.

### Informationen zu dieser Task

Die für die Integration der Unternehmenssuche in WebSphere Portal erforderlichen Dateien werden in der Datei `es.wp6.install.jar` bereitgestellt. Wenn Sie diese Datei entpacken, werden die folgenden Dateien extrahiert:

- `ESSearchPortlet.war`
- `ESPACTServer.ear`
- `esapi.jar`
- `es.search.provider.jar`
- `es.security.jar`
- Scriptdateien, Stapeldateien sowie XML- und JACL-Dateien, die für die Installation erforderlich sind

### Vorgehensweise

Gehen Sie wie folgt vor, um die Unternehmenssuche in einem WebSphere Portal-Clustersystem der Version 6 zu integrieren:

1. Kopieren Sie die Datei `es.wp6.install.jar` vom Server für die Unternehmenssuche auf jeden Knoten des Clusters, auf dem WebSphere Portal installiert ist, und entpacken Sie dann die Datei mit dem Java-Befehl `JAR` (oder dem Befehl `TAR`).
2. Optional: Wenn Sie die Integration mit WebSphere Portal Document Manager (PDM) unterstützen wollen, führen Sie auf jedem Knoten des Clusters einen der folgenden Schritte aus:
  - Führen Sie das WebSphere Information Integrator Content Edition-Installationsprogramm aus, wählen Sie die Option für die Installation eines einzelnen Connectors aus, und installieren Sie den PDM-Connector auf dem WebSphere Portal-Server.
  - Erstellen Sie die WebSphere Information Integrator Content Edition-Verzeichnisstruktur auf dem WebSphere Portal-Server, und kopieren Sie die Dateien einer vorhandenen WebSphere Information Integrator Content Edition-Installation auf den WebSphere Portal-Server, wobei `CE_ROOT` das Installationsstammverzeichnis von WebSphere Information Integrator Content Edition angibt:

```
CE_ROOT/lib/vbr.jar
CE_ROOT/ejb/vbr_pdm.jar
CE_ROOT/war/services.war
CE_ROOT/vbr_services.properties
```

3. Führen Sie an einer Eingabeaufforderung den Befehl **wp6\_cluster\_copyFiles.bat** (unter Windows) bzw. den Befehl **wp6\_cluster\_copyFiles.sh** (unter AIX, Linux oder Solaris) auf jedem Knoten des Clusters aus, auf dem WebSphere Portal installiert ist. Die folgenden Beispiele zeigen Optionen aus Gründen der Lesbarkeit in separaten Zeilen. Sie müssen die Optionen mit dem Befehl angeben:

```
wp6_cluster_copyFiles.bat
-WPSDir "C:\\Programme\\IBM\\WebSphere\\PortalServer"
-WSPProfileDir "C:\\Programme\\IBM\\WebSphere\\AppServer\\profiles\\wp_profile"
```

```
wp6_cluster_copyFiles.sh
-WASDir /opt/IBM/WebSphere/AppServer
-WPSDir /opt/IBM/WebSphere/PortalServer
-WSPProfileDir /opt/IBM/WebSphere/AppServer/profiles/wp_profile
```

4. Führen Sie an einer Eingabeaufforderung den Befehl **wp6\_cluster\_install.bat** (unter Windows) bzw. den Befehl **wp6\_cluster\_install.sh** (unter AIX, Linux oder Solaris) aus. Das folgende Beispiel zeigt Optionen aus Gründen der Lesbarkeit in separaten Zeilen. Sie müssen die Optionen mit dem Befehl angeben:

```
wp6_cluster_install.bat
-WPSClusterName eigener_cluster
-WSPProfileDir "C:\\Programme\\IBM\\WebSphere\\profiles\\wp_profile"
-WASDir "C:\\Programme\\IBM\\WebSphere\\AppServer"
-WASUser wpsbind -WASPassword wpsbind
-WPSDir "C:\\Programme\\IBM\\WebSphere\\PortalServer"
-WPSUser wpsadmin -WASPassword wpsadmin
-WPSHost "portalserver.ibm.com"
-webServerName web-server1
-webServerNodeName knoten1
-IICEDir "C:\\Programme\\IBM\\Content Edition"
```

#### **WPSClusterName**

Der Name des Clusters, in dem WebSphere Portal installiert ist.

#### **WSPProfileDir**

Der vollständig qualifizierte Pfad für das WebSphere Portal-Profilverzeichnis. Der Standardpfad ist `/usr/IBM/WebSphere/AppServer/profiles/wp_profile` auf AIX-Systemen, `/opt/IBM/WebSphere/AppServer/profiles/wp_profile` auf Linux- oder Solaris-Systemen und `C:\\Programme\\IBM\\WebSphere\\profiles\\wp_profile` auf Windows-Systemen.

#### **WASDir**

Der vollständig qualifizierte Pfad für das WebSphere Application Server-Stammverzeichnis. Er ist nur auf AIX-, Linux- und Solaris-Systemen erforderlich. Der Standardstammverzeichnispfad ist `/usr/IBM/WebSphere/AppServer` auf AIX-Systemen bzw. `/opt/IBM/WebSphere/AppServer` auf Linux- oder Solaris-Systemen.

#### **WASUser**

Der Benutzername für den WebSphere Application Server-Benutzer mit Verwaltungsaufgaben. Er ist nur erforderlich, wenn die globale Sicherheit in WebSphere Application Server aktiviert ist.

#### **WASPassword**

Das Kennwort für den WebSphere Application Server-Benutzer mit Verwaltungsaufgaben, falls angegeben.

**WPSDir**

Der vollständig qualifizierte Pfad für das WebSphere Portal-Installationsverzeichnis.

**WPSUser**

Der Benutzername für den WebSphere Portal-Benutzer mit Verwaltungsaufgaben.

**WSPassword**

Das Kennwort für den angegebenen WebSphere Portal-Benutzer mit Verwaltungsaufgaben.

**WPSHost**

Der Hostname und die Portnummer des WebSphere Portal-Servers.

**webServerName**

Der Name der Web-Server-Definition, zu der WebSphere Portal gehört.

**webServerNodeName**

Der Name des WebSphere Application Server-Knotens, zu dem die Web-Server-Definition gehört.

**IICEDir**

Der vollständig qualifizierte Pfad für das Installationsverzeichnis von WebSphere Information Integrator Content Edition. Er ist nur erforderlich, wenn Sie den Portal Document Manager-Connector bereits auf dem WebSphere Portal-Server konfiguriert haben.

5. Nachdem das Script beendet wurde, öffnen Sie einen Web-Browser, und melden Sie sich an der WebSphere-Administrationskonsole auf Ihrem Network Deployment-Server an. Die Adresse ist normalerweise `http://hostname:9060/ibm/console`.
6. Erweitern Sie den Abschnitt **Server**, und wählen Sie **Web-Server** aus.
7. Wählen Sie das Auswahlkästchen neben Ihrem Web-Server aus, und klicken Sie anschließend den Knopf **Plug-in generieren** an.
8. Wählen Sie das Auswahlkästchen neben Ihrem Web-Server aus, und klicken Sie anschließend den Knopf **Plug-in weitergeben** an.
9. Melden Sie sich von der Administrationskonsole ab.
10. Aktualisieren Sie das Portlet für die Unternehmenssuche, um den Suchserver anzugeben:
  - a. Melden Sie sich mit der Portal-Administrator-ID und dem Portal-Kennwort an WebSphere Portal an.
  - b. Klicken Sie **Verwaltung** in der linken unteren Ecke an.
  - c. Klicken Sie **Portlet-Verwaltung** im Navigationsbereich auf der linken Seite an, und klicken Sie anschließend **Portlets** an.
  - d. Ändern Sie die Option **Suchen nach in Der Titel enthält**.
  - e. Geben Sie Unternehmenssuche in das Feld **Suchen** ein, und klicken Sie den Knopf **Suchen** an.
  - f. Wenn neue Symbole auf der rechten Seite angezeigt werden, klicken Sie das Schraubenschlüsselsymbol an, um das Suchportlet für die Unternehmenssuche zu konfigurieren.
  - g. Modifizieren Sie in der Liste mit den Portletparametern die folgenden Parameter:

**hostname**

Geben Sie den vollständig qualifizierten Hostnamen eines Suchservers für die Unternehmenssuche an.

**port** Geben Sie die Portnummer an, die WebSphere Application Server auf dem Suchserver für die Unternehmenssuche verwendet. Der Standardwert ist 80. (Der Standardwert für die SSL-Kommunikation ist 443.)

**benutzername**

Geben Sie einen Benutzernamen an, der in einer WebSphere Application Server-Benutzerregistry gültig ist, wenn die globale Sicherheit auf dem Suchserver in WebSphere Application Server aktiviert ist.

**kennwort**

Wenn Sie einen WebSphere Application Server-Benutzernamen angegeben haben, geben Sie das entsprechende Kennwort an.

**protokoll**

Geben Sie das Protokoll an, das für die Kommunikation zwischen WebSphere Portal und dem Suchserver verwendet wird. Der Standardwert ist HTTP. Geben Sie HTTPS an, wenn Sie SSL verwenden.

**truststore**

Geben Sie den vollständig qualifizierten Pfad (mit dem Dateinamen) für den SSL-Zertifikatsspeicher an, wenn Sie SSL verwenden.

**trustkennwort**

Geben Sie das Kennwort für die angegebene Truststore-Datei an, wenn Sie SSL (Secure Sockets Layer) verwenden.

**ssoCookieName**

Geben Sie den Namen des Cookies an, das die SSO-Tokenzeichenfolge (Single Sign-on) enthält. Der Standardwert ist LtpaToken.

**proxyHost**

Geben Sie den vollständig qualifizierten Hostnamen eines Proxy-Servers an, wenn ein Proxy-Server erforderlich ist, um auf den Suchserver für die Unternehmenssuche zuzugreifen.

**proxyPort**

Geben Sie die Portnummer für den Proxy-Server an, wenn Sie einen Proxy-Server angegeben haben.

**proxyUser**

Wenn der Proxy-Server eine Basisauthentifizierung erfordert, geben Sie einen Benutzernamen an, der für die Anmeldung am Proxy-Server verwendet werden soll.

**proxyPassword**

Wenn Sie einen Benutzernamen für den Proxy-Server angegeben haben, geben Sie das entsprechende Kennwort an.

h. Klicken Sie **OK** an, um Ihre Änderungen zu speichern.

## Entfernen der Unternehmenssuche aus einem WebSphere Portal-Clustersystem

Zum Entfernen der Unternehmenssuche aus einem WebSphere Portal-Clustersystem der Version 6 verwenden Sie das Script `wp6_cluster_uninstall`.

**Informationen zu dieser Task**

Wenn Sie die Unternehmenssuche aus WebSphere Portal entfernen, werden die Portletparameter nicht gespeichert, die Sie als Teil des Installationsprozesses für das Portlet der Unternehmenssuche angegeben haben.

Wenn Sie das Script starten, stoppt das Script den WebSphere Portal-Server. Nachdem die Software der Unternehmenssuche entfernt wurde, startet das Script den WebSphere Portal-Server erneut.

### Vorgehensweise

Gehen Sie wie folgt vor, um die Unternehmenssuche von einem WebSphere Portal-Clustersystem der Version 6 zu entfernen:

1. Führen Sie an einer Eingabeaufforderung den Befehl **wp6\_cluster\_uninstall.bat** (unter Windows) bzw. den Befehl **wp6\_cluster\_uninstall.sh** (unter AIX, Linux oder Solaris) auf einem der Knoten des Clusters aus. Das folgende Beispiel zeigt Parameter aus Gründen der Lesbarkeit in separaten Zeilen. Sie müssen die Parameter mit dem Befehl angeben:

```
wp6_cluster_uninstall.bat
-WPSClusterName eigener_cluster
-WSPProfileDir "C:\\Programme\\IBM\\WebSphere\\profiles\\wp_profile"
-WASDir "C:\\Programme\\IBM\\WebSphere\\AppServer"
-WASUser wpsbind
-WASPassword wpsbind
-WPSDir "C:\\Programme\\IBM\\WebSphere\\PortalServer"
-WPSUser wpsadmin
-WSPassword wpsadmin
-WPSHost "portalserver.ibm.com"
-webServerName web-server1
-webServerNodeName knoten1
```

#### **WPSClusterName**

Der Name des Clusters, in dem WebSphere Portal installiert ist.

#### **WSPProfileDir**

Der vollständig qualifizierte Pfad für das WebSphere Portal-Profilverzeichnis. Der Standardpfad ist /usr/IBM/WebSphere/AppServer/profiles/wp\_profile auf AIX-Systemen, /opt/IBM/WebSphere/AppServer/profiles/wp\_profile auf Linux- oder Solaris-Systemen und C:\\Programme\\IBM\\WebSphere\\profiles\\wp\_profile auf Windows-Systemen.

#### **WASDir**

Der vollständig qualifizierte Pfad für das WebSphere Application Server-Stammverzeichnis. Er ist nur auf AIX-, Linux- und Solaris-Systemen erforderlich. Der Standardstammverzeichnispfad ist /usr/IBM/WebSphere/AppServer auf AIX-Systemen bzw. /opt/IBM/WebSphere/AppServer auf Linux- oder Solaris-Systemen.

#### **WASUser**

Der Benutzername für den WebSphere Application Server-Benutzer mit Verwaltungsaufgaben. Er ist nur erforderlich, wenn die globale Sicherheit in WebSphere Application Server aktiviert ist.

#### **WASPassword**

Das Kennwort für den WebSphere Application Server-Benutzer mit Verwaltungsaufgaben, falls angegeben.

#### **WPSDir**

Der vollständig qualifizierte Pfad für das WebSphere Portal-Installationsverzeichnis.

**WPSUser**

Der Benutzername für den WebSphere Portal-Benutzer mit Verwaltungsaufgaben.

**WSPassword**

Das Kennwort für den angegebenen WebSphere Portal-Benutzer mit Verwaltungsaufgaben.

**WPSHost**

Der Hostname und die Portnummer des WebSphere Portal-Servers.

**webServerName**

Der Name der Web-Server-Definition, zu der WebSphere Portal gehört.

**webServerNodeName**

Der Name des WebSphere Application Server-Knotens, zu dem die Web-Server-Definition gehört.

2. Nachdem das Script beendet wurde, öffnen Sie einen Web-Browser, und melden Sie sich an der WebSphere-Administrationskonsole auf Ihrem Network Deployment-Server an. Die Adresse ist normalerweise `http://hostname:9060/ibm/console`.
3. Erweitern Sie den Abschnitt **Server**, und wählen Sie **Web-Server** aus.
4. Wählen Sie das Auswahlkästchen neben Ihrem Web-Server aus, und klicken Sie anschließend den Knopf **Plug-in generieren** an.
5. Wählen Sie das Auswahlkästchen neben Ihrem Web-Server aus, und klicken Sie anschließend den Knopf **Plug-in weitergeben** an.
6. Melden Sie sich von der Administrationskonsole ab.



---

## Migration von WebSphere Portal in die Unternehmenssuche

Die Unternehmenssuche stellt einen Migrationsassistenten zur Verfügung, mit dem Sie Objektgruppen und regelbasierte Taxonomien aus IBM WebSphere Portal in die Unternehmenssuche migrieren können.

In der Unternehmenssuche wird eine Taxonomie als *Kategoriebaum* bezeichnet. Nach der Migration einer Taxonomie bearbeiten Sie mit der Administrationskonsole für die Unternehmenssuche den Kategoriebaum und die Kategorisierungsregeln. Nach der Migration einer Objektgruppe verwenden Sie die Administrationskonsole für die Verwaltung der Objektgruppe.

Wenn Sie Taxonomien und Objektgruppen migrieren möchten, führen Sie den Migrationsassistenten auf dem Indexserver für die Unternehmenssuche aus.

---

### Migration einer Objektgruppe aus WebSphere Portal

Wenn Sie Objektgruppen und regelbasierte Taxonomien aus WebSphere Portal in die Unternehmenssuche migrieren möchten, bereiten Sie die betreffenden Objektgruppen in WebSphere Portal vor, und rufen Sie anschließend den Migrationsassistenten auf, um sie zu migrieren.

#### Vorbereitung

Wenn Sie Taxonomien und Objektgruppen migrieren wollen, müssen Sie zuerst die Taxonomiedateien migrieren, bevor Sie diese Prozedur zum Migrieren von Objektgruppen verwenden. Durch diese Vorgehensweise wird sichergestellt, dass die migrierten Kategorisierungsregeln mit den migrierten Objektgruppen zusammenarbeiten.

#### Vorgehensweise

Gehen Sie wie folgt vor, um eine Objektgruppe (und optional die Taxonomie) aus WebSphere Portal in die Unternehmenssuche zu migrieren:

1. Stoppen Sie in WebSphere Portal Search Engine sämtliche Crawlerprozesse innerhalb der Objektgruppen, die Sie migrieren möchten, und bestätigen Sie alle anstehenden Dokumente, oder weisen Sie sie zurück. (Die Unternehmenssuche unterstützt nicht das Konzept der anstehenden Dokumente).
2. Verwenden Sie für jede Objektgruppe, die Sie migrieren möchten, die Portlets von Portal Search Engine, um die Einstellungen in XML-Dateien zu exportieren.
3. Wenn der Indexserver für die Unternehmenssuche auf einem separaten Server installiert ist, kopieren Sie die exportierten XML-Dateien in den Indexserver.
4. Melden Sie sich am Indexserver für die Unternehmenssuche als Administrator für die Unternehmenssuche an. Diese Benutzer-ID wurde bei der Installation von OmniFind Enterprise Edition angegeben.
5. Wechseln Sie in das Installationsverzeichnis der Unternehmenssuche:

```
UNIX: cd $ES_INSTALL_ROOT/bin
Windows: cd %ES_INSTALL_ROOT%\bin
```

6. Zum Migrieren von Objektgruppen mit aktivierter Sicherheit geben Sie den folgenden Befehl ein, um den Migrationsassistenten zu starten, und klicken Sie **Weiter** an.

UNIX: ./eswpsmigrate.sh  
Windows: eswpsmigrate.bat

7. Zum Inaktivieren der Sicherheit auf Objektgruppenebene für die Objektgruppen, die Sie migrieren, geben Sie den folgenden Befehl ein, um den Migrationsassistenten zu starten, und klicken Sie **Weiter** an.

UNIX: ./eswpsmigrate.sh disable.security  
Windows: eswpsmigrate.bat disable.security

8. Wählen Sie **Sucheinstellungen von Portal Search Engine nach WebSphere Portal migrieren** aus, und klicken Sie **Weiter** an.

9. Suchen Sie das Verzeichnis, das die exportierten Portal Search Engine-Konfigurationsdateien enthält, wählen Sie die Dateien aus, die Sie migrieren möchten, und klicken Sie **Weiter** an. Die ausgewählten Konfigurationsdateien werden analysiert und geprüft.

10. Geben Sie die folgenden Informationen für jede Objektgruppe an, und klicken Sie anschließend **Weiter** an, um mit der Migration von Objektgruppen in die Unternehmenssuche zu beginnen:

- Name der Objektgruppe, die Sie für die Unternehmenssuche verwenden möchten
- Die Bedingung, nach der der Dokumentrang für die Objektgruppe bestimmt wird. Der statische Rangfolgefaktor kann "Keine" sein, auf Dokumentdaten basieren, oder auf der Anzahl Links zu Webdokumenten von anderen Webdokumenten basieren.
- Der Typ der Kategorisierung, den Sie für diese Objektgruppe verwenden möchten. Wenn Sie keinen angeben, werden keine Taxonomieinformationen in die Unternehmenssuche migriert. Wenn Sie regelbasierte Kategorien auswählen, wird die Taxonomie zusammen mit der Objektgruppe in die Unternehmenssuche migriert.

Wenn Fehler während der Migration auftreten, sehen Sie sich die Datei MigrationWizard.log an, die sich in dem Verzeichnis befindet, in dem der Migrationsassistent installiert ist.

Über die Administrationskonsole für die Unternehmenssuche können Sie jetzt weitere Einstellungen für die migrierten Objektgruppen konfigurieren.

**Voraussetzung:** Wenn Sie Web-Crawler-Merkmale für eine migrierte Objektgruppe konfigurieren, müssen Sie eine E-Mail-Adresse für den Empfang von Kommentaren über den Crawler sowie einen Benutzeragentenamen angeben (Unterstützung erhalten Sie, wenn Sie bei der Konfiguration der Web-Crawler-Merkmale **Hilfe** anklicken).

11. Starten Sie für die migrierten Objektgruppen die Prozesse für die Crawler-suche, das Parsing und die Indexierung über die Administrationskonsole für die Unternehmenssuche.
12. Nachdem Sie festgelegt haben, dass die migrierte Objektgruppe in der Unternehmenssuche durchsucht werden kann, können Sie die ursprüngliche Objektgruppe in Portal Search Engine löschen.
13. Optional: Als WebSphere Portal-Administrator können Sie die folgenden Schritte ausführen, wenn Sie es dem Benutzer ermöglichen möchten, die migrierte Objektgruppe über ein Portal in WebSphere Portal zu durchsuchen.

- a. Implementieren Sie das Portlet für die Unternehmenssuche in Ihrer WebSphere Portal-Installation.

In einem WebSphere Portal-Servercluster sollte dies auf dem Server geschehen, auf dem der Bereitstellungsmanager von WebSphere Application Server installiert ist. Der Bereitstellungsmanager gibt das Portlet für die Unternehmenssuche an die anderen Server im WebSphere Portal-Servercluster weiter.

- b. Fügen Sie das Portlet für die Unternehmenssuche in die geeigneten Portalseiten ein.

In WebSphere Portal wird die Zugriffssteuerung für das Suchportlet durch die Zugriffsmöglichkeiten auf bestimmte Seiten und Portlets gesteuert. Obwohl die Objektgruppeneinstellungen migriert werden, muss das Portlet manuell vom WebSphere Portal-Serveradministrator positioniert werden.

---

## Einstellungen für migrierte Objektgruppen

Wenn Sie Objektgruppen aus IBM WebSphere Portal migrieren, richtet der Migrationsassistent Standardeinstellungen für Objektgruppen und Crawler ein.

Wenn dieselben Einstellungen in Portal Search Engine-Objektgruppen und in Objektgruppen für die Unternehmenssuche vorhanden sind, verwendet der Migrationsassistent die Portal Search Engine-Einstellung, wenn er die Objektgruppe in die Unternehmenssuche migriert. Für die Einstellungen, die nur in der Unternehmenssuche vorhanden sind, verwendet der Assistent die Einstellungen, die Sie angeben, während Sie die Objektgruppe oder die Standardeinstellungen für Objektgruppen in die Unternehmenssuche migrieren.

### **Einstellungen, die in Portal Search Engine und in der Unternehmenssuche vorhanden sind**

Der Migrationsassistent migriert die folgenden Einstellungen für jede Objektgruppe, die Sie migrieren:

- Die Portal Search Engine-Sites innerhalb der Portal Search Engine-Objektgruppe
- Die Objektgruppensprache
- Die Taxonomie (oder den Kategoriebaum) und die Regeln für die regelbasierten Kategorien, falls die Objektgruppe für die Unternehmenssuche regelbasierte Kategorisierung verwendet.

Jede Portal Search Engine-Site in einer Objektgruppe wird in einem Web-Crawler für die Unternehmenssuche konsolidiert. Der Migrationsassistent migriert die folgenden Crawlereinstellungen:

- Die Start-URLs
- Die Anzahl paralleler Crawlersuchprozesse
- Die Crawlersuchtiefe
- Das Zeitlimit (in Sekunden) für das Abrufen eines Dokuments
- Den Standardzeichensatz
- Regeln für die Crawlersuche in Websites (einschließen oder ausschließen)

### **Einstellungen, die nur in der Unternehmenssuche vorhanden sind**

Wenn Sie eine Objektgruppe migrieren, geben Sie Informationen zu der Objektgruppe an. Der Migrationsassistent migriert diese Einstellungen und verwendet die

Standardeinstellungen für Objektgruppen in der Unternehmenssuche, um die jeweilige Objektgruppe zu konfigurieren, die Sie migrieren.

Sie können die Objektgruppen- und Web-Crawlerkonfigurationen modifizieren, indem Sie die Administrationskonsole für die Unternehmenssuche verwenden. Die in den runden Klammern angezeigten Werte sind die Standardeinstellungen für die migrierten Daten.

- Der Objektgruppenname
- Die statische Rangfolgestrategie für das Dokument
- Der Typ der verwendeten Kategorisierung (Regelbasiert oder Keine)
- Ob der Suchcache verwendet werden soll und wie viele Antworten auf Suchanforderungen der Suchcache aufnehmen kann (Ja, 5000)
- Ob die Suchantwortzeiten überwacht und ein Alert abgesetzt werden soll, wenn der Grenzwert überschritten ist (Ja, 5 Sekunden)
- Ob Zugriffssteuerungen verwendet werden sollen (Nein)
- Zeitplan für die Deltaindexerstellung
- Zeitplan für die Hauptindexerstellung
- Die Protokolldetailebene (Alle Nachrichten)

Außerdem erstellt der Migrationsassistent auch die folgenden Einstellungen für alle Web-Crawler:

- Crawlername
- Crawlerbeschreibung
- Maximale Seitenlänge
- Dokumentsicherheitseinstellungen
- Die Multipurpose Internet Mail Extensions (MIME-Typen) für das Dokument, das durchsucht werden muss, falls zutreffend für den Datenquellentyp

Überprüfen Sie vor dem Starten eines gerade migrierten Web-Crawlers alle Crawlermerkmale und Crawlerbereichseinstellungen, und stellen Sie sicher, dass alle erforderlichen Werte angegeben sind (Pflichtfelder sind mit einem roten Stern markiert). Stellen Sie vor allem sicher, dass Sie eine E-Mail-Adresse für den Empfang von Kommentaren über den Crawler sowie einen Benutzeragentennamen für den Crawler angeben. Unterstützung erhalten Sie, indem Sie beim Konfigurieren der Web-Crawler-Merkmale **Hilfe** anklicken.

---

## Protokolldatei des Migrationsassistenten

Der Migrationsassistent schreibt alle Nachrichten in die Datei `WpsMigratorLog.log` im Verzeichnis, in dem der Migrationsassistent installiert ist.

Die Protokolldatei `WpsMigratorLog.log` enthält für jede migrierte Objektgruppe die Werte aller Einstellungen, die aus WebSphere Portal Search Engine gelesen wurden und gibt an, ob diese Einstellungen in Objektgruppen für die Unternehmenssuche importiert wurden.

---

## Befehle, Rückkehrcodes und Sitzungs-IDs der Unternehmenssuche

Sie können Befehle verwenden, um Probleme zu diagnostizieren, den Status der verschiedenen Teile des Systems zu ermitteln, Sitzungen zu starten und zu stoppen oder um das System zu starten und zu stoppen.

In einer Installation mit mehreren Servern können Sie die Befehle über jeden Server in Ihrem System ausführen. Sie sollten jedoch die Befehle über den Indexserver ausführen. Der Indexserver oder Controller-Server kann auf Informationen auf allen anderen Servern im System zugreifen.

Die meisten Befehle weisen folgende Formate auf:

```
esadmin befehlsname argumente
esadmin sitzungs-id aktion -option
```

Geben Sie `esadmin help` ein, um weitere Informationen zu allen Befehlen zu erhalten. Geben Sie `esadmin aktion help` ein, um weitere Informationen zu einem bestimmten Befehl zu erhalten.

### esadmin-Befehle für die Unternehmenssuche

Geben Sie die folgenden Befehle in einer Zeile ein.

*Tabelle 9. esadmin-Befehle für die Unternehmenssuche*

Befehl	Beschreibung
<code>esadmin system startall</code>	<p>Startet die Komponenten für die Unternehmenssuche auf allen Servern für die Unternehmenssuche einschließlich des Web-Servers, der Anwendung ESSearchServer und der Informationszentrale auf den Suchservern sowie die Crawlersitzungen auf dem Crawler-Server und die Indexsitzungen auf dem Indexserver. Startet die allgemeine Übertragungsschicht (Common Communication layer, CCL) nur auf dem lokalen Server. Sie müssen die allgemeine Übertragungsschicht auf jedem fernen Server für die Unternehmenssuche manuell stoppen und erneut starten, um die allgemeine Übertragungsschicht erneut zu starten.</p> <p><b>Beispielbefehl:</b> <code>esadmin system startall</code></p>
<code>esadmin system stopall</code>	<p>Stoppt die Komponenten für die Unternehmenssuche auf allen Servern für die Unternehmenssuche einschließlich des Web-Servers, der Anwendung ESSearchServer und der Informationszentrale auf den Suchservern sowie die Crawlersitzungen auf dem Crawler-Server und die Indexsitzungen auf dem Indexserver. Stoppt die allgemeine Übertragungsschicht nur auf dem lokalen Server. Sie müssen die allgemeine Übertragungsschicht auf jedem fernen Server für die Unternehmenssuche manuell stoppen und erneut starten, um die allgemeine Übertragungsschicht erneut zu starten.</p> <p><b>Beispielbefehl:</b> <code>esadmin system stopall</code></p>

Tabelle 9. **esadmin**-Befehle für die Unternehmenssuche (Forts.)

Befehl	Beschreibung
esadmin system checkall	Überprüft den Status aller Komponenten für die Unternehmenssuche auf allen Servern für die Unternehmenssuche.  <b>Beispielbefehl:</b> esadmin system checkall
esadmin <i>crawlersitzungs-id</i> start	Startet eine Crawlersitzung. Dieser Befehl startet keine Crawleraktivität.  <b>Beispielbefehl:</b> esadmin col1.WEB1.esadmin start  <b>Beispiele für Nachrichten und Rückkehrcodes:</b> FFQC5310I WEB2Crawler1 (sid: col1.WEB1.esadmin) ist nicht aktiv. FFQC5314I Ergebnis: 0
esadmin <i>crawlersitzungs-id</i> startCrawl	Startet die Crawler suche.  <b>Beispielbefehl:</b> esadmin col3.DB21.esadmin startCrawl  <b>Beispiele für Nachrichten und Rückkehrcodes:</b> FFQC5303I DB2Crawler1 (sid: col3.DB21.esadmin) ist bereits aktiv. PID: 23650 FFQC5314I Ergebnis: 0
esadmin <i>crawlersitzungs-id</i> pause	Hält die Crawler suche an.  <b>Beispielbefehl:</b> esadmin col3.DB21.esadmin pause  <b>Beispiele für Nachrichten und Rückkehrcodes:</b> FFQC5303I DB2Crawler1 (sid: col3.DB21.esadmin) ist bereits aktiv. PID: 23650 FFQC5314I Ergebnis: 0
esadmin <i>crawlersitzungs-id</i> resume	Setzt die Crawler suche fort.  <b>Beispielbefehl:</b> esadmin col3.DB21.esadmin resume  <b>Beispiele für Nachrichten und Rückkehrcodes:</b> FFQC5303I DB2Crawler1 (sid: col3.DB21.esadmin) ist bereits aktiv. PID: 23650 FFQC5314I Ergebnis: 0
esadmin <i>crawlersitzungs-id</i> stopCrawl	Stoppt die Crawler suche.  <b>Beispielbefehl:</b> esadmin col3.DB21.esadmin stopCrawl  <b>Beispiele für Nachrichten und Rückkehrcodes:</b> FFQC5303I DB2Crawler1 (sid: col3.DB21.esadmin) ist bereits aktiv. PID: 23650 FFQC5314I Ergebnis: 0

Tabelle 9. **esadmin**-Befehle für die Unternehmenssuche (Forts.)

Befehl	Beschreibung
esadmin <i>crawlersitzungs-id</i> stop	<p>Stoppt eine Crawlersitzung.</p> <p><b>Beispielbefehl:</b> esadmin col3.DB21.esadmin stop</p> <p><b>Beispiele für Nachrichten und Rückkehrcodes:</b> FFQC5303I DB2Crawler1 (sid: col3.DB21.esadmin) ist bereits aktiv. PID: 23650 FFQC5314I Ergebnis: 0</p>
esadmin <i>crawlersitzungs-id</i> getCrawlerStatus	<p>Ruft den Status eines Crawlers ab. Die zurückgegebenen Informationen sind davon abhängig, ob es sich bei dem Crawler um einen Web-Crawler oder um einen Crawler für alle anderen Datenquellen handelt.</p> <p><b>Beispiel für einen Web-Crawler:</b> esadmin col1.WEB1.esadmin getCrawlerStatus</p> <p><b>Mögliche Rückkehrcodes und Nachrichten für einen Web-Crawler:</b> FFQC5303I WebCrawler1 (sid: col1.WEB1.esadmin) ist bereits aktiv. PID: 23650</p> <p><b>Beispiel für einen Nicht-Web-Crawler:</b> esadmin col3.DB21.esadmin getCrawlerStatus</p> <p><b>Mögliche Rückkehrcodes und Nachrichten für einen Nicht-Web-Crawler:</b> FFQC5303I db2crawler (sid: db2col.DB2_96945) ist bereits aktiv. PID: 5936</p> <p>Weitere Informationen zu zurückgegebenen Statusnachrichten finden Sie in „Detaillierte Informationen zu Statusbefehlen“ auf Seite 406.</p>
esadmin <i>dscrawlersitzungs-id</i> getCrawlSpaceStatus esadmin <i>web-crawler-sitzungs-id</i> getCrawlStatus -selections wert	<p>Ruft den allgemeinen Status des Crawlerbereichs für jeden anderen Crawler als den Web-Crawler ab.</p> <p><b>Beispielbefehl:</b> esadmin col3.DB21.esadmin getCrawlSpaceStatus</p> <p><b>Beispiele für Nachrichten und Rückkehrcodes:</b> FFQC5303I DB2Crawler1 (sid: col3.DB21.esadmin) ist bereits aktiv. PID: 23650</p> <p>Ruft den allgemeinen Status des Crawlerbereichs für den Web-Crawler ab.</p> <p><b>Beispielbefehl:</b> esadmin col1.WEB1.esadmin getCrawlStatus</p> <p>Weitere Informationen zu zurückgegebenen Statusnachrichten finden Sie in „Detaillierte Informationen zu Statusbefehlen“ auf Seite 406.</p>

Tabelle 9. **esadmin**-Befehle für die Unternehmenssuche (Forts.)

Befehl	Beschreibung
<pre>esadmin dscrawlersitzungs-id getCrawlSpaceStatusDetail -ts zielservers-id  esadmin web-crawler-sitzungs-id getCrawlDetailsPerSite -url zeichenfolge -selections anzahl -threshold anzahl</pre>	<p>Ruft den detaillierten Status des Crawlerbereichs für jeden anderen Crawler als den Web-Crawler ab. Wenn Sie die Zielserveroption nicht angeben, werden Daten für alle Zielserver zurückgegeben. Beispiel: Wenn der DB2-Crawler die Datenbanken FOUNTAIN und SAMPLE durchsucht und Sie die Zielserveroption nicht angeben, wird der Status aller Tabellen in den Datenbanken FOUNTAIN und SAMPLE zurückgegeben.</p> <p><b>Beispielbefehl:</b></p> <pre>esadmin col3.DB21.esadmin getCrawlSpaceStatusDetail -ts FOUNTAIN</pre> <p><b>Beispiele für Nachrichten und Rückkehrcodes:</b></p> <pre>FFQC5303I DB2Crawler1 (sid: col3.DB21.esadmin) ist bereits aktiv. PID: 23650</pre> <p>Ruft den detaillierten Status des Crawlerbereichs für den Web-Crawler ab.</p> <p><b>Beispielbefehl:</b></p> <pre>esadmin col1.WEB1.esadmin getCrawlDetailsPerSite</pre> <p>Weitere Informationen zu zurückgegebenen Statusnachrichten finden Sie in „Detaillierte Informationen zu Statusbefehlen“ auf Seite 406.</p>
<pre>esadmin monitor getCollectionParserMonitorStatus -cid objektgruppen-id</pre>	<p>Ruft den Parserstatus ab.</p> <p><b>Beispielbefehl:</b></p> <pre>esadmin monitor getCollectionParserMonitorStatus -cid coll</pre> <p><b>Beispiele für Nachrichten und Rückkehrcodes:</b></p> <pre>FFQC5303I Monitor (knoten1) (sid: monitor) ist bereits aktiv. PID: 12543</pre> <p>Weitere Informationen zu zurückgegebenen Statusnachrichten finden Sie in „Detaillierte Informationen zu Statusbefehlen“ auf Seite 406.</p>
<pre>esadmin startMain -cid objektgruppen-id</pre>	<p>Startet die Hauptindexerstellung.</p> <p><b>Beispielbefehl:</b></p> <pre>esadmin startMain -cid coll</pre> <p><b>Beispiele für Nachrichten und Rückkehrcodes:</b></p> <pre>FFQC5303I Controller (knoten1) (sid: controller) ist bereits aktiv. PID: 25917 FFQC5314I Ergebnis: 1117671147056</pre>
<pre>esadmin startDelta -cid objektgruppen-id</pre>	<p>Startet eine Deltaindexerstellung.</p> <p><b>Beispielbefehl:</b></p> <pre>esadmin startDelta -cid coll</pre> <p><b>Beispiele für Nachrichten und Rückkehrcodes:</b></p> <pre>FFQC5303I Controller (knoten1) (sid: controller) ist bereits aktiv. PID: 4548 FFQC5314I Ergebnis: 1117670603408</pre>

Tabelle 9. **esadmin**-Befehle für die Unternehmenssuche (Forts.)

Befehl	Beschreibung
<pre>esadmin controller startIndexBuild -cid <i>objektgruppen-id</i> -buildType <i>typ</i> -detectChanges</pre>	<p>Starten Sie eine Haupt- oder Deltaindexerstellung, und geben Sie an, dass die Erstellung nur ausgeführt werden soll, wenn Änderungen ermittelt werden, die auf den Index angewendet werden müssen.</p> <p><b>Beispielbefehl:</b></p> <pre>esadmin controller startIndexBuild -cid col_1 -buildType main -detectChanges</pre>
<pre>esadmin monitor getCollectionIndexMonitorStatus -cid <i>objektgruppen-id</i> -buildType [main   delta] -numrecords <i>letzte-n-datensätze</i></pre>	<p>Ruft den Status einer Deltaindexerstellung ab. Die Option <code>numrecords</code> zeigt die letzten <i>N</i> Statusdatensätze der Indexerstellung. Wenn <code>numrecords</code> nicht angegeben wird, wird der Status für die letzten 20 Indexerstellung zurückgegeben.</p> <p><b>Beispielbefehl:</b></p> <pre>esadmin monitor getCollectionIndexMonitorStatus -cid col1 -buildType main -numrecords 4</pre> <p><b>Beispiele für Nachrichten und Rückkehrcodes:</b></p> <pre>FFQC5303I Monitor (knoten1) (sid: monitor) ist bereits aktiv. PID: 12649</pre> <p>Weitere Informationen zu zurückgegebenen Statusnachrichten finden Sie in „Detaillierte Informationen zu Statusbefehlen“ auf Seite 406.</p>
<pre>esadmin startSearch -cid <i>objektgruppen-id</i></pre>	<p>Startet die Suchserverprozesse.</p> <p><b>Beispielbefehl:</b></p> <pre>esadmin startSearch -cid col1</pre> <p><b>Beispiele für Nachrichten und Rückkehrcodes:</b></p> <pre>FFQC5303I Controller (knoten1) (sid: controller) ist bereits aktiv. PID: 25917 FFQC5314I Ergebnis: 0</pre>
<pre>esadmin stopSearch -cid <i>objektgruppen-id</i></pre>	<p>Stoppt die Suchserverprozesse.</p> <p><b>Beispielbefehl:</b></p> <pre>esadmin stopSearch -cid col1</pre> <p><b>Beispiele für Nachrichten und Rückkehrcodes:</b></p> <pre>FFQC5303I Controller (knoten1) (sid: controller) ist bereits aktiv. PID: 15292 FFQC5314I Ergebnis: 0</pre>

Tabelle 9. **esadmin**-Befehle für die Unternehmenssuche (Forts.)

Befehl	Beschreibung
<pre>esadmin monitor getCollectionSearchMonitorStatus -cid objektgruppen-id esadmin suchmanager-sitzungs-id getStatus -cid objektgruppen-id</pre>	<p>Ruft den Status des Suchservers ab.</p> <p><b>Beispielbefehl:</b></p> <pre>esadmin monitor getCollectionSearchMonitorStatus -cid coll</pre> <p><b>Beispiele für Nachrichten und Rückkehrcodes:</b></p> <pre>FFQC5303I Monitor (knoten1) (sid: monitor) ist bereits aktiv. PID: 12649</pre> <p>Gibt detaillierte Statusinformationen zum Suchindex für eine Objektgruppe auf einem bestimmten Server zurück. Es gibt eine Suchmanagersitzung pro Suchserver. Jede Suchmanagersitzung ist für das Überwachen und Ausführen der Suchindizes auf einem bestimmten Suchserver verantwortlich.</p> <p><b>Beispielbefehl:</b></p> <pre>esadmin searchmanager.node1 getStatus -cid coll</pre> <p><b>Beispiele für Nachrichten und Rückkehrcodes:</b></p> <pre>FFQC5303I Suchmanager (knoten1) (sid: searchmanager.node1) ist bereits aktiv. PID: 15711 FFQC5314I Ergebnis: PID=18390 CacheHits=3 QueryRate=1 Port=44008 SessionId=coll.runtime.node1 CacheHitRate=0.333 ResponseTime=70 Status=1 SessionName=coll.runtime.node1.1</pre> <p>Weitere Informationen zu zurückgegebenen Statusnachrichten finden Sie in „Detaillierte Informationen zu Statusbefehlen“.</p>

## Detaillierte Informationen zu Statusbefehlen

Manche Befehle können umfassende Informationen zurückgeben. In diesem Abschnitt werden die Informationen beschrieben, die für den Status des Crawlers und des Crawlerbereichs zurückgegeben werden können. In der Tabelle aus dem Abschnitt „esadmin-Befehle für die Unternehmenssuche“ auf Seite 401 sind mögliche zurückgegebene Informationen von jedem **esadmin**-Befehl angegeben. In diesem Abschnitt werden die zurückgegebenen Informationen aus den folgenden Befehlen beschrieben:

- Web-Crawler-Status
- Nicht-Web-Crawler-Status
- Status des Crawlerbereichs des Web-Crawlers
- Status des Crawlerbereichs von Nicht-Web-Crawlern
- Detaillierter Status des Crawlerbereichs des Web-Crawlers
- Detaillierter Status des Crawlerbereichs von Nicht-Web-Crawlern
- Parserstatus
- Status der Indexerstellung
- Status des Suchservers

- Detaillierter Status des Suchservers

**Web-Crawler-Status:** Wenn Sie den Befehl zum Abrufen des Web-Crawler-Status ausführen, gibt der Befehl Informationen im XML-Dokumentformat zurück. Die folgenden Informationen können durch den Statusbefehl für den Web-Crawler zurückgegeben werden:

```
FFQC5314I Ergebnis: <?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 der folgenden Tabelle wird jedes XML-Element und seine möglichen Attribute beschrieben, die durch den Statusbefehl für den Web-Crawler zurückgegeben werden:

Tabelle 10. Informationen zum Web-Crawler-Status

Element	Attribute	Beschreibung
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 Value	<ul style="list-style-type: none"> <li>• Zeichenfolge (Englisch) „Not started“: Die Crawlersitzung ist vorhanden, aber sie hat noch keine Startnachricht zum Verarbeiten von Dokumenten empfangen.</li> <li>• „Started“: Der Crawler wird gestartet.</li> <li>• „Running“: Der Crawler hat die Initialisierung und den Start beendet und führt gerade eine Crawlersuche aktiv aus.</li> <li>• „Paused“: Der Crawler wurde aufgefordert, die aktive Crawlersuche auszusetzen, aber nicht zu beenden.</li> <li>• „Stopping“: Der Crawler hat das Stoppsignal empfangen und wird gestoppt.</li> <li>• „Error“: Der Crawler befindet sich in einem Status, in dem keine Recovery durchgeführt werden kann. Er muss gestoppt und erneut gestartet werden, um die Crawlersuche fortzusetzen.</li> </ul>	Informationen zur Aktivität des Crawlers.

Tabelle 10. Informationen zum Web-Crawler-Status (Forts.)

Element	Attribute	Beschreibung	
CrawlerThreadState	State	Zeichenfolge (Englisch)	Aktivität des Crawler-Threads. In diesem Feld wird die Aktivität des/der Threads gezeigt.
ActiveBucket	<ul style="list-style-type: none"> <li>URL: Zeichenfolge (URL-Angabe) Das Protokoll, der Host und der Port, dessen URL-Adressen gerade durchsucht werden.</li> <li>NumActURLs: Ganze Zahl (positiv) Die Anzahl URL-Adressen im Bucket, als er für die Crawlersuche zur Verfügung gestellt wurde (aktiviert).</li> <li>NumProcURLs: Ganze Zahl (nicht negativ) Die Anzahl URL-Adressen aus dem Bucket, die bisher verarbeitet, d. h. durchsucht oder zurückgewiesen wurden.</li> <li>TimeRem: Ganze Zahl Die Anzahl Sekunden bis zur Zeitlimit-überschreitung des Bucket.</li> <li>Duration: Ganze Zahl (nicht negativ) Die Anzahl Sekunden, seitdem der Bucket aktiviert wurde.</li> </ul>	Die aktuelle Aktivität einer angegebenen Website.	
CrawlRate	Value: Ganze Zahl (nicht negativ)	Der Messwert des Crawler-Durchsatzes.	
	Seiten, die pro Sekunde durchsucht werden (alle Buckets kombiniert).		
RecentlyCrawledURL	URL: Zeichenfolge (URL-Angabe)	Eine Seite, die kürzlich durchsucht wurde.	
	Zeichenfolge, die ein Protokoll, einen Host, einen Port und eine Datei angibt, die durchsucht wurde.		
NumURLsThisSession	Value: Ganze Zahl (nicht negativ)	Die Anzahl URL-Adressen, die durchsucht wurden, seitdem diese Instanz des Crawlers (Prozess) die Crawlersuche begonnen hat.	

**Nicht-Web-Crawler-Status:** Wenn Sie den Befehl zum Abrufen des Crawlerstatus für einen Nicht-Web-Crawler ausführen, gibt der Befehl Informationen im XML-Dokumentformat zurück. Die folgenden Informationen können durch den Befehl **getCrawlerStatus** für Nicht-Web-Crawler zurückgegeben werden:

```
FFQC5314I Ergebnis: <?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 den folgenden Tabellen werden die XML-Elemente und -Attribute für jeden Crawler für die Unternehmenssuche (mit Ausnahme des Web-Crawlers) beschrieben. Diese Informationen werden mit dem Statusbefehl des Crawlers zurückgegeben.

*Tabelle 11. Crawlerstatusinformationen für die NNTP-, DB2-, JDBC-Datenbank- und Notes-Crawler*

<b>Element- und Attributname</b>	<b>NNTP-Crawler</b>	<b>DB2- und JDBC-Datenbankcrawler</b>	<b>Notes-Crawler</b>
Status	Status (0, 1, 2, -1)	Status (0, 1, 2, -1)	Status (0, 1, 2, -1)
StatusMessage	Status: 0 - Inaktiv, 1 - Aktiv, 2 - Angehalten, -1 - Fehler	Status: 0 - Inaktiv, 1 - Aktiv, 2 - Angehalten, -1 - Fehler	Status: 0 - Inaktiv, 1 - Aktiv, 2 - Angehalten, -1 - Fehler
NumberOfServers	Die Anzahl NNTP-Server im Crawlerbereich.	Die Anzahl Datenbanken im Crawlerbereich.	Die Anzahl Datenbanken im Crawlerbereich.
NumberOfCompletedServers	Die Anzahl durchsuchter NNTP-Server.	Die Anzahl durchsuchter Datenbanken.	Die Anzahl durchsuchter Datenbanken.
NumberOfTargets	Die Anzahl Newsgroups im Crawlerbereich.	Die Anzahl Datenbanken im Crawlerbereich.	Die Anzahl Sichten und Ordner im Crawlerbereich.
NumberOfCompletedTargets	Die Anzahl durchsuchter Newsgroups.	Die Anzahl durchsuchter Tabellen.	Die Anzahl durchsuchter Sichten und Ordner.
NumberOfCompletedRecords	Die Anzahl durchsuchter Artikel.	Die Anzahl durchsuchter Datensätze.	Die Anzahl durchsuchter Dokumente.
RunningThreads	Die Anzahl Crawler-Threads.	Die Anzahl Crawler-Threads.	Die Anzahl Crawler-Threads.

*Tabelle 12. Crawlerstatusinformationen für die Exchange Server-, DB2 Content Manager- und Content Edition-Crawler*

<b>Element- und Attributname</b>	<b>Exchange Server-Crawler</b>	<b>DB2 Content Manager-Crawler</b>	<b>Content Edition-Crawler</b>
Status	Status (0, 1, 2, -1)	Status (0, 1, 2, -1)	Status (0, 1, 2, -1)
StatusMessage	Status: 0 - Inaktiv, 1 - Aktiv, 2 - Angehalten, -1 - Fehler	Status: 0 - Inaktiv, 1 - Aktiv, 2 - Angehalten, -1 - Fehler	Status: 0 - Inaktiv, 1 - Aktiv, 2 - Angehalten, -1 - Fehler
NumberOfServers	Die Anzahl Exchange Server-Server im Crawlerbereich.	Die Anzahl Content Manager-Server im Crawlerbereich.	Die Anzahl Repositories im Crawlerbereich.
NumberOfCompletedServers	Die Anzahl durchsuchter Exchange Server-Server.	Die Anzahl durchsuchter Content Manager-Server.	Die Anzahl durchsuchter Repositories.
NumberOfTargets	Die Anzahl Unterordner im Crawlerbereich.	Die Anzahl Elementtypen im Crawlerbereich.	Die Anzahl Klassen im Crawlerbereich.
NumberOfCompletedTargets	Die Anzahl durchsuchter Unterordner.	Die Anzahl durchsuchter Elementtypen.	Die Anzahl durchsuchter Elementklassen.
NumberOfCompletedRecords	Die Anzahl durchsuchter Dokumente.	Die Anzahl durchsuchter Dokumente.	Die Anzahl durchsuchter Dokumente.
RunningThreads	Die Anzahl Crawler-Threads.	Die Anzahl Crawler-Threads.	Die Anzahl Crawler-Threads.

Tabelle 13. Crawlerstatusinformationen für die QuickPlace-, Domino Document Manager-, UNIX-Dateisystem- und Windows-Dateisystemcrawler

Element- und Attributname	QuickPlace-Crawler	Domino Document Manager-Crawler	Crawler für das UNIX- und Windows-Dateisystem
Status	Status (0, 1, 2, -1)	Status (0, 1, 2, -1)	Status (0, 1, 2, -1)
StatusMessage	Status: 0 - Inaktiv, 1 - Aktiv, 2 - Angehalten, -1 - Fehler	Status: 0 - Inaktiv, 1 - Aktiv, 2 - Angehalten, -1 - Fehler	Status: 0 - Inaktiv, 1 - Aktiv, 2 - Angehalten, -1 - Fehler
NumberOfServers	Die Anzahl Bereiche im Crawlerbereich.	Die Anzahl Bibliotheken im Crawlerbereich.	Fester Wert 1.
NumberOfCompletedServers	Die Anzahl durchsuchter Bereiche.	Die Anzahl durchsuchter Bibliotheken.	0 oder 1, wenn alle Unterverzeichnisse durchsucht werden.
NumberOfTargets	Die Anzahl Datenbanken für Bereiche und Räume im Crawlerbereich.	Die Anzahl Aktenschrankdateien im Crawlerbereich.	Die Anzahl Unterverzeichnisse im Crawlerbereich.
NumberOfCompletedTargets	Die Anzahl durchsuchter Datenbanken für Bereiche und Räume.	Die Anzahl durchsuchter Aktenschrankdateien.	Die Anzahl durchsuchter Unterverzeichnisse.
NumberOfCompletedRecords	Die Anzahl durchsuchter Dokumente.	Die Anzahl durchsuchter Dokumente.	Die Anzahl durchsuchter Dateien.
RunningThreads	Die Anzahl Crawler-Threads.	Die Anzahl Crawler-Threads.	Die Anzahl Crawler-Threads.

Tabelle 14. Crawlerstatusinformationen für die WebSphere Portal- und Web Content Management-Crawler

Element- und Attributname	WebSphere Portal-Crawler	Web Content Management-Crawler
Status	Status (0, 1, 2, -1)	Status (0, 1, 2, -1)
StatusMessage	Status: 0 - Inaktiv, 1 - Aktiv, 2 - Angehalten, -1 - Fehler	Status: 0 - Inaktiv, 1 - Aktiv, 2 - Angehalten, -1 - Fehler
NumberOfServers	Die Anzahl Server im Crawlerbereich.	Die Anzahl Sites im Crawlerbereich.
NumberOfCompletedServers	Die Anzahl durchsuchter Server.	Die Anzahl durchsuchter Sites.
NumberOfTargets	Die Anzahl Server im Crawlerbereich.	Die Anzahl Sites im Crawlerbereich.
NumberOfCompletedTargets	Die Anzahl durchsuchter Server.	Die Anzahl durchsuchter Sites.
NumberOfCompletedRecords	Die Anzahl durchsuchter Dokumente.	Die Anzahl durchsuchter Dokumente.
RunningThreads	Die Anzahl Crawler-Threads.	Die Anzahl Crawler-Threads.

**Status für den Crawlerbereich für den Web-Crawler:** Wenn Sie den Befehl zum Abrufen des Status für den Crawlerbereich für einen Web-Crawler ausführen, gibt der Befehl Informationen im XML-Dokumentformat zurück. Die folgenden Informationen können durch den Statusbefehl für den Web-Crawler zurückgegeben werden:

Tabelle 15. Auswahlmaskenwerte für den Statusbefehl des Crawlerbereichs für den Web-Crawler

Maskenbit	Auswahl
1	Anzahl Seiten im Rohdatenspeicher
2	Anzahl aufgespürter Sites
4	Anzahl Sites mit DNS
8	Anzahl Sites ohne DNS
16	Anzahl aufgespürter URL-Adressen
32	Anzahl eindeutiger, gespeicherter Seiten
64	Anzahl durchsuchter URL-Adressen
128	Anzahl nicht durchsuchter URL-Adressen
256	Anzahl überfälliger URL-Adressen
512	Verteilung des HTTP-Statuscodes

Alle Werte geben kumulative Gesamtanzahlen für alle Sitzungen an, die die aktuelle interne Datenbank verwenden:

```
<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>
```

Die Rückgabedaten enthalten bestimmte oder alle (möglicherweise auch keine) folgenden Elemente:

Tabelle 16. Crawlerbereichsstatusinformationen des Web-Crawlers

Element	Attribute	Beschreibung
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>	Informationen zum kumulativen Status der Crawler suche (alle Sitzungen), die schnell abgerufen werden können.

Tabelle 16. Crawlerbereichsstatusinformationen des Web-Crawlers (Forts.)

Element	Attribute	Beschreibung
NumPagesInRDS	Value: Nicht negative ganze Zahl  Anzahl Seiten, die sich gegenwärtig im Zwischenspeicherungsbereich des Rohdatenspeichers (RDS) befinden (nur von diesem Crawler)	Auslastung des Rohdatenspeichers (RDS) (nur durch Daten von diesem Crawler).
NumSitesDiscovered	Value: Nicht negative ganze Zahl  Anzahl Hosts, die durch die Crawlersuche (oder anhand von Einstiegspunkten) aufgespürt wurden.	Eine Kennzahl für den Geltungsbereich der Domäne des Crawlers, die durchsucht werden soll (Hostanzahl).
NumSitesWithDNS	Value: Nicht negative ganze Zahl  Anzahl Hosts mit zugeordneten IP-Adressen (vom Crawler im Hintergrund aufgelöst)	Eine Kennzahl, wie effektiv der Crawler IP-Adressen für Hosts abrufen kann, die durch DNS-Namen in den URL-Adressen aufgespürt werden.
NumSitesWithoutDNS	Value: Nicht negative ganze Zahl  Anzahl Hosts ohne zugeordnete IP-Adressen (vom Crawler im Hintergrund aufgelöst)	Eine Kennzahl, wie effektiv der Crawler IP-Adressen für Hosts abrufen kann, die durch DNS-Namen in den URL-Adressen aufgespürt werden.
NumURLsDiscovered	Value: Nicht negative ganze Zahl  Anzahl eindeutiger URL-Adressen, die durch den Crawler aufgerufen wurden.	Eine Kennzahl für den Geltungsbereich der Domäne des Crawlers, die durchsucht werden soll (URL-Anzahl).
NumUniquePagesSaved	Value: Nicht negative ganze Zahl  Anzahl eindeutiger Seiten, die zur weiteren Verarbeitung durch andere Komponenten für die Unternehmenssuche in den Rohdatenspeicher (RDS) geschrieben wurden.	Der Beitrag dieses Crawlers zur Größe des Index.
NumURLsCrawled	Value: Nicht negative ganze Zahl  Anzahl eindeutiger URL-Adressen, die vom Crawler durchsucht wurden.	Eine Kennzahl dafür, ob der Crawler Daten verarbeiten kann (End-to-End). Diese Zahl unterscheidet sich von der Anzahl in den Rohdatenspeicher geschriebener Seiten, weil nicht alle durchsuchten Seiten in den Rohdatenspeicher geschrieben werden.
NumURLsOverdue	Value: Nicht negative ganze Zahl  Anzahl eindeutiger URL-Adressen, die erneut durchsucht werden können.	Eine Kennzahl dafür, ob der Crawler den Web-Space durchqueren kann.

**Status für den Crawlerbereich für Nicht-Web-Crawler:** Wenn Sie den Befehl zum Abrufen des Status für den Crawlerbereich für einen Nicht-Web-Crawler ausführen, gibt der Befehl Informationen im XML-Dokumentformat zurück. Die folgenden Informationen können durch den Befehl **getCrawlSpaceStatus** für Nicht-Web-Crawler zurückgegeben werden:

```
FFQC5314I Ergebnis: <?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 den folgenden Tabellen werden die XML-Elemente und -Attribute für jeden Crawler für die Unternehmenssuche (mit Ausnahme des Web-Crawlers) beschrieben. Diese Informationen werden mit dem Statusbefehl für den Crawlerbereich zurückgegeben. Für Notes-Crawler: Wenn die Aggregationsstufe 0 ist, besteht 'Server@Name' aus servername + datenbankname. Wenn die Aggregationsstufe 1 ist, besteht 'Server@Name' aus servername + verzeichnisname.

Tabelle 17. Crawlerbereichsstatusinformationen für die NNTP-, DB2-, JDBC-Datenbank- und Notes-Crawler

Element- und Attributname	NNTP-Crawler	DB2- und JDBC-Datenbankcrawler	Notes-Crawler
Server@Name	Name des News-Servers	Datenbankname	Datenbankname oder Verzeichnisname
Server/Status	Status: (0, 1, 2, 3, 4, 5, -1) <ul style="list-style-type: none"> <li>• 0: Keine Crawlersuche ausgeführt</li> <li>• 1: Crawlersuche wird ausgeführt</li> <li>• 2: Abgeschlossen (nicht terminiert)</li> <li>• 3: Wartestatus</li> <li>• 4: Angehalten</li> <li>• 5: Terminiert</li> <li>• -1: Fehler</li> </ul>	Status (0, 1, 2, 3, 4, 5, -1) <ul style="list-style-type: none"> <li>• 0: Keine Crawlersuche ausgeführt</li> <li>• 1: Crawlersuche wird ausgeführt</li> <li>• 2: Abgeschlossen (nicht terminiert)</li> <li>• 3: Wartestatus</li> <li>• 4: Angehalten</li> <li>• 5: Terminiert</li> <li>• -1: Fehler</li> </ul>	Status (0, 1, 2, 3, 4, 5, -1) <ul style="list-style-type: none"> <li>• 0: Keine Crawlersuche ausgeführt</li> <li>• 1: Crawlersuche wird ausgeführt</li> <li>• 2: Abgeschlossen (nicht terminiert)</li> <li>• 3: Wartestatus</li> <li>• 4: Angehalten</li> <li>• 5: Terminiert</li> <li>• -1: Fehler</li> </ul>
Server/StatusMessage	<ul style="list-style-type: none"> <li>• 0: Keine Crawlersuche ausgeführt</li> <li>• 1: Crawlersuche wird ausgeführt</li> <li>• 2: Abgeschlossen (nicht terminiert)</li> <li>• 3: Wartestatus</li> <li>• 4: Angehalten</li> <li>• 5: Terminiert</li> <li>• -1: Fehler</li> </ul>	<ul style="list-style-type: none"> <li>• 0: Keine Crawlersuche ausgeführt</li> <li>• 1: Crawlersuche wird ausgeführt</li> <li>• 2: Abgeschlossen (nicht terminiert)</li> <li>• 3: Wartestatus</li> <li>• 4: Angehalten</li> <li>• 5: Terminiert</li> <li>• -1: Fehler</li> </ul>	<ul style="list-style-type: none"> <li>• 0: Keine Crawlersuche ausgeführt</li> <li>• 1: Crawlersuche wird ausgeführt</li> <li>• 2: Abgeschlossen (nicht terminiert)</li> <li>• 3: Wartestatus</li> <li>• 4: Angehalten</li> <li>• 5: Terminiert</li> <li>• -1: Fehler</li> </ul>
Server/NumberOfTargets	Die Anzahl Newsgroups im Crawlerbereich.	Die Anzahl Datenbanken im Crawlerbereich.	Die Anzahl Sichten und Ordner oder Verzeichnisse im Crawlerbereich.
Server/NumberOfCompletedTargets	Die Anzahl durchsuchter Newsgroups.	Die Anzahl durchsuchter Tabellen.	Die Anzahl durchsuchter Sichten und Ordner oder Verzeichnisse.
Server/NumberOfErrors	Die Anzahl Fehler.	Die Anzahl Fehler.	Die Anzahl Fehler
Server/StartTime	Die Startzeit, falls zutreffend.	Die Startzeit, falls zutreffend.	Die Startzeit, falls zutreffend.
Server/EndTime	Die Endzeit, falls zutreffend.	Die Endzeit, falls zutreffend.	Die Endzeit, falls zutreffend.

Tabelle 17. Crawlerbereichsstatusinformationen für die NNTP-, DB2-, JDBC-Datenbank- und Notes-Crawler (Forts.)

Element- und Attributname	NNTP-Crawler	DB2- und JDBC-Datenbankcrawler	Notes-Crawler
Server/ScheduleConfigured	0, 1, 2 <ul style="list-style-type: none"> <li>0: Der Crawler ist entsprechend der Crawlerkonfigurationsdateien nicht für die Terminierung konfiguriert.</li> <li>1: Der Crawler ist für die Terminierung konfiguriert, aber die Terminierung wurde für die Sitzung inaktiviert.</li> <li>2: Der Crawler ist für die Terminierung konfiguriert, und die Terminierung ist für die Sitzung aktiviert.</li> </ul>	0, 1, 2 <ul style="list-style-type: none"> <li>0: Der Crawler ist entsprechend der Crawlerkonfigurationsdateien nicht für die Terminierung konfiguriert.</li> <li>1: Der Crawler ist für die Terminierung konfiguriert, aber die Terminierung wurde für die Sitzung inaktiviert.</li> <li>2: Der Crawler ist für die Terminierung konfiguriert, und die Terminierung ist für die Sitzung aktiviert.</li> </ul>	0, 1, 2 <ul style="list-style-type: none"> <li>0: Der Crawler ist entsprechend der Crawlerkonfigurationsdateien nicht für die Terminierung konfiguriert.</li> <li>1: Der Crawler ist für die Terminierung konfiguriert, aber die Terminierung wurde für die Sitzung inaktiviert.</li> <li>2: Der Crawler ist für die Terminierung konfiguriert, und die Terminierung ist für die Sitzung aktiviert.</li> </ul>
Server/ScheduleTime	Planzeit, falls zutreffend.	Planzeit, falls zutreffend.	Planzeit, falls zutreffend.
Server/TotalTime	Die Gesamtzeit, falls zutreffend.	Die Gesamtzeit, falls zutreffend.	Die Gesamtzeit, falls zutreffend.
Server/AggregationLevel	0: Der Crawler durchsucht Dokumente im normalen Modus.	0: Der Crawler durchsucht Dokumente im normalen Modus.	0, 1: <ul style="list-style-type: none"> <li>0: Der Notes-Crawler durchsucht Dokumente im normalen Modus. (Die übrigen Crawler außer dem Notes-Crawler geben immer 0 zurück.)</li> <li>1: Der Notes-Crawler durchsucht Dokumente im Verzeichnismodus.</li> </ul>

Tabelle 18. Crawlerbereichsstatusinformationen für die Exchange Server-, DB2 Content Manager- und Content Edition-Crawler

Element- und Attributname	Exchange Server-Crawler	DB2 Content Manager-Crawler	Content Edition-Crawler
Server@Name	Exchange Server-Servername.	DB2 Content Manager-Server.	Repositoryname.
Server/Status	Status (0, 1, 2, 3, 4, 5, -1) <ul style="list-style-type: none"> <li>0: Keine Crawlersuche ausgeführt</li> <li>1: Crawlersuche wird ausgeführt</li> <li>2: Abgeschlossen (nicht terminiert)</li> <li>3: Wartestatus</li> <li>4: Angehalten</li> <li>5: Terminiert</li> <li>-1: Fehler</li> </ul>	Status (0, 1, 2, 3, 4, 5, -1) <ul style="list-style-type: none"> <li>0: Keine Crawlersuche ausgeführt</li> <li>1: Crawlersuche wird ausgeführt</li> <li>2: Abgeschlossen (nicht terminiert)</li> <li>3: Wartestatus</li> <li>4: Angehalten</li> <li>5: Terminiert</li> <li>-1: Fehler</li> </ul>	Status (0, 1, 2, 3, 4, 5, -1) <ul style="list-style-type: none"> <li>0: Keine Crawlersuche ausgeführt</li> <li>1: Crawlersuche wird ausgeführt</li> <li>2: Abgeschlossen (nicht terminiert)</li> <li>3: Wartestatus</li> <li>4: Angehalten</li> <li>5: Terminiert</li> <li>-1: Fehler</li> </ul>

Tabelle 18. Crawlerbereichsstatusinformationen für die Exchange Server-, DB2 Content Manager- und Content Edition-Crawler (Forts.)

Element- und Attributname	Exchange Server-Crawler	DB2 Content Manager-Crawler	Content Edition-Crawler
Server/StatusMessage	<ul style="list-style-type: none"> <li>• 0: Keine Crawlersuche ausgeführt</li> <li>• 1: Crawlersuche wird ausgeführt</li> <li>• 2: Abgeschlossen (nicht terminiert)</li> <li>• 3: Wartestatus</li> <li>• 4: Angehalten</li> <li>• 5: Terminiert</li> <li>• -1: Fehler</li> </ul>	<ul style="list-style-type: none"> <li>• 0: Keine Crawlersuche ausgeführt</li> <li>• 1: Crawlersuche wird ausgeführt</li> <li>• 2: Abgeschlossen (nicht terminiert)</li> <li>• 3: Wartestatus</li> <li>• 4: Angehalten</li> <li>• 5: Terminiert</li> <li>• -1: Fehler</li> </ul>	<ul style="list-style-type: none"> <li>• 0: Keine Crawlersuche ausgeführt</li> <li>• 1: Crawlersuche wird ausgeführt</li> <li>• 2: Abgeschlossen (nicht terminiert)</li> <li>• 3: Wartestatus</li> <li>• 4: Angehalten</li> <li>• 5: Terminiert</li> <li>• -1: Fehler</li> </ul>
Server/NumberOfTargets	Die Anzahl Unterordner im Crawlerbereich.	Die Anzahl Elementtypen im Crawlerbereich.	Die Anzahl Elementklassen im Crawlerbereich.
Server/NumberOfCompletedTargets	Die Anzahl durchsuchter Unterordner.	Die Anzahl durchsuchter Elementtypen.	Die Anzahl durchsuchter Elementklassen.
Server/NumberOfErrors	Die Anzahl Fehler.	Die Anzahl Fehler.	Die Anzahl Fehler.
Server/StartTime	Die Startzeit, falls zutreffend.	Die Startzeit, falls zutreffend.	Die Startzeit, falls zutreffend.
Server/EndTime	Die Endzeit, falls zutreffend.	Die Endzeit, falls zutreffend.	Die Endzeit, falls zutreffend.
Server/ScheduleConfigured	0, 1, 2 <ul style="list-style-type: none"> <li>• 0: Der Crawler ist entsprechend der Crawlerkonfigurationsdateien nicht für die Terminierung konfiguriert.</li> <li>• 1: Der Crawler ist für die Terminierung konfiguriert, aber die Terminierung wurde für die Sitzung inaktiviert.</li> <li>• 2: Der Crawler ist für die Terminierung konfiguriert, und die Terminierung ist für die Sitzung aktiviert.</li> </ul>	0, 1, 2 <ul style="list-style-type: none"> <li>• 0: Der Crawler ist entsprechend der Crawlerkonfigurationsdateien nicht für die Terminierung konfiguriert.</li> <li>• 1: Der Crawler ist für die Terminierung konfiguriert, aber die Terminierung wurde für die Sitzung inaktiviert.</li> <li>• 2: Der Crawler ist für die Terminierung konfiguriert, und die Terminierung ist für die Sitzung aktiviert.</li> </ul>	0, 1, 2 <ul style="list-style-type: none"> <li>• 0: Der Crawler ist entsprechend der Crawlerkonfigurationsdateien nicht für die Terminierung konfiguriert.</li> <li>• 1: Der Crawler ist für die Terminierung konfiguriert, aber die Terminierung wurde für die Sitzung inaktiviert.</li> <li>• 2: Der Crawler ist für die Terminierung konfiguriert, und die Terminierung ist für die Sitzung aktiviert.</li> </ul>
Server/ScheduleTime	Planzeit, falls zutreffend.	Planzeit, falls zutreffend.	Planzeit, falls zutreffend.
Server/TotalTime	Die Gesamtzeit, falls zutreffend.	Die Gesamtzeit, falls zutreffend.	Die Gesamtzeit, falls zutreffend.
Server/AggregationLevel	0: Der Crawler durchsucht Dokumente im normalen Modus.	0: Der Crawler durchsucht Dokumente im normalen Modus.	0: Der Crawler durchsucht Dokumente im normalen Modus.

Tabelle 19. Crawlerbereichsstatusinformationen für die QuickPlace-, Domino Document Manager-, UNIX-Dateisystem und Windows-Dateisystemcrawler

Element- und Attributname	QuickPlace-Crawler	Domino Document Manager-Crawler	Crawler für das UNIX- und Windows-Dateisystem
Server@Name	Verzeichnis für Bereiche	Bibliotheksdatenbank	Der feste Wert ist localhost.
Server/Status	Status (0, 1, 2, 3, 4, 5, -1) <ul style="list-style-type: none"> <li>• 0: Keine Crawlersuche ausgeführt</li> <li>• 1: Crawlersuche wird ausgeführt</li> <li>• 2: Abgeschlossen (nicht terminiert)</li> <li>• 3: Wartestatus</li> <li>• 4: Angehalten</li> <li>• 5: Terminiert</li> <li>• -1: Fehler</li> </ul>	Status (0, 1, 2, 3, 4, 5, -1) <ul style="list-style-type: none"> <li>• 0: Keine Crawlersuche ausgeführt</li> <li>• 1: Crawlersuche wird ausgeführt</li> <li>• 2: Abgeschlossen (nicht terminiert)</li> <li>• 3: Wartestatus</li> <li>• 4: Angehalten</li> <li>• 5: Terminiert</li> <li>• -1: Fehler</li> </ul>	Status (0, 1, 2, 3, 4, 5, -1) <ul style="list-style-type: none"> <li>• 0: Keine Crawlersuche ausgeführt</li> <li>• 1: Crawlersuche wird ausgeführt</li> <li>• 2: Abgeschlossen (nicht terminiert)</li> <li>• 3: Wartestatus</li> <li>• 4: Angehalten</li> <li>• 5: Terminiert</li> <li>• -1: Fehler</li> </ul>
Server/StatusMessage	<ul style="list-style-type: none"> <li>• 0: Keine Crawlersuche ausgeführt</li> <li>• 1: Crawlersuche wird ausgeführt</li> <li>• 2: Abgeschlossen (nicht terminiert)</li> <li>• 3: Wartestatus</li> <li>• 4: Angehalten</li> <li>• 5: Terminiert</li> <li>• -1: Fehler</li> </ul>	<ul style="list-style-type: none"> <li>• 0: Keine Crawlersuche ausgeführt</li> <li>• 1: Crawlersuche wird ausgeführt</li> <li>• 2: Abgeschlossen (nicht terminiert)</li> <li>• 3: Wartestatus</li> <li>• 4: Angehalten</li> <li>• 5: Terminiert</li> <li>• -1: Fehler</li> </ul>	<ul style="list-style-type: none"> <li>• 0: Keine Crawlersuche ausgeführt</li> <li>• 1: Crawlersuche wird ausgeführt</li> <li>• 2: Abgeschlossen (nicht terminiert)</li> <li>• 3: Wartestatus</li> <li>• 4: Angehalten</li> <li>• 5: Terminiert</li> <li>• -1: Fehler</li> </ul>
Server/NumberOfTargets	Die Anzahl Datenbanken für Bereiche und Räume im Crawlerbereich.	Die Anzahl Aktenschrankdateien im Crawlerbereich.	Die Anzahl Unterverzeichnisse im Crawlerbereich.
Server/NumberOfCompletedTargets	Die Anzahl durchsuchter Datenbanken für Bereiche und Räume.	Die Anzahl durchsuchter Aktenschrankdateien.	Die Anzahl Unterverzeichnisse im Crawlerbereich.
Server/NumberOfErrors	Die Anzahl Fehler.	Die Anzahl Fehler.	Die Anzahl Fehler.
Server/StartTime	Die Startzeit, falls zutreffend.	Die Startzeit, falls zutreffend.	Die Startzeit, falls zutreffend.
Server/EndTime	Die Endzeit, falls zutreffend.	Die Endzeit, falls zutreffend.	Die Endzeit, falls zutreffend.

Tabelle 19. Crawlerbereichsstatusinformationen für die QuickPlace-, Domino Document Manager-, UNIX-Dateisystem und Windows-Dateisystemcrawler (Forts.)

Element- und Attributname	QuickPlace-Crawler	Domino Document Manager-Crawler	Crawler für das UNIX- und Windows-Dateisystem
Server/ScheduleConfigured	0, 1, 2 <ul style="list-style-type: none"> <li>• 0: Der Crawler ist entsprechend der Crawlerkonfigurationsdateien nicht für die Terminierung konfiguriert.</li> <li>• 1: Der Crawler ist für die Terminierung konfiguriert, aber die Terminierung wurde für die Sitzung inaktiviert.</li> <li>• 2: Der Crawler ist für die Terminierung konfiguriert, und die Terminierung ist für die Sitzung aktiviert.</li> </ul>	0, 1, 2 <ul style="list-style-type: none"> <li>• 0: Der Crawler ist entsprechend der Crawlerkonfigurationsdateien nicht für die Terminierung konfiguriert.</li> <li>• 1: Der Crawler ist für die Terminierung konfiguriert, aber die Terminierung wurde für die Sitzung inaktiviert.</li> <li>• 2: Der Crawler ist für die Terminierung konfiguriert, und die Terminierung ist für die Sitzung aktiviert.</li> </ul>	0, 1, 2 <ul style="list-style-type: none"> <li>• 0: Der Crawler ist entsprechend der Crawlerkonfigurationsdateien nicht für die Terminierung konfiguriert.</li> <li>• 1: Der Crawler ist für die Terminierung konfiguriert, aber die Terminierung wurde für die Sitzung inaktiviert.</li> <li>• 2: Der Crawler ist für die Terminierung konfiguriert, und die Terminierung ist für die Sitzung aktiviert.</li> </ul>
Server/ScheduleTime	Planzeit, falls zutreffend.	Planzeit, falls zutreffend.	Planzeit, falls zutreffend.
Server/TotalTime	Die Gesamtzeit, falls zutreffend.	Die Gesamtzeit, falls zutreffend.	Die Gesamtzeit, falls zutreffend.
Server/AggregationLevel	0: Der Crawler durchsucht Dokumente im normalen Modus.	0: Der Crawler durchsucht Dokumente im normalen Modus.	0: Der Crawler durchsucht Dokumente im normalen Modus.

Tabelle 20. Crawlerbereichsstatusinformationen für die WebSphere Portal- und Web Content Management-Crawler

Element- und Attributname	WebSphere Portal-Crawler	Web Content Management-Crawler
Server@Name	WebSphere Portal-Server	Web Content Management-Ausgangs-URL für die Suche
Server/Status	Status (0, 1, 2, 3, 4, 5, -1) <ul style="list-style-type: none"> <li>• 0: Keine Crawlersuche ausgeführt</li> <li>• 1: Crawlersuche wird ausgeführt</li> <li>• 2: Abgeschlossen (nicht terminiert)</li> <li>• 3: Wartestatus</li> <li>• 4: Angehalten</li> <li>• 5: Terminiert</li> <li>• -1: Fehler</li> </ul>	Status (0, 1, 2, 3, 4, 5, -1) <ul style="list-style-type: none"> <li>• 0: Keine Crawlersuche ausgeführt</li> <li>• 1: Crawlersuche wird ausgeführt</li> <li>• 2: Abgeschlossen (nicht terminiert)</li> <li>• 3: Wartestatus</li> <li>• 4: Angehalten</li> <li>• 5: Terminiert</li> <li>• -1: Fehler</li> </ul>
Server/StatusMessage	<ul style="list-style-type: none"> <li>• 0: Keine Crawlersuche ausgeführt</li> <li>• 1: Crawlersuche wird ausgeführt</li> <li>• 2: Abgeschlossen (nicht terminiert)</li> <li>• 3: Wartestatus</li> <li>• 4: Angehalten</li> <li>• 5: Terminiert</li> <li>• -1: Fehler</li> </ul>	<ul style="list-style-type: none"> <li>• 0: Keine Crawlersuche ausgeführt</li> <li>• 1: Crawlersuche wird ausgeführt</li> <li>• 2: Abgeschlossen (nicht terminiert)</li> <li>• 3: Wartestatus</li> <li>• 4: Angehalten</li> <li>• 5: Terminiert</li> <li>• -1: Fehler</li> </ul>

Tabelle 20. Crawlerbereichsstatusinformationen für die WebSphere Portal- und Web Content Management-Crawler (Forts.)

Element- und Attributname	WebSphere Portal-Crawler	Web Content Management-Crawler
Server/NumberOfTargets	Die Anzahl Server im Crawlerbereich.	Die Anzahl Sites im Crawlerbereich.
Server/NumberOfCompletedTargets	Die Anzahl durchsuchter Server.	Die Anzahl durchsuchter Sites.
Server/NumberOfErrors	Die Anzahl Fehler.	Die Anzahl Fehler.
Server/StartTime	Die Startzeit, falls zutreffend.	Die Startzeit, falls zutreffend.
Server/EndTime	Die Endzeit, falls zutreffend.	Die Endzeit, falls zutreffend.
Server/ScheduleConfigured	0, 1, 2 <ul style="list-style-type: none"> <li>• 0: Der Crawler ist entsprechend der Crawlerkonfigurationsdateien nicht für die Terminierung konfiguriert.</li> <li>• 1: Der Crawler ist für die Terminierung konfiguriert, aber die Terminierung wurde für die Sitzung inaktiviert.</li> <li>• 2: Der Crawler ist für die Terminierung konfiguriert, und die Terminierung ist für die Sitzung aktiviert.</li> </ul>	0, 1, 2 <ul style="list-style-type: none"> <li>• 0: Der Crawler ist entsprechend der Crawlerkonfigurationsdateien nicht für die Terminierung konfiguriert.</li> <li>• 1: Der Crawler ist für die Terminierung konfiguriert, aber die Terminierung wurde für die Sitzung inaktiviert.</li> <li>• 2: Der Crawler ist für die Terminierung konfiguriert, und die Terminierung ist für die Sitzung aktiviert.</li> </ul>
Server/ScheduleTime	Planzeit, falls zutreffend.	Planzeit, falls zutreffend.
Server/TotalTime	Die Gesamtzeit, falls zutreffend.	Die Gesamtzeit, falls zutreffend.
Server/AggregationLevel	0: Der Crawler durchsucht Dokumente im normalen Modus.	0: Der Crawler durchsucht Dokumente im normalen Modus.

**Detaillierter Status für den Crawlerbereich für den Web-Crawler:** Wenn Sie den Befehl zum Abrufen des detaillierten Status für den Crawlerbereich für einen Web-Crawler ausführen, gibt der Befehl Informationen im XML-Dokumentformat zurück. Die folgenden Informationen können durch den detaillierten Statusbefehl für den Crawlerbereich zurückgegeben werden:

Tabelle 21. Auswahlmaskenwerte für den detaillierten Statusbefehl des Crawlerbereichs für den Web-Crawler

Maskenbit	Auswahl
1	Anzahl Seiten im Rohdatenspeicher
2	Anzahl aufgespürter Sites
4	Anzahl Sites mit DNS
8	Anzahl Sites ohne DNS
16	Anzahl aufgespürter URL-Adressen
32	Anzahl eindeutiger, gespeicherter Seiten
64	Anzahl durchsuchter URL-Adressen
128	Anzahl nicht durchsuchter URL-Adressen
256	Anzahl überfälliger URL-Adressen
512	Verteilung des HTTP-Statuscodes

Beispiel für zurückgegebene Informationen:

```

<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 der folgenden Tabelle wird jedes Feld beschrieben, das für den Status für den detaillierten Crawlerbereich des Web-Crawlers zurückgegeben wird:

*Tabelle 22. Detaillierte Crawlerbereichsstatusinformationen für den Web-Crawler*

Element	Attribute	Beschreibung
CrawlDetailsPerSite	<ul style="list-style-type: none"> <li>LastActivationTime:</li> <li>LastActivationDuration:</li> <li>IPAddressList:</li> <li>RobotsContent:</li> <li>HTTPCodeDist:</li> </ul>	Schnell abrufbare Informationen zum detaillierten Status einer Site.
Site	URL	URL der Stammseite der Site.
NumURLsDiscovered	Value	Die Anzahl URL-Adressen, die von der Site aufgespürt wurden.
NumURLsOverdue	Value	Die Anzahl URL-Adressen, die von der Site erneut durchsucht werden können.
NumURLsCrawled	Value	Die Anzahl URL-Adressen, die für die Site durchsucht wurden.
NumURLsUncrawled	Value	Die Anzahl URL-Adressen, die für die Site noch nicht durchsucht wurden.

Tabelle 22. Detaillierte Crawlerbereichsstatusinformationen für den Web-Crawler (Forts.)

Element	Attribute	Beschreibung
NumURLsOverdueBy	Threshold, Value: Ganze Zahl (positiv oder negativ)  Der Wert steht für die Anzahl URL-Adressen, die erneut durchsucht werden können. Der Schwellenwert gibt die Zeit an, die bis zum erneuten Durchsuchen der URL-Adressen vergangen ist. Der Schwellenwert wird als Anzahl Sekunden relativ zur aktuellen Uhrzeit gemessen. Ein negativer Schwellenwert bedeutet, dass eine erneute Crawlersuche für die URL-Adresse überfällig ist. Ein positiver Schwellenwert bedeutet, dass eine erneute Crawlersuche für die URL-Adressen bevorsteht.	Die Anzahl URL-Adressen, die vor einigen Sekunden erneut durchsucht hätten werden können oder die nach der angegebenen Anzahl Sekunden für die erneute Crawlersuche auswählbar werden.
NumURLsActivated	Value	Die Anzahl URL-Adressen, die während der letzten Überprüfung dieser Site in den Speicher geladen und den Crawler-Threads zur Verfügung gestellt wurden.
LastActivationTime	Value	Die Anzahl Sekunden, seitdem die URL-Adressen dieser Site zuletzt in den Speicher geladen wurden.
LastActivationDuration	Value	Die Anzahl Sekunden, in denen die URL-Adressen dieser Site zuletzt im Speicher waren und den Crawler-Threads zur Verfügung standen.
IPAddressList	IPAddress	Alle bekannten IP-Adressen für den Server-Host dieser Site.
IPAddress	Value	IPv4-Adresse des Server-Hosts der Site in Punktschreibweise.
RobotsContent	Text	Text aus der Datei robots.txt, falls darin Text vorhanden ist.
HTTPCodeDist	HTTPCode	Verteilung von HTTP-Codes aus versuchten Downloads dieser Site.
HTTPCode	Code: Ganze Zahl  Ein HTTP-Statuscode oder ein anderer interner Code.	Häufigkeit, mit der ein bestimmter HTTP-Statuscode während des Durchsuchens dieser Site aufgetreten ist.

**Detaillierter Status für den Crawlerbereich für Nicht-Web-Crawler:** Wenn Sie den Befehl zum Abrufen des detaillierten Status für den Crawlerbereich für Nicht-Web-Crawler ausführen, gibt der Befehl Informationen im XML-Dokumentformat zurück. Die folgenden Informationen können durch den Befehl **getCrawlSpaceStatusDetail** für Nicht-Web-Crawler zurückgegeben werden:

```
FFQC5314I Ergebnis: <?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>

```

Tabelle 23. Detaillierte Crawlerbereichsstatusinformationen für die NNTP-, DB2-, JDBC-Datenbank- und Notes-Crawler

Element- und Attributname	NNTP-Crawler	DB2- und JDBC-Datenbankcrawler	Notes-Crawler
Target@Name	Name der Newsgroup	Tabellenname	Sicht- oder Ordnername
Target@CrawlType	Nicht zutreffend.	0,1 (DB2); 0 (JDBC-Datenbank) <ul style="list-style-type: none"> <li>0: Aktive Crawlersuche (normal)</li> <li>1: Passive Crawlersuche (DB2-Ereignisveröffentlichung)</li> </ul>	0
Target/Status	Status: (0, 1, 2, 3, 4, -1) <ul style="list-style-type: none"> <li>0: Keine Crawlersuche ausgeführt</li> <li>1: Crawlersuche wird ausgeführt</li> <li>2: Abgeschlossen (nicht terminiert)</li> <li>3: Wartestatus</li> <li>4: Angehalten</li> <li>-1: Fehler</li> </ul>	Status: (0, 1, 2, 3, 4, -1) <ul style="list-style-type: none"> <li>0: Keine Crawlersuche ausgeführt</li> <li>1: Crawlersuche wird ausgeführt</li> <li>2: Abgeschlossen (nicht terminiert)</li> <li>3: Wartestatus</li> <li>4: Angehalten</li> <li>-1: Fehler</li> </ul>	Status: (0, 1, 2, 3, 4, -1) <ul style="list-style-type: none"> <li>0: Keine Crawlersuche ausgeführt</li> <li>1: Crawlersuche wird ausgeführt</li> <li>2: Abgeschlossen (nicht terminiert)</li> <li>3: Wartestatus</li> <li>4: Angehalten</li> <li>-1: Fehler</li> </ul>
Target/StatusMessage	<ul style="list-style-type: none"> <li>0: Keine Crawlersuche ausgeführt</li> <li>1: Crawlersuche wird ausgeführt</li> <li>2: Abgeschlossen (nicht terminiert)</li> <li>3: Wartestatus</li> <li>4: Angehalten</li> <li>-1: Fehler</li> </ul>	<ul style="list-style-type: none"> <li>0: Keine Crawlersuche ausgeführt</li> <li>1: Crawlersuche wird ausgeführt</li> <li>2: Abgeschlossen (nicht terminiert)</li> <li>3: Wartestatus</li> <li>4: Angehalten</li> <li>-1: Fehler</li> </ul>	<ul style="list-style-type: none"> <li>0: Keine Crawlersuche ausgeführt</li> <li>1: Crawlersuche wird ausgeführt</li> <li>2: Abgeschlossen (nicht terminiert)</li> <li>3: Wartestatus</li> <li>4: Angehalten</li> <li>-1: Fehler</li> </ul>
Target/NumberOf Records	Die letzte Artikelnummer auf dem Server.	Die Anzahl durchsuchter Datensätze.	Die Anzahl durchsuchter Dokumente.
Target/NumberOf CompletedRecords	Die Anzahl durchsuchter Artikel.	Die Anzahl durchsuchter Datensätze.	Die Anzahl durchsuchter Dokumente.
Target/NumberOf InsertedRecords	Die Anzahl neu veröffentlichter Artikel.	Die Anzahl eingefügter Datensätze.	Die Anzahl eingefügter Datensätze.
Target/NumberOf UpdatedRecords	Nicht zutreffend.	Die Anzahl aktualisierter Datensätze.	Die Anzahl aktualisierter Datensätze.
Target/NumberOf DeletedRecords	Nicht zutreffend.	Die Anzahl gelöschter Datensätze.	Die Anzahl gelöschter Datensätze.
Target/StartTime	Das Datum und die Uhrzeit, zu dem bzw. der Crawler zuletzt gestartet wurde.	Das Datum und die Uhrzeit, zu dem bzw. der Crawler zuletzt gestartet wurde.	Das Datum und die Uhrzeit, zu dem bzw. der Crawler zuletzt gestartet wurde.

Tabelle 23. Detaillierte Crawlerbereichsstatusinformationen für die NNTP-, DB2-, JDBC-Datenbank- und Notes-Crawler (Forts.)

Element- und Attributname	NNTP-Crawler	DB2- und JDBC-Datenbankcrawler	Notes-Crawler
Target/EndTime	Das Datum und die Uhrzeit, zu dem bzw. der die Crawlersuche abgeschlossen wurde.	Das Datum und die Uhrzeit, zu dem bzw. der die Crawlersuche abgeschlossen wurde.	Das Datum und die Uhrzeit, zu dem bzw. der die Crawlersuche abgeschlossen wurde.
Target/TotalTime	Die Zeit, die der Crawler für die Crawlersuche benötigt hat.	Die Zeit, die der Crawler für die Crawlersuche benötigt hat.	Die Zeit, die der Crawler für die Crawlersuche benötigt hat.
Target/AggregationLevel	0: Der Crawler durchsucht Dokumente im normalen Modus.	0: Der Crawler durchsucht Dokumente im normalen Modus.	0, 1: <ul style="list-style-type: none"> <li>• 0: Der Crawler durchsucht Dokumente im normalen Modus.</li> <li>• 1: Der Crawler durchsucht Dokumente im Verzeichnismodus.</li> </ul>
Target/LastUpdatedTime	Nicht zutreffend.	Die Zeit der letzten Aktualisierung: <ul style="list-style-type: none"> <li>• 0: Aktive Crawlersuche (normal)</li> <li>• 1: Passive Crawlersuche (DB2-Ereignisveröffentlichung)</li> </ul>	Nicht zutreffend.
Target/LastResetTime	Nicht zutreffend.	Statistikdaten zur letzten Zurücksetzung: <ul style="list-style-type: none"> <li>• 0: Aktive Crawlersuche (normal)</li> <li>• 1: Passive Crawlersuche (DB2-Ereignisveröffentlichung)</li> </ul>	Nicht zutreffend.

Tabelle 24. Detaillierte Crawlerbereichsstatusinformationen für die Exchange Server-, DB2 Content Manager- und Content Edition-Crawler

Element- und Attributname	Exchange Server-Crawler	DB2 Content Manager-Crawler	Content Edition-Crawler
Target@Name	Name des Unterordners	Name des Elementtyps	Name der Elementklasse
Target@CrawlType	0	0	0
Target/Status	Status (0, 1, 2, 3, 4, -1) <ul style="list-style-type: none"> <li>• 0: Keine Crawlersuche ausgeführt</li> <li>• 1: Crawlersuche wird ausgeführt</li> <li>• 2: Abgeschlossen (nicht terminiert)</li> <li>• 3: Wartestatus</li> <li>• 4: Angehalten</li> <li>• -1: Fehler</li> </ul>	Status (0, 1, 2, 3, 4, -1) <ul style="list-style-type: none"> <li>• 0: Keine Crawlersuche ausgeführt</li> <li>• 1: Crawlersuche wird ausgeführt</li> <li>• 2: Abgeschlossen (nicht terminiert)</li> <li>• 3: Wartestatus</li> <li>• 4: Angehalten</li> <li>• -1: Fehler</li> </ul>	Status (0, 1, 2, 3, 4, -1) <ul style="list-style-type: none"> <li>• 0: Keine Crawlersuche ausgeführt</li> <li>• 1: Crawlersuche wird ausgeführt</li> <li>• 2: Abgeschlossen (nicht terminiert)</li> <li>• 3: Wartestatus</li> <li>• 4: Angehalten</li> <li>• -1: Fehler</li> </ul>

Tabelle 24. Detaillierte Crawlerbereichsstatusinformationen für die Exchange Server-, DB2 Content Manager- und Content Edition-Crawler (Forts.)

Element- und Attributname	Exchange Server-Crawler	DB2 Content Manager-Crawler	Content Edition-Crawler
Target/StatusMessage	<ul style="list-style-type: none"> <li>• 0: Keine Crawlersuche ausgeführt</li> <li>• 1: Crawlersuche wird ausgeführt</li> <li>• 2: Abgeschlossen (nicht terminiert)</li> <li>• 3: Wartestatus</li> <li>• 4: Angehalten</li> <li>• -1: Fehler</li> </ul>	<ul style="list-style-type: none"> <li>• 0: Keine Crawlersuche ausgeführt</li> <li>• 1: Crawlersuche wird ausgeführt</li> <li>• 2: Abgeschlossen (nicht terminiert)</li> <li>• 3: Wartestatus</li> <li>• 4: Angehalten</li> <li>• -1: Fehler</li> </ul>	<ul style="list-style-type: none"> <li>• 0: Keine Crawlersuche ausgeführt</li> <li>• 1: Crawlersuche wird ausgeführt</li> <li>• 2: Abgeschlossen (nicht terminiert)</li> <li>• 3: Wartestatus</li> <li>• 4: Angehalten</li> <li>• -1: Fehler</li> </ul>
Target/NumberOf Records	Nicht zutreffend.	Nicht zutreffend.	Nicht zutreffend.
Target/NumberOf CompletedRecords	Die Anzahl durchsuchter Dokumente.	Die Anzahl durchsuchter Dokumente.	Die Anzahl durchsuchter Dokumente.
Target/NumberOf InsertedRecords	Die Anzahl eingefügter Datensätze.	Die Anzahl eingefügter Datensätze.	Die Anzahl eingefügter Datensätze.
Target/NumberOf UpdatedRecords	Nicht zutreffend.	Die Anzahl aktualisierter Datensätze.	Die Anzahl aktualisierter Datensätze.
Target/NumberOf DeletedRecords	Nicht zutreffend.	Die Anzahl gelöschter Datensätze.	Die Anzahl gelöschter Datensätze.
Target/StartTime	Das Datum und die Uhrzeit, zu dem bzw. der Crawler zuletzt gestartet wurde.	Das Datum und die Uhrzeit, zu dem bzw. der Crawler zuletzt gestartet wurde.	Das Datum und die Uhrzeit, zu dem bzw. der Crawler zuletzt gestartet wurde.
Target/EndTime	Das Datum und die Uhrzeit, zu dem bzw. der die Crawlersuche abgeschlossen wurde.	Das Datum und die Uhrzeit, zu dem bzw. der die Crawlersuche abgeschlossen wurde.	Das Datum und die Uhrzeit, zu dem bzw. der die Crawlersuche abgeschlossen wurde.
Target/TotalTime	Die Zeit, die der Crawler für die Crawlersuche benötigt hat.	Die Zeit, die der Crawler für die Crawlersuche benötigt hat.	Die Zeit, die der Crawler für die Crawlersuche benötigt hat.
Target/AggregationLevel	0: Der Crawler durchsucht Dokumente im normalen Modus.	0: Der Crawler durchsucht Dokumente im normalen Modus.	0: Der Crawler durchsucht Dokumente im normalen Modus.
Target/LastUpdatedTime	Nicht zutreffend.	Nicht zutreffend.	Nicht zutreffend.
Target/LastResetTime	Nicht zutreffend.	Nicht zutreffend.	Nicht zutreffend.

Tabelle 25. Detaillierte Crawlerbereichsstatusinformationen für die QuickPlace-, Domino Document Manager-, UNIX-Dateisystem und Windows-Dateisystemcrawler

Element- und Attributname	QuickPlace-Crawler	Domino Document Manager-Crawler	Crawler für das UNIX- und Windows-Dateisystem
Target@Name	Name der Datenbank für Bereiche oder Räume	Name der Datenbank für Aktenschrankdateien	Unterverzeichnisname
Target@CrawlType	0	0	0

Tabelle 25. Detaillierte Crawlerbereichsstatusinformationen für die QuickPlace-, Domino Document Manager-, UNIX-Dateisystem und Windows-Dateisystemcrawler (Forts.)

Element- und Attributname	QuickPlace-Crawler	Domino Document Manager-Crawler	Crawler für das UNIX- und Windows-Dateisystem
Target/Status	Status (0, 1, 2, 3, 4, -1) <ul style="list-style-type: none"> <li>• 0: Keine Crawlersuche ausgeführt</li> <li>• 1: Crawlersuche wird ausgeführt</li> <li>• 2: Abgeschlossen (nicht terminiert)</li> <li>• 3: Wartestatus</li> <li>• 4: Angehalten</li> <li>• -1: Fehler</li> </ul>	Status (0, 1, 2, 3, 4, -1) <ul style="list-style-type: none"> <li>• 0: Keine Crawlersuche ausgeführt</li> <li>• 1: Crawlersuche wird ausgeführt</li> <li>• 2: Abgeschlossen (nicht terminiert)</li> <li>• 3: Wartestatus</li> <li>• 4: Angehalten</li> <li>• -1: Fehler</li> </ul>	Status (0, 1, 2, 3, 4, -1) <ul style="list-style-type: none"> <li>• 0: Keine Crawlersuche ausgeführt</li> <li>• 1: Crawlersuche wird ausgeführt</li> <li>• 2: Abgeschlossen (nicht terminiert)</li> <li>• 3: Wartestatus</li> <li>• 4: Angehalten</li> <li>• -1: Fehler</li> </ul>
Target/StatusMessage	<ul style="list-style-type: none"> <li>• 0: Keine Crawlersuche ausgeführt</li> <li>• 1: Crawlersuche wird ausgeführt</li> <li>• 2: Abgeschlossen (nicht terminiert)</li> <li>• 3: Wartestatus</li> <li>• 4: Angehalten</li> <li>• -1: Fehler</li> </ul>	<ul style="list-style-type: none"> <li>• 0: Keine Crawlersuche ausgeführt</li> <li>• 1: Crawlersuche wird ausgeführt</li> <li>• 2: Abgeschlossen (nicht terminiert)</li> <li>• 3: Wartestatus</li> <li>• 4: Angehalten</li> <li>• -1: Fehler</li> </ul>	<ul style="list-style-type: none"> <li>• 0: Keine Crawlersuche ausgeführt</li> <li>• 1: Crawlersuche wird ausgeführt</li> <li>• 2: Abgeschlossen (nicht terminiert)</li> <li>• 3: Wartestatus</li> <li>• 4: Angehalten</li> <li>• -1: Fehler</li> </ul>
Target/NumberOf Records	Nicht zutreffend.	Nicht zutreffend.	Nicht zutreffend.
Target/NumberOf CompletedRecords	Die Anzahl durchsuchter Dokumente.	Die Anzahl durchsuchter Dokumente.	Die Anzahl durchsuchter Dateien.
Target/NumberOf InsertedRecords	Die Anzahl eingefügter Datensätze.	Die Anzahl eingefügter Datensätze.	Die Anzahl eingefügter Datensätze.
Target/NumberOf UpdatedRecords	Die Anzahl aktualisierter Datensätze.	Die Anzahl aktualisierter Datensätze.	Die Anzahl aktualisierter Datensätze.
Target/NumberOf DeletedRecords	Die Anzahl gelöschter Datensätze.	Die Anzahl gelöschter Datensätze.	Die Anzahl gelöschter Datensätze.
Target/StartTime	Das Datum und die Uhrzeit, zu dem bzw. der Crawler zuletzt gestartet wurde.	Das Datum und die Uhrzeit, zu dem bzw. der Crawler zuletzt gestartet wurde.	Das Datum und die Uhrzeit, zu dem bzw. der Crawler zuletzt gestartet wurde.
Target/EndTime	Das Datum und die Uhrzeit, zu dem bzw. der die Crawlersuche abgeschlossen wurde.	Das Datum und die Uhrzeit, zu dem bzw. der die Crawlersuche abgeschlossen wurde.	Das Datum und die Uhrzeit, zu dem bzw. der die Crawlersuche abgeschlossen wurde.
Target/TotalTime	Die Zeit, die der Crawler für die Crawlersuche benötigt hat.	Die Zeit, die der Crawler für die Crawlersuche benötigt hat.	Die Zeit, die der Crawler für die Crawlersuche benötigt hat.
Target/AggregationLevel	0: Der Crawler durchsucht Dokumente im normalen Modus.	0: Der Crawler durchsucht Dokumente im normalen Modus.	0: Der Crawler durchsucht Dokumente im normalen Modus.
Target/LastUpdatedTime	Nicht zutreffend.	Nicht zutreffend.	Nicht zutreffend.
Target/LastResetTime	Nicht zutreffend.	Nicht zutreffend.	Nicht zutreffend.

Tabelle 26. Detaillierte Crawlerbereichsstatusinformationen für die WebSphere Portal- und Web Content Management-Crawler

Element- und Attributname	WebSphere Portal-Crawler	Web Content Management-Crawler
Target@Name	Name des WebSphere Portal-Servers	Die Ausgangs-URL für die Suche, die die Site darstellt
Target@CrawlType	0	0
Target/Status	Status: (0, 1, 2, 3, 4, -1) <ul style="list-style-type: none"> <li>• 0: Keine Crawler suche ausgeführt</li> <li>• 1: Crawler suche wird ausgeführt</li> <li>• 2: Abgeschlossen (nicht terminiert)</li> <li>• 3: Wartestatus</li> <li>• 4: Angehalten</li> <li>• -1: Fehler</li> </ul>	Status: (0, 1, 2, 3, 4, -1) <ul style="list-style-type: none"> <li>• 0: Keine Crawler suche ausgeführt</li> <li>• 1: Crawler suche wird ausgeführt</li> <li>• 2: Abgeschlossen (nicht terminiert)</li> <li>• 3: Wartestatus</li> <li>• 4: Angehalten</li> <li>• -1: Fehler</li> </ul>
Target/StatusMessage	<ul style="list-style-type: none"> <li>• 0: Keine Crawler suche ausgeführt</li> <li>• 1: Crawler suche wird ausgeführt</li> <li>• 2: Abgeschlossen (nicht terminiert)</li> <li>• 3: Wartestatus</li> <li>• 4: Angehalten</li> <li>• -1: Fehler</li> </ul>	<ul style="list-style-type: none"> <li>• 0: Keine Crawler suche ausgeführt</li> <li>• 1: Crawler suche wird ausgeführt</li> <li>• 2: Abgeschlossen (nicht terminiert)</li> <li>• 3: Wartestatus</li> <li>• 4: Angehalten</li> <li>• -1: Fehler</li> </ul>
Target/NumberOf Records	Nicht zutreffend.	Nicht zutreffend.
Target/NumberOf CompletedRecords	Die Gesamtzahl durchsuchter Datensätze.	Die Gesamtzahl durchsuchter Datensätze.
Target/NumberOf InsertedRecords	Die Anzahl eingefügter Datensätze.	Die Anzahl eingefügter Datensätze.
Target/NumberOf UpdatedRecords	Die Anzahl aktualisierter Datensätze.	Die Anzahl aktualisierter Datensätze.
Target/NumberOf DeletedRecords	Die Anzahl gelöschter Datensätze.	Die Anzahl gelöschter Datensätze.
Target/StartTime	Das Datum und die Uhrzeit, zu dem bzw. der Crawler zuletzt gestartet wurde.	Das Datum und die Uhrzeit, zu dem bzw. der Crawler zuletzt gestartet wurde.
Target/EndTime	Das Datum und die Uhrzeit, zu dem bzw. der die Crawler suche abgeschlossen wurde.	Das Datum und die Uhrzeit, zu dem bzw. der die Crawler suche abgeschlossen wurde.
Target/TotalTime	Die Zeit, die der Crawler für die Crawler suche benötigt hat.	Die Zeit, die der Crawler für die Crawler suche benötigt hat.
Target/AggregationLevel	0: Der Crawler durchsucht Dokumente im normalen Modus.	0: Der Crawler durchsucht Dokumente im normalen Modus.
Target/LastUpdatedTime	Nicht zutreffend.	Nicht zutreffend.
Target/LastResetTime	Nicht zutreffend.	Nicht zutreffend.

**Parserstatus:** Wenn Sie den Befehl zum Abrufen des Parserstatus ausführen, gibt der Befehl Informationen im XML-Dokumentformat zurück. Die folgenden Informationen können durch den Befehl für den Parserstatus zurückgegeben werden:

```
FFQC5314I Ergebnis:
<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>
</Monitor>

```

In der folgenden Tabelle werden die XML-Elemente für Informationen beschrieben, die durch den Befehl für den Parserstatus zurückgegeben werden:

Tabelle 27. Elemente für den Befehl für den Parserstatus

Element	Beschreibung
Status	<ul style="list-style-type: none"> <li>• 0: Die Parsersitzung für diese Objektgruppe wurde gestoppt.</li> <li>• 1: Die Parsersitzung für diese Objektgruppe ist aktiv.</li> </ul>
State	<p>Möglicher Status: Initializing, Idle, Restart, Parsing, Stopped, Paused, Resuming, NoParserServiceIsAvailable.</p> <p>Der Status <b>Initializing</b> gibt an, dass der Parser gerade gestartet wird und sein Status initialisiert wird.</p> <p>Der Status <b>Idle</b> gibt an, dass der Parser <i>N</i> Minuten im Ruhemodus ist und darauf wartet, dass mehr Dokumente von den Crawlern in dieser Objektgruppe empfangen werden. Die Standardruhezeit beträgt 300 Sekunden.</p> <p>Der Status <b>Restart</b> gibt an, dass der Parser darauf wartet, dass die JVM für die Syntexanalyse und das Aufbereiten erneut gestartet wird. Die JVM für die Syntexanalyse und das Aufbereiten wird in einer separaten Sitzung ausgeführt und befindet sich dort, wo Dokumente endgültig verarbeitet werden.</p> <p>Der Status <b>Parsing</b> gibt an, dass der Parser Dokumente verarbeitet.</p> <p>Der Status <b>Paused</b> gibt an, dass der Parser durch die Indexerstellungssitzung für diese Objektgruppe angehalten wurde.</p> <p>Der Status <b>Resuming</b> gibt an, dass der Parser durch die Indexerstellungssitzung für diese Objektgruppe aus dem Status <b>Paused</b> in den Status <b>Parsing</b> versetzt wurde.</p> <p>Der Status <b>NoParserServiceIsAvailable</b> gibt an, dass keine JVMs für die Syntexanalyse und das Aufbereiten verfügbar sind, um die Dokumente für diese Objektgruppe zu verarbeiten. Dieser Status bedeutet, dass alle JVMs für die Syntexanalyse und für das Aufbereiten von anderen Objektgruppen verwendet werden.</p>
NumberOfDocsToBeIndexed	Die Anzahl Dokumente im Speicher für diese Objektgruppe. Diese Anzahl schließt auch Dokumente ein, die für das Löschen bei der nächsten Indexerstellung markiert wurden
ParseRate	Die Syntexanalyserate in Dokumenten pro Sekunde.
ParseRateMBPerHour	Die Syntexanalyserate in MB pro Stunde.
NumberOfCpmThreads	Die Anzahl CPM-Threads, die von der JVM für die Syntexanalyse und das Aufbereiten zum Verarbeiten von Dokumenten für diese Objektgruppe verwendet werden.

Tabelle 27. Elemente für den Befehl für den Parserstatus (Forts.)

Element	Beschreibung
ParserServiceSession	Der Name der JVM für die Syntaxanalyse und das Aufbereiten, die die Dokumente für diese Objektgruppe verarbeitet. Dieses Feld ist nur verfügbar, wenn sich der Parser im Status <b>Parsing</b> befindet.
Name	Der Name des Crawlers.
Crawlerid	Die für diesen Crawler vom System erstellte ID.
Type	Der Crawlertyp (Web-, NNTP-, DB2-Crawler usw.)
ParserStatus	<ul style="list-style-type: none"> <li>• 0: Die Dokumente aus diesem Crawler werden zurzeit nicht syntaktisch analysiert (die Parsersitzung ist gestoppt).</li> <li>• 1: Die Dokumente aus diesem Crawler werden zurzeit syntaktisch analysiert (die Parsersitzung ist aktiv).</li> </ul>
NumberOfDocsAlreadyParsed	Die Anzahl syntaktisch analysierter Dokumente aus diesem Crawler.

**Status der Indexerstellung:** Wenn Sie den Befehl zum Abrufen des Status der Indexerstellung ausführen, gibt der Befehl Informationen im XML-Dokumentformat zurück. Die folgenden Informationen können durch den Statusbefehl für die Indexerstellung zurückgegeben werden:

```
<?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 der folgenden Tabelle wird jedes XML-Element für Informationen beschrieben, die durch den Statusbefehl für die Indexerstellung zurückgegeben werden:

Tabelle 28. Elemente für den Statusbefehl für die Indexerstellung

Element	Beschreibung
IndexStatusId	Die Indexstatus-ID.
StartTime	Die Zeit in Sekunden seit 1970, zu der dieser Index erstellt wurde. Zur Berechnung des aktuellen Zeitpunkts, den diese Zeitangabe darstellt, verwenden Sie die Formel January 1, 1970 %2B StartTime. Wenn Sie weitere Informationen zur Epoche benötigen, rufen Sie <a href="http://de.wikipedia.org/wiki/Unixzeit">http://de.wikipedia.org/wiki/Unixzeit</a> auf.

Tabelle 28. Elemente für den Statusbefehl für die Indexerstellung (Forts.)

Element	Beschreibung
Progress	Der Prozentsatz der Fertigstellung dieser Indexerstellung.
CurrentPhase	<ul style="list-style-type: none"> <li>• 1: Phase des erneuten Schreibens im Speicher</li> <li>• 2: Globale Analysephase</li> <li>• 3: Indexstellungsphase</li> </ul>
TotalPhase	Die Anzahl Phasen für diese Indexerstellung. Dieser Wert ist zurzeit 3.
IndexCopyProgress	Der Prozentsatz der Fertigstellung dieser Indexkopie. Der Indexkopierprozess kopiert den erstellten Index vom Indexerstellungsserver auf die Suchserver.
CurrentServer	Der Suchserver, auf den die Indexkopie den Index kopiert.
TotalServer	Die Anzahl Suchserver, auf die der Index kopiert wird.
IndexCopyTime	Die Gesamtzeit für das Kopieren des Index auf alle Suchserver.
IndexBuildTime	Die Gesamtzeit für alle Phasen der Indexerstellung.
Status	<ul style="list-style-type: none"> <li>• 0: Index erstellen und kopieren</li> <li>• -1: Anforderung für Indexerstellung fehlgeschlagen</li> <li>• 1: Indexerstellung, Kopiervorgang oder beides in Bearbeitung</li> </ul>
JobID	Eine eindeutige ID, die mit jeder Anforderung einer Indexerstellung verknüpft ist.
MessagesAvailable	Ein Boolescher Wert, der angibt, ob (im Falle einer Störung) Fehlernachrichten verfügbar sind.
StopTime	Die Endzeit für die Indexerstellung (alle Phasen) und die Indexkopie.
TotalTime	Der Zeitraum zwischen Startzeit und Stoppzeit.
NumberOfDocuments	Die Anzahl Dokumente im Index.
CurrentIndexWildcardSupport	Die Platzhalterzeicheneinstellungen, die für die nächste Indexerstellung verwendet werden sollen. Mögliche Werte sind <b>None</b> , <b>QueryExpansion</b> oder <b>IndexExpansion</b> .
ScheduleStatus	<ul style="list-style-type: none"> <li>• 0, falls kein Zeitplan für diese Objektgruppe und diesen Indextyp aktiviert ist.</li> <li>• 1, falls ein Zeitplan für diese Objektgruppe und diesen Indextyp aktiviert ist.</li> </ul>
ScheduledTimeEnabled	Die Zeit in Sekunden seit 1970, zu der die nächste Indexerstellung für diese Objektgruppe und diesen Indextyp ausgeführt wird. Zur Berechnung des aktuellen Zeitpunkts, den diese Zeitangabe darstellt, verwenden Sie die Formel January 1, 1970 %2B ScheduledTimeEnabled. Wenn Sie weitere Informationen zur Epoche benötigen, rufen Sie <a href="http://de.wikipedia.org/wiki/Unixzeit">http://de.wikipedia.org/wiki/Unixzeit</a> auf.

**Status des Suchservers:** Wenn Sie den Befehl zum Abrufen des Status des Suchservers ausführen, gibt der Befehl Informationen im XML-Dokumentformat zurück. Die folgenden Informationen können durch den Statusbefehl für den Suchserver zurückgegeben werden:

```
FFQC5314I Ergebnis: <?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 der folgenden Tabelle werden die XML-Elemente für Informationen beschrieben, die durch den Statusbefehl für den Suchserver zurückgegeben werden:

*Tabelle 29. Elemente für den Statusbefehl für den Suchserver*

Element	Beschreibung
SearchStatusName	Der Name und die ID für die Suchmanagersitzung, die den Suchindex für diese Objektgruppe überwacht und verwaltet.
HostName	Der Hostname des Servers, auf dem der Suchindex ausgeführt wird.
Status	<ul style="list-style-type: none"> <li>• 0, falls der Suchindex für diese Objektgruppe nicht aktiv ist.</li> <li>• 1, falls der Suchindex für diese Objektgruppe aktiv ist.</li> </ul>

**Detaillierter Status für den Suchserver:** Der Befehl, zum Status des Suchservers zurückzukehren, kann die folgenden Informationen zurückgeben:

```
FFQC5303I Suchmanager (knoten1) (sid: searchmanager.node1)
ist bereits aktiv. PID: 15711
FFQC5314I Ergebnis: 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 der folgenden Tabelle werden die Elemente in den Informationen beschrieben, die durch den Befehl für den detaillierten Status für den Suchserver zurückgegeben werden:

*Tabelle 30. Elemente für den Befehl für den detaillierten Status für den Suchserver*

Element	Beschreibung
CacheHits	Die Anzahl Ergebnisse, die aus dem Suchcache abgerufen wurden.
QueryRate	Die Anzahl im letzten Zeitintervall empfangener Abfragen. Standardmäßig beträgt das Zeitintervall 5 Minuten.
Port	Die Portnummer, die vom Suchindex zum Überwachen oder Empfangen von Abfragen verwendet wird.
SessionId	Die Sitzungs-ID für den Suchindex dieser Objektgruppe.
CacheHitRate	Die Anzahl Ergebnisse, die aus dem Suchcache abgerufen wurden, als Prozentsatz aller Suchergebnisse.
ResponseTime	Die durchschnittliche Antwortzeit in Millisekunden für das angegebene Zeitintervall. (Die Standardeinstellung ist 5 Minuten.)
Status	<ul style="list-style-type: none"> <li>• 0, falls der Suchindex für diese Objektgruppe nicht aktiv ist.</li> <li>• 1, falls der Suchindex für diese Objektgruppe aktiv ist.</li> </ul>
SessionName	Der Sitzungsname für den Suchindex dieser Objektgruppe.

## Rückkehrcodes für die esadmin-Befehle

Die folgenden Codes können für die **esadmin**-Befehle zurückgegeben werden:

Tabelle 31. Rückkehrcodes für die **esadmin**-Befehle

Code	Name	Beschreibung
0	CODE_ERROR_NONE	Der Befehl wurde erfolgreich ausgeführt.
102	CODE_ERROR_INSTANTIATION_EXCEPTION	Beim Instanzieren des Handlers für den Befehl ist ein Fehler aufgetreten.
103	CODE_ERROR_ACCESS_EXCEPTION	Beim Instanzieren des Handlers für den Befehl ist ein Fehler aufgrund eines unzulässigen Zugriffs aufgetreten.
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	

## Abrufen von Sitzungs-IDs

Verwenden Sie den Befehl `esadmin check`, um eine Liste mit Komponenten für die Unternehmenssuche mit ihren entsprechenden Sitzungs-IDs anzuzeigen. In der folgenden Tabelle wird eine Liste mit allgemeinen Sitzungen, deren IDs sowie mit dem Server, auf dem sie sich befinden, und dem Status der Sitzung angezeigt.

Tabelle 32. Beispiele für Sitzungsnamen, Ursprungsserver, Sitzungs-IDs und Sitzungsstatus

Sitzung	Server, auf dem die Sitzung aktiv ist	Sitzungs-ID	Sitzungsstatus
configmanager	Indexserver	10433	Gestartet
controller	Indexserver	10464	Gestartet
customcommunication	Indexserver	Nicht zutreffend	Nicht zutreffend
discovery	Indexserver	10649	Gestartet
monitor	Indexserver	10682	Gestartet
parserservice	Indexserver	10718	Gestartet
resource.node1	Indexserver	10759	Gestartet
samplecpp	Indexserver	10827	Gestartet
sampletest	Indexserver	10857	Gestartet
scheduler	Indexserver	10889	Gestartet
searchmanager.node1	Indexserver	10927	Gestartet
utilities.node1	Indexserver	10384	Gestartet

### Zugehörige Konzepte

„Überwachen der Aktivitäten der Unternehmenssuche“ auf Seite 327

„Durchführen von Backup und Restore für ein System für die Unternehmenssuche“ auf Seite 363



Nachrichten für die Unternehmenssuche



Nachrichten für die Unternehmenssuche

### Zugehörige Tasks

„Überwachen von Crawlern“ auf Seite 330  
„Starten eines Systems für die Unternehmenssuche“ auf Seite 319  
„Stoppen eines Systems für die Unternehmenssuche“ auf Seite 321  
„Verwalten der Suchserver im Standalone-Modus“ auf Seite 324



---

## Groß-/Kleinschreibung bei der Unternehmenssuche

Die Komponenten der Unternehmenssuche, wie z. B. Abfragesyntax, Quick Links, Feldnamen usw., behandeln die Groß-/Kleinschreibung unterschiedlich.

### Abfragesyntax

Bei der Suche wird die Groß-/Kleinschreibung außer in den folgenden Fällen nicht beachtet:

#### XML-Elementnamen und -Attributnamen

Groß-/Kleinschreibung muss beachtet werden. Bei Begriffen und Attributwerten wird auch in XML-Abfragen die Groß-/Kleinschreibung nicht beachtet. Als Beispiel hierfür dient das folgende Dokument:

```
<book>
 <Author>
 <Name>Ferdinand</Name>
 <Contact Type="eMail">ferdi@nand.org</Contact Type>
 <Contact Type="Phone">+1 408 876 4242</Contact Type>
 </Author>
</book>
```

Bei den folgenden Abfragen wird das Dokument nicht zurückgegeben:

- @xmlxp::'author[Name ftcontains ("Ferdinand")]'
- @xmlxp::'//contact[@type="eMail"]'
- @xmlf2::'<author><name>Ferdinand</name></author>'
- @xmlf2::'<CONTACT TYPE="email">ferdi</contact>'

Bei den folgenden Abfragen wird das Dokument jedoch zurückgegeben:

- @xmlxp::'Author[Name ftcontains ("ferdinand")]'
- @xmlxp::'//Contact[@Type="email"]'
- @xmlf2::'<Author><Name>ferdinand</Author><Name>'
- @xmlf2::'<Contact Type="email">ferdi</Contact>'

#### Zugriffssteuerungslisten (Access Control Lists, ACLs)

Groß-/Kleinschreibung muss beachtet werden.

#### URLs in den Begriffen docid: und samegroupas:

Groß-/Kleinschreibung muss beachtet werden. Allerdings muss die Groß-/Kleinschreibung bei URL-Bestandteilen in den Begriffen site: oder url: nicht beachtet werden. Zum Beispiel in einem Dokument mit der URL-Adresse <http://www.here.com/HR/>:

- Bei der Abfrage docid:<http://www.here.com/hr> wird das Dokument nicht zurückgegeben.
- Bei den Abfragen url:hr und url:HERE wird das Dokument zurückgegeben.
- Bei der Abfrage site:HERE.com wird das Dokument zurückgegeben.

#### Kategorie-IDs und Taxonomie-IDs

Groß-/Kleinschreibung muss nicht beachtet werden. Beispielsweise spielt in der Abfrage taxonomy\_id::category\_id die Groß-/Kleinschreibung von taxonomy\_id und category\_id keine Rolle. Die Abfrage RuleBased::c42

stimmt mit der Kategorie c42 in der regelbasierten Taxonomie und ebenfalls mit rulebased::C42 überein. Sie können nicht nach Kategorienname, aber nach Kategorie-ID suchen.

#### **Bereiche**

Groß-/Kleinschreibung muss nicht beachtet werden. Beispielsweise gibt sowohl Scope:RESEARCH als auch scope::research Dokumente aus dem Bereich Research zurück.

#### **Begriffe mit Platzhalterzeichen**

Groß-/Kleinschreibung muss nicht beachtet werden. Beispielsweise entspricht der Begriff Fer\*n\*d dem Begriff fer\*n\*d.

#### **Feldnamen**

Groß-/Kleinschreibung muss nicht beachtet werden. Beispielsweise sind die Abfragen Title:Expenses, TITLE:expenses und title:expenses äquivalent. Bei allen Feldnamen braucht die Groß-/Kleinschreibung nicht beachtet zu werden, auch wenn sie aus einer XML-Zuordnungsdatei stammen. Es ist allerdings möglich, dass externe Quellen entsprechend ihrer eigenen Semantik die Groß-/Kleinschreibung bei Feldnamen beachten.

#### **Quick Links**

Groß-/Kleinschreibung muss nicht beachtet werden.

#### **Regelbasierte Kategorisierungsfunktion**

Bei URL-Regeln muss die Groß-/Kleinschreibung beachtet werden, aber bei Dokumentinhaltsregeln ist dies nicht der Fall.

#### **Ausgeblendete URIs und Boostdefinitionen auf Basis von URI-Mustern**

Bei URIs muss die Groß-/Kleinschreibung beachtet werden, aber bei ausgeblendeten URI-Gruppenamen ist dies nicht der Fall. Sie können nicht zwei ausgeblendete URI-Gruppenamen definieren, die sich nur durch die Groß-/Kleinschreibung unterscheiden.

#### **Feldnamen**

Groß-/Kleinschreibung muss nicht beachtet werden. Diese Regel gilt für Feldnamen in Abfragen, XML-Zuordnungen, Boostklassendefinitionen und allen anderen Verwaltungsschnittstellen, in denen Feldnamen angegeben werden. Wenn Sie in der Schnittstelle einen Feldnamen in Großbuchstaben oder mit gemischter Groß-/Kleinschreibung konfigurieren, wird er vom System möglicherweise in Kleinbuchstaben geändert, sodass er in Kleinbuchstaben angezeigt wird, wenn Sie die Konfiguration das nächste Mal anzeigen. Es ist auch möglich, dass eine externe Quelle bei Feldnamen die Groß-/Kleinschreibung beachtet.

#### **Verzeichnisse**

Groß-/Kleinschreibung muss nicht beachtet werden. Diese Regel gilt für Synonymverzeichnisse, Verzeichnisse von Stoppwörtern, Verzeichnisse von Rechtschreibvorschlägen und Verzeichnisse von Boostwörtern.

#### **Objektgruppennamen und -IDs**

Groß-/Kleinschreibung muss beachtet werden. Wenn Sie einen Objektgruppennamen oder eine Objektgruppen-ID in der Verwaltungs- oder Such-API angeben, muss diese Angabe exakt mit der Groß-/Kleinschreibung der Objektgruppe mit diesem Namen übereinstimmen. Obwohl bei Objektgruppen-IDs die Groß-/Kleinschreibung beachtet werden muss, können Sie keine Objektgruppen-IDs angeben, die sich nur durch die Groß-/Kleinschreibung unterscheiden. Diese Einschränkung gilt auch für die Namen und IDs von Crawlern und Datenquellen.

### **Suchanwendungen**

Groß-/Kleinschreibung muss beachtet werden. Sie können nicht zwei Anwendungsnamen oder -IDs angeben, die sich nur durch die Groß-/Kleinschreibung unterscheiden.



---

## Dokumentation für die Unternehmenssuche

Die Dokumentation zu OmniFind Enterprise Edition steht im PDF- oder HTML-Format zur Verfügung.

Das Installationsprogramm von OmniFind Enterprise Edition installiert automatisch die Informationszentrale, die HTML-Versionen der Dokumentation für die Unternehmenssuche umfasst. Bei einer Installation auf mehreren Servern wird die Informationszentrale auf beiden Suchservern installiert. Wenn Sie die Informationszentrale nicht installieren, wird beim Anklicken von **Hilfe** die Informationszentrale auf einer IBM Website geöffnet.

Wechseln Sie in das Verzeichnis `ES_INSTALL_ROOT/docs/ländereinstellung/pdf`, um installierte Versionen der PDF-Dokumente anzuzeigen. Wenn Sie beispielsweise Dokumente in englischer Sprache suchen, wechseln Sie in das Verzeichnis `ES_INSTALL_ROOT/docs/en_US/pdf`.

Über die Site mit der Dokumentation zu OmniFind Enterprise Edition Version 8.5 können Sie auf die PDF-Versionen der Dokumentation in allen verfügbaren Sprachen zugreifen.

Über die Unterstützungssite für OmniFind Enterprise Edition können Sie außerdem auf Produktdownloads, Fixpacks, technische Hinweise und die Informationszentrale zugreifen.

In der folgenden Tabelle ist die verfügbare Dokumentation mit Dateinamen und Speicherposition aufgeführt.

*Tabelle 33. Dokumentation für die Unternehmenssuche*

<b>Titel</b>	<b>Dateiname</b>	<b>Speicherposition</b>
Informationszentrale		<a href="http://publib.boulder.ibm.com/infocenter/discover/v8r5/">http://publib.boulder.ibm.com/infocenter/discover/v8r5/</a>
<i>Installationshandbuch für die Unternehmenssuche</i>	iiysi.pdf	ES_INSTALL_ROOT/docs/ländereinstellung/pdf/
<i>Schnelleinstieg</i> (Dieses Dokument ist in Englisch, Französisch und Japanisch auch als Hardcopy verfügbar.)	OmniFindEE850_qsg_zweibuchstabige_ländereinstellung.pdf	ES_INSTALL_ROOT/docs/ländereinstellung/pdf/
<i>Verwaltung der Unternehmenssuche</i>	iiysa.pdf	ES_INSTALL_ROOT/docs/ländereinstellung/pdf/
<i>Programming Guide and API Reference for Enterprise Search</i>	iiysp.pdf	ES_INSTALL_ROOT/docs/en_US/pdf/
<i>Fehlerbehebung und Nachrichtenreferenz</i>	iiysm.pdf	ES_INSTALL_ROOT/docs/ländereinstellung/pdf/
<i>Integration der Textanalyse</i>	iiyst.pdf	ES_INSTALL_ROOT/docs/ländereinstellung/pdf/
<i>Plug-in für Google Desktop Search</i>	iiysg.pdf	ES_INSTALL_ROOT/docs/ländereinstellung/pdf/



---

## Funktionen zur behindertengerechten Bedienung

Funktionen zur behindertengerechten Bedienung helfen Menschen mit Behinderungen, wie z. B. eingeschränkte Beweglichkeit oder eingeschränktes Sehvermögen, erfolgreich mit Softwareprodukten zu arbeiten.

IBM hat sich zum Ziel gesetzt, Produkte bereitzustellen, auf die jeder - unabhängig von Alter oder Behinderung - zugreifen kann.

### Funktionen zur behindertengerechten Bedienung

Die folgende Liste enthält die wichtigsten Funktionen zur behindertengerechten Bedienung in OmniFind Enterprise Edition:

- Ausschließliche Bedienung über die Tastatur
- Häufig von Sprachausgabeprogrammen verwendete Schnittstellen

Die Informationszentrale von OmniFind Enterprise Edition und die zugehörigen Veröffentlichungen sind für die behindertengerechte Bedienung aktiviert. Die Funktionen zur behindertengerechten Bedienung der Informationszentrale werden unter [http://publib.boulder.ibm.com/infocenter/discover/v8r5m0/topic/com.ibm.classify.nav.doc/dochome/accessibility\\_info.htm](http://publib.boulder.ibm.com/infocenter/discover/v8r5m0/topic/com.ibm.classify.nav.doc/dochome/accessibility_info.htm) beschrieben.

### Navigation über die Tastatur

Dieses Produkt verwendet Microsoft Windows-Standardnavigationstasten.

Sie können auch mithilfe der folgenden Direktaufrufe über die Tastatur im Installationsprogramm von OmniFind Enterprise Edition navigieren und die Schritte ausführen.

*Tabelle 34. Direktaufrufe über die Tastatur für das Installationsprogramm*

Aktion	Direktaufruf
Hervorheben eines Radioknopfs	Pfeiltaste
Auswählen eines Radioknopfs	Tabulatortaste
Hervorheben eines Druckknopfs	Tabulatortaste
Auswählen eines Druckknopfs	Eingabetaste
Wechseln zum nächsten oder vorhergehenden Fenster oder Ausführen eines Abbruchs	Heben Sie einen Druckknopf durch Drücken der Tabulatortaste hervor, und drücken Sie die Eingabetaste.
Inaktivieren des aktiven Fensters	Strg + Alt + Esc

### Schnittstelleninformationen

Die Benutzerschnittstellen für die Administrationskonsole, die Mustersuchanwendung und die Anpassungsfunktion für die Suchanwendung sind browserbasierte Schnittstellen, die Sie über Microsoft Internet Explorer oder Mozilla FireFox anzeigen können. Eine Liste der Direktaufrufe über die Tastatur und andere Funktionen zur behindertengerechten Bedienung von Internet Explorer bzw. FireFox finden Sie in der Onlinehilfe des jeweiligen Browsers.

## **Zugehörige Informationen zur behindertengerechten Bedienung**

Sie können die Veröffentlichungen für OmniFind Enterprise Edition mithilfe von Adobe Acrobat Reader im Adobe PDF-Format anzeigen. Die PDFs werden auf einer CD bereitgestellt, die zum Lieferumfang des Produkts gehört. Alternativ können Sie unter <http://www.ibm.com/support/docview.wss?rs=63&uid=swg27010938> auf die Informationen zugreifen.

## **IBM und behindertengerechte Bedienung**

Weitere Informationen zum Engagement von IBM hinsichtlich der behindertengerechten Bedienung finden Sie im IBM Human Ability and Accessibility Center.

---

## Glossar der Begriffe für die Unternehmenssuche

Dieses Glossar enthält Begriffe, die in den Schnittstellen für die Unternehmenssuche und in der zugehörigen Dokumentation verwendet werden.

### **Abschließendes Zeichen (Trailing Character)**

Ein Zeichen an der letzten Position in einem Wort.

### **Administrator für die Unternehmenssuche (Enterprise Search Administrator)**

Eine Verwaltungsrolle, mit der ein Benutzer das gesamte System für die Unternehmenssuche verwalten kann.

### **Allgemeine Analysestruktur (CAS - Common Analysis Structure)**

Eine Struktur, die den Inhalt und die Metadaten eines Dokuments sowie alle von einer Textanalysesteuerkomponente erstellten Analyseergebnisse speichert. Der gesamte Datenaustausch während der Dokumentanalyse wird mithilfe der allgemeinen Analysestruktur gehandhabt.

### **Allgemeine Übertragungsschicht (CCL - Common Communication Layer)**

Die Kommunikationsinfrastruktur, die die verschiedenen Komponenten (Controller, Parser, Crawler, Indexserver) von OmniFind Enterprise Edition verbindet.

### **Analyseergebnisse (Analysis Results)**

Die von Annotatoren erstellten Informationen. Analyseergebnisse werden in eine Datenstruktur geschrieben, die *allgemeine Analysestruktur* genannt wird. Analyseergebnisse, die von den benutzerdefinierten Textanalysesteuerkomponenten (Annotatoren) erstellt werden, können für die Suche verfügbar gemacht werden, indem Sie sie in den Index für die Unternehmenssuche aufnehmen.

### **Analysesteuerkomponente (Analysis Engine)**

Siehe Textanalysesteuerkomponente.

### **Annotation**

Informationen zu einer Textpassage. Eine Annotation (ergänzender Kommentar) könnte beispielsweise darauf hinweisen, dass eine kurze Textpassage für einen Unternehmensnamen steht. In UIMA (Unstructured Information Management Architecture) ist eine Annotation eine besondere Merkmalstruktur.

### **Annotator**

Eine Softwarekomponente, die bestimmte linguistische Analysetasks ausführt und Annotationen (ergänzende Kommentare) erstellt und erfasst. Der Annotator (Kommentatorfunktion) ist die Analyselogikkomponente einer Analysesteuerkomponente.

### **Annotator für reguläre Ausdrücke (Regular Expression Annotator)**

Eine Softwarekomponente, die Entitäten oder Informationseinheiten, wie z. B. Produktnummern, in einem Textdokument erkennt. Dies geschieht anhand von regulären Ausdrücken, die die genauen Muster beschreiben, nach denen im Dokumenttext gesucht wird. Entspricht einer der regulären Ausdrücke Teilen des Dokumenttexts, erstellt der Annotator für reguläre Ausdrücke die entsprechenden Annotationen mit der Übereinstimmung oder einem Teil davon. Diese Ausdrücke mit Annotationen werden dann unter Verwendung einer Indexzuordnungsdatei im Index für die Unter-

nehmenssuche oder unter Verwendung einer Datenbankzuordnungsdatei in einer JDBC-fähigen Datenbank gespeichert.

**Archiv der Verarbeitungssteuerkomponente (Processing Engine Archive)**

Eine komprimierte .pear-Archivdatei, die eine UIMA-Analysesteuerkomponente (Unstructured Information Management Architecture) sowie sämtliche Ressourcen enthält, die zu ihrer Nutzung für eine benutzerdefinierte Analyse in der Unternehmenssuche erforderlich sind.

**Aufbereitung (Tokenization)**

Die Syntaxanalyse der Eingabe in Token.

**Aufspürfunktion (Discoverer)**

Eine Crawlerfunktion, die feststellt, welche Datenquellen dem Crawler zum Abrufen von Informationen zur Verfügung stehen.

**Aus Warteschlange entfernen (Dequeue)**

Einträge aus einer Warteschlange entfernen.

**Basisannotatoren für die Unternehmenssuche (Enterprise Search Base Annotators)**

Eine Gruppe von Standardtextanalysesteuerkomponenten, die in der Unternehmenssuche verwendet werden, um Standarddokumentanalysen zu verarbeiten.

**Bediener (Operator)**

Ein Benutzer der Unternehmenssuche, der über die Berechtigung zum Beobachten, Starten und Stoppen von Prozessen auf Objektgruppenebene verfügt.

**Begriffsextraktion (Concept Extraction)**

Eine Textanalysefunktion, die signifikante Vokabularelemente (z. B. Personen, Orte oder Produkte) in Textdokumenten identifiziert und eine Liste dieser Elemente erstellt. Siehe auch Themenextraktion.

**Benutzeragent (User Agent)**

Eine Anwendung, die das Internet durchsucht und Informationen zu sich selbst auf den besuchten Sites hinterlässt. In der Unternehmenssuche ist der Web-Crawler ein Benutzeragent.

**Benutzerdefinierte Steuerkomponente für Textanalyse (Custom Text Analysis Engine)**

Eine Steuerkomponente für Textanalyse, die mithilfe des UIMA Software-Development-Kits (Unstructured Information Management Architecture SDK) erstellt wird und der Gruppe von Standardtextanalysesteuerkomponenten für die Unternehmenssuche (auch Basisannotatoren für die Unternehmenssuche genannt) hinzugefügt werden kann. Siehe auch Textanalysesteuerkomponente.

**Berechtigungs nachweis (Credential)**

Während der Authentifizierung zugewiesene detaillierte Informationen, die den Benutzer, gegebenenfalls vorhandene Gruppenzuordnungen und andere sicherheitsrelevante Identitätsattribute beschreiben. Mit Berechtigungs nachweisen lassen sich eine Vielzahl von Services ausführen, wie z. B. Berechtigung, Prüfung und Delegierung. Die Anmeldeinformationen (Benutzer-ID und Kennwort) für einen Benutzer sind Berechtigungs nachweise, mit denen der Benutzer auf ein Konto zugreifen kann.

**Bereich (Place)**

Ein virtueller Ort, der im Portal angezeigt wird und in dem sich Einzelpersonen sowie Gruppen zur Zusammenarbeit online treffen. In einem Por-

tal verfügt jeder Benutzer über einen eigenen Bereich für seine persönliche Aufgaben, und darüber hinaus haben Einzelpersonen und Gruppen Zugang zu einer Reihe gemeinsam genutzter Bereiche, die allgemein zugängliche oder eingeschränkte Bereiche sein können. Siehe auch Lotus QuickPlace-Bereich.

**Bibliothek (Library)**

Ein Systemobjekt, das anderen Objekten als Verzeichnis dient. Siehe auch Domino Document Manager-Bibliothek.

**Boolesche Suche (Boolean Search)**

Eine Suche, bei der mehrere Suchbegriffe mithilfe von Operatoren wie AND, NOT und OR kombiniert werden.

**Boostklasse (Boost Class)**

Ein Objekt, das Spezifikationen enthält, mit denen der relative Rang eines Dokuments in den Suchergebnissen beeinflusst werden kann.

**Boostwort (Boost Word)**

Ein Wort, mit dem der relative Rang eines Dokuments in den Suchergebnissen beeinflusst werden kann. Bei der Abfrageverarbeitung erhält ein Dokument, das ein Boostwort enthält, möglicherweise einen höheren oder niedrigeren Rang, je nachdem welche Bewertung für das Wort vordefiniert wurde.

**Crawler**

Ein Softwareprogramm, das Dokumente aus Datenquellen abrufen und Informationen zusammenstellt, mit denen Suchindizes erstellt werden können.

**Crawlerbereich (Crawl Space)**

Eine bestimmten Mustern (wie z. B. URL-Adressen (Uniform Resource Locators), Datenbanknamen, Dateisystempfaden, Domännennamen und IP-Adressen) entsprechende Gruppe von Quellen, die ein Crawler liest, um Elemente zum Indexieren abzurufen.

**Datenquelle (Data Source)**

Jedes Datenrepository, aus dem Dokumente abgerufen werden können, z. B. das Internet, relationale und nicht relationale Datenbanken sowie Content-Management-Systeme.

**Datenquellentyp (Data Source Type)**

Eine Zusammenfassung von Datenquellen nach dem Protokoll, mit dem auf die Daten zugegriffen wird.

**Datenspeicher (Data Store)**

Eine Datenstruktur, in der Dokumente in syntaktisch analysierter Form gespeichert werden.

**Definierter Name (Distinguished Name)**

Der Name, der einen Eintrag in einem Verzeichnis eindeutig identifiziert. Ein definierter Name besteht aus durch Kommata getrennten Attribut-Wert-Paaren. Außerdem eine Gruppe von Name-Wert-Paaren (z. B. CN=Name der Person und C=Land oder Region), die eine Entität in einem digitalen Zertifikat eindeutig identifizieren.

**Deltaindexerstellung (Delta Index Build)**

Das Hinzufügen neuer Informationen zu einem vorhandenen Index in einem System für die Unternehmenssuche. Gegensatz zu Hauptindexerstellung.

**Diakritisches Zeichen (Diacritic)**

Eine Markierung, die eine Änderung im phonetischen Wert eines Zeichens oder einer Zeichenkombination angibt.

**Dokumentobjektmodell (Document Object Model)**

Ein System, bei dem ein gegliedertes Dokument (z. B. eine XML-Datei) in Form einer Baumstruktur mit Objekten angezeigt wird, auf die über das Programm zugegriffen werden kann und die aktualisierbar sind.

**Domino Document Manager-Aktenschrankdatei (Domino Document Manager Cabinet)**

Eine Domino Document Manager-Datenbank, die zum Organisieren von Dokumenten verwendet wird. Aktenschrankdateien enthalten Domino-Datenbanken.

**Domino Document Manager-Bibliothek (Domino Document Manager Library)**

Eine Domino Document Manager-Datenbank, die den Einstiegspunkt in Domino Document Manager bildet.

**Domino Internet Inter-ORB Protocol (DIIOP)**

Eine Server-Task, die auf dem Server ausgeführt wird und mit dem Domino-Object-Request-Broker zusammenarbeitet, um eine Kommunikation zwischen Java-Applets, die mit Notes-Java-Klassen erstellt werden, und dem Domino-Server zu ermöglichen. Browserbenutzer und Domino-Server führen die Kommunikation und den Austausch von Objektdaten über DIIOP aus.

**Dynamische Rangfolge (Dynamic Ranking)**

Ein Rangfolgetyp, bei dem die Begriffe in der Abfrage in Hinblick auf die durchsuchten Dokumente analysiert werden, um die Rangfolge der Ergebnisse zu ermitteln. Siehe auch textbasierte Bewertung. Vergleiche statische Rangfolge.

**Dynamische Zusammenfassung (Dynamic Summarization)**

Eine Art der Zusammenfassung, bei der die Suchbegriffe hervorgehoben werden und die Suchergebnisse Ausdrücke enthalten, die die Konzepte des gesuchten Dokuments am besten darstellen. Vergleiche statische Zusammenfassung.

**Escapezeichen (Escape Character)**

Ein Zeichen, das eine spezielle Bedeutung für mindestens ein nachfolgendes Zeichen unterdrückt oder auswählt.

**Externe Datenquellen (External Data Source)**

Eine Datenquelle für die Föderation, die nicht von OmniFind Enterprise Edition durchsucht, syntaktisch analysiert oder indexiert wird. Das Durchsuchen von externen Datenquellen wird an die Anwendungsprogrammierschnittstelle für die Abfrage dieser Datenquellen delegiert.

**Feld (Field)**

Ein Bereich, in den eine bestimmte Kategorie von Daten oder Steuerinformationen eingegeben wird.

**Feldspezifische Suche (Fielded Search)**

Eine auf ein bestimmtes Feld beschränkte Abfrage.

**Ferner Föderator (Remote Federator)**

Ein Serverföderator, der eine Föderation für eine Gruppe durchsuchbarer Objekte ausführt.

**Föderation (Federation)**

Der Prozess des Kombinierens von Benennungssystemen, wodurch es dem

zusammengefassten System ermöglicht wird, zusammengesetzte Namen zu verarbeiten, die alle Benennungssysteme umfassen.

**Föderierte Suche (Federated Search)**

Eine Suchfunktionalität, die das Durchsuchen mehrerer Suchservices ermöglicht und eine konsolidierte Liste mit Suchergebnissen zurückgibt.

**Freiformatsuche (Free Text Search)**

Eine Suche, in der der Suchbegriff als unformatierter Text dargestellt wird.

**Hauptindexerstellung (Main Index Build)**

Der Prozess des Erstellens des gesamten Index bei der Unternehmenssuche. Gegensatz zu Deltaindexerstellung.

**Hybridsuche (Hybrid Search)**

Eine Kombination aus Boolescher Suche und Freiformatsuche.

**Identitätsmanagement (Identity Management)**

Eine Gruppe von Anwendungsprogrammierschnittstellen für die Unternehmenssuche, die den Zugriff auf sichere Daten steuern und es Benutzern ermöglichen, eine Objektgruppe zu durchsuchen, ohne eine Benutzer-ID und ein Kennwort für jedes Repository in der Objektgruppe angeben zu müssen.

**Index** Siehe Volltextindex.

**Indexierungswarteschlange (Index Queue)**

Eine Liste von zu verarbeitenden Anforderungen für die Haupt- und Deltaindexerstellung.

**Informationsextraktion (Information Extraction)**

Eine Art der Begriffsextraktion, bei der signifikante Vokabularelemente (z. B. Namen, Begriffe und Ausdrücke) in Textdokumenten automatisch erkannt werden.

**In Warteschlange stellen (Enqueue)**

Eine Nachricht oder einen Eintrag in eine Warteschlange einfügen.

**IP-Adresse (IP Address)**

Eine eindeutige Adresse für ein Gerät oder eine logische Einheit in einem Netz, das den IP-Standard verwendet.

**Java Database Connectivity (JDBC)**

Ein Industriestandard für datenbankunabhängige Konnektivität zwischen der Java-Plattform und einer großen Reihe von Datenbanken. Die JDBC-Schnittstelle bietet eine API auf Aufrufebene für SQL-Datenbankzugriffe.

**JavaScript**

Eine Web-Scripting-Sprache, die in Browsern und auf Web-Servern verwendet wird.

**JavaServer Pages (JSP)**

Eine Servertechnologie zur Scripterstellung, die es ermöglicht, Java-Code dynamisch in Webseiten (HTML-Dateien) einzubetten und diesen auszuführen, wenn die Seite bereitgestellt wird, um einem Client dynamischen Inhalt zurückzugeben.

**Java Virtual Machine (JVM)**

Softwareimplementierung eines Prozessors, die kompilierten Java-Code (Applets und Anwendungen) ausführt.

**Katakana**

Ein Zeichensatz aus Symbolen, die in einem der beiden gebräuchlichen

phonetischen Alphabete der japanischen Sprache verwendet werden. Dieser dient in erster Linie zum phonetischen Schreiben von Fremdwörtern.

**Kategoriebaum (Category Tree)**

Eine Hierarchie von Kategorien.

**Klitik (Clitic)**

Ein Wort, das syntaktisch eigenständig ist, phonetisch aber mit einem anderen Wort zusammenhängt. Ein Klitik kann mit dem Wort, an das es angelehnt ist, zusammengeschrieben oder davon getrennt geschrieben werden. Typische Beispiele für in der englischen Sprache vorkommende Klitika sind der hintere Teil einer Zusammenfügung (*wouldn't* oder *you're*).

**Komponentenpfad (Feature Path)**

Ein Pfad, über den auf den Wert eines Merkmals in einer UIMA-Merkmalstruktur (Unstructured Information Management Architecture) zugegriffen wird.

**Lemma**

Die Grundform eines Worts. Lemmata spielen vor allem in stark flektierten Sprachen wie dem Tschechischen eine wichtige Rolle.

**Lexikalische Affinität (Lexical Affinity)**

Die Beziehung von Suchbegriffen in einem Dokument, die in ihrer Bedeutung in engem Zusammenhang stehen. Mit der lexikalischen Affinität wird die Relevanz eines Ergebnisses berechnet.

**Ligatur (Ligature)**

Mindestens zwei Zeichen, die so miteinander verbunden werden, dass sie ein einzelnes Zeichen bilden. Bei ff und ffi handelt es sich beispielsweise um Zeichen, die als Ligaturen dargestellt werden können.

**Lightweight Directory Access Protocol (LDAP)**

Ein offenes Protokoll, das mithilfe von TCP/IP Zugriff auf Verzeichnisse ermöglicht, die ein X.500-Modell unterstützen, und das nicht die Ressourcenanforderungen des komplexeren X.500 Directory Access Protocol (DAP) aufweist. LDAP kann beispielsweise verwendet werden, um Personen, Organisationen und andere Ressourcen in einem Internet- oder Intranetverzeichnis zu suchen.

**Linguistische Suche (Linguistic Search)**

Eine Art der Suche, bei der ein Dokument mit auf ihre Grundformen reduzierten Begriffen durchsucht, abgerufen und indexiert (Beispiel: *Mäuse* wird als *Maus* indexiert) oder mit ihrer Grundform erweitert wird (wie bei zusammengesetzten Wörtern).

**Linkanalyse (Link Analysis)**

Ein Verfahren, das auf der Analyse von Hyperlinks zwischen Dokumenten basiert und mit dem festgestellt wird, welche Seiten in der Objektgruppe für Benutzer von Bedeutung sind.

**Lokaler Föderator (Local Federator)**

Ein Clientobjekt in einer Anwendung für die Unternehmenssuche, das über die SIAPs (Search and Index APIs) erstellt wird und es Benutzern ermöglicht, eine Gruppe von heterogenen Objektgruppen zu durchsuchen und eine einheitliche Gruppe von Suchergebnissen zu erhalten.

**Lotus QuickPlace-Bereich (Lotus QuickPlace Place)**

Ein von Lotus QuickPlace bereitgestellter Arbeitsbereich im Web, der geographisch weit verteilten Teilnehmern die Möglichkeit bietet, zusammen an

Projekten zu arbeiten und online in einem strukturierten und sicheren Arbeitsbereich miteinander zu kommunizieren.

**Lotus QuickPlace-Raum (Lotus QuickPlace Room)**

Ein partitionierter Bereich in einem Lotus QuickPlace-Bereich, der ausschließlich berechtigten Mitgliedern vorbehalten ist, die eine gemeinsame Aufgabe und die Notwendigkeit zur Zusammenarbeit verbindet.

**Merkmalstruktur (Feature Structure)**

Die zugrunde liegende Datenstruktur, die dem Ergebnis der Textanalyse entspricht. Die Merkmalstruktur hat eine Attribut-Wert-Struktur. Jede Merkmalstruktur ist von einem bestimmten Typ, wobei jeder Typ, ähnlich wie eine Java-Klasse, über eine angegebene Gruppe gültiger Merkmale oder Attribute verfügt.

**MIME-Typ (MIME Type)**

Ein Internetstandard zur Angabe des Typs eines Objekts, das über das Internet übertragen wird.

**N-Gram-Segmentierung (N-Gram Segmentation)**

Eine Analysemethode, bei der nicht wie bei der Unicode-basierten Segmentierung mit Leerzeichen Wörter durch eine Leerstelle begrenzt sind, sondern sich überlappende Folgen einer bestimmten Anzahl Zeichen als ein Wort betrachtet werden.

**No-Follow-Anweisung (No-Follow Directive)**

Eine Anweisung auf einer Webseite, die Roboter (z. B. den Web-Crawler) anweisen, den Links auf dieser Seite nicht zu folgen.

**No-Index-Anweisung (No-Index Directive)**

Eine Anweisung auf einer Webseite, die Roboter (z. B. den Web-Crawler) anweisen, den Inhalt dieser Seite nicht in den Index einzuschließen.

**Notes Remote Procedure Call (NRPC)**

Ein für die gesamte Notes-zu-Notes-Kommunikation verwendeter Kommunikationsmechanismus von Lotus Notes.

**Objektgruppe (Collection)**

Eine Gruppe von Datenquellen und Optionen für die Crawlersuche, die Syntaxanalyse, das Indexieren und das Durchsuchen dieser Datenquellen.

**Parametrische Suche (Parametric Search)**

Eine Art der Suche, bei der Objekte gesucht werden, die einen numerischen Wert oder ein numerisches Attribut (wie z. B. Datumsangaben, ganze Zahlen oder andere numerische Datentypen in einem angegebenen Bereich) enthalten.

**Parser** Ein Programm, das Dokumente interpretiert, die dem Datenspeicher für die Unternehmenssuche hinzugefügt werden. Der Parser extrahiert Informationen aus den Dokumenten und bereitet sie für Indexierungs-, Such- und Abrufvorgänge vor.

**Parser-Service (Parser Service)**

Der Service für die Unternehmenssuche, der die gesamte Syntaxanalyse für Dokumente und die Verarbeitung der Textanalysen in Dokumentobjektgruppen handhabt. Zu jedem Zeitpunkt wird mindestens ein Parser-Service ausgeführt.

**Parsertreiber (Parser Driver)**

Ein Service für die Unternehmenssuche, der dem Parser-Service Dokumente zuführt. Pro Objektgruppe gibt es einen Parsertreiber. Der Parser-

treiberservice einer Objektgruppe entspricht dem Parser der Objektgruppe in der Administrationskonsole für die Unternehmenssuche.

**Platzhalterzeichen (Masking Character)**

Ein Zeichen, das optionale Zeichen am Anfang, in der Mitte und am Ende eines Suchbegriffs darstellt. Mit Platzhalterzeichen werden normalerweise Varianten eines Begriffs in einem Index gesucht.

**Platzhalterzeichen (Wildcard Character)**

Ein Zeichen, das optionale Zeichen am Anfang, in der Mitte oder am Ende eines Suchbegriffs darstellt.

**Popularitätsrangfolge (Popular Ranking)**

Ein Rangfolgetyp, der die vorhandene Rangfolge eines Dokuments gemäß der Popularität des Dokuments erhöht.

**Portal Document Manager (PDM)**

Ermöglicht es Benutzern, ein einzelnes zentrales Dokumentrepository für elektronisches Teamwork zu verwenden. Administratoren haben die Möglichkeit, ihre Dokumente effizient zu verwalten und zu steuern, wie Benutzer mit Informationen interagieren.

**Privater Anwender der allgemeinen Analysestruktur (Common Analysis Structure Consumer)**

Ein privater Anwender führt die abschließende Verarbeitung der Analyseergebnisse durch, die in der allgemeinen Analysestruktur gespeichert sind. Beispielsweise indexiert ein privater Anwender den Inhalt der allgemeinen Analysestruktur in einer Suchmaschine oder füllt eine relationale Datenbank mit bestimmten Analyseergebnissen.

**Protokoll zum Sperren von Websitebereichen für Robots (Robots Exclusion Protocol)** Ein Protokoll mit dem Website-Administratoren durchsuchende Robots anweisen können, welche Bereiche ihrer Site vom Robot nicht besucht werden soll.

**Proxy-Server (Proxy Server)**

Ein Server, der als Mittler für HTTP-Webanforderungen von einer Anwendung oder von einem Web-Server fungiert. Ein Proxy-Server wird als Ersatzsystem für die Content-Server im Unternehmen verwendet.

**Quick Link**

Eine Zuordnung zwischen einem Uniform-Resource-Identifier (URI) und Schlüsselwörtern oder Ausdrücken.

**Rangfolge (Ranking)**

Die Zuordnung eines ganzzahligen Werts zu jedem Dokument in den Suchergebnissen einer Abfrage. Die Reihenfolge der Dokumente in den Suchergebnissen basiert auf der Relevanz für die Abfrage. Eine höhere Einstufung in der Rangfolge bedeutet eine größere Übereinstimmung. Siehe auch dynamische Rangfolge und statische Rangfolge.

**Raum (Room)**

Ein Programm, mit dem Benutzer die Möglichkeit haben, von anderen Benutzern zu lesende Dokumente zu erstellen, auf Kommentare anderer Personen zu antworten und den Status sowie den Endtermin eines Projekts anzuzeigen. Außerdem können die Benutzer mit anderen im selben Raum befindlichen Personen chatten. Siehe auch Lotus QuickPlace-Raum.

**Reduktion auf Grundform (Lemmatization)**

Ein Prozess, der die Grundform und verschiedene grammatische Formen eines Worts angibt. Wenn Sie beispielsweise nach 'Maus' suchen, werden

auch Dokumente gefunden, die das Wort 'Mäuse' enthalten; bei einer Suche nach dem Wort 'gehen' werden auch Dokumente gefunden, die die Wörter 'geht', 'ging' oder 'gegangen' enthalten.

#### **Regelbasierte Kategorie (Rule-Based Category)**

Durch Regeln erstellte Kategorien, die angeben, welche Dokumente welchen Kategorien zugeordnet werden. Sie können beispielsweise Regeln definieren, mit denen Dokumente, die bestimmte Wörter enthalten oder nicht enthalten oder die einem URI-Muster (Uniform-Resource-Identifier) entsprechen, bestimmten Kategorien zugeordnet werden.

#### **Rohdatenspeicher (Raw Data Store)**

Eine Datenstruktur, in der durchsuchte Dokumente gespeichert werden, bevor Sie an den Parser gesendet werden. Crawler schreiben in den Rohdatenspeicher, und der Parser liest aus dem Rohdatenspeicher. Wenn Dokumente syntaktisch analysiert wurden, werden sie aus dem Rohdatenspeicher entfernt. Der Rohdatenspeicher darf nicht mit dem Datenspeicher verwechselt werden.

#### **Schlüsseldatenbankdatei (Key Database File)**

Siehe Schlüsselring.

#### **Schlüsselring (Key Ring)**

Im Bereich der IT-Sicherheit eine Datei, die öffentliche Schlüssel, private Schlüssel, Trusted Roots und Zertifikate enthält. Siehe auch Schlüssel-speicherdatei.

#### **Schlüsselspeicherdatei (Keystore File)**

Ein Schlüsselring mit öffentlichen Schlüsseln, die als Unterzeichner-zertifikate gespeichert werden, und mit privaten Schlüsseln, die in persönlichen Zertifikaten gespeichert werden.

#### **Secure Sockets Layer (SSL)**

Ein Sicherheitsprotokoll zur Gewährleistung von Datenschutz bei der Kommunikation. Mit SSL können Client-/Serveranwendungen in einer Weise kommunizieren, die das Ausspionieren, die Manipulation von Daten während der Übertragung und das Fälschen von Nachrichten verhindert.

#### **Segmentierung (Segmentation)**

Die Unterteilung von Text in bestimmte lexikalische Einheiten. Die nicht auf Wörterbüchern basierende Verarbeitung schließt Leerzeichen und N-Gram-Segmentierung ein, während die auf Wörterbüchern basierende Unterstützung die Wort-, Satz- und Absatzsegmentierung sowie eine Reduktion auf die Grundform umfasst.

#### **Seite mit detaillierten Fehlerhinweisen (Soft Error Page)**

Ein Typ einer Webseite, die Informationen dazu bereitstellt, weshalb die angeforderte Webseite nicht zurückgegeben werden kann. Beispielsweise kann der HTTP-Server statt eines einfachen Statuscodes eine Seite zurückgeben, auf der der Statuscode detailliert beschrieben wird.

#### **Seite mit Einstiegspunktliste (Seed List Page)**

Eine XML-Seite in WebSphere Portal, die Links zu den in einem Portal verfügbaren Seiten enthält. Crawler verwenden die Einstiegspunktliste, um die zu durchsuchenden Dokumente zu identifizieren. Die Seite mit der Einstiegspunktliste enthält außerdem Metadaten, die mit den durchsuchten Dokumenten im Index für die Unternehmenssuche gespeichert werden.

#### **Semantische Suche (Semantic Search)**

Ein Typ der Schlüsselwortsuche, der die linguistische Analyse und die Kontextanalyse umfasst. Siehe auch Textanalyse.

**Servlet**

Ein Java-Programm, das auf einem Web-Server ausgeführt wird und die Funktionalität des Servers erweitert, indem es aufgrund von Web-Client-Anforderungen dynamischen Inhalt generiert. Servlets werden gewöhnlich verwendet, für Datenbanken eine Verbindung zum Web herzustellen.

**Shingle**

Eine Zeichenfolge mit fortlaufenden Token (Wörter), die aus einem Satz stammen. Beispielsweise können dem Satz "Dies ist ein sehr kurzer Satz." die folgenden 3-Wort-Shingles (bzw. Trigramme) entnommen werden:

Dies ist ein  
ist ein sehr  
ein sehr kurzer  
sehr kurzer Satz

Shingles können in der statistischen Linguistik verwendet werden. Wenn beispielsweise zwei verschiedene Text viele gemeinsame Shingles aufweisen, besteht wahrscheinlich ein gewisser Zusammenhang zwischen den Texten.

**Sicherheitstoken (Security Token)**

Informationen zu Identität und Sicherheit, mit denen der Zugriff auf Dokumente in einer Objektgruppe berechtigt wird. Verschiedene Datenquellentypen unterstützen verschiedene Sicherheitstokentypen. Beispiele: Benutzerrollen, Benutzer-IDs, Gruppen-IDs und andere Informationen für die Datenzugriffssteuerung.

**Spracherkennung (Language Identification)**

Eine Suchfunktion der Unternehmenssuche, die die Sprache eines Dokuments bestimmt.

**Stammbildung (Stemming)**

Siehe Wortstammbildung.

**Start-URL (Start Uniform Resource Locator)**

Der Ausgangspunkt einer Crawlersuche.

**Statische Rangfolge (Static Ranking)**

Ein Rangfolgetyp, bei dem Faktoren der eingestuften Dokumente (z. B. das Datum, die Anzahl der Links, die auf das Dokument verweisen, usw.) den Rang erhöhen. Vergleiche dynamische Rangfolge.

**Statische Zusammenfassung (Static Summarization)**

Ein Zusammenfassungstyp, bei dem die Suchergebnisse eine angegebene, gespeicherte Zusammenfassung aus dem Dokument enthalten. Vergleiche dynamische Zusammenfassung.

**Steuerkomponente für Textanalyse (Text Analysis Engine)**

Eine Softwarekomponente, deren Aufgabe es ist, Kontext und semantischen Inhalt in Text aufzufinden und darzustellen.

**Stoppwortentfernung (Stop Word Removal)**

Das Entfernen von Stoppwörtern aus der Abfrage, damit allgemeine Wörter ignoriert und auf diese Weise gezieltere Ergebnisse zurückgegeben werden.

**Stoppwort (Stop Word)**

Ein häufig verwendetes Wort (z. B. *der*, *ein*, *und*), das von einer Suchanwendung ignoriert wird.

**Suchanwendung (Search Application)**

Ein Programm bei der Unternehmenssuche, das Abfragen verarbeitet, den Index durchsucht, die Suchergebnisse zurückgibt und die Quelldokumente abrufen.

**Suchcache (Search Cache)**

Ein Puffer, der die Daten und Ergebnisse vorangegangener Suchanforderungen enthält.

**Suche mit Begriffsgewichtung (Weighted Term Search)**

Eine Abfrage, bei der bestimmten Begriffen größere Bedeutung beigemessen wird.

**Suche nach grober Übereinstimmung (Fuzzy Search)**

Eine Suche, bei der Wörter zurückgegeben werden, deren Schreibweise der des Suchbegriffs ähnlich ist.

**Suchergebnisse (Search Results)**

Eine Liste der Dokumente, die der Suchanforderung entsprechen.

**Suchindexdateien (Search Index Files)**

Gruppe von Dateien, in der ein Index in der Suchmaschine gespeichert wird.

**Suchmaschine (Search Engine)**

Ein Programm, das eine Suchanforderung annimmt und eine Dokumentenliste an den Benutzer zurückgibt.

**Synonymverzeichnis (Synonym Dictionary)**

Ein Wörterverzeichnis, das es Benutzern ermöglicht, eine Objektgruppe auch nach Synonymen ihrer Abfragebegriffe zu durchsuchen.

**Sprache XPath (XML Path) (XML Path Language (XPath))**

Eine Sprache, die konzipiert wurde, um Teile der XML-Quelldaten für die Verwendung mit XML-Technologien wie z. B. XSLT-, XQuery- und XML-Parsern eindeutig anzugeben oder zu adressieren. Die Sprache XPath ist ein Standard des World Wide Web Consortium.

**Taxonomie (Taxonomy)**

Eine auf Ähnlichkeiten basierende Klassifikation von Objekten zu Gruppen. In der Unternehmenssuche fasst eine Taxonomie Daten zu Kategorien und Unterkategorien zusammen. Siehe auch Kategoriebaum.

**Textanalyse (Text Analysis)**

Das Extrahieren der Semantik und anderer Informationen aus Text, um die Abrufbarkeit von Daten in einer Objektgruppe zu verbessern. Siehe auch Semantische Suche.

**Textbasierte Bewertung (Text-Based Scoring)**

Die Zuordnung eines ganzzahligen Werts zu einem Dokument, der die Relevanz des Dokuments in Bezug auf die Abfragebegriffe anzeigt. Ein hoher ganzzahliger Wert zeigt eine große Übereinstimmung mit der Abfrage an. Siehe auch dynamische Rangfolge.

**Text mit freiem Format (Free-Form Text)**

Unstrukturierter Text, der aus Worten oder Sätzen besteht.

**Textsegmentierung (Text Segmentation)**

Siehe Segmentierung.

**Themenextraktion (Theme Extraction)**

Eine Art der Begriffsextraktion, bei der signifikante Vokabularelemente in

Textdokumenten automatisch erkannt werden, um das Thema eines Dokuments zu extrahieren. Siehe auch Begriffsextraktion.

**Token** Die Basistexteinheiten, die von der Unternehmenssuche indexiert werden. Token können aus den Wörtern einer Sprache oder aus anderen Texteinheiten bestehen, die sich für das Indexieren eignen.

**Tokenizer**

Ein Textsegmentierungsprogramm, das Text überprüft und ermittelt, wann und ob eine Zeichenfolge als Token erkannt werden kann.

**Typsystem (Type System)**

Das Typsystem definiert die Objekttypen (Komponentenstrukturen), die eine Textanalysesteuerkomponente in einem Dokument erkennen kann. Das Typsystem definiert alle möglichen Komponentenstrukturen in Form von Typen und Komponenten. Sie können eine beliebige Anzahl unterschiedlicher Typen in einem Typsystem definieren. Ein Typsystem ist domänen- und anwendungsspezifisch.

**Überwachungsbeauftragter (Monitor)**

Ein Benutzer der Unternehmenssuche, der über die Berechtigung zum Beobachten von Prozessen auf Objektgruppenebene verfügt.

**Unicode-basierte Segmentierung mit Leerzeichen (Unicode-Based White Space Segmentation)**

Ein Aufbereitungsverfahren, bei dem mittels Unicode-Zeichenmerkmalen zwischen Token und Trennzeichen unterschieden wird.

**Uniform-Resource-Identifier (URI)**

Eine kompakte Zeichenfolge, die eine abstrakte oder physische Ressource angibt.

**Uniform-Resource-Locator (URL)**

Die eindeutige Adresse einer Informationsressource, auf die in einem Netz wie dem Internet zugegriffen werden kann. Die URL enthält den abgekürzten Namen des Protokolls, mit dem auf die Informationsquelle zugegriffen wird, sowie die Informationen, mit denen das Protokoll die Informationsquelle lokalisiert.

**Unstructured Information Management Architecture (UIMA)**

Eine IBM Architektur, die ein Framework zur Implementierung von Systemen zur Analyse unstrukturierter Daten definiert.

**Verknüpfte Suche (Proximity Search)**

Eine Textsuche, die ein Ergebnis zurückgibt, wenn mindestens zwei übereinstimmende Begriffe in einem bestimmten Abstand voneinander auftreten, z. B. im selben Satz oder Absatz.

**Verwaltungsrolle (Administrative Role)**

Eine Klassifizierung eines Benutzers, die die Zugriffsberechtigungen eines Benutzers festlegt.

**Volltextindex (Full-Text Index)**

Eine Datenstruktur, die auf Datenelemente verweist, damit eine Suche Dokumente finden kann, die die Abfragebegriffe enthalten.

**Web-Crawler (Web Crawler)**

Ein Crawlertyp, der das World Wide Web durchsucht, indem er ein Webdokument abrufen und den Links in diesem Dokument folgt.

**Wortstambildung (Word Stemming)**

Ein Prozess der linguistischen Normalisierung, in dem die Varianten eines

Worts auf eine allgemeine Form reduziert werden. Beispielsweise werden Wörter wie *Speicherung*, *speichernd* und *gespeichert* zu *speich-* reduziert.

**Zeichennormalisierung (Character Normalization)**

Ein Prozess, bei dem die Varianten eines Zeichens (z. B. Großschreibung und diakritische Zeichen) auf eine gemeinsame Form reduziert werden.

**Zeilenvorschubzeichen (Newline Character)**

Ein Steuerzeichen, das bewirkt, dass die Druck- oder Anzeigeposition einer Zeile nach unten versetzt wird.

**Zertifikat (Certificate)**

Im Bereich der IT-Sicherheit ein digitales Dokument, mit dem der Identität des Zertifikatinhabers ein öffentlicher Schlüssel angefügt wird, um eine Authentifizierung des Zertifikatinhabers zu ermöglichen. Zertifikate werden von einer Zertifizierungsstelle ausgestellt und digital signiert.

**Zertifizierungsstelle (Certificate Authority)**

Ein anerkanntes Fremdunternehmen, das digitale Zertifikate ausstellt, die verwendet werden, um digitale Signaturen und öffentlich/private Schlüsselpaare zu erstellen. Die Zertifizierungsstelle gewährleistet die Identität der Einzelpersonen, denen das eindeutige Zertifikat erteilt wird.

**Zugriffssteuerungsliste (ACL - Access Control List)**

Im Bereich der IT-Sicherheit eine Liste, die einem Objekt zugeordnet ist und in der alle Personen, die auf das Objekt zugreifen können, und deren Zugriffsberechtigungen angegeben werden.

**Zusammenfassung (Summarization)**

Das Einfügen von nicht redundanten Sätzen in Suchergebnissen, die den Inhalt eines Dokuments kurz beschreiben. Siehe auch dynamische Zusammenfassung und statische Zusammenfassung.



---

## Bemerkungen und Marken

---

### Bemerkungen

Die vorliegenden Informationen wurden für Produkte und Services entwickelt, die auf dem deutschen Markt angeboten werden.

Möglicherweise bietet IBM die in dieser Dokumentation beschriebenen Produkte, Services oder Funktionen in anderen Ländern nicht an. Informationen über die gegenwärtig im jeweiligen Land verfügbaren Produkte und Services sind beim IBM Ansprechpartner erhältlich. Hinweise auf IBM Lizenzprogramme oder andere IBM Produkte bedeuten nicht, dass nur Programme, Produkte oder Services von IBM verwendet werden können. An Stelle der IBM Produkte, Programme oder Services können auch andere ihnen äquivalente Produkte, Programme oder Services verwendet werden, solange diese keine gewerblichen oder anderen Schutzrechte der IBM verletzen. Die Verantwortung für den Betrieb von Fremdprodukten, Fremdprogrammen und Fremdservices liegt beim Kunden.

Für in diesem Handbuch beschriebene Erzeugnisse und Verfahren kann es IBM Patente oder Patentanmeldungen geben. Mit der Auslieferung dieses Handbuchs ist keine Lizenzierung dieser Patente verbunden. Lizenzanforderungen sind schriftlich an folgende Adresse zu richten (Anfragen an diese Adresse müssen auf Englisch formuliert werden):

IBM Director of Licensing  
IBM Europe, Middle East & Africa  
Tour Descartes  
2, avenue Gambetta  
92066 Paris La Defense  
France

Trotz sorgfältiger Bearbeitung können technische Ungenauigkeiten oder Druckfehler in dieser Veröffentlichung nicht ausgeschlossen werden. Die Angaben in diesem Handbuch werden in regelmäßigen Zeitabständen aktualisiert. Die Änderungen werden in Überarbeitungen oder in Technical News Letters (TNLs) bekannt gegeben. IBM kann ohne weitere Mitteilung jederzeit Verbesserungen und/oder Änderungen an den in dieser Veröffentlichung beschriebenen Produkten und/oder Programmen vornehmen.

Verweise in diesen Informationen auf Websites anderer Anbieter dienen lediglich als Benutzerinformationen und stellen keinerlei Billigung des Inhalts dieser Websites dar. Das über diese Websites verfügbare Material ist nicht Bestandteil des Materials für dieses IBM Produkt. Die Verwendung dieser Websites geschieht auf eigene Verantwortung.

Werden an IBM Informationen eingesandt, können diese beliebig verwendet werden, ohne dass eine Verpflichtung gegenüber dem Einsender entsteht.

Lizenznehmer des Programms, die Informationen zu diesem Produkt wünschen mit der Zielsetzung: (i) den Austausch von Informationen zwischen unabhängig voneinander erstellten Programmen und anderen Programmen (einschließlich des vorliegenden Programms) sowie (ii) die gemeinsame Nutzung der ausgetauschten Informationen zu ermöglichen, wenden sich an folgende Adresse:

IBM Corporation  
J46A/G4  
555 Bailey Avenue  
San Jose, CA 95141-1003  
USA

Die Bereitstellung dieser Informationen kann unter Umständen von bestimmten Bedingungen - in einigen Fällen auch von der Zahlung einer Gebühr - abhängig sein.

Die Lieferung des im Dokument aufgeführten Lizenzprogramms sowie des zugehörigen Lizenzmaterials erfolgt auf der Basis der IBM Rahmenvereinbarung bzw. der Allgemeinen Geschäftsbedingungen von IBM, der IBM Internationalen Nutzungsbedingungen für Programmpakete oder einer äquivalenten Vereinbarung.

Alle in diesem Dokument enthaltenen Leistungsdaten stammen aus einer kontrollierten Umgebung. Die Ergebnisse, die in anderen Betriebsumgebungen erzielt werden, können daher erheblich von den hier erzielten Ergebnissen abweichen. Einige Daten stammen möglicherweise von Systemen, deren Entwicklung noch nicht abgeschlossen ist. Eine Gewährleistung, dass diese Daten auch in allgemein verfügbaren Systemen erzielt werden, kann nicht gegeben werden. Darüber hinaus wurden einige Daten unter Umständen durch Extrapolation berechnet. Die tatsächlichen Ergebnisse können abweichen. Benutzer dieses Dokuments sollten die entsprechenden Daten in ihrer spezifischen Umgebung prüfen.

Alle Informationen zu Produkten anderer Anbieter stammen von den Anbietern der aufgeführten Produkte, deren veröffentlichten Ankündigungen oder anderen allgemein verfügbaren Quellen. IBM hat diese Produkte nicht getestet und kann daher keine Aussagen zu Leistung, Kompatibilität oder anderen Merkmalen machen. Fragen zu den Leistungsmerkmalen von Produkten anderer Anbieter sind an den jeweiligen Anbieter zu richten.

Die oben genannten Erklärungen bezüglich der Produktstrategien und Absichtserklärungen von IBM stellen die gegenwärtige Absicht der IBM dar, unterliegen Änderungen oder können zurückgenommen werden, und repräsentieren nur die Ziele der IBM.

Alle von IBM angegebenen Preise sind empfohlene Richtpreise und können jederzeit ohne weitere Mitteilung geändert werden. Händlerpreise können u. U. von den hier genannten Preisen abweichen.

Diese Veröffentlichung dient nur zu Planungszwecken. Die in dieser Veröffentlichung enthaltenen Informationen können geändert werden, bevor die beschriebenen Produkte verfügbar sind.

Diese Veröffentlichung enthält Beispiele für Daten und Berichte des alltäglichen Geschäftsablaufes. Sie sollen nur die Funktionen des Lizenzprogrammes illustrieren; sie können Namen von Personen, Firmen, Marken oder Produkten enthalten. Alle diese Namen sind frei erfunden; Ähnlichkeiten mit tatsächlichen Namen und Adressen sind rein zufällig.

#### COPYRIGHTLIZENZ:

Diese Veröffentlichung enthält Musteranwendungsprogramme, die in Quellsprache geschrieben sind. Sie dürfen diese Musterprogramme kostenlos kopieren, ändern und verteilen, wenn dies zu dem Zweck geschieht, Anwendungs-

programme zu entwickeln, zu verwenden, zu vermarkten oder zu verteilen, die mit der Anwendungsprogrammierschnittstelle konform sind, für die diese Musterprogramme geschrieben werden. Diese Beispiele wurden nicht unter allen denkbaren Bedingungen getestet. Daher kann IBM die Zuverlässigkeit, Wartungsfreundlichkeit oder Funktion dieser Programme weder zusagen noch gewährleisten.

Kopien oder Teile der Musterprogramme bzw. daraus abgeleiteter Code müssen folgenden Copyrightvermerk beinhalten:

© (Name Ihrer Firma) (Jahr). Teile des vorliegenden Codes wurden aus Musterprogrammen der IBM Corp. abgeleitet. © Copyright IBM Corp. \_Jahr/Jahre angeben\_. Alle Rechte vorbehalten.

Für Teile des vorliegenden Produkts gilt Folgendes:

- Oracle® Outside In Content Access, Copyright © 1992, 2008, Oracle. Alle Rechte vorbehalten.
- IBM XSLT-Prozessor Lizenziertes Material - Eigentum der IBM © Copyright IBM Corporation, 1999-2008. Alle Rechte vorbehalten.

## Marken

Informationen zu IBM Marken finden Sie unter <http://www.ibm.com/legal/copytrade.shtml>.

Die folgenden Namen sind Marken oder eingetragene Marken anderer Unternehmen:

Adobe, Acrobat, Portable Document Format (PDF), PostScript und alle auf Adobe basierenden Marken sind Marken oder eingetragene Marken der Adobe Systems Incorporated in den USA und/oder anderen Ländern.

Intel, das Intel-Logo, Intel Inside, das Intel Inside-Logo, Intel Centrino, das Intel Centrino-Logo, Celeron, Intel Xeon, Intel SpeedStep, Itanium und Pentium sind Marken oder eingetragene Marken der Intel Corporation oder deren Tochtergesellschaften in den USA oder anderen Ländern.

Java und alle auf Java basierenden Marken und Logos sind Marken von Sun Microsystems, Inc. in den USA und/oder anderen Ländern.

Linux ist eine Marke von Linus Torvalds in den USA und/oder anderen Ländern.

Microsoft, Windows, Windows NT und das Windows-Logo sind Marken der Microsoft Corporation in den USA und/oder anderen Ländern.

UNIX ist eine eingetragene Marke von The Open Group in den USA und anderen Ländern.

Weitere Unternehmens-, Produkt- oder Servicenamen können Marken anderer Hersteller sein.



---

# Index

## A

Abfrageerweiterung  
  Auswirkung auf Indexerstellungzeit 202  
  Auswirkung auf Indexgröße 202  
  Beschreibung 200  
Abfragen mit vielen Ergebnissen  
  Beschreibung 232  
  Standardboostfaktoren 235  
Abfragen mit wenigen Ergebnissen  
  Beschreibung 232  
  Standardboostfaktoren 235  
Abfrageprotokollkonfiguration 359  
Abfrageprüfung 287  
Administrationskonsole  
  anmelden 19  
  Beschreibung 9  
  Schnittstelle 15  
  Taskübersicht 15  
Administrator für die Unternehmenssuche  
  Beschreibung 280  
  Kennwort auf einem Einzelserver ändern 20  
  Kennwort auf mehreren Servern ändern 21  
  Rollenkonfiguration 281  
Administratorkennwort  
  auf einem Einzelserver ändern 20  
  auf mehreren Servern ändern 21  
AdminLinkBarInclude.jsp, Datei 378  
Adressenregeln für Web-Crawler 101  
AIX, Betriebssystem  
  Content Edition-Crawler, Konfiguration 50  
  DB2 Content Manager-Crawler, Konfiguration 63  
  DB2-Crawler, Konfiguration 54  
  Domino Document Manager-Crawler, Konfiguration 83  
  Ereignisveröffentlichung, Konfiguration 54  
  Notes-Crawler, Konfiguration 83  
  QuickPlace-Crawler, Konfiguration 83  
Aktive Websites überwachen 331, 333  
Aktivieren der Indexzeitpläne 189  
Alerts  
  Beschreibung 351  
  durchsuchte Dokumente 352  
  E-Mail-Empfang 357  
  E-Mail-Optionen 352, 354  
  freier Speicherbereich auf Servern 354  
  Indexbegrenzungen 327  
  indexierte Dokumente 352  
  Objektgruppenebene 327, 352  
  SMTP-Server, Konfiguration 356  
  Suchantwortzeit 352  
  Systemebene 354  
Alerts zu freiem Speicherbereich 354

Allgemeine Analysestrukturen  
  allgemeinen Datenbanken zuordnen 160  
  Beschreibung 154  
  dem Index zuordnen 159  
  XML-Elemente zuordnen 158  
Alphabetische Schriften 181  
Ankertextanalyse  
  Beschreibung 277  
  Dokumente indexieren 296  
  globale Analyse 295  
  Objektgruppensicherheit 295  
Anmelden an der Administrationskonsole 19  
Annotatoren 154  
Anpassungsfunktion für die Suchanwendung  
  config.properties, Datei 261  
  starten 261  
Antwortzeitprotokoll überwachen 343  
Anzeigen  
  gelöschte Dokumentprotokoll-dateien 349  
  Protokolldateien 360  
  URI-Details 328  
APIs  
  Beschreibung 11  
  Search and Index 11, 239  
Archivdateien  
  Crawlersuche 129  
  unterstützte Formate 129  
  URI-Formate 130  
ASCII-Parser 170  
Ausgeblendete Suchergebnisse  
  Beschreibung 205  
  konfigurieren 206  
  Sicherheitseinschränkungen 317  
Ausgeblendete URIs  
  Beschreibung 205  
  konfigurieren 206  
  Sicherheitseinschränkungen 317  
Authentifizierung  
  Beschreibung 280  
  für Unternehmensanwendungen inaktivieren 315  
Automatische Erkennung  
  Codepages 183  
  Sprachen 182

## B

Backup-Scripts  
  ausführen 364  
  Beschreibung 363  
Balkendiagramme  
  häufigste Ergebnisse 265  
  Java-Klassen für die häufigsten Ergebnisse 265  
banner.jspf, Datei 385  
banner\_searchControl.jspf, Datei 385

## Bearbeiten

Anwendungen für Datenempfängsfunktion 125  
Crawlerbereiche 44  
Crawlermerkmale 44  
Objektgruppen 37  
Suchanwendungsmerkmale 242, 260  
Befehle, Unternehmenssuche 401  
Benutzeragenten 97  
Benutzerdefinierte Textanalyse  
  allgemeine Analysestruktur dem Index zuordnen 159  
  allgemeine Analysestruktur einer relationalen Datenbank zuordnen 160  
  Analyseergebnisse dem Index zuordnen 159  
  Analyseergebnisse einer relationalen Datenbank zuordnen 160  
  Beschreibung 154  
  Textanalysesteuerkomponenten 156, 157  
  XML-Elemente zuordnen 158  
Benutzerprofile  
  Beschreibung 291  
  Konfiguration 294  
Berechtigung, Beschreibung 280  
Bereiche  
  Beschreibung 203  
  durchsuchen 203  
  erstellen 204  
  URI-Formate 130  
  URI-Muster 203, 204  
Boostfaktoren  
  Boostklassenkonfiguration 232, 235  
  für Boostklassen 233, 235  
  für URI-Muster 230, 231  
  für Verzeichnisse von Boostwörtern 227  
Boostklassen  
  Abfragen mit vielen Ergebnissen 232, 235  
  Abfragen mit wenigen Ergebnissen 232, 235  
  Beschreibung 232  
  Erkennung von doppelten Dokumenten 232  
  Felder zuordnen 233  
  Konfiguration 233, 235  
  Standardwerte 235  
bos.iocp.rte, Modul 88

## C

ccl.properties, Datei 26  
CCLServer\_date.log, Datei 25  
Chinesisch  
  N-Gram-Segmentierung 184  
  Zeilenvorschubzeichen entfernen 185  
Cluster  
  WebSphere Portal 389

- Codepages
    - automatische Erkennung 183
    - unterstützt 183
  - config.properties, Datei 300, 301
    - anpassen 261
    - bearbeiten 260
    - klonen 263
    - Merkmalbeschreibungen 242
  - Content Edition-Crawler
    - Direktmodus 48
    - in der Solaris-Betriebsumgebung konfigurieren 50
    - Konfiguration 46
    - Servermodus 49
    - unter dem Betriebssystem AIX konfigurieren 50
    - unter dem Betriebssystem Linux konfigurieren 50
    - unter Windows konfigurieren 51
    - URI-Formate 130
  - Cookies für Web-Crawler-Suche
    - Beschreibung 112
    - Format 112
    - konfigurieren 113
  - cookies.ini, Datei
    - Beschreibung 112
    - Format 112
    - konfigurieren 113
  - crawl.rules, Datei 114
  - Crawler
    - Anfangswerte 43
    - Anwendungen für Datenempfangsfunktion 125
    - Archivdateien 129
    - Basiswerte 41
    - Beschreibung 4
    - Content Edition 46, 48, 49
    - Crawlerbereiche bearbeiten 44
    - Crawlermerkmale bearbeiten 44
    - Crawlertypen kombinieren 41
    - DB2 52
    - DB2 Content Manager 62
    - Domino Document Manager 66
    - Einstiegspunktliste 93
    - erstellen 43
    - Exchange Server 69, 307
    - JDBC-Datenbank 70, 72, 74
    - Konfigurationsüberblick 41
    - löschen 45
    - NNTP 78
    - Notes 79, 82
    - Plug-ins 127
    - QuickPlace 89
    - Sicherheit auf Dokumentenebene 284
    - Sicherheit auf Dokumentenebene aktivieren 41
    - Standardmigrationseinstellungen 399
    - Systemstatus 330
    - terminieren 41, 46
    - überwachen 330
    - UNIX-Dateisystem 95
    - Unterstützung für externe 125
    - URI-Formate 130
    - Web 96
    - Web Content Management 118, 122
    - WebSphere Portal 120, 122
    - Windows-Dateisystem 123
  - Crawler, Beschreibung der Seite 15
  - Crawler für das Windows-Dateisystem
    - Konfiguration 123
    - Konfiguration der Sicherheit auf Dokumentenebene 311
    - URI-Formate 130
  - Crawler für Einstiegspunktlisten
    - Integration in WebSphere Portal 373, 387
    - Konfiguration 93
    - URI-Formate 130
  - Crawler für UNIX-Dateisysteme
    - Konfiguration 95
    - URI-Formate 130
  - Crawler-Plug-ins 127
  - crawler\_rdb\_plugin.xml, Datei 74
  - Crawler-Server
    - starten 319, 330
    - stoppen 321, 330
  - Crawlerbereich
    - Alerts 352
    - bearbeiten 44
    - Beschreibung 4
    - Web-Crawler-Konfiguration 101
  - Crawlermerkmale
    - bearbeiten 44
    - Beschreibung 4
  - Crawlertypen
    - Basiswerte 41
    - in Objektgruppe kombinieren 41
  - Crawlerverlaufsbericht
    - Beschreibung 331
    - erstellen 334
    - HTTP-Statuscodebericht 334
    - Sitebericht 334
- ## D
- Dateierweiterungen
    - aus Web-Crawler-Bereichen ausschließen 101
    - unterstützte Stellt-Parser 176
    - von Objektgruppenparsern unterstützt 168, 169
    - von Stellt-Parsern unterstützt 173
  - Datenempfangsfunktion
    - erneut starten 125, 346
    - konfigurieren 125
    - überwachen 346
  - Datenfluss, System für die Unternehmensuche 13
  - Datenquellentypen
    - CA-Datacom-Datenbanken 52
    - Content Edition-Repositorys 46, 48, 49
    - DB2 Content Manager-Elementtypen 62
    - DB2-Datenbanken 52, 70, 273
    - DB2 für iSeries-Datenbanken 52
    - DB2 für z/OS 52
    - Domino Document Manager-Datenbanken 66
    - Exchange Server, allgemein zugängliche Ordner 69
    - IMS-Datenbanken 52
    - Informix-Datenbanken 52
    - JDBC-Datenbanken 70, 72, 74, 273
  - Datenquellentypen (*Forts.*)
    - Lotus Quickr für Domino 89
    - Lotus Quickr für WebSphere Portal 93
    - NNTP-Newsgruppen 78
    - Notes-Datenbanken 79, 82
    - Oracle-Datenbanken 52, 70, 273
    - QuickPlace-Datenbanken 89
    - relationale Datenbanken 52
    - Software AG Adabas-Datenbanken 52
    - SQL Server-Datenbanken 52, 70
    - Sybase-Datenbanken 52
    - UNIX-Dateisysteme 95
    - Unterstützung für externe 2, 11
    - von der Unternehmensuche unterstützt 2
    - VSAM-Datenbanken 52
    - Web Content Management-Sites 118
    - Websites 96
    - WebSphere-Portal-Sites 120
    - Windows-Dateisysteme 123
  - Datumsangaben für durchsuchte Dokumente
    - für Web-Crawler konfigurieren 117
  - DB2 Content Manager-Crawler
    - in der Solaris-Betriebsumgebung konfigurieren 63
    - Konfiguration 62
    - unter dem Betriebssystem AIX konfigurieren 63
    - unter dem Betriebssystem Linux konfigurieren 63
    - unter Windows konfigurieren 65
    - URI-Formate 130
  - DB2-Crawler
    - Ereignisveröffentlichung 52
    - Ereignisveröffentlichung, Konfiguration 54, 55
    - Konfiguration 52
    - unter AIX konfigurieren 54
    - unter Linux konfigurieren 54
    - unter Solaris konfigurieren 54
    - unter Windows konfigurieren 55
    - URI-Formate 130
    - WebSphere II Classic Federation 61
    - WebSphere II Event Publisher Edition-Konfiguration 56
    - WebSphere MQ-Installation unter AIX 54
    - WebSphere MQ-Installation unter Linux 54
    - WebSphere MQ-Installation unter Solaris 54
    - WebSphere MQ-Installation unter Windows 55
    - WebSphere MQ-Konfiguration 59
  - DB2-Datenbanken
    - als externe Quelle zugreifen 273
    - mit DB2-Crawlern zugreifen 52
    - mit JDBC-Datenbankcrawlern zugreifen 70
  - Default.jsp, Datei 378, 385
  - Deltaindizes
    - Änderungen feststellen 192
    - Beschreibung 7, 187

- Deltaindices (*Forst.*)
    - gleichzeitig ablaufende Erstellungen 190
    - terminieren 188
  - DIIO-Protokoll, Crawler-konfiguration 87
  - Directory Assistance, Konfiguration 310
  - Direktmodus, Content Edition-Repositories 48
  - Discovery 4
  - distinctRecentQueryCheck, Parameter 345
  - Dokumentation
    - HTML 437
    - PDF 437
    - suchen 437
  - Dokumentinhalt, Beschreibung 194
  - Dokumentrang
    - Boostklassen 232, 235
    - für Objektgruppe aktivieren 35
    - in migrierten Objektgruppen 397
    - Standardwerte wiederherstellen 225
    - statisch 224
    - URI-Muster 230, 231
    - Verzeichnisse von Boostwörtern 230
  - Dokumentrangfolge
    - Standardwerte wiederherstellen 225
  - Dokumenttypen
    - erkennen 165
    - für Parser-Services 168, 169
    - für Stellant-Parser 173
    - Syntaxanalyse 167
    - von Stellant-Parsern unterstützt 176
  - Dokumentüberwachung
    - aktivieren 347
    - Berichte 348
    - Beschreibung 346
    - inaktivieren 347
    - Protokolldateien 349
    - Protokolldateikonfiguration 347
  - Dokumentzusammenfassungen
    - anpassen 218
    - Merkmale bearbeiten 219
  - Domänenregeln für Web-Crawler 101
  - Domino-Benutzerkonfiguration, Quick-Place-Crawler 309
  - Domino Document Manager-Crawler in der Solaris-Betriebsumgebung konfigurieren 83
  - IOCP-Konfiguration 88
  - Konfiguration 66
  - Konfiguration für DIIO-Protokoll 87
  - NRPC-Protokoll 83, 85
  - unter dem Betriebssystem AIX konfigurieren 83
  - unter dem Betriebssystem Linux konfigurieren 83
  - unter Windows konfigurieren 85
  - URI-Formate 130
  - Dublin Core-Elemente 152
  - Durchsuchen
    - HTML-Dokumente 152
    - Kategorien 142
    - Objektgruppen 194
    - Quick Links 220
    - XML-Dokumente 148, 149, 158
  - Dynamische Rangfolge 223
  - Dynamische Zusammenfassung 217
- ## E
- E-Mail-Benachrichtigungen
    - Alerts 357
    - Nachrichten 357
    - SMTP-Server, Konfiguration 356
  - EAR-Dateien
    - ESAdmin, Anwendung 315
    - ESSearchApplication, Anwendung 315
    - ESSearchServer, Anwendung 315
  - Einzelbytedcodierung 183
  - Entfernen von Leerraum 185
  - Entfernen von Zeilenvorschubzeichen 185
  - Ereignisveröffentlichung
    - Beschreibung 52
    - DB2-Crawler, Konfiguration 56, 59
    - in der Solaris-Betriebsumgebung konfigurieren 54
    - unter dem Betriebssystem AIX konfigurieren 54
    - unter dem Betriebssystem Linux konfigurieren 54
    - unter Windows konfigurieren 55
  - Erkennung von doppelten Dokumenten
    - Beschreibung 199, 277
    - Boostklassenkonfiguration 232
    - globale Analyse 199, 282
    - Sicherheit aktivieren 282
  - Erstellen
    - Bereiche 204
    - Crawler 43
    - HTML-Suchfelder 152
    - Objektgruppen 33, 35
    - Quick Links 220
    - regelbasierte Kategorien 146
    - Web-Crawler-Berichte 334
    - XML-Suchfelder 149
  - es.cfg, Datei 20, 21, 26, 28, 269, 300, 301
  - ES\_INSTALL\_ROOT, Beschreibung 20, 21
  - ES\_NODE\_ROOT, Beschreibung 20, 21
  - es.search.provider.jar, Datei 381, 390
  - es.security.jar, Datei 375, 381, 390
  - es\_special\_field.default\_field, reserviertes Feld 235
  - es\_special\_field.default\_metadata\_field, reserviertes Feld 235
  - es\_special\_field.regular\_text, reserviertes Feld 235
  - es.wp5.install.jar, Datei 375
  - es.wp6.install.jar, Datei 381, 390
  - ESAdmin, Anwendung
    - anmelden 19
    - Sicherheit inaktivieren 315
  - esadmin, Befehl 401
  - esadmin startSearch, Befehl 324
  - esadmin stopIndex, Befehl 193
  - esadmin stopSearch, Befehl 324
  - esadmin system startall, Befehl 401
  - esadmin system stopall, Befehl 401
  - esapi.jar, Datei 375, 381, 390
  - esbackup.bat, Script 364
  - esbackup.sh, Script 364
  - eschangeproxypw, Befehl 271
  - eschangepw, Script 20, 21
  - eschangetrustpw, Befehl 269
  - eschangewaspw, Befehl
    - Konfiguration mit einem Server 300
    - Konfiguration mit mehreren Servern 301
  - escrcm.sh, Script 63
  - escrcm.vbs, Script 65
  - escrdb2.sh, Script 54
  - escrdb2.vbs, Script 55
  - escrnote.sh, Script 83
  - escrnote.vbs, Script 85
  - escrvbr.sh, Script 50
  - escrvbr.vbs, Script 51
  - ESPACServer.ear, Datei 375, 381, 390
  - esrestore.bat, Script 365
  - esrestore.sh, Script 365
  - ESSearchAdapter.ear, Datei 375
  - ESSearchApplication, Anwendung
    - config.properties, Datei 242, 260
    - Sicherheit inaktivieren 315
    - starten 268
  - ESSearchPortlet.war, Datei 375, 381, 390
  - ESSearchRegistrationPortlet.war, Datei 375
  - ESSearchServer, Anwendung
    - erneut starten 260, 261
    - Sicherheit inaktivieren 315
  - Exchange Server-Crawler
    - Konfiguration 69
    - sichere Dokumente 307
    - URI-Formate 130
  - Externe Crawler
    - Anwendungen für Datenempfängsfunktion 125
    - konfigurieren 125
  - Externe Quellen
    - Beschreibung 273
    - Konfiguration 273
    - Suchanwendungen zuordnen 276
    - Suchanwendungs-ID, Sicherheit 283
- ## F
- Fehlernachrichten
    - E-Mail-Empfang 354, 357
    - gelöschte Dokumentprotokolldateien anzeigen 349
    - Protokolldateien anzeigen 360
    - SMTP-Server, Konfiguration 356
  - Felder zu Boostklassen zuordnen 233
  - Feldspezifische Suche
    - Beschreibung 194
    - Zeichenfolgesortierung 194
  - Firewalls, Exchange Server-Dokumente durchsuchen 307
  - Föderierte Objektgruppen 33
  - followindex.rules, Datei
    - Beschreibung 115
    - konfigurieren 116
  - Formularbasierte Authentifizierung 109, 110
  - Freiformatsuche, Beschreibung 194
  - Funktionen zur behindertengerechten Bedienung für dieses Produkt 439

## G

- Gelöschte Dokumente
  - Berichte 348
  - Beschreibung 346
  - Protokolldateien 349
  - Protokolldateikonfiguration 347
- Gleichzeitig ablaufende Indexerstellung 190
- global.rules, Datei 114
- Globale Analyse
  - Ankertextanalyse 277, 295
  - Beschreibung 7
  - Erkennung von doppelten Dokumenten 199, 277, 282
- Globale WebSphere-Sicherheit
  - Anpassungsfunktion für die Suchanwendung 261
  - inaktivieren 315
  - Suchanwendungsmerkmale 260
- Globaler Web-Crawler-Bereich 114

## H

- Häufige Abfragen
  - berechnen 345
- Häufige Abfragen überwachen 343
- Häufigste Ergebnisse
  - Balkendiagramme 265
- Hauptindizes
  - Änderungen feststellen 192
  - Beschreibung 7, 187
  - gleichzeitig ablaufende Erstellungen 190
  - terminieren 188
- HTML-Dokumentation für die Unternehmensuche 437
- HTML-Dokumente
  - durchsuchen 152
  - Durchsuchen 152
  - Ersetzungsregeln 171, 172
  - Syntaxanalyse 171, 172
- HTML-Ersetzungsregeln 171, 172
- HTML-Suchfelder
  - Beschreibung 152
  - Dublin Core-Elemente 152
  - Elemente zuordnen 152
  - erstellen 152
- HTTP-Basisauthentifizierung 109
- HTTP-Proxy-Server 111
- HTTP-Statuscodes
  - von Web-Crawlern empfangen 334
  - Web-Crawler-Bericht 334
- HTTPS, Suchserverkonfiguration 269

## I

- I/O Completion Port-Modul, Crawlerkonfiguration 88
- Identitätsmanagement
  - Benutzerprofile 291
  - Benutzersicherheitskontext 288
  - Beschreibung 288
  - Gruppenextraktion 288
  - inaktivieren 288
  - Konfiguration 294
  - SSO-Unterstützung 292
- Identitätsmanagement (*Forts.*)
  - XML-Abfragezeichenfolge 288
- Ideographische Schriften 181
- Inaktivieren der Indexzeitpläne 189
- Index, Beschreibung der Seite 15
- Indexerstellungen
  - Änderungen feststellen 192
  - Beschreibung 187
  - gleichzeitig ablaufend 190
  - parallel 190
  - starten 341
  - startIndexBuild, Befehl 192
  - stoppen 193, 341, 342
  - Systemstatus 342
  - terminieren 189
- Indexerweiterung
  - Auswirkung auf Indexerstellungszeit 202
  - Auswirkung auf Indexgröße 202
  - Beschreibung 200
- Indexierungswarteschlange 342
- Indexserver
  - starten 319
  - stoppen 321
- Indizes
  - Alerts 352
  - Änderungen feststellen 192
  - Ankertext 296
  - aus der Warteschlange löschen 342
  - ausgeblendete URIs 194, 205, 206
  - Auswirkung von Platzhalterzeichen 202
  - Bereiche 194, 203
  - Beschreibung 7, 187
  - gleichzeitig ablaufende Erstellungen 190
  - parallele Erstellungen 190
  - Platzhalterzeichen 194, 200, 203
  - startIndexBuild, Befehl 192
  - terminieren 188
  - überwachen 341, 342
  - URI-Formate 130
  - URIs entfernen 194, 207
  - Zeitplan aktivieren 189, 341
  - Zeitplan inaktivieren 189, 341
  - Zeitplanänderung 189
- Indizes erstellen 187
- Integration in WebSphere Portal
  - Beschreibung 373
  - Clustersystem 389
  - es.wp5.install.jar, Datei 375
  - es.wp6.install.jar, Datei 381, 390
  - Lotus Quickr 373
  - Setup-Scripts 374
  - Web Content Management 373
  - wp5\_install, Script 375
  - wp6\_cluster\_install, Script 390
  - wp6\_install, Script 381
- Intervalle für Web-Crawler für erneute Suche 106
- IOCP, Crawlerkonfiguration 88
- IP-Adressen
  - IPv6, Unterstützung 29
  - Loopback-Adapter 28
  - Unterstützung für doppelte 28
- IP-Adressen ändern 26
- IP-Adressenregeln für Web-Crawler 101

IPv6, Protokoll 29

## J

- Japanisch
  - N-Gram-Segmentierung 184
  - Zeilenvorschubzeichen entfernen 185
- Java-Connector für DB2 Content Manager 63, 65
- JDBC, externe Quellen
  - bearbeiten 273
  - JDBC-Treiber 273
  - Konfiguration 273
  - löschen 273
- JDBC-Datenbankcrawler
  - Konfiguration 70
  - mehrere Tabellen durchsuchen 72, 74
  - Plug-in zum Durchsuchen von mehreren Tabellen 72, 74
  - unterstützte Treiber 70
  - URI-Formate 130
- JDBC-Treiber
  - für externe JDBC-Quellen 273
  - für JDBC-Datenbankcrawler 70

## K

- Kategoriebäume
  - aus WebSphere Portal migrieren 397
  - Beschreibung 145
- Kategorien
  - aus WebSphere Portal migrieren 397
  - Beschreibung 142
  - durchsuchen 142
  - erstellen 146
  - Kategoriebäume 145
  - Kategorisierungstyp 145
  - regelbasiert 143, 145
  - Unterkategorien verschachteln 145
  - URI-Formate 130
- Kategorisierungsregeln
  - Dokumentinhalt 143, 146
  - konfigurieren 146
  - URI-Muster 143, 146
- Kategorisierungstyp
  - auswählen 35, 145
  - regelbasiert 143
- Kennwort, Administrator für die Unternehmensuche 20, 21
- Kennwortgeschützte Websites 109
  - formularbasierte Authentifizierung 110
  - HTTP-Basisauthentifizierung 109
- Klonen
  - Crawler 41
  - Suchanwendungen 263
- Komplexe alphabetische Schriften 181
- Komprimierte Dateien
  - Crawlersuche 129
  - URI-Formate 130
- Konfiguration des Loopback-Adapters 28
- Koreanisch
  - Analyse zusammengesetzter Begriffe 162
  - N-Gram-Segmentierung 184

## L

- Ländereinstellungen
  - durchsuchen 181
  - Syntaxanalyse 181
- LDAP, externe Quellen
  - bearbeiten 273
  - konfigurieren 273
  - löschen 273
- LDAP-Benutzerregistry 298
- Letzte Abfragen
  - berechnen 345
- Letzte Abfragen überwachen 343
- Linguistische Unterstützung
  - benutzerdefinierte Textanalyse 154
  - Codepage-Erkennung 183
  - Entfernen von Leerraum 185
  - Ländereinstellungen 181
  - N-Gram-Segmentierung 184
  - native XML-Suche 163
  - semantische Suche 154, 163
  - Sprachencodes 181
  - Sprachenerkennung 182
  - Synonymverzeichnisse 211
  - Verzeichnisse von Boostwörtern 227
  - Verzeichnisse von Stoppwörtern 213
- Linux, Betriebssystem
  - Content Edition-Crawler, Konfiguration 50
  - DB2 Content Manager-Crawler, Konfiguration 63
  - DB2-Crawler, Konfiguration 54
  - Domino Document Manager-Crawler, Konfiguration 83
  - Ereignisveröffentlichung, Konfiguration 54
  - Notes-Crawler, Konfiguration 83
  - QuickPlace-Crawler, Konfiguration 83
  - Solaris-Betriebsumgebung
    - Ereignisveröffentlichung, Konfiguration 54
- Löschen
  - Crawler 45
  - Indizes aus der Warteschlange 342
  - Objektgruppen 38
- Lotus Domino-Domänen 308
- Lotus Domino-Trusted Server 308
- Lotus Notes
  - Plug-in-Installation 371
  - Suchsymbolleiste 371
  - Unternehmenssuche, Integration 371
  - Update-Site für das Plug-in 371
- Lotus Quickr
  - Crawlerkonfiguration für Einstiegspunktlisten 93
  - Integration in WebSphere Portal 373, 387
  - QuickPlace-Crawler, Konfiguration 89

## M

- Maximales Intervall für erneute Suche 106
- Mehrbytecodierung 183

Metadatenfelder, Balkendiagramme für die häufigsten Ergebnisse 265

- Migration
  - Objektgruppen 397
  - regelbasierte Taxonomie 397
- Migrationsassistent
  - Beschreibung 397
  - Objektgruppen 397
  - Protokolldatei 400
  - regelbasierte Taxonomien 397
  - Standardcrawlereinstellungen 399
  - Standardobjektgruppeneinstellungen 399
  - starten 397
- MIME-Typen in Web-Crawler-Bereiche einschließen 101
- Minimales Intervall für erneute Suche 106
- Mustersuchanwendung
  - Beschreibung 12, 240
  - config.properties, Datei 242, 260
  - HTTPS-Umsetzung 269
  - klonen 263
  - Sicherheit inaktivieren 315
  - SSL-Umsetzung 269
  - Standardimplementierung 260
  - Suchfunktionen 239, 240

## N

- N-Gram-Segmentierung 184
- Native XML-Suche 163
- newHtmlTagReplacement, Parameter 171
- NNTP-Crawler konfigurieren 78
- no-follow, Anweisung
  - Beschreibung 115
  - konfigurieren 116
- no-index, Anweisung
  - Beschreibung 115
  - konfigurieren 116
- nodes.ini, Datei 26, 269
- Notes-Crawler
  - Feldzuordnungsregeln 82
  - in der Solaris-Betriebsumgebung konfigurieren 83
  - IOCP-Konfiguration 88
  - Konfiguration 79
  - Konfiguration der Sicherheit auf Dokumentenebene 308
  - Konfiguration für DIOP-Protokoll 87
  - Lotus Domino-Trusted Server 308
  - NRPC-Protokoll 83, 85
  - Prüfung aktueller Berechtigungsnachweise 308
  - Tipps zur Nutzung 82
  - unter dem Betriebssystem AIX konfigurieren 83
  - unter dem Betriebssystem Linux konfigurieren 83
  - unter Windows konfigurieren 85
  - URI-Formate 130
- NRPC-Protokoll, Crawlerkonfiguration 83, 85

## O

- Objektgruppen
  - Ankertextsicherheit 295
  - aus WebSphere Portal migrieren 397
  - bearbeiten 37
  - Beschreibung 3
  - durchsuchen 194
  - Entwurf 33
  - Erkennung von doppelten Dokumenten 199
  - Ermitteln der ID 39
  - Erstellungsmöglichkeiten 33
  - Föderation von Datenquellen 33
  - Größe schätzen 35
  - löschen 38
  - MigrationWizard.log, Datei 397
  - mit dem Objektgruppenassistenten erstellen 33
  - mit der Sicht 'Objektgruppen' erstellen 35
  - Ressourcen schätzen 327
  - Sicherheit 282
  - Sicherheit für doppelte Dokumente 282
  - Standardmigrationseinstellungen 399
  - Suchanwendungen zuordnen 240
  - Suchanwendungs-ID, Sicherheit 283
  - Suchserver 209
  - Syntaxanalyse 141
  - Systemstatus 328
  - überwachen 328
  - Zugriffssteuerung auf Dokumentenebene umgehen 316
- Objektgruppen, Sicht
  - Beschreibung 15
  - Objektgruppen erstellen 35
- Objektgruppen-ID, Syntaxregeln 35
- Objektgruppen-IDs 39
- Objektgruppenadministrator
  - Beschreibung 280
  - Rollenkonfiguration 281
- Objektgruppenassistent 33
- OmniFind Enterprise Edition
  - Administrationskonsole 9
  - APIs 11
  - Befehle 401
  - Crawler-Server 4
  - Datenflussdiagramm 13
  - Indexserver 7
  - Integration in Lotus Notes 371
  - Integration in WebSphere Portal 373
  - IP-Adressen ändern 26
  - IPv6, Protokollunterstützung 29
  - Kennwort auf einem Einzelserver ändern 20
  - Kennwort auf mehreren Servern ändern 21
  - Komponenten 3
  - Parser 5
  - Portnummernkonfiguration 25
  - Rückkehrcodes 401
  - Sitzungs-IDs 401
  - Suchanwendungen 12
  - Suchserver 8
  - Überblick 1
  - Unterstützung für doppelte IP-Adressen 28

- Operator
  - Beschreibung 280
  - Rollenkonfiguration 281
- Oracle-Datenbanken
  - als externe Quelle zugreifen 273
  - mit DB2-Crawlern zugreifen 52
  - mit JDBC-Datenbankcrawlern zugreifen 70

## P

- Parallele Indexerstellungen 190
- Parameterfelder
  - Beschreibung 194
  - numerische Sortierung 194
- Parser
  - Analyse zusammengesetzter Begriffe 162
  - ASCII-Parser 170
  - Beschreibung 5, 141
  - Codepage-Erkennung 183
  - Dateien ohne Erweiterung 170
  - Datenanalysetasks 5
  - Dokumentformaterkennung 165
  - Dokumenttypen für Parser-Services 168, 169
  - Dokumenttypen für Stellent-Parser 173
  - Dokumenttypen syntaktisch analysieren 167
  - Entfernen von Leerraum 185
  - Entfernen von Zeilenvorschubzeichen 185
  - HTML-Ersetzungsregeln 171, 172
  - N-Gram-Segmentierung 184
  - native XML-Suche 163
  - Parsertypauswahl 165
  - Sprachenerkennung 182
  - starten 340
  - stoppen 340
  - Systemstatus 340
  - Threads 162
  - überwachen 340
  - unbekannte Dokumenttypen 170
  - unterstützte Sprachen 181
  - unterstützte Stellent-Dokumenttypen 176
  - Verarbeitung auf linguistischer Basis 181
- Parser-Server
  - Threadkonfiguration 162
- parserdriver.collection.properties, Datei 171
- parserTypes.cfg, Datei 165, 168, 170
- PDF-Dokumentation für die Unternehmenssuche 437
- Platzhalterzeichen
  - Abfrageerweiterung 200, 203
  - in Abfragen 200
  - Indexerweiterung 200, 202, 203
- Plug-in für Lotus Notes
  - Installation 371
  - Update-Site 371
- Plug-in für mehrere strukturierte Tabellen 72, 74
- Plug-ins
  - JDBC-Datenbankcrawler 72, 74

- Plug-ins (*Forts.*)
  - mehrere strukturierte Tabellen durchsuchen 72, 74
- Plug-ins für Crawler 127
- Portlets
  - aus WebSphere Portal 5.1 entfernen 380
  - aus WebSphere Portal-Clustern entfernen 393
  - aus WebSphere Portal Version 6 entfernen 388
  - Beschreibung 373
  - Integration in WebSphere Portal 5.1 375
  - Integration in WebSphere Portal 6 381, 387
  - Integration in WebSphere Portal-Cluster 390
  - Konfiguration für Lotus Quickr-Quellen 387
  - Unternehmenssuche 373
- Portnummer, Unternehmenssuche 25
- Präfixregeln für Web-Crawler 101
- Protokoll, Beschreibung der Seite 15
- Protokoll zum Sperren von Websitebereichen für Robots
  - Identifikation von Benutzern 97
  - Web-Crawler-Konformität 98
- Protokolldateien
  - Abfrageprotokolle 359
  - anzeigen 360
  - Beschreibung 351
  - Bewertungsstufen 354
  - E-Mail-Optionen 357
  - filtern 360
  - für Dokumentüberwachung 347
  - gelöschte Dokumente anzeigen 349
  - Größenkonfiguration 359
  - maximale Größe 354
  - Migrationsassistent 400
  - rollieren 354
  - SMTP-Server, Konfiguration 356
  - Standardposition 351
  - überwachen 349, 360
- Proxy-Server 111
  - Suchserverkonfiguration 271
- Prüfung aktueller Berechtigungs-nachweise 287, 308, 311

## Q

- Quick Links
  - Beschreibung 220
  - durchsuchen 220
  - erstellen 220
  - URI-Formate 130
- QuickPlace-Crawler
  - Directory Assistance, Konfiguration 310
  - Domino-Benutzerkonfiguration 309
  - in der Solaris-Betriebsumgebung konfigurieren 83
  - IOCP-Konfiguration 88
  - Konfiguration 89
  - Konfiguration für DIIOP-Protokoll 87

- QuickPlace-Crawler (*Forts.*)
  - Local User (Lokaler Benutzer), Sicherheit 309
  - NRPC-Protokoll 83, 85
  - unter dem Betriebssystem AIX konfigurieren 83
  - unter dem Betriebssystem Linux konfigurieren 83
  - unter Windows konfigurieren 85
  - URI-Formate 130

## R

- Rangfolge von Suchergebnissen
  - Beschreibung 223
  - Boostklassen 232, 233, 235
  - dynamisch 223
  - runtime.properties, Datei 225
  - Standardwerte wiederherstellen 225
  - statisch 224, 225
  - textbasierte Bewertung 223
  - URI-Muster 230, 231
  - Verzeichnisse von Boostwörtern 230
- Regelbasierte Kategorien
  - Beschreibung 143
  - erstellen 146
  - Kategorisierungstyp auswählen 145
- Regelbasierte Taxonomie aus WebSphere Portal migrieren 397
- removeCjNewlineChars, Option 185
- removeCjNewlineCharsMode, Option 185
- Restore-Scripts
  - ausführen 365
  - Beschreibung 363
- robots.txt, Dateien
  - Identifikation von Benutzern 97
  - Web-Crawler-Konformität 98
- Rückkehrcodes, Unternehmenssuche 401
- runtime-generic.properties, Datei 219, 345

## S

- Schlüsselspeicherdateien 269
- Schlüsselwörter in Quick Links 220
- Scripts
  - esbackup.bat 364
  - esbackup.sh 364
  - escrcm.sh 63
  - escrcm.vbs 65
  - escrdb2.sh 54
  - escrdb2.vbs 55
  - escrnote.sh 83
  - escrnote.vbs 85
  - escrvbr.sh 50
  - escrvbr.vbs 51
  - esrestore.bat 365
  - esrestore.sh 365
  - startcl 365
- Search and Index API 11, 239
- SearchBarInclude.jsp, Datei 378
- Seiten mit detaillierten Fehlerhinweisen, Web-Crawler 107
- Semantische Suche 154, 158, 163

- Server für die Unternehmenssuche
  - IP-Adressen ändern 26
  - IPv6, Protokollunterstützung 29
  - Unterstützung für doppelte IP-Adressen 28
- Servermodus, Content Edition-Repositories 49
- Setup-Scripts
  - WebSphere Portal 374
- SI-API (Search and Index API) 11, 239
- siapi.jar, Datei 375
- Sicherheit
  - Ankertextanalyse 295
  - ausgeblendete Suchergebnisse 317
  - Authentifizierung 280, 315
  - Benutzerprofile 291
  - Beschreibung 277
  - Crawler-Plug-ins 127
  - Dokumentenebene 284, 286, 287, 294, 316
  - Erkennung von doppelten Dokumenten 282
  - für die Unternehmenssuche aktivieren 297
  - für eine Objektgruppe aktivieren 277
  - für Objektgruppe aktivieren 35
  - für Unternehmensanwendung inaktivieren 315
  - global, WebSphere Application Server 297, 298
  - globale WebSphere-Sicherheit 315
  - HTTPS-Konfiguration für Suche 269
  - Identitätsmanagement 288, 294
  - Konfiguration mit einem Server 300
  - Konfiguration mit mehreren Servern 301
  - LDAP-Benutzerregistry 298
  - Lotus Domino-Dokumente 308
  - Objektgruppenebene 282, 316
  - SSL-Konfiguration für Suche 269
  - SSO-Unterstützung 292
  - Suchanwendungs-IDs 283
  - Verwaltungsrollen 281
  - Windows-Domänen 311
  - Zugriffssteuerung 280
  - Zugriffssteuerung auf Dokumentenebene umgehen 316
- Sicherheit, Beschreibung der Sicht 15
- Sicherheit 'Local User', QuickPlace-Crawler 309
- Sicherheit auf Dokumentenebene
  - aktuellen Berechtigungsnachweis prüfen 287
  - Benutzerprofile 291
  - Beschreibung 277, 284
  - Crawler-Plug-ins 127
  - Crawlerkonfiguration 41
  - Echtzeitprüfung 287
  - Ergebnisse für das Nachfiltern 284
  - Ergebnisse für das Vorfiltern 284
  - für Lotus Domino-Dokumente 308
  - für Windows-Dateisysteme 311
  - Identitätsmanagement 288, 294
  - indexierte Zugriffssteuerung 286
  - Lotus Domino-Dokumente 308
  - Sicherheitstoken 286
  - SSO-Unterstützung 292
- Sicherheit auf Objektgruppenebene
  - aktivieren 35
  - Ankertextanalyse 295
  - Beschreibung 277, 282
  - Erkennung von doppelten Dokumenten 282
  - Suchanwendungs-IDs 283
- Sicherheitstoken
  - Crawlerkonfiguration 286
  - für eine Objektgruppe inaktivieren 316
  - Sicherheit auf Dokumentenebene 286
- Sitedetailberichte
  - Beschreibung 331
  - erstellen 334
- Sitzungs-IDs, Unternehmenssuche 401
- SMTP-Server, Konfiguration 356
- Solaris, Betriebssystem
  - DB2-Crawler, Konfiguration 54
- Solaris-Betriebsumgebung
  - Content Edition-Crawler, Konfiguration 50
  - DB2 Content Manager-Crawler, Konfiguration 63
  - Domino Document Manager-Crawler, Konfiguration 83
  - Notes-Crawler, Konfiguration 83
  - QuickPlace-Crawler, Konfiguration 83
- Sortierbare Felder
  - numerische Sortierung 194
  - Zeichenfolgesortierung 194
- Sprachen
  - automatische Erkennung 182
  - Codes mit zwei Zeichen 181
  - durchsuchen 181
  - unterstützt 181, 182
- SQL Server-Datenbanken
  - mit DB2-Crawlern zugreifen 52
  - mit JDBC-Datenbankcrawlern zugreifen 70
- SSL, Suchserverkonfiguration 269
- SSO-Sicherheit
  - Identitätsmanagement 292
  - Konfiguration 294
- Standardsuchanwendung 260
- Start-URL-Adressen für Web-Crawler 101, 106
- startccl, Script 365
- Starten
  - Anpassungsfunktion für die Suchanwendung 261
  - Crawler-Server 330
  - Datenempfangsfunktion 346
  - Indexerstellungen 341
  - Migrationsassistent 397
  - Parser 340
  - Server für die Unternehmenssuche 319
  - Suchanwendungen 268
  - Suchserver 324, 343
- startIndexBuild, Befehl 192
- Statische Rangfolge
  - Beschreibung 224
  - für Objektgruppe aktivieren 35
  - in migrierten Objektgruppen 397
- Stellent-Parser
  - Beschreibung 165
  - Dokumenttypen syntaktisch analysieren 167
  - Dokumenttypen zuordnen 173
  - Standarddokumenttypen 176
- stellent.properties, Datei 173
- stellenttypes.cfg, Datei 173
- stellentTypes.cfg, Datei 165
- Stoppen
  - Crawler-Server 330
  - Indexerstellungen 341, 342
  - Parser 340
  - Server für die Unternehmenssuche 319, 321
  - Suchserver 324, 343
- Suchantwortzeit
  - Alerts 352
  - überwachen 343
- Suchanwendungen
  - anpassen 261
  - benutzerdefiniert 239
  - Beschreibung 12
  - externen Quellen zuordnen 276
  - Muster 239, 240
  - Objektgruppen zuordnen 240
  - Sicherheit auf Objektgruppenebene 283
  - starten 268
  - Suchanwendungs-IDs 283
  - zugreifen 268
- Suchanwendungen anpassen 260, 261
- Suchanwendungs-IDs 283
- Suchcache
  - Beschreibung 210
  - konfigurieren 210
- Suche, Beschreibung der Seite 15
- Suchergebnisfelder, Beschreibung 194
- Suchergebnisse
  - ausblenden 205, 206, 317
  - Beschreibung 223
  - Boostklassenkonfiguration 232, 233, 235
  - dynamische Rangfolge 223
  - dynamische Zusammenfassung 217
  - gruppieren 205, 206
  - Konfiguration von URI-Mustern 230
  - nachfiltern 284
  - Platzhalterzeichen 200
  - Platzhalterzeichenerweiterung 203
  - Rangfolge 231
  - statische Rangfolge 224
  - textbasierte Bewertung 223
  - vorfiltern 284
  - Zusammenfassungen 218, 219
  - Zusammenfassungen anpassen 218, 219
- Suchfelder für vollständige Übereinstimmung, Beschreibung 194
- Suchleiste, WebSphere Portal
  - Version 5.1, Umleitung zu Unternehmenssuche 378
  - Version 6, Umleitung zu Unternehmenssuche 385
- Suchoptionen
  - Dokumentinhalt 194
  - feldspezifische Suche 194

- Suchoptionen (*Forts.*)
  - Freiformatsuche 194
  - für Suchergebnisse 194
  - parametrische Suche 194
  - sortierbare Felder 194
  - vollständige Übereinstimmung 194
- Suchportlet, Implementierung
  - es.wp5.install.jar, Datei 375
  - es.wp6.install.jar, Datei 381, 390
  - wp5\_install, Script 375
  - wp5\_uninstall, Script 380
  - wp6\_cluster\_install, Script 390
  - wp6\_cluster\_uninstall 393
  - wp6\_install, Script 381
  - wp6\_uninstall, Script 388
- Suchrate überwachen 334
- Suchserver
  - Antwortzeitprotokoll 343
  - Berechnung von Abfragezählern 345
  - Beschreibung 8, 209
  - häufige Abfragen 343, 345
  - HTTPS-Konfiguration 269
  - letzte Abfragen 343, 345
  - Proxy-Server-Konfiguration 271
  - SSL-Konfiguration 269
  - starten 319, 324, 343
  - stoppen 321, 324, 343
  - Suchcache 210
  - Synonymverzeichnisse 211, 212
  - Synonymverzeichnisse zuordnen 213
  - Systemstatus 343
  - überwachen 343, 345
  - Verzeichnisse von Boostwörtern 227
  - Verzeichnisse von Boostwörtern zuordnen 229, 230
  - Verzeichnisse von Stoppwörtern 213
  - Verzeichnisse von Stoppwörtern zuordnen 215, 216
  - Wörterverzeichnisse erneut implementieren 216
- Suchzentrale, WebSphere Portal
  - Beschreibung 373
  - Version 6, Integration in Unternehmenssuche 384
- Synonymverzeichnisse
  - Beschreibung 211
  - dem System hinzufügen 212
  - einer Objektgruppe zuordnen 213
  - erneut implementieren 216
- Syntaxanalyse, Beschreibung der Seite 15
- System, Beschreibung der Sicht 15
- System-Backup 363, 364
- Systemressourcen
  - schätzen 327
- Systemressourcen schätzen 327
- Systemrestore 363, 365
- Systemstatus
  - Crawler 330
  - Indexerstellungen 342
  - Objektgruppen 328
  - Parser 340
  - Suchserver 343
  - Web-Crawler 331

## T

- TAR-Dateien
  - Crawlersuche 129
  - URI-Formate 130
- Taskübersicht, Administrationskonsole 15
- Taxonomien aus WebSphere Portal migrieren 397
- Terminieren
  - Crawler 41, 46
  - Indexerstellungen 188, 189
- Textanalyse
  - allgemeine Analysestrukturen 159, 160
  - Textanalysesteuerkomponenten 156, 157
  - XML-Elemente zuordnen 158
- Textanalysesteuerkomponenten
  - allgemeine Analysestruktur relationalen Datenbanken zuordnen 160
  - Analyseergebnisse dem Index zuordnen 159
  - Analyseergebnisse relationalen Datenbanken zuordnen 160
  - Beschreibung 154
  - dem System hinzufügen 156
  - Objektgruppen zuordnen 157
  - XML-Elemente zuordnen 158
- Textbasierte Bewertung 223
- Textverarbeitung
  - allgemeine Analysestrukturen 154
  - Annotatoren 154
  - Textanalysesteuerkomponenten 154
- Threddetails überwachen 331
- Threads
  - Parser 162
  - Web-Crawler 332
- Trusted Server, Konfiguration 308

## U

- Überwachen
  - aktive Web-Crawler-Sites 333
  - Antwortzeitprotokoll 343
  - Beschreibung 280
  - Crawler 330
  - Datenempfangsfunktion 346
  - Details von Web-Crawler-Threads 332
  - gelöschte Dokumente 348, 349
  - häufige Abfragen 343, 345
  - letzte Abfragen 343, 345
  - Objektgruppen 328
  - Parser 340
  - Protokolldateien 360
  - Rollenkonfiguration 281
  - Suchrate des Web-Crawlers 334
  - Suchserver 343, 345
  - Unternehmenssuche 327
  - URI-Details 328
  - Web-Crawler 331
- Überwachung, Beschreibung der Sicht 15
- UIMA
  - allgemeine Analysestruktur dem Index zuordnen 159

## UIMA (*Forts.*)

- allgemeine Analysestruktur relationalen Datenbanken zuordnen 160
- allgemeine Analysestrukturen 159, 160
- Analyseergebnisse dem Index zuordnen 159
- Analyseergebnisse relationalen Datenbanken zuordnen 160
- Beschreibung 154
- Objektgruppen zuordnen 157
- Textanalysesteuerkomponenten dem System hinzufügen 156
- XML-Elemente zuordnen 158
- Unbekannte Dokumenttypen 170
- Unicode-Codierung 183
- Unternehmensanwendungen
  - ESAdmin, Anwendung 315
  - ESSearchApplication, Anwendung 315
  - ESSearchServer, Anwendung 315
- Unternehmenssuche
  - Administrationskonsole 9
  - APIs 11
  - aus Backup wiederherstellen 365
  - Backup-Scripts 363
  - Befehle 401
  - Crawler-Server 4, 41
  - Datenflussdiagramm 13
  - Indexserver 7, 187
  - Integration in Lotus Notes 371
  - Integration in WebSphere Portal 373
  - Komponenten 3
  - Parser 5, 141
  - Portnummernkonfiguration 25
  - Protokolldateien 351
  - Restore-Scripts 363
  - Rückkehrcodes 401
  - Server starten 319
  - Server stoppen 319, 321
  - Sicherheit 277
  - Sicherheit auf Dokumentenebene 284
  - Sicherheit auf Objektgruppenebene 282
  - sichern 364
  - Sitzungs-IDs 401
  - Suchanwendungen 12
  - Suchserver 8, 209
  - Suchserver starten 324
  - Suchserver stoppen 324
  - Überblick 1
  - überwachen 327
  - URI-Formate 130
  - Verwaltungsrollen 280
- Unternehmenssuche sichern 363, 364
- Unternehmenssuche wiederherstellen 363
- Unterstützung für JavaScript in Web-Crawlern 100
- URI-Details
  - gelöschte Dokumente 348
  - überwachen 328
- URIs
  - aus dem Index entfernen 207
  - Beeinflussung von statischen Bewertungen 230, 231
  - Bereiche 203, 204

- URIs (*Forts.*)
  - Details anzeigen 328
  - Formate in der Unternehmenssuche 130
  - in Suchergebnissen ausgeblendet 205, 206
  - Kategorisierungsregeln 143, 146
  - Quick Links 220
- URIs aus dem Index entfernen 207
- URL-Adressen so schnell wie möglich aufrufen 106
- URL-Adressen so schnell wie möglich erneut aufrufen 106
- URL-Pfadtiefe 101
- USC-Zeichenfolge 288

## V

- vbr\_access\_services.jar, Datei 50, 51
- Verwaltungsrollen
  - Administrator für die Unternehmenssuche 280, 281
  - Beschreibung 280
  - konfigurieren 281
  - Objektgruppenadministrator 280, 281
  - Operator 280, 281
  - überwachen 280, 281
- Verzeichnisse von Boostwörtern
  - Beschreibung 227
  - dem System hinzufügen 229
  - einer Objektgruppe zuordnen 230
  - erneut implementieren 216
- Verzeichnisse von Stoppwörtern
  - Beschreibung 213
  - dem System hinzufügen 215
  - einer Objektgruppe zuordnen 216
  - erneut implementieren 216

## W

- Web Content Management
  - Integration in WebSphere Portal 373
- Web Content Management-Crawler
  - Konfiguration 118
  - Site-URL-Adressen kopieren 122
  - URI-Formate 130
- Web-Crawler
  - aktive Sites 331, 333
  - Benutzeragenten 97
  - Berichte erstellen 334
  - Cookieformat 112
  - Cookiekonfiguration 113
  - Cookies 112
  - Crawlerbereich begrenzen 101
  - Crawlerregeln 101
  - Crawlerverlauf 331
  - Datumsangaben für durchsuchte Dokumente 117
  - followindex.rules, Datei 115, 116
  - globaler Crawlerbereich 114
  - Intervalle für erneute Suche 106
  - kennwortgeschützte Websites 109, 110
  - Konfiguration 96
  - no-follow, Anweisung 115, 116
  - no-index, Anweisung 115, 116

- Web-Crawler (*Forts.*)
  - Proxy-Server 111
  - robots.txt, Dateien 97, 98
  - Seiten mit detaillierten Fehlerhinweisen 107
  - Sitedetails 331
  - Start-URL-Adressen 101, 106
  - Suchrate 334
  - Systemstatus 331
  - Threaddetails 331, 332
  - überwachen 331
  - Unterstützung für JavaScript 100
  - URL-Adressen so schnell wie möglich aufrufen 106
  - URL-Status 331
  - zuletzt durchsuchte URLs 331
- Web-Crawler-Bereich begrenzen 101
- WebSphere Application Server-Benutzer
  - Kennwort für Konfiguration mit einem Server 300
  - Kennwort für Konfiguration mit mehreren Servern 301
- WebSphere II Classic Federation 61
- WebSphere II Event Publisher Edition, Konfiguration des DB2-Crawlers 56
- WebSphere MQ, Konfiguration des Crawler-Servers 54, 55
- WebSphere MQ, Konfiguration des DB2-Crawlers 59
- WebSphere Portal
  - Clustersystem 389
  - Kategoriebaummigration 397
  - Objektgruppenmigration 397
  - Setup-Scripts für die Unternehmenssuche 374
  - Standardmigrationseinstellungen 399
  - Suchleiste, Beschreibung 373
  - Suchzentrale, Beschreibung 373
  - Taxonomiemigration 397
  - Unternehmenssuche, Integration 373
  - Version 5.1, Entfernen der Unternehmenssuche 380
  - Version 5.1, Integrationsscripts 375
  - Version 5.1, Suchleistenkonfiguration 378
  - Version 6, Entfernen der Unternehmenssuche 388
  - Version 6, Integrationsscripts 381
  - Version 6, Konfiguration der Suchzentrale 384
  - Version 6, Suchleistenkonfiguration 385
- WebSphere Portal-Cluster
  - Integrationsrichtlinien 389
  - Integrationsscripts 390
  - Unternehmenssuche entfernen 393
- WebSphere Portal-Crawler
  - Konfiguration 120
  - URI-Formate 130
- Windows
  - IPv6, Protokollunterstützung 29
- Windows-Betriebssystem
  - Content Edition-Crawler, Konfiguration 51
  - Crawlerkonfiguration 85
  - DB2 Content Manager-Crawler, Konfiguration 65

- Windows-Betriebssystem (*Forts.*)
  - DB2-Crawler, Konfiguration 55
  - Ereignisveröffentlichung, Konfiguration 55
- Windows-Domänen 311
- wp5\_install, Script 375
- wp5\_uninstall, Script 380
- wp6\_cluster\_install, Script 390
- wp6\_cluster\_uninstall, Script 393
- wp6\_install, Script 381
- wp6\_uninstall, Script 388
- WpsMigratorLog.log, Protokolldatei 400

## X

- XML-Abfragesyntax, nativ 163
- XML-Dokumente
  - durchsuchen 149
  - native XML-Suche 163
- XML-Elemente
  - der allgemeinen Analysestruktur zuordnen 158
  - durchsuchen 148, 158
  - Suchfeldern zuordnen 149
- XML-Fragmente, native XML-Suche 163
- XML-Suchfelder
  - Beschreibung 148, 158
  - Elemente zuordnen 148, 149, 158
  - erstellen 149
- XPath, native XML-Suche 163

## Z

- Zeichenfolge für den Benutzersicherheitskontext 288
- Zugriffssteuerung
  - aktuelle Benutzerprüfung 287
  - Anforderungen für Lotus Domino 308
  - Anforderungen für Windows-Dateisysteme 311
  - Beschreibung 280
  - für eine Objektgruppe inaktivieren 316
  - Identitätsmanagement 288, 291
  - Sicherheit auf Dokumentenebene 286
  - SSO-Sicherheit 292
- Zuletzt durchsuchte URLs überwachen 331
- Zuordnen
  - allgemeine Analysestruktur relationalen Datenbanken zuordnen 160
  - allgemeine Analysestrukturen dem Index 159
  - Analyseergebnisse relationalen Datenbanken zuordnen 160
  - Felder zu Boostklassen 233
  - HTML-Suchfelder 152
  - XML-Elemente der allgemeinen Analysestruktur zuordnen 158
  - XML-Suchfelder 149
- Zusammenfassungen
  - anpassen 218, 219
  - dynamische 217
- Zusammengesetzte Begriffe, Syntaxanalyse 162





**IBM**



**Java**<sup>™</sup>  
**COMPATIBLE**

SC12-3432-04

