

IBM WebSphere Information Integrator
OmniFind Edition



Verwaltung der Unternehmenssuche

Version 8.3

IBM WebSphere Information Integrator
OmniFind Edition



Verwaltung der Unternehmenssuche

Version 8.3

Hinweis

Vor Verwendung dieser Informationen und des darin beschriebenen Produkts sollten Sie die allgemeinen Informationen unter "Bemerkungen" lesen.

Dieses Dokument enthält proprietäre Informationen von IBM. Sie werden mit einer Lizenzvereinbarung zur Verfügung gestellt und durch Urheberrechtsgesetze geschützt. Die Informationen in dieser Veröffentlichung enthalten keine Produktgarantien.

Sie können IBM Veröffentlichungen online oder über Ihren lokalen IBM Ansprechpartner bestellen.

- Veröffentlichungen bestellen Sie online über das IBM Publications Center unter der Internetadresse www.ibm.com/shop/publications/order.
- Ihren lokalen IBM Ansprechpartner finden Sie über die Internetsite "IBM Directory of Worldwide Contacts" unter der Internetadresse www.ibm.com/planetwide.

Werden an IBM Informationen eingesandt, können diese beliebig verwendet werden, ohne dass eine Verpflichtung gegenüber dem Einsender entsteht.

Diese Veröffentlichung ist eine Übersetzung des Handbuchs
IBM WebSphere Information Integrator OmniFind Edition Administering Enterprise Search,
IBM Form SC18-9283-02,
herausgegeben von International Business Machines Corporation, USA

© Copyright International Business Machines Corporation 2005
© Copyright IBM Deutschland GmbH 2005

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
November 2005

Inhaltsverzeichnis

Informationen zu diesen Themen	vii	
Zielgruppe	vii	
Was ist Unternehmenssuche?	1	
Von der Unternehmenssuche unterstützte Daten-		
quellentypen	2	
Komponentenübersicht für die Unternehmenssuche	3	
Crawler für die Unternehmenssuche	4	
Parser für die Unternehmenssuche	5	
Indizes für die Unternehmenssuche.	7	
Suchserver für die Unternehmenssuche	8	
Administrationskonsole für die Unternehmens-		
suche	9	
Überwachen eines Systems für die Unter-		
nehmenssuche	10	
Protokolldateien für die Unternehmenssuche	10	
Anpassen der Unternehmenssuche.	11	
Mustersuchanwendung für die Unternehmens-		
suche	12	
Der Datenfluss für die Unternehmenssuche.	13	
Übersicht über die Verwaltung der		
Unternehmenssuche	15	
Anmelden an der Administrationskonsole	19	
Ändern des Kennworts des Administrators für die		
Unternehmenssuche in einer Einzelserver-		
konfiguration.	20	
Ändern des Kennworts des Administrators für die		
Unternehmenssuche in einer Konfiguration mit		
mehreren Servern	22	
Ändern der Portnummer für das System für die		
Unternehmenssuche	26	
Objektgruppen für die Unternehmens-		
suche	29	
Erstellen einer Objektgruppe mit dem Objekt-		
gruppenassistenten	30	
Erstellen einer Objektgruppe mit der Sicht 'Objekt-		
gruppen'	31	
Bearbeiten einer Objektgruppe	33	
Löschen einer Objektgruppe.	34	
Crawlerverwaltung für die Unter-		
nehmenssuche	37	
Erstellen eines Crawlers	40	
Bearbeiten von Crawlermerkmalen	41	
Bearbeiten eines Crawlerbereichs	41	
Löschen eines Crawlers	42	
Content Edition-Crawler	42	
Zugriff auf WebSphere II Content Edition-Repositorys im Servermodus	44	
Zugriff auf WebSphere II Content Edition-Repositorys im Direktmodus	44	
Konfigurieren des Crawler-Servers unter UNIX für WebSphere II Content Edition	45	
Konfigurieren des Crawler-Servers unter Windows für WebSphere II Content Edition	46	
DB2-Crawler	47	
Konfigurieren von WebSphere Information Integrator Event Publisher Edition für DB2-Crawler	49	
Konfigurieren von WebSphere MQ für DB2-Crawler	52	
Konfigurieren des Crawler-Servers unter UNIX für die Verwendung der Ereignisveröffentlichung.	54	
Konfigurieren des Crawler-Servers unter Windows für die Verwendung der Ereignisveröffentlichung.	55	
DB2 Content Manager-Crawler	56	
Konfigurieren des Crawler-Servers unter UNIX für DB2 Content Manager	57	
Konfigurieren des Crawler-Servers unter Windows für DB2 Content Manager	58	
Domino Document Manager-Crawler.	60	
Exchange Server-Crawler.	61	
Zugriffsprüfung zum Schutz von Exchange Server-Dokumenten.	62	
NNTP-Crawler	63	
Notes-Crawler	63	
Tipps für das Durchsuchen von Lotus Domino-Datenbanken	66	
Konfigurieren des Crawler-Servers unter UNIX für das Durchsuchen von Lotus Domino-Quellen.	67	
Konfigurieren des Crawler-Servers unter Windows für das Durchsuchen von Lotus Domino-Quellen.	69	
Konfigurieren von Servern, die das Protokoll DIIOP verwenden	70	
Konfigurieren des I/O Completion Ports unter AIX für das Durchsuchen von Lotus Domino-Quellen.	71	
QuickPlace-Crawler	73	
Konfigurieren des QuickPlace-Servers für die Verwendung der Sicherheit "Local User" (Lokaler Benutzer)	75	
Konfigurieren von Directory Assistance auf einem QuickPlace-Server	76	
Crawler für das UNIX-Dateisystem	77	
Web-Crawler	77	
Benutzeragentenkonfiguration	78	
Unterstützung für JavaScript	79	
Regeln zum Begrenzen des Web-Crawler-Bereichs	80	
Testen von URL-Verbindungen mit dem Web-Crawler	84	
Intervalleinstellungen für erneute Suche im Web-Crawler	85	
Optionen für das Aufrufen von URL-Adressen mit dem Web-Crawler.	85	
Vorgehensweise des Web-Crawlers bei Seiten mit detaillierten Fehlerhinweisen	86	

Unterstützung für die Crawlersuche auf sicheren Websites	87
Über Proxy-Server bereitgestellte Websites	90
Cookieverwaltung	91
Konfiguration eines globalen Web-Crawler-Bereichs	92
Die Anweisungen 'no-follow' und 'no-index'	94
Außerkräftsetzen der Anweisungen 'no-follow' und 'no-index' auf Webseiten	95
WebSphere Portal-Crawler	96
Implementieren der Unternehmensanwendung für den WebSphere Portal-Crawler.	97
Kopieren der zu durchsuchenden URL-Adresse von WebSphere Portal	98
Crawler für das Windows-Dateisystem	99
Konfigurieren der Unterstützung für Anwendungen der Datenempfangsfunktion	100
Benutzerdefinierte Crawler-Plug-ins	101
URI-Formate in einem Index für die Unternehmenssuche	103

Verwaltung von Parsern für die Unternehmenssuche. 111

Funktionsweise von Kategorien	112
Regelbasierte Kategorien.	113
Modellbasierte Kategorien	116
Kategoriebäume	116
Auswählen des Kategorisierungstyps	117
Konfigurieren von Kategorien	118
Funktionsweise von XML-Suchfeldern	120
XML-Suchfelder	120
Zuordnen von XML-Elementen zu Suchfeldern	121
Funktionsweise von HTML-Suchfeldern	122
HTML-Suchfelder	123
Zuordnen von HTML-Metadatenelementen zu Suchfeldern	123
Benutzerdefinierte Textverarbeitung	124
Hinzufügen von Textanalysesteuerkomponenten zum System	126
Zuordnen einer Textanalysesteuerkomponente zu einer Objektgruppe	127
Zuordnen von XML-Elementen zu einer allgemeinen Analysestruktur	128
Zuordnen einer allgemeinen Analysestruktur zum Index	129
Zuordnen einer allgemeinen Analysestruktur zu JDBC-Tabellen	130
Konfigurieren von Threads für den Parser-Service	131
Aktivieren der erweiterten Analyse für zusammengesetzte Begriffe	132
Aktivieren der Unterstützung für eine native XML-Suche	133
Linguistische Analyse von Dokumenten in chinesischer, japanischer und koreanischer Sprache	134
N-Gram-Segmentierung	134
Entfernen von Zeilenvorschubzeichen, die nicht zum ASCII-Zeichenbereich gehören	134
Dokumenttypen für Objektgruppenparser und Stel- lent-Sitzungen	135
Zuordnen von Dokumenttypen zu einem Objektgruppenparser	136

Standardserviceregeln für Objektgruppenparser	137
Zuordnen von Dokumententypen zu einer Stel- lent-Sitzung	138
Standardsyntaxanalyseregeln für Stel- lent-Sitzungen.	140

Indexverwaltung für die Unternehmenssuche. 143

Terminieren von Indexerstellungen	144
Ändern des Indexzeitplans	145
Aktivieren und Inaktivieren der Indexzeitpläne	145
Konfigurieren von gleichzeitig ablaufenden Indexerstellungen	146
Optionen zum Beeinflussen der durchsuchbaren Sicht des Index	147
Indizierte Optionen für das Durchsuchen von Dokumenten	148
Platzhalterzeichen in Abfragen	150
Konfigurieren von Optionen für Platzhalter- zeichen in Abfragen	152
Bereiche	153
Konfigurieren von Bereichen	154
Ausgeblendete URIs	155
Ausblenden von URIs in den Suchergebnissen	156
Entfernen von URIs aus dem Index	157

Verwaltung des Suchservers für die Unternehmenssuche 159

Suchcaches	160
Konfigurieren eines Suchcaches	160
Benutzerdefinierte Synonymverzeichnisse	161
Hinzufügen von Synonymverzeichnissen zum System	162
Zuordnen eines Synonymverzeichnisses zu einer Objektgruppe	163
Benutzerdefinierte Verzeichnisse von Stoppwör- tern	163
Hinzufügen von Verzeichnissen von Stopp- wörtern zum System	164
Zuordnen eines Verzeichnisses von Stopp- wörtern zu einer Objektgruppe	165
Dynamische Zusammenfassung	166
Anpassen von Dokumentzusammenfassungen in der Administrationskonsole.	166
Anpassen von Dokumentzusammenfassungen durch Bearbeitung von Merkmalen	167
Funktionsweise von Quick Links	168
Quick Links	168
Konfigurieren von Quick Links	168

Dokumentrangfolge bei der Unternehmenssuche. 171

Textbasierte Bewertung	171
Statische Rangfolge	172
Benutzerdefinierte Verzeichnisse von Boostwör- tern	172
Hinzufügen von Verzeichnissen von Boostwör- tern zum System	174
Zuordnen eines Verzeichnisses von Boostwör- tern zu einer Objektgruppe.	174
Auf URI-Mustern basierende Dokumentrangfolge	175

	Beeinflussen der Bewertungen von Dokumenten, die mit URI-Mustern übereinstimmen . . .	176		Installieren des Registrierungsportlets für die Unternehmenssuche für die Suchzentrale . . .	232
	Auf Boostklassen basierende Dokumentrangfolge	177			
	Zuordnen von Feldern zu Boostklassen . . .	179			
	Konfigurieren von Boostfaktoren für Boostklassen	180			
	Standardwerte für Boostklassen	181			
	Suchanwendungen für die Unternehmenssuche.	183			
	Zuordnen von Suchanwendungen zu Objektgruppen	184			
	Funktionen der Mustersuchanwendung	184			
	Merkmale von Mustersuchanwendungen	186			
	Bearbeiten der Merkmale der Mustersuchanwendung	195			
	Zugreifen auf die Mustersuchanwendung	196			
	Aktivieren der Sicherheit für die Mustersuchanwendung	197			
	Externe Quellen für die Unternehmenssuche.	201			
	Hinzufügen von externen Quellen zum System	201			
	Zuordnen von Suchanwendungen zu externen Quellen	202			
	Sicherheit für die Unternehmenssuche	205			
	Verwaltungsrollen	206			
	Konfigurieren von Benutzern mit Verwaltungsaufgaben	208			
	Authentifizierung und Zugriffssteuerung	208			
	Inaktivieren der Sicherheit für eine Unternehmensanwendung in WebSphere Application Server	210			
	Sicherheit auf Objektgruppenebene	211			
	Analyse auf doppelte Dokumente	211			
	Ankertextanalyse	212			
	Indexieren von Ankertext in Links zu verbotenen Dokumenten	213			
	Sicherheit mit Suchanwendungs-IDs	214			
	Sicherheit auf Dokumentenebene	214			
	Prüfung durch gespeicherte Sicherheitstoken	215			
	Prüfung aktueller Berechtigungsnachweise während der Abfrageverarbeitung	216			
	Umsetzung der Sicherheit auf Dokumentenebene für Dokumente des Windows-Dateisystems	219			
	Umsetzung der Sicherheit auf Dokumentenebene für Lotus Domino-Dokumente	221			
	Inaktivieren der Sicherheit auf Dokumentenebene	223			
	Integration der Unternehmenssuche in WebSphere Portal	225			
	Sicherheit auf Dokumentenebene mit Portal Search Engine	227			
	Implementieren des Suchportlets	228			
	Konfigurieren des Suchportlets von WebSphere Portal für die Unternehmenssuche	229			
	Installieren des Adapters für die Unternehmenssuche für die Suchzentrale	231			
	Migration von WebSphere Portal auf die Unternehmenssuche.	235			
	Modellbasierte Taxonomien aus WebSphere Portal migrieren	235			
	Migration einer Objektgruppe aus WebSphere Portal	237			
	Einstellungen für migrierte Objektgruppen	239			
	Protokolldatei des Migrationsassistenten	241			
	Starten und Stoppen der Server für die Unternehmenssuche.	243			
	Starten der Server für die Unternehmenssuche	243			
	Stoppen der Server für die Unternehmenssuche	245			
	Überwachen der Aktivitäten der Unternehmenssuche.	249			
	Schätzen der Anzahl Dokumente in einer Objektgruppe	249			
	Überprüfen der Verfügbarkeit von Systemressourcen	250			
	Überwachen einer Objektgruppe	251			
	Anzeigen von Details zu einer URI	252			
	Überwachen von Crawlern	254			
	Anzeigen von Details zur Web-Crawler-Aktivität	256			
	Details von Web-Crawler-Threads	257			
	Aktive Web-Crawler-Sites	257			
	Suchrate des Web-Crawlers	258			
	Erstellen von Web-Crawler-Berichten	259			
	Web-Crawler-HTTP-Rückkehrcodes	260			
	Überwachen des Parsers	264			
	Überwachen der Indexaktivität für eine Objektgruppe	265			
	Überwachen der Indexierungswarteschlange für die Unternehmenssuche	266			
	Überwachen der Suchserver	267			
	Überwachen der Datenempfangsfunktion	268			
	Dokumentüberwachung	269			
	Konfigurieren von Protokolldateien für die Dokumentüberwachung	270			
	Anzeigen von Berichten über gelöschte Dokumente	270			
	Protokolldateien und Alerts für die Unternehmenssuche	273			
	Alerts	273			
	Konfigurieren von Alerts auf Objektgruppenebene	274			
	Konfigurieren von Alerts auf Systemebene	276			
	Konfigurieren von Protokolldateien	277			
	Konfigurieren von SMTP-Serverinformationen	278			
	Empfangen von E-Mail zu protokollierten Nachrichten	279			
	Anzeigen von Protokolldateien	281			

Durchführen von Backup und Restore für ein System für die Unternehmenssuche.	283	Glossar der Begriffe für die Unternehmenssuche.	321
Sichern des Systems für die Unternehmenssuche	284	Zugreifen auf Informationen zu WebSphere Information Integration . .	331
Wiederherstellung des Systems für die Unternehmenssuche	284	Kommentare zur Dokumentation . . .	333
Wiederherstellen von Systemdateien für die Unternehmenssuche auf neuen Servern	285	Kontaktaufnahme mit IBM	335
 Befehle, Rückkehrcodes und Sitzungs-IDs der Unternehmenssuche . .	287	Marken	337
Dokumentation zur Unternehmenssuche.	317	Index	341
WebSphere II OmniFind Edition - Behindertengerechte Bedienung . . .	319		

Informationen zu diesen Themen

Beachten Sie die folgenden Informationen, wenn Sie ein IBM WebSphere Information Integrator OmniFind Edition-System der Version 8.3 verwalten.

WebSphere Information Integrator OmniFind Edition (WebSphere II OmniFind Edition) stellt eine Technologie bereit, die *Unternehmenssuche* genannt wird. Die Komponenten für die Unternehmenssuche werden bei der Installation von WebSphere II OmniFind Edition installiert. Der Begriff *Unternehmenssuche* wird in der gesamten Dokumentation zu WebSphere II OmniFind Edition verwendet, mit Ausnahme von Verweisen auf Installationspfade und Bezeichnungen von Produktpaketeten.

In der Dokumentation zur Verwaltung der Unternehmenssuche werden die folgenden Themen behandelt:

- Einführung in die Konzepte, Komponenten und Funktionen für die Unternehmenssuche
- Anweisungen zum Erstellen, Überwachen und Verwalten von Objektgruppen
- Informationen zur syntaktischen Analyse von Daten und zu den Möglichkeiten zum Anpassen von Syntaxanalysen zur Optimierung von Such- und Abrufvorgängen
- Informationen zur Indexierung von syntaktisch analysierten Daten und zu den Möglichkeiten zum Verwalten von Indexierungsvorgängen
- Informationen dazu, wie die Suchserver Suchergebnisse suchen, einstufen und zurückgeben
- Informationen zur Zuordnung von benutzerdefinierten Suchanwendungen zu Objektgruppen
- Überblick über die verschiedenen Sicherheitsstufen, die in der Unternehmenssuche verfügbar sind
- Informationen zur Integration der Unternehmenssuche in IBM WebSphere Portal
- Anweisungen zur Migration der WebSphere Portal-Taxonomien und -Objektgruppen in Kategoriebäume und Objektgruppen für die Unternehmenssuche
- Anweisungen für das Sichern und Wiederherstellen des Systems
- Anweisungen für das Erstellen und Anzeigen von Protokolldateien
- Anweisungen für das Verwenden von Befehlen für die Unternehmenssuche und für das Interpretieren der zurückgegebenen Informationen

Zielgruppe

Diese Informationen richten sich an Systemadministratoren und Systembediener, die für das Erstellen, Überwachen und Verwalten der Objektgruppen für die Unternehmenssuche verantwortlich sind.

Verwenden Sie diese Informationen, um Objektgruppen zu erstellen, Inhalt für Objektgruppen auszuwählen und Optionen zu konfigurieren, die den Inhalt für Suchvorgänge aktivieren. Sie können diese Informationen ferner verwenden, um Objektgruppen und die Systemaktivität zu überwachen, Benutzer als Administratoren für die Unternehmenssuche zu registrieren und Suchanwendungen Objektgruppen sowie externe durchsuchbare Quellen zuzuordnen.

Zur effektiven Nutzung dieser Informationen müssen Sie mit Webanwendungen vertraut sein und über Erfahrung mit den Datenquellen verfügen, die Sie durchsuchen möchten.

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. Das System für die Unternehmenssuche bietet schnelle Abfrageantwortzeiten und eine konsolidierte, nach Rang sortierte Ergebnismenge, mit der Sie die benötigten Informationen problemlos finden können.

Die Komponenten für die Unternehmenssuche, die mit IBM WebSphere Information Integrator OmniFind Edition (WebSphere II OmniFind Edition) installiert werden, erfassen Informationen aus Ihrem gesamten Unternehmen und stellen diese Informationen für Suchvorgänge zur Verfügung. 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

WebSphere II OmniFind Edition unterstützt IBM Unstructured Information Management Architecture (UIMA), um sicherzustellen, dass Benutzer die gesuchten Informationen auch finden. UIMA ist ein offenes Framework, das eine allgemeine Standardschnittstelle für die Textanalyse definiert. Dank einer umfassenden Textanalyse 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 die 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 geändert haben, nachdem ein Dokument und die zugehörigen Sicherheitstoken indiziert wurden.

Zugehörige Konzepte

Sicherheit für die Unternehmenssuche

"Integration der benutzerdefinierten Textanalyse" in "Integration der Textanalyse"

"Unstructured Information Management Architecture (UIMA) - Übersicht" in "Integration der Textanalyse"

Von der Unternehmenssuche unterstützte Datenquellentypen

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

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

- IBM DB2 Content Manager-Elementtypen (Dokumente, Ressourcen und Elemente)
- IBM DB2 Universal Database-Datenbanken
- IBM Domino Document Manager-Datenbanken (früher Domino.Doc)
- IBM Lotus Notes-Datenbanken
- IBM Lotus QuickPlace-Datenbanken
- Allgemein zugängliche Ordner von Microsoft Exchange Server
- Microsoft Windows-Dateisysteme
- NNTP-Newsgroups (NNTP - Network News Transfer Protocol)
- UNIX-Dateisysteme
- IBM WebSphere Information Integrator Content Edition-Repositorys: Elementklassen von Documentum, FileNet Panagon Content Services, FileNet P8 Content Manager, Hummingbird Document Management (DM), OpenText Livelink und Portal Document Manager (PDM)
- IBM WebSphere Information Integrator-Kurznamenstabellen für Datenbanken von IBM DB2 Universal Database für z/OS, IBM Informix, Microsoft SQL Server, Oracle und Sybase
- IBM WebSphere Portal-Sites
- 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:

- JDBC-Datenbanktabellen (Java Database Connectivity) - (Nur IBM DB2 UDB- (DB2 Universal Database) und Oracle-Datenbanken). Für jede Tabelle in einer JDBC-Datenbank 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 finden Sie auf der WebSphere Information Integrator OmniFind Edition-Website.

Zugehörige Konzepte

Externe Quellen für die Unternehmenssuche

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 WebSphere Information Integrator OmniFind 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 für die Unternehmenssuche

Verwaltung von Parsern für die Unternehmenssuche

Indexverwaltung für die Unternehmenssuche

Verwaltung des Suchservers für die Unternehmenssuche

Crawler für die Unternehmenssuche

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

Die Crawlerkomponente, die zusammen mit WebSphere Information Integrator OmniFind 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, so dass 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 zusammenzustellen. 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 Crawler suche entsprechend der Änderungshäufigkeit der von dem jeweiligen Crawler erfassten Daten festlegen.

Die Crawler für Web-, WebSphere Portal- und NNTP-Quellen (Network News Transfer Protocol) werden kontinuierlich ausgeführt. Nachdem Sie angegeben haben, welche URL-Adressen (URL - Uniform Resource Locator) oder NNTP-Newsgroups Sie durchsuchen wollen, kehrt der Crawler regelmäßig zurück, um zu prüfen, ob die Daten neu sind oder sie geändert wurden. Sie können andere Crawler-typen manuell starten oder stoppen, oder Sie können Zeitpläne für die Crawler suche konfigurieren. Wenn Sie einen Crawler terminieren, geben Sie an, wann er das erste Mal ausgeführt werden soll und wie oft er die Datenquellen 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 Crawler suche 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 für die Unternehmenssuche

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 WebSphere Information Integrator OmniFind 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 Stellent for IBM WebSphere Information Integrator OmniFind Edition Outside In Viewer Technology 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.
- Wendet die Syntaxanalyseregeln an, die Sie für die Objektgruppe angegeben haben. Bei der Parserkonfiguration können Sie Folgendes 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, sofern zutreffend, die gewünschte Kategorie aus. Wenn Sie *regelbasierte* Kategorien verwenden, 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.

Wenn Sie *modellbasierte* Kategorien verwenden, werden Dokumente modellbasierten Kategorien zugeordnet, die auf Ihrem WebSphere Portal-System vorhanden sind. Für diese Option muss WebSphere Portal auf

dem Indexserver für die Unternehmenssuche installiert sein. Sie müssen außerdem die Kategorisierungstools in WebSphere Portal verwenden, um die Kategorien zu verwalten.

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 (wie z. B. 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.

- 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.

Zugehörige Konzepte

Funktionsweise von Kategorien

XML-Suchfelder

HTML-Suchfelder

"Integration der benutzerdefinierten Textanalyse" in "Integration der Textanalyse"

"Textanalyse innerhalb der Unternehmenssuche" in "Integration der Textanalyse"

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:

Reorganisieren des Index

Wenn ein Index reorganisiert wird, wird der gesamte Index erneut erstellt, so dass die Struktur optimal organisiert ist. Die Indexierungsprozesse lesen alle Daten, die von Crawlern erfasst und vom Parser analysiert wurden.

Aktualisieren des Index

Beim Aktualisieren eines Index werden dem Index seit der letzten Indexreorganisation von Crawlern erfasste Informationen hinzugefügt.

Wenn Sie Indexoptionen für eine Objektgruppe konfigurieren, geben Sie Zeitpläne für das Reorganisieren und Aktualisieren des Index an. Die Häufigkeit der Indexreorganisation und -aktualisierung hängt von Ihren Systemressourcen und davon ab, ob die indexierten Quellen statischen oder dynamischen Inhalt aufweisen.

Terminieren Sie eine häufige Indexaktualisierung, um so sicherzustellen, dass neue Informationen verfügbar sind. Terminieren Sie regelmäßig eine Indexreorganisation, 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 Indexreorganisation starten, nachdem die Daten erneut durchsucht und syntaktisch analysiert wurden, anstatt darauf zu warten, dass die Indexreorganisation zum terminierten Zeitpunkt startet.

Zur Steuerung der Ressourcennutzung können Sie steuern, wie viele Objektgruppen gleichzeitig Indexierungsprozesse gemeinsam nutzen und Indexerstellungsanforderungen übergeben können. Mit der gleichzeitigen Indexerstellung stellen Sie sicher, dass die Reorganisation eines sehr großen Index nicht die Aktualisierung anderer Indizes blockiert. Die Indexerstellung kann ein ressourcenintensiver Prozess sein. Daher müssen Sie bei Großsystemen die Systembelastung überwachen, um die Indexreorganisations- und Aktualisierungsraten 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, so dass 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 für die Unternehmenssuche

Platzhalterzeichen in Abfragen

Bereiche

Ausgeblendete URIs

Auf URI-Mustern basierende Dokumentrangfolge

Zugehörige Tasks

Terminieren von Indexerstellungen

Konfigurieren von gleichzeitig ablaufenden Indexerstellungen

Entfernen von URIs aus dem Index

Überwachen der Indexaktivität für eine Objektgruppe

Überwachen der Indexierungwarteschlange 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 WebSphere Information Integrator OmniFind 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 WebSphere Information Integrator OmniFind 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 spezifisch 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

Übersicht über die Verwaltung der 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 WebSphere Information Integrator OmniFind Edition installiert und mindestens eine Objektgruppe erstellt haben, können Sie detaillierte Statistikdaten für jede Hauptaktivität (Crawlersuche, Syntexanalyse, 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 überwachten 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 der Server 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. Fehlernachrichten 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 Fehlernachrichten, sobald diese auftreten, auswählen.

Wenn Sie Protokolldateien anzeigen, wählen Sie die Protokolldatei aus, die Sie anzeigen wollen (der Dateiname enthält Informationen dazu, wann die Datei erstellt wurde und welche Komponente die Nachrichten ausgegeben 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 für die Unternehmenssuche

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 WebSphere Information Integrator OmniFind 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.

Datenempfangsfunktions-API

Empfangen Sie mit dieser API Daten von externen Crawlern. Die externen Crawler können eine Verbindung zur Datenempfangsfunktion für die Unternehmenssuche herstellen, und dann einer Objektgruppe Daten hinzufügen bzw. Daten aus einer Objektgruppe entfernen.

Crawler-Plug-ins

Mit APIs für Web- und Nicht-Web-Crawler fügen Sie Dokumenten Metadaten hinzu, während 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

"Search and index API overview" in "Programming Guide and API Reference for Enterprise Search"

"Data listener" in "Programming Guide and API Reference for Enterprise Search"

"Crawler-Plug-ins" in "Programming Guide and API Reference for Enterprise Search"

Zugehörige Tasks

Konfigurieren der Unterstützung für Anwendungen der Datenempfangsfunktion

Mustersuchanwendung für die Unternehmenssuche

Sie können die Mustersuchanwendung für die Unternehmenssuche in der vorliegenden Form oder als Schablone für die Entwicklung benutzerdefinierter Suchanwendungen verwenden.

Bei der Installation von WebSphere Information Integrator OmniFind 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.

Wenn Sie die Sicherheit für eine Objektgruppe und das WebSphere II OmniFind Edition-Identitätsmanagement für das System aktivieren, können Benutzer Profile erstellen, wenn sie die Suchanwendung verwenden. In dem Benutzerprofil werden die Berechtigungsnachweise gespeichert, die Benutzer bei der Anmeldung an verschiedenen Domänen angeben.

Während der Abfrageverarbeitung verwenden die Suchprozesse die gespeicherten Berechtigungsnachweise (die in einem sicheren von WebSphere II OmniFind Edition verwalteten Speicher verschlüsselt werden), um zu ermitteln, ob ein Benutzer zum Durchsuchen der sicheren Domänen berechtigt ist. Sind für eine Domäne keine Berechtigungsnachweise vorhanden oder nicht gültig, werden die Dokumente dieser Domäne aus den Suchergebnissen ausgeschlossen.

Wenn Sie weitere Informationen zur Verwendung einer Mustersuchanwendung wünschen, klicken Sie **Hilfe** an, während Sie die Anwendung verwenden. 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

"Search and index API overview" in "Programming Guide and API Reference for Enterprise Search"

Zugehörige Tasks

Zugreifen auf die Mustersuchanwendung

Bearbeiten der Merkmale der Mustersuchanwendung

Aktivieren der Sicherheit für die Mustersuchanwendung

Der Datenfluss für die Unternehmenssuche

Die Komponenten für die Unternehmenssuche, die Sie zusammen mit WebSphere Information Integrator OmniFind 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.

Mit Hilfe 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.

|

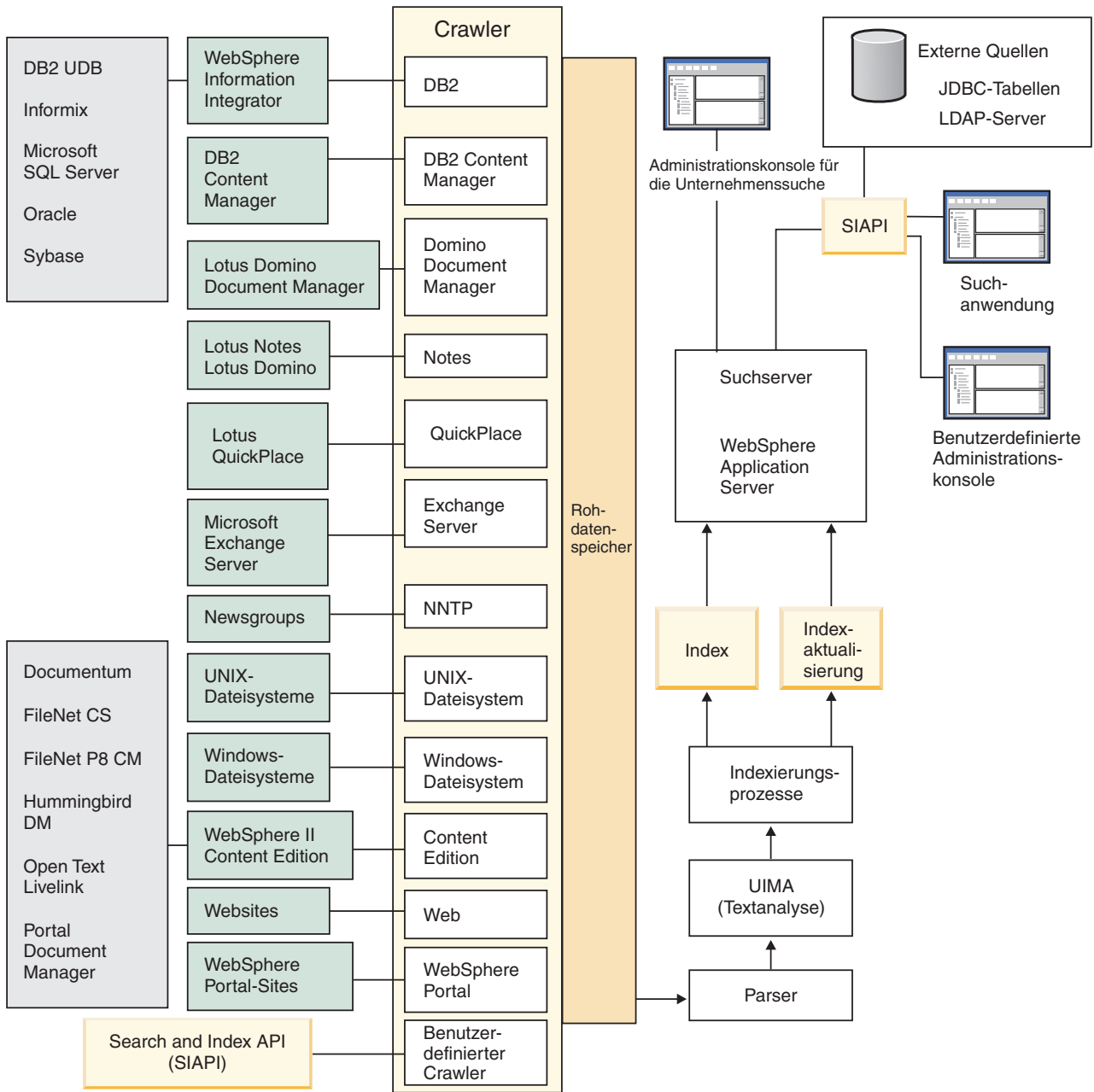


Abbildung 1. Datenfluss durch ein System für die Unternehmenssuche

Übersicht über die Verwaltung der 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.

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 im Web-Browser den Knopf **Aktualisieren** oder **Zurück** anklicken, kann es zu inkonsistenten Ergebnissen und zu Datenverlust kommen. Wenn Sie Objektgruppen konfigurieren und zur vorherigen Seite zurückkehren oder Informationen in der Administrationskonsole aktualisieren müssen, klicken Sie die Knöpfe **Zurück** und **Aktualisieren** in der Administrationskonsole für die Unternehmenssuche an Stelle der Knöpfe **Zurück** und **Aktualisieren** des Web-Browsers an.

- 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 Dokumentenebene 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, so dass diese effektiv durchsucht werden können:
 - Sie können angeben, ob XML-Dokumente syntaktisch analysiert werden sollen, so dass sie mit einer nativen XML-Suche durchsucht werden können.
 - Sie können Dokumente Kategorien zuordnen, so dass 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, so dass 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.
 - 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 das Reorganisieren und Aktualisieren des Index. Terminieren Sie häufige Indexstellungen, so dass 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, so dass Benutzer nur einen bestimmten Teil der Objektgruppe an Stelle der gesamten Dokumente im Index durchsuchen können.
 - Sie können Suchergebnisse ausblenden, so dass alle Dokumente derselben Quelle in den Suchergebnissen zusammen gruppiert und 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 sie syntaktisch analysiert werden können. Parser müssen die durchsuchten Daten analysieren, bevor sie indiziert werden können, und der Index muss aktualisiert oder reorganisiert werden, bevor die Suchserver mit der Verarbeitung der Suchanforderung beginnen 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. Nach-

dem 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 indiziert wurden.

Sicht 'System'

Wenn Sie die Rolle eines Administrators für die Unternehmenssuche haben, haben Sie durch Anklicken von **System** in der Funktionsleiste die folgenden Möglichkeiten (Administratoren für die Objektgruppe, Bediener und Überwachungsbeauftragte können auf diese Sicht nur zugreifen, wenn ihnen ein Administrator für die Unternehmenssuche die entsprechende Berechtigung erteilt):

- Überprüfen der Verfügbarkeit von Systemressourcen
- Konfigurieren von Clientanwendungen der Datenempfangsfunktion, so dass diese Objektgruppen aktualisieren können
- 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 Aktualisierungs- und Reorganisationsanforderungen für eine einzelne Objektgruppe gleichzeitig 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, so dass 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 IBM 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.

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 für die Unternehmenssuche“ auf Seite 37

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

„Überwachen der Aktivitäten der Unternehmenssuche“ auf Seite 249

Wenn Sie System- und Objektgruppenaktivitäten überwachen, können Sie den Status verschiedener Prozesse anzeigen, potenzielle Probleme aufspüren oder Konfigurationseinstellungen anpassen, um die Leistung zu verbessern.

Zugehörige Tasks

„Starten der Server für die Unternehmenssuche“ auf Seite 243

Damit Benutzer eine Objektgruppe durchsuchen können, müssen Sie die Systemprozesse und anschließend die Server für die Crawlersuche, die Syntaxanalyse, die Indexierung und die Suche für die Objektgruppe starten.

„Stoppen der Server für die Unternehmenssuche“ auf Seite 245

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.

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.

„Erstellen einer Objektgruppe mit der Sicht 'Objektgruppen'“ auf Seite 31

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.

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 WebSphere II OmniFind 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 206

Bei der Unternehmenssuche wird das Konzept der Rollen verwendet, um den Zugriff auf verschiedene Funktionen in der Administrationskonsole zu steuern.

Zugehörige Tasks

„Starten der Server für die Unternehmenssuche“ auf Seite 243

Damit Benutzer eine Objektgruppe durchsuchen können, müssen Sie die Systemprozesse und anschließend die Server für die Crawlersuche, die Syntaxanalyse, die Indexierung und die Suche für die Objektgruppe starten.

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`.

Vorbereitung

Die ID und das Kennwort für den Administrator für die Unternehmenssuche müssen auf Ihrem Betriebssystem gültig sein und für den Zugriff auf und die Konfiguration von DB2 Universal Database berechtigt sein.

Informationen zu dieser Task

Das Kennwort für die erste ID des Administrators für die Unternehmenssuche wird bei der Installation von WebSphere II OmniFind 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.

Da das Script `eschangepw` im Verzeichnis `ES_INSTALL_ROOT/bin` installiert wird, können Sie es überall im System ausführen.

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.
2. Stoppen Sie den Server, indem Sie den folgenden Befehl eingeben: `esadmin stop`
3. Öffnen Sie die Administrationskonsole von WebSphere Application Server, und stoppen Sie die Unternehmensanwendungen `server1` und `ESSearchServer`.
4. Ändern Sie das Systemkennwort für die Benutzer-ID des Administrators für die Unternehmenssuche, indem Sie den entsprechenden Befehl des Betriebssystems UNIX oder die Microsoft Windows-Funktion zur Kennwortänderung verwenden.
5. Führen Sie das folgende Script aus. Dabei ist *neuer_wert* das Kennwort, das Sie in Schritt 4 angegeben haben:

Betriebssystem	Befehl
UNIX	<code>eschangepw.sh neuer_wert</code>
Windows	<code>eschangepw neuer_wert</code>

6. Starten Sie in der Administrationskonsole von WebSphere Application Server die Anwendungen `server1` und `ESSearchServer`.
7. Starten Sie die allgemeine Übertragungsschicht (CCL = Common Communication Layer) von WebSphere II OmniFind Edition erneut, indem Sie die folgenden Befehle eingeben:

Betriebssystem	Befehle
UNIX	<code>stopccl.sh</code> , anschließend <code>startccl.sh -bg</code>
Windows-Eingabeaufforderung	<code>stopccl</code> , anschließend <code>startccl</code>

Betriebssystem	Befehle
Verwaltungstool für Windows-Services	<ol style="list-style-type: none"> 1. Öffnen Sie das Windows-Fenster Dienste. 2. Klicken Sie WebSphere Information Integrator OmniFind Edition mit der rechten Maustaste an, und wählen Sie Beenden aus. 3. Klicken Sie erneut WebSphere Information Integrator OmniFind Edition mit der rechten Maustaste an, und wählen Sie Eigenschaften aus. 4. Klicken Sie die Indexzunge Anmelden an. 5. Ändern Sie das Kennwort, und geben Sie <i>neuer_wert</i> an. Klicken Sie anschließend OK an. 6. Klicken Sie erneut WebSphere Information Integrator OmniFind Edition mit der rechten Maustaste an, und wählen Sie Starten aus.

8. Starten Sie die Unternehmenssuche erneut, indem Sie den folgenden Befehl eingeben: `esadmin start`.

Zugehörige Verweise

„Befehle, Rückkehrcodes und Sitzungs-IDs der Unternehmenssuche“ auf Seite 287

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.

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.

Vorbereitung

Die ID und das Kennwort für den Administrator für die Unternehmenssuche müssen auf Ihrem Betriebssystem gültig sein und für den Zugriff auf und die Konfiguration von DB2 Universal Database berechtigt sein.

Das Kennwort für den Administrator für die Unternehmenssuche muss auf allen Computern identisch sein, die zu dieser Installation von WebSphere II OmniFind Edition gehören.

Informationen zu dieser Task

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

Zur Änderung des Kennworts und zur Verbreitung der Änderung im gesamten System für die Unternehmenssuche müssen Sie das Script `eschangepw` auf jedem Computer in Ihrer Konfiguration mit mehreren Servern 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.

Da das Script `eschangepw` im Verzeichnis `ES_INSTALL_ROOT/bin` installiert wird, können Sie es überall im System ausführen.

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 auf dem Computer, auf dem Sie den Indexserver installiert haben, als Administrator für die Unternehmenssuche an.
2. Starten Sie den Systemwartungsmodus, indem Sie diesen Befehl eingeben:
`esadmin system maintenance`
3. Ändern Sie das Systemkennwort für die Benutzer-ID des Administrators für die Unternehmenssuche, indem Sie den entsprechenden Befehl des Betriebssystems UNIX oder die Microsoft Windows-Funktion zur Kennwortänderung verwenden.
4. Führen Sie das folgende Script aus. Dabei ist *neuer_wert* das Kennwort, das Sie in Schritt 3 angegeben haben:

Betriebssystem	Befehl
UNIX	<code>eschangepw.sh neuer_wert</code>
Windows	<code>eschangepw neuer_wert</code>

5. Starten Sie die allgemeine Übertragungsschicht (CCL = Common Communication Layer) von WebSphere II OmniFind Edition erneut, indem Sie die folgenden Befehle auf dem Indexserver eingeben:

Betriebssystem	Befehle
UNIX	<code>stopccl.sh</code> , anschließend <code>startccl.sh -bg</code>
Windows-Eingabeaufforderung	<code>stopccl</code> , anschließend <code>startccl</code>

Betriebssystem	Befehle
Verwaltungstool für Windows-Services	<ol style="list-style-type: none"> 1. Öffnen Sie das Windows-Fenster Dienste. 2. Klicken Sie WebSphere Information Integrator OmniFind Edition mit der rechten Maustaste an, und wählen Sie Beenden aus. 3. Klicken Sie erneut WebSphere Information Integrator OmniFind Edition mit der rechten Maustaste an, und wählen Sie Eigenschaften aus. 4. Klicken Sie die Indexzunge Anmelden an. 5. Ändern Sie das Kennwort, und geben Sie <i>neuer_wert</i> an. Klicken Sie anschließend OK an. 6. Klicken Sie erneut WebSphere Information Integrator OmniFind Edition mit der rechten Maustaste an, und wählen Sie Starten aus.

6. Melden Sie sich auf dem Computer, auf dem Sie den Crawler-Server installiert haben, als Administrator für die Unternehmenssuche an.
7. Ändern Sie das Systemkennwort für die Benutzer-ID des Administrators für die Unternehmenssuche, indem Sie den entsprechenden Befehl des Betriebssystems UNIX oder die Microsoft Windows-Funktion zur Kennwortänderung verwenden. Dieses Kennwort muss mit dem Kennwort übereinstimmen, das Sie in Schritt 3 auf Seite 23 angegeben haben.
8. Führen Sie das folgende Script aus. Dabei ist *neuer_wert* das Kennwort, das Sie in Schritt 3 auf Seite 23 angegeben haben:

Betriebssystem	Befehl
UNIX	<code>eschangepw.sh neuer_wert</code>
Windows	<code>eschangepw neuer_wert</code>

9. Starten Sie die allgemeine Übertragungsschicht auf dem Crawler-Server erneut, indem Sie die folgenden Befehle eingeben:

Betriebssystem	Befehle
UNIX	<code>stopccl.sh</code> , anschließend <code>startccl.sh -bg</code>
Windows-Eingabeaufforderung	<code>stopccl</code> , anschließend <code>startccl</code>

Betriebssystem	Befehle
Verwaltungstool für Windows-Services	<ol style="list-style-type: none"> 1. Öffnen Sie das Windows-Fenster Dienste. 2. Klicken Sie WebSphere Information Integrator OmniFind Edition mit der rechten Maustaste an, und wählen Sie Beenden aus. 3. Klicken Sie erneut WebSphere Information Integrator OmniFind Edition mit der rechten Maustaste an, und wählen Sie Eigenschaften aus. 4. Klicken Sie die Indexzunge Anmelden an. 5. Ändern Sie das Kennwort, und geben Sie <i>neuer_wert</i> an. Klicken Sie anschließend OK an. 6. Klicken Sie erneut WebSphere Information Integrator OmniFind Edition mit der rechten Maustaste an, und wählen Sie Starten aus.

10. Melden Sie sich auf einem der Computer, auf dem Sie einen Suchserver installiert haben, als Administrator für die Unternehmenssuche an.
11. Stoppen Sie den IBM HTTP-Web-Server. (Dadurch wird verhindert, dass der Netzdispatcher Abfragen an diesen Server weiterleitet.)
12. Stoppen Sie die allgemeine Übertragungsschicht auf dem Suchserver, indem Sie die folgenden Befehle eingeben:

Betriebssystem	Befehle
UNIX	stopccl.sh
Windows-Eingabeaufforderung	stopccl
Verwaltungstool für Windows-Services	<ol style="list-style-type: none"> 1. Öffnen Sie das Windows-Fenster Dienste. 2. Klicken Sie WebSphere Information Integrator OmniFind Edition mit der rechten Maustaste an, und wählen Sie Beenden aus.

13. Öffnen Sie die Administrationskonsole von WebSphere Application Server, und stoppen Sie die Anwendungen server1 und ESSearchServer.
14. Ändern Sie das Systemkennwort für die Benutzer-ID des Administrators für die Unternehmenssuche, indem Sie den entsprechenden Befehl des Betriebssystems UNIX oder die Microsoft Windows-Funktion zur Kennwortänderung verwenden. Dieses Kennwort muss mit dem Kennwort übereinstimmen, das Sie in Schritt 3 auf Seite 23 angegeben haben.
15. Führen Sie das folgende Script aus. Dabei ist *neuer_wert* das Kennwort, das Sie in Schritt 3 auf Seite 23 angegeben haben:

Betriebssystem	Befehl
UNIX	eschangepw.sh <i>neuer_wert</i>
Windows	eschangepw <i>neuer_wert</i>

16. Starten Sie die allgemeine Übertragungsschicht auf dem Suchserver erneut, indem Sie die folgenden Befehle eingeben:

Betriebssystem	Befehle
UNIX	startccl.sh -bg
Windows-Eingabeaufforderung	startccl
Verwaltungstool für Windows-Services	<ol style="list-style-type: none"> 1. Öffnen Sie das Windows-Fenster Dienste. 2. Klicken Sie WebSphere Information Integrator OmniFind Edition mit der rechten Maustaste an, und wählen Sie Eigenschaften aus. 3. Klicken Sie die Indexzunge Anmelden an. 4. Ändern Sie das Kennwort, und geben Sie <i>neuer_wert</i> an. Klicken Sie anschließend OK an. 5. Klicken Sie erneut WebSphere Information Integrator OmniFind Edition mit der rechten Maustaste an, und wählen Sie Starten aus.

17. Wiederholen Sie die Schritte 10 auf Seite 25 bis 16 auf dem zweiten Suchserver.
18. Starten Sie den IBM HTTP-Web-Server.
19. Starten Sie in der Administrationskonsole von WebSphere Application Server die Unternehmensanwendungen server1 und ESsearchServer.
20. Melden Sie sich auf dem Computer, auf dem Sie den Indexserver installiert haben, als Administrator für die Unternehmenssuche an, und geben Sie den folgenden Befehl ein, um die Unternehmenssuche zu starten: esadmin start
21. Melden Sie sich an der Administrationskonsole für die Unternehmenssuche an, überwachen Sie eine Objektgruppe, klicken Sie die Seite **Suche** an, und klicken Sie anschließend **Stoppen** und **Starten** an, um die Suchserver erneut zu starten.

Wiederholen Sie diesen Schritt für jede Objektgruppe in Ihrem System für die Unternehmenssuche.

Zugehörige Verweise

„Befehle, Rückkehrcodes und Sitzungs-IDs der Unternehmenssuche“ auf Seite 287

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.

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 WebSphere II OmniFind Edition. (In einer Konfiguration mit mehreren Servern wird auf allen Servern dieselbe Portnummer angegeben.)

Kann die Portnummer nicht verwendet werden (weil diese beispielsweise einem anderen Produkt auf demselben Server zugeordnet ist), wird infolge 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 - Ausnahmsbedingungsnachricht:
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. Geben Sie den folgenden Befehl ein, um das System für die Unternehmenssuche zu stoppen:
esadmin stop
2. Wechseln Sie zu dem Server, auf dem die Portnummer geändert werden muss, und stoppen Sie die allgemeine Übertragungsschicht (CCL = Common Communication Layer) durch Eingabe der folgenden Befehle:

Betriebssystem	Befehle
UNIX	stopccl.sh,
Windows-Eingabeaufforderung	stopccl
Verwaltungstool für Windows-Services	<ol style="list-style-type: none"> 1. Öffnen Sie das Windows-Fenster Dienste. 2. Klicken Sie WebSphere Information Integrator OmniFind Edition mit der rechten Maustaste an, und wählen Sie Beenden aus.

3. Bearbeiten Sie die Datei \$ES_NODE_ROOT/nodeinfo/es.cfg (unter UNIX) bzw. die Datei %ES_NODE_ROOT%\nodeinfo\es.cfg (unter Windows), 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
```

4. Starten Sie die allgemeine Übertragungsschicht erneut, indem Sie die folgenden Befehle eingeben:

Betriebssystem	Befehle
UNIX	startccl.sh -bg
Windows-Eingabeaufforderung	startccl
Verwaltungstool für Windows-Services	<ol style="list-style-type: none"> 1. Öffnen Sie das Windows-Fenster Dienste. 2. Klicken Sie WebSphere Information Integrator OmniFind Edition mit der rechten Maustaste an, und wählen Sie Starten aus.

| 5. Wechseln Sie zum Indexserver und folgen Sie den Anweisungen in Schritt 2 auf
| Seite 27, um die allgemeine Übertragungsschicht zu stoppen.

| 6. Bearbeiten Sie die Datei \$ES_NODE_ROOT/nodeinfo/es.cfg (unter UNIX) bzw. die
| Datei %ES_NODE_ROOT%\nodeinfo\es.cfg (unter Windows).

| a. Suchen Sie das folgende Merkmal, wobei *computername* für den Namen des
| Servers steht, auf dem die Portnummer in Schritt 3 auf Seite 27 geändert
| wurde. Die Angabe *N* im Merkmal *nodeN* steht für einen Wert, der den Ser-
| ver angibt.

| `nodeN.destination=computername`

| b. Suchen Sie das folgende Untermerkmal, geben Sie hier dieselbe Port-
| nummer an, die Sie im Schritt 3 auf Seite 27 für den Server angegeben
| haben, und speichern und schließen Sie die Datei:

| `nodeN.port=neue_portnummer`

| 7. Folgen Sie den Anweisungen in Schritt 4 auf Seite 27, um die allgemeine
| Übertragungsschicht erneut zu starten.

| 8. Geben Sie den folgenden Befehl ein, um das System für die Unternehmens-
| suche erneut zu starten:

| `esadmin start`

| Nach der Ausführung dieses Befehls wird die neue Portnummer auf allen Ser-
| vern für die Unternehmenssuche aktualisiert.

| **Zugehörige Verweise**

| „Befehle, Rückkehrcodes und Sitzungs-IDs der Unternehmenssuche“ auf Seite
| 287

| 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.

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 251

Sie können allgemeine Informationen zum Status jeder Komponente in einer Objektgruppe anzeigen oder Optionen auswählen, um detaillierte Informationen zu einzelnen Komponenten und URIs anzuzeigen.

„Migration einer Objektgruppe aus WebSphere Portal“ auf Seite 237

Wenn Sie Objektgruppen 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.

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.

Informationen zu dieser Task


Während Sie eine Objektgruppe erstellen, können Sie diese im Entwurfszustand speichern. Solange sie sich noch im Entwurfszustand 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 mit Hilfe 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 der geschätzten Größe der Objektgruppe ist 1 Million 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.

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 Indizierung und das Durchsuchen der Objektgruppe starten. Sie können zum Durchsuchen der neuen Objektgruppe die Mustersuchanwendung verwenden, bis Sie den Benutzern die Objektgruppe mit Ihren benutzerdefinierten Suchanwendungen verfügbar machen wollen.

Zugehörige Konzepte

„Übersicht über die Verwaltung der Unternehmenssuche“ auf Seite 15
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.

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 Reorganisieren und Aktualisieren 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 die Objektgruppe, die Sie bearbeiten wollen, in der Objektgruppenliste, und klicken Sie  **Bearbeiten** an.
3. Nehmen Sie auf den folgenden Seiten die gewünschten Änderungen vor:

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.

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 außerdem Felder Boostklassen zuordnen, um die Rangfolge von Dokumenten zu beeinflussen.

Index Geben Sie Zeitpläne zum Reorganisieren 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.

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

Crawlerverwaltung für die Unternehmenssuche

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

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 Crawleruche konfigurieren.

Indem Sie einen Crawler kopieren, können Sie die Auslastung der Crawleruche auf mehrere Crawler verteilen, die dieselben Crawlerregeln verwenden. Sie könnten z. B. einen Notes-Crawler kopieren, weil Sie dieselben Merkmale und Crawlerregeln 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, so dass 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.

Sicherheit auf Dokumentebene

Wenn Sie bei der Erstellung einer Objektgruppe die Sicherheit für diese Objektgruppe aktivieren, können Sie Optionen für die Sicherheit auf Dokumentebene konfigurieren. Jeder Crawler kann den von ihm durchsuchten Dokumenten Sicherheitstoken zuordnen. Wenn Sie sich bei der Konfiguration des Crawlers für die Sicherheit auf Dokumentebene 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 an Stelle oder zusätzlich zu der Prüfung erfolgen, die auf Sicherheitstoken im Index für die Unternehmenssuche basiert.

Terminierung von Crawlern

Crawler, die Sie für Web-, NNTP- und WebSphere Portal-Quellen erstellen, werden gleichzeitig ausgeführt. Nachdem Sie solche Crawler gestartet haben, müssen Sie diese in der Regel nicht stoppen, sofern Sie die Konfiguration des Crawlers nicht ändern.

Geben Sie für alle anderen Crawlertypen einen Zeitplan für die Crawler suche 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 Daten-

quellen 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 Crawler suche ausgeführt werden soll. Sie können eine vollständige Crawler suche aller Dokumente im Crawlerbereich, eine Crawler suche, die alle Aktualisierungen (neue, geänderte und gelöschte Dokumente) im Crawlerbereich beinhaltet, oder eine Crawler suche terminieren, bei der nur neue und geänderte Dokumente berücksichtigt werden. Eine vollständige Crawler suche dauert am längsten. Werden bei einer Crawler suche gelöschte Dokumente entfernt, dauert diese länger als eine Crawler suche, bei der gelöschte Dokumente ignoriert werden.

Beim Bearbeiten des Crawlerbereichs eines Crawlers können Sie einen zweiten Zeitplan für die Crawler suche 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.

Zugehörige Konzepte

„Übersicht über die Verwaltung der Unternehmenssuche“ auf Seite 15
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.

„Sicherheit auf Dokumentebene“ auf Seite 214

Wenn die Sicherheit bei der Erstellung einer Objektgruppe aktiviert ist, können Sie Sicherheitsmaßnahmen auf Dokumentebene konfigurieren. Die Sicherheit auf Dokumentebene sorgt dafür, dass Benutzer, die Objektgruppen durchsuchen, nur auf die Dokumente zugreifen können, für deren Anzeige sie die Berechtigung haben.

Zugehörige Tasks

„Überwachen von Crawlern“ auf Seite 254

Sie können allgemeine Informationen zum Status der einzelnen Crawler in der Objektgruppe anzeigen oder Optionen auswählen, um detaillierte Informationen zur Crawleraktivität anzuzeigen.

Zugehörige Verweise

„URI-Formate in einem Index für die Unternehmenssuche“ auf Seite 103

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.

Erstellen eines Crawlers

Beim Erstellen eines Crawlers geben Sie an, welchen Crawlertyp Sie erstellen wollen. Mit Hilfe 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 mit Hilfe 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 Sie diesen reorganisieren.

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

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 Repositorys von Documentum, FileNet Panagon Content Services, FileNet P8 Content Manager, Hummingbird Document Management (DM), OpenText Livelink und Portal Document Manager (PDM) durchsuchen.

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.

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 Repositorys zuzugreifen.
- Definieren eines Zeitplans für das Durchsuchen der Repositorys.
- 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 Repositorys 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. Mit Hilfe dieser Daten können Suchanwendungen die Zugriffssteuerung auf Basis der Zugriffssteuerungslisten oder Sicherheitstoken erzwingen.

Für Documentum-, FileNet Panagon Content Services- und Portal Document Manager-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 Berechtigungsnachweise 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 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. Mit Hilfe dieser Methode können Sie die Vorteile von J2EE-Anwendungsserverumgebungen 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 im Lieferumfang von WebSphere Information Integrator OmniFind Edition enthalten ist, ermöglicht dem Content Edition-Crawler Zugriff auf die Repositorys auf dem Server.

Führen Sie die Ihrer Umgebung entsprechende Task aus, bevor Sie mit der Administrationskonsole für die Unternehmenssuche einen Content Edition-Crawler so konfigurieren, dass dieser im Servermodus arbeitet:

- „Konfigurieren des Crawler-Servers unter UNIX für WebSphere II Content Edition“ auf Seite 45.
- „Konfigurieren des Crawler-Servers unter Windows für WebSphere II Content Edition“ auf Seite 46.

Zugriff auf WebSphere II Content Edition-Repositorys im Direktmodus

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

Im Direktmodus verwendet der Crawler einen WebSphere Information Integrator Content Edition-Connector, der bei der Installation von WebSphere II OmniFind Edition auf dem Crawler-Server installiert wird.

Vorgehensweise

Gehen Sie wie folgt vor, um das System so zu konfigurieren, dass im Direktmodus auf Repositorys zugegriffen werden 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.bat` (unter UNIX) bzw. der Datei `ice_installationsroot\bin\Admin.bat` (unter Windows) das Java-Systemmerkmal `vbr.as.operationMode=direct` hinzu.
3. Starten Sie die WebSphere Information Integrator Content Edition-Administrationskonsole im Direktmodus, und konfigurieren Sie den Connector für den WebSphere II OmniFind Edition-Crawler-Server. (Anweisungen hierzu finden Sie in der WebSphere Information Integrator Content Edition-Dokumentation.)

4. Wählen Sie die Option für den Direktmodus aus, wenn Sie zur Konfiguration des Content Edition-Crawlers die WebSphere II OmniFind Edition-Administrationskonsole verwenden.

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

Wenn Sie WebSphere II OmniFind 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. Mit Hilfe 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 WebSphere II OmniFind 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 WebSphere II OmniFind 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 WebSphere II OmniFind 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 Bedienung:
3. Stoppen Sie das System für die Unternehmenssuche und alle Sitzungen auf der allgemeinen Übertragungsschicht (CCL = Common Communication Layer) der Unternehmenssuche, und starten Sie sie erneut:

```
esadmin stop
stopccl.sh
startccl.sh -bg
esadmin start
```

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 WebSphere II OmniFind 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. Mit Hilfe 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 WebSphere II OmniFind 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 WebSphere II OmniFind 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 WebSphere II OmniFind 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 WebSphere II OmniFind Edition angegeben).

- b. Starten Sie das folgende Script, das im Verzeichnis %ES_INSTALL_ROOT%\bin installiert wurde, und folgen Sie der Bedienerführung:
escrvbr.vbs
3. Stoppen Sie das System für die Unternehmenssuche und alle Sitzungen auf der allgemeinen Übertragungsschicht (CCL = Common Communication Layer) der Unternehmenssuche, und starten Sie sie erneut:
 - a. Stoppen Sie das System für die Unternehmenssuche an einer Eingabeaufforderung:
esadmin stop
 - b. Wählen Sie **Start** → **Programme** → **Verwaltung** → **Dienste** aus, und starten Sie anschließend den Service "IBM WebSphere Information Integrator OmniFind Edition" erneut.
 - c. Starten Sie das System für die Unternehmenssuche an einer Eingabeaufforderung:
esadmin start
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 *iice_installationsroot\lib*. Dabei steht *iice_installationsroot* für das WebSphere Information Integrator Content Edition-Installationsverzeichnis auf dem Crawler-Server.

DB2-Crawler

Sie verwenden den DB2-Crawler, um IBM DB2 Universal Database-Datenbanken in eine Objektgruppe einzuschließen. Sie können mit dem DB2-Crawler auch Kurznamentabellen einschließen, die Sie für IBM DB2 Universal Database für z/OS-, IBM Informix-, Oracle- und Microsoft SQL Server-Datenbanken erstellen.

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.

Als Tabellen für die Crawleruche sollten Sie Datenbanktabellen, Kurznamentabellen oder Sichten auswählen. Der DB2-Crawler unterstützt keine verknüpften Tabellen.

Ereignisveröffentlichung

Wenn Sie WebSphere Information Integrator Event Publisher Edition verwenden und die zu durchsuchenden Datenbanken Veröffentlichungswarteschlangenmasken zuordnen, kann der DB2-Crawler Aktualisierungen der Datenbanktabellen mit Hilfe 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 52.
- „Konfigurieren von WebSphere Information Integrator Event Publisher Edition für DB2-Crawler“ auf Seite 49.
- „Konfigurieren des Crawler-Servers unter UNIX für die Verwendung der Ereignisveröffentlichung“ auf Seite 54.
- „Konfigurieren des Crawler-Servers unter Windows für die Verwendung der Ereignisveröffentlichung“ auf Seite 55.

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.

- Auswählen der zu durchsuchenden Datenbank.
- 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.

- Wählen Sie die Tabellen aus, die durchsucht werden sollen, wenn in einer Ereignisveröffentlichungswarteschlange Aktualisierungen zu diesen Tabellen veröffentlicht werden, und geben Sie Informationen an, 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 Dokumentebene. Wenn die Sicherheit bei der Erstellung der Objektgruppe aktiviert war, kann der Crawler Sicherheitsdaten Dokumenten im Index zuordnen. Mit Hilfe 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 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:

Betriebssystem	Befehl
UNIX	db2rc
Windows	Wählen Sie Start → IBM DB2-Replikationszentrale aus.

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 den Knopf 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 Nachrichteninhaltenstyp 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-Skripts 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-Skripts 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 Veröffentlichen 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-Skripts 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-Skripts 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:
`asncap 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 UDB, 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 Universal Database (DB2 UDB), von WebSphere MQ und von WebSphere II OmniFind 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.

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:
`runmqtsr -m QM1 -t TCP -p 1414 &`
3. Geben Sie die folgenden Befehle ein, um einen DB2 UDB-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:

ren. 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
```

Konfigurieren des Crawler-Servers unter UNIX für die Verwendung der Ereignisveröffentlichung

Wenn Sie WebSphere II OmniFind Edition auf einem Computer installieren, auf dem IBM AIX, Linux oder die Solaris-Betriebsumgebung ausgeführt wird, und den DB2-Crawler so konfigurieren, dass dieser die Ereignisveröffentlichung verwendet, müssen Sie ein Script ausführen, um den Crawler-Server zu konfigurieren. Mit Hilfe des Scripts kann der Crawler auf WebSphere MQ-Warteschlangenmanager und -Warteschlangen zugreifen.

Informationen zu dieser Task

Der DB2-Crawler verwendet die WebSphere MQ 5.3-Module für Java Messaging, um auf WebSphere MQ-Warteschlangenmanager und -Warteschlangen zuzugreifen. Sie müssen diese Module auf dem Crawler-Server installieren.

Nach der Installation der WebSphere MQ-Module führen Sie auf dem Crawler-Server ein Setup-Script aus, das von WebSphere II OmniFind Edition bereitgestellt wird, um sicherzustellen, dass der DB2-Crawler die Ereignisveröffentlichung verwenden kann.

Vorgehensweise

Gehen Sie wie folgt vor, um den Crawler-Server für die Verwendung der Ereignisveröffentlichung zu konfigurieren:

1. 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 WebSphere II OmniFind 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 und alle Sitzungen auf der allgemeinen Übertragungsschicht (CCL = Common Communication Layer) der Unternehmenssuche, und starten Sie sie erneut:


```
esadmin stop
stopccl.sh
startccl.sh -bg
esadmin start
```

Konfigurieren des Crawler-Servers unter Windows für die Verwendung der Ereignisveröffentlichung

Wenn Sie WebSphere II OmniFind Edition auf einem Microsoft Windows-Computer installieren und den DB2-Crawler so konfigurieren, dass dieser die Ereignisveröffentlichung verwendet, müssen Sie ein Script ausführen, um den Crawler-Server zu konfigurieren. Mit Hilfe des Scripts kann der Crawler auf WebSphere MQ-Warteschlangenmanager und -Warteschlangen zugreifen.

Informationen zu dieser Task

Der DB2-Crawler verwendet die WebSphere MQ 5.3-Module für Java Messaging, um auf WebSphere MQ-Warteschlangenmanager und -Warteschlangen zuzugreifen. Sie müssen diese Module auf dem Crawler-Server installieren.

Nach der Installation der WebSphere MQ-Module führen Sie auf dem Crawler-Server ein Setup-Script aus, das von WebSphere II OmniFind Edition bereitgestellt wird, um sicherzustellen, dass der DB2-Crawler die Ereignisveröffentlichung verwenden kann.

Vorgehensweise

Gehen Sie wie folgt vor, um den Crawler-Server für die Verwendung der Ereignisveröffentlichung zu konfigurieren:

1. 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 WebSphere II OmniFind 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 und alle Sitzungen auf der allgemeinen Übertragungsschicht (CCL = Common Communication Layer) der Unternehmenssuche, und starten Sie sie erneut:
 - a. Stoppen Sie das System für die Unternehmenssuche an einer Eingabeaufforderung:

```
esadmin stop
```
 - b. Wählen Sie **Start** → **Programme** → **Verwaltung** → **Dienste** aus, und starten Sie anschließend den Service "IBM WebSphere Information Integrator OmniFind Edition" erneut.
 - c. Starten Sie das System für die Unternehmenssuche an einer Eingabeaufforderung:

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 WebSphere Information Integrator OmniFind 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“ auf Seite 57.
- „Konfigurieren des Crawler-Servers unter Windows für DB2 Content Manager“ auf Seite 58.

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 Dokumentenebene. Wenn die Sicherheit bei der Erstellung der Objektgruppe aktiviert war, kann der Crawler Sicherheitsdaten Dokumenten im Index zuordnen. Mit Hilfe 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 Content Manager

Wenn Sie WebSphere II OmniFind 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. Mit Hilfe 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 höher auf dem Crawler-Server installieren. Nach der Installation des Connectors führen Sie auf dem Crawler-Server ein Setup-Script aus, das von WebSphere II OmniFind 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/sql1lib/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
```

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 SConnectDisconnect 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 und alle Sitzungen auf der allgemeinen Übertragungsschicht (CCL = Common Communication Layer) der Unternehmenssuche, und starten Sie sie erneut:

```
esadmin stop
stopccl.sh
startccl.sh -bg
esadmin start
```

Konfigurieren des Crawler-Servers unter Windows für DB2 Content Manager

Wenn Sie WebSphere II OmniFind Edition auf einem Microsoft Windows-Computer installieren, müssen Sie ein Script ausführen, um den Crawler-Server zu konfigurieren. Mit Hilfe 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 Con-

tent Version 8.2 oder höher auf dem Crawler-Server installieren. Nach der Installation des Connectors führen Sie auf dem Crawler-Server ein Setup-Script aus, das von WebSphere II OmniFind 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
```

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

```
cmbenv81.bat  
cd installationsverzeichnis_für_Information_Integrator_for_Content  
  \samples\java\icm  
javac *.java  
java SConnectDisconnect icm-dbname cm-admin-id cm-admin-kennwort
```

4. 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.vbs
5. Stoppen Sie das System für die Unternehmenssuche und alle Sitzungen auf der allgemeinen Übertragungsschicht (CCL = Common Communication Layer) der Unternehmenssuche, und starten Sie sie erneut:

- a. Stoppen Sie das System für die Unternehmenssuche an einer Eingabeaufforderung:

```
esadmin stop
```
- b. Wählen Sie **Start** → **Programme** → **Verwaltung** → **Dienste** aus, und starten Sie anschließend den Service "IBM WebSphere Information Integrator OmniFind Edition" erneut.
- c. Starten Sie das System für die Unternehmenssuche an einer Eingabeaufforderung:

```
esadmin start
```

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 Protokoll NRPC (Notes Remote Procedure Call) verwendet, müssen Sie ein Script auf dem Crawler-Server ausführen. Dieses Script, das im Lieferumfang von WebSphere Information Integrator OmniFind 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 Protokoll DIIOP (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.

Wenn WebSphere II OmniFind 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 67.
- „Konfigurieren des Crawler-Servers unter Windows für das Durchsuchen von Lotus Domino-Quellen“ auf Seite 69.
- „Konfigurieren von Servern, die das Protokoll DIIOP verwenden“ auf Seite 70.
- „Konfigurieren des I/O Completion Ports unter AIX für das Durchsuchen von Lotus Domino-Quellen“ auf Seite 71.

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.
- 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 Dokumentenebene. Wenn die Sicherheit bei der Erstellung der Objektgruppe aktiviert war, kann der Crawler Sicherheitsdaten Dokumenten im Index zuordnen. Mit Hilfe 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.

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, so dass 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 Dokumententypen 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. Mit Hilfe 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.

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.

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 Newsgroups auf jedem Server aus, die Sie durchsuchen wollen. Außerdem können Sie Muster für die Newsgroups angeben, die Sie ausschließen wollen. 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 von Mustern zum Einschließen von Newsgroups in den Crawlerbereich und Angeben von Mustern zum Ausschließen von bestimmten Newsgroups aus dem Crawlerbereich.
- 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. Mit Hilfe 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.

Crawler-Server-Konfiguration

Wenn ein Lotus Notes-Server, den Sie durchsuchen wollen, das Protokoll NRPC (Notes remote procedure call) verwendet, müssen Sie ein Script auf dem Crawler-Server ausführen. Dieses Script, das im Lieferumfang von WebSphere Information Integrator OmniFind 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 Protokoll DIIOP (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.

Wenn WebSphere II OmniFind 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 67.
- „Konfigurieren des Crawler-Servers unter Windows für das Durchsuchen von Lotus Domino-Quellen“ auf Seite 69.
- „Konfigurieren von Servern, die das Protokoll DIIOP verwenden“ auf Seite 70.
- „Konfigurieren des I/O Completion Ports unter AIX für das Durchsuchen von Lotus Domino-Quellen“ auf Seite 71.

Sicherheit auf Dokumentenebene

Wenn die Objektgruppensicherheit aktiviert ist und ein Lotus Notes-Server, den Sie durchsuchen wollen, das Protokoll NRPC 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“ auf Seite 222.
- Aktivieren der globalen Sicherheit in WebSphere Application Server und Konfigurieren der Suchanwendung für die Verwendung der Sicherheit. Durch diesen Schritt wird sichergestellt, dass Benutzer zur Angabe von Berechtigungsnachweisen aufgefordert werden, wenn sie versuchen, die Suchanwendung zu verwenden. Mit Hilfe dieser Berechtigungsnachweise können die Suchserver die Berechtigung eines Benutzers auf Lotus Notes-Dokumente zuzugreifen prüfen.

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.
- Auswählen der zu durchsuchenden Datenbanken oder Verzeichnisse.
- 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. Mit Hilfe 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.

Zugehörige Konzepte

„Umsetzung der Sicherheit auf Dokumentenebene für Lotus Domino-Dokumente“ auf Seite 221

Wenn der Lotus Notes-Server, der durchsucht werden soll, das Protokoll NRPC (Notes Remote Procedure Call) verwendet, müssen Sie den Crawler-Server entsprechend konfigurieren, damit die Zugriffssteuerungen auf Dokumentenebene umgesetzt werden können.

Zugehörige Tasks

„Konfigurieren der Lotus Domino-Trusted Server, um Benutzerberechtigungen zu prüfen“ auf Seite 222

Die zu durchsuchenden Domino-Server müssen als Lotus Domino-Trusted Ser-

ver konfiguriert werden, um die Sicherheit für Dokumente umzusetzen, die von einem Notes-Crawler durchsucht wurden, der das Protokoll NRPC (Notes Remote Procedure Call) verwendet.

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 1. 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.

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

Wenn Sie WebSphere II OmniFind Edition auf einem Computer installieren, auf dem IBM AIX, Linux oder die Solaris-Betriebsumgebung ausgeführt wird, und Server durchsuchen möchten, die das Protokoll NRPC (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 Protokoll NRPC verwenden, nutzen Domino-Bibliotheken als Client. Sie installieren diese Bibliotheken, indem Sie Lotus Domino Server Version 6.0.2 oder höher auf dem Crawler-Server installieren. Nach der Installation der Domino-Bibliotheken führen Sie auf dem Crawler-Server ein Setup-Script aus, das von WebSphere II OmniFind 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 notes und die Gruppe notes auf dem Crawler-Server:
 - a. Melden Sie sich als Root an:

```
su - root
```
 - b. Fügen Sie einen Benutzer hinzu:

```
useradd notes
```
 - c. Fügen Sie für diesen Benutzer ein Kennwort hinzu:

```
passwd notes
```

Sie werden zur Änderung des Kennworts aufgefordert.
2. Installieren Sie Lotus Domino Server auf dem Crawler-Server:
 - a. Legen Sie die Domino Server-CD Version 6.0.2 oder höher ein, und hängen Sie diese 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.
3. Führen Sie auf dem Crawler-Server das von WebSphere II OmniFind Edition bereitgestellte Setup-Script aus:
 - a. Melden Sie sich als Administrator für die Unternehmenssuche an (diese Benutzer-ID wurde während der Installation von WebSphere II OmniFind 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 und alle Sitzungen auf der allgemeinen Übertragungsschicht (CCL = Common Communication Layer) der Unternehmenssuche, und starten Sie sie erneut:


```
esadmin stop
stopccl.sh
startccl.sh -bg
esadmin start
```

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

Wenn Sie WebSphere II OmniFind Edition auf einem Microsoft Windows-Computer installieren und Server durchsuchen möchten, die das Protokoll NRPC (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 Protokoll NRPC verwenden, arbeiten mit Lotus Notes-Clientbibliotheken. Sie installieren diese Bibliotheken, indem Sie Lotus Notes Release 6.0.2 oder höher auf dem Crawler-Server installieren. Nach der Installation der Lotus Notes-Clientbibliotheken führen Sie auf dem Crawler-Server ein Setup-Script aus, das von WebSphere II OmniFind Edition bereitgestellt wird, um sicherzustellen, dass die Crawler mit den Lotus Notes-Clientbibliotheken 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 Lotus Notes-CD Release 6.0.2 oder höher 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 Notes-Dokumentation, wenn Sie Unterstützung benötigen.
3. Führen Sie auf dem Crawler-Server das von WebSphere II OmniFind Edition bereitgestellte Setup-Script aus:
 - a. Melden Sie sich mit der ID des Administrators für die Unternehmenssuche an (diese Benutzer-ID wurde während der Installation von WebSphere II OmniFind 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 und alle Sitzungen auf der allgemeinen Übertragungsschicht (CCL = Common Communication Layer) der Unternehmenssuche, und starten Sie sie erneut:
 - a. Stoppen Sie das System für die Unternehmenssuche an einer Eingabeaufforderung:
esadmin stop
 - b. Wählen Sie **Start** → **Programme** → **Verwaltung** → **Dienste** aus, und starten Sie anschließend den Service "IBM WebSphere Information Integrator OmniFind Edition" erneut.
 - c. Starten Sie das System für die Unternehmenssuche an einer Eingabeaufforderung:
esadmin start

Konfigurieren von Servern, die das Protokoll DIIOOP verwenden

Zum Durchsuchen von Servern, die das Protokoll DIIOOP (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 DIIOOP- und HTTP-Tasks ausführen.

Vorgehensweise

Gehen Sie wie folgt vor, um Server zu konfigurieren, die das Protokoll DIIOOP 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 Protokoll DIIOP 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 mit Hilfe dieser Berechtigungsnachweise auf den Server zu.
3. Starten Sie die DIIOP-Task auf dem Server erneut.

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_wwwjmtt.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.

QuickPlace-Crawler

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

Crawler-Server-Konfiguration

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

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

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

Wenn WebSphere II OmniFind 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 67.
- „Konfigurieren des Crawler-Servers unter Windows für das Durchsuchen von Lotus Domino-Quellen“ auf Seite 69.
- „Konfigurieren von Servern, die das Protokoll DIIOP verwenden“ auf Seite 70.

- „Konfigurieren des QuickPlace-Servers für die Verwendung der Sicherheit "Local User" (Lokaler Benutzer)" auf Seite 75.
- „Konfigurieren von Directory Assistance auf einem QuickPlace-Server" auf Seite 76.
- „Konfigurieren des I/O Completion Ports unter AIX für das Durchsuchen von Lotus Domino-Quellen" auf Seite 71.

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 Räume in den für die Crawleruche ausgewählten Bereichen oder nur bestimmte Räume durchsucht werden sollen.

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.
- 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. Mit Hilfe 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.

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 verwendenden 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/bereichsname/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 mit Hilfe 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** (Berechtigungsnachweise 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 Verzeichnisverwaltungsdatenbank) 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.

Crawler für das UNIX-Dateisystem

Wenn Sie Dokumente einschließen möchten, die in einer Objektgruppe für die Unternehmenssuche in UNIX-Dateisystemen gespeichert sind, müssen Sie einen Crawler für das UNIX-Dateisystem konfigurieren.

Sie können mit dem Crawler für das UNIX-Dateisystem eine beliebige Anzahl UNIX-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 UNIX-Dateisystemquellen verwenden (der Crawler für das UNIX-Dateisystem wird in der Liste der verfügbaren Crawlertypen nicht angezeigt).

Wenn Sie einen UNIX-Dateisystem-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 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. Mit Hilfe 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.

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 Crawleruche 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. Mit Hilfe 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

Der Web-Crawler hält das Protokoll zum Ausschluss von Robots ein. Wenn Sie eine Website durchsuchen möchten, die dieses Protokoll 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 Ausschluss von Robots ist nicht verbindlich. Der Web-Crawler für die Unternehmenssuche hält es jedoch ein.

- 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 Ausschluss von 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 Ausschluss von 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 Ausschluss von 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.

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 Dokument-

inhalt (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, wird vom Web-Crawler nicht erfasst und somit nicht indiziert.

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. Crawlersuche 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 Crawlersuche 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 Crawlersuche beginnen soll
- Drei Typen von Crawlersuchregeln: 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

Crawlersuchregeln 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
```

Das erste Zeichen im Ziel kann ein Stern sein, der bewirkt, dass die Regel auf jeden Hostnamen angewendet wird, der mit dem Rest des Musters übereinstimmt. Sie können z. B. angeben, dass keine Domänen, die mit dem folgenden Muster übereinstimmen, durchsucht werden:

```
forbid domain *.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 Dateismounts 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 (nur IPv4) und die Netzmaske als das Ziel angeben. Beispiel:

```
allow address 9.0.0.0 255.0.0.0
```

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.

Eine nützliche Regel, die Sie als letzte Adresse in Ihrer Liste mit Regeln einschließen sollten, ist `forbid address 0.0.0.0 0.0.0.0`. 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).

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 Rückkehrcode 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 Crawlerforschungsprozesses 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 Domänennamen 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.

Die Adressenregeln enden normalerweise mit der folgenden Regel, die verhindert, dass der Crawler Websites außerhalb des Unternehmensnetzes durchsucht.

```
forbid 0.0.0.0 0.0.0.0
```

- 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 Crawler-suchtiefe 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 Crawler-suche startet. Diese URL-Adressen werden jedesmal in die Crawler-suche 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 Crawler-suche nicht bei einer angegebenen Start-URL-Adresse beginnen, wenn die Crawlersuchregeln nicht zulassen, dass diese URL-Adresse durchsucht wird.

Testen von URL-Verbindungen mit dem Web-Crawler

Nach der Angabe von URL-Adressen für den für die Crawler-suche zu verwenden- den Web-Crawler, können Sie die Konfiguration der Crawlersuchregeln testen.

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 auf Grund 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-Rückkehrcode enthalten (der angibt, ob eine Crawler-suche 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, so dass sie beim nächsten Crawlerstart durchsucht werden kann.

Stellen Sie sicher, dass die Crawlersuchregeln 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 Crawlersuchregel vorhanden sein, die zulässt, dass die URL-Adresse durchsucht werden kann.

Die neu durchsuchten Daten werden für die Suche verfügbar, wenn der Index das nächste Mal reorganisiert oder aktualisiert 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 Rückkehrcode besteht. Der Rückkehrcode 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. An Stelle eines Headers mit der Angabe des Problems empfängt der Crawler beispielsweise eine Seite mit einem normalen Fehler und den Rückkehrcode 200, der angibt, dass eine gültige HTML-Seite erfolgreich heruntergeladen wurde. Diese heruntergeladene Seite mit einem normalen Fehler bezieht sich jedoch nicht auf die angeforderte URL, und ihr Inhalt ist jedes Mal fast identisch, wenn sie an Stelle 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 Rückkehrcode 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 (Rückkehrcode 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 Rückkehrcode 404 (die Seite wurde nicht gefunden).

Tabelle 2. Beispiel einer Seite mit einem normalen Fehler

URL-Muster	Titelmuster	Inhaltsmuster	HTTP-Rückkehrcode
<code>http://www.mysite.com/hr/*</code>	Sorry, the page*	*	404

Sie können mehrere Einträge für dieselbe Website erstellen, um verschiedene Rückkehrcodes zu handhaben. Die einzelnen Rückkehrcodes 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*

Leistungseinfluss

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 Ersatzrückkehrcode 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 ange-

ben, damit Seiten auf den kennwortgeschützten Websites durchsucht werden können. Sie können auch Optionen für die manuelle Konfiguration von Cookiedateien 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 Rückkehrcode 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, so dass der berechtigte Benutzer auf weitere eingeschränkt zugängliche Inhalte Zugriff hat, ohne die Berechtigungsnachweise bei jeder Anforderung angeben zu müssen.

Nachdem die Identität eines Benutzers festgestellt wurde, tauschen der Web-Server und der HTTP-Benutzeragent in der Regel Token aus, so genannte *Cookies*, so dass 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 Crawleruche 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-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.

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.

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 Crawleruche ohne Proxy-Server wird eine URL-Anforderung direkt an den Host gesendet. Bei einer Crawleruche 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 verwenden soll, wenn Seiten in der Domäne des Proxy-Servers durchsucht werden sollen. Sie benötigen die folgenden Informationen, bevor Sie dem Crawlerbereich einen Proxy-Server hinzufügen:

Proxy-Server-Domänen

Die Domänen, die vom Proxy-Server bereitgestellt werden. Sie können einen Stern (*) als Platzhalterzeichen verwenden. Beispiel: * entspricht allen Domänen, die von diesem Proxy-Server bereitgestellt werden, und *.resource.com entspricht allen Domänen, die mit resource.com enden.

Einschränkung: IP-Adressenregeln können nicht zum Durchsuchen eines Proxy-Servers angegeben werden, da die IP-Adresse des Proxy-Servers die einzige IP-Adresse ist, 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 Rückkehrcode 760 klassifiziert werden (dadurch wird angegeben, dass sie durch den Web-Space verboten sind).

Hostname oder IP-Adresse des Proxy-Servers

Der DNS-Hostname oder die IP-Adresse des Proxy-Servers mit Punkten als Trennzeichen.

Portnummer des Proxy-Servers

Die Nummer des TCP/IP-Ports, an dem der Proxy-Server für HTTP-Proxy-Anforderungen empfangsbereit ist.

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 gene-

rischeren 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 auf Grund 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 Datenschutzeinschränkungen 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=SLNSIDFNLSIDNFLSIDNFLSNL;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 WebSphere II OmniFind 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. auf Grund 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-Rückkehrcodes 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. Auf Grund 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 auf Grund 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 auf Grund einer lokalen Präfix- oder Adressenregel aus der Datenbank des Crawlers ausgeschlossen wird, wird der URL der Rückkehrcode 760 zugeordnet, und sie wird aus dem Index entfernt. Die URL wird nicht mehr durchsucht.
2. Wenn eine URL auf Grund einer lokalen Domänenregel aus der Datenbank des Crawlers ausgeschlossen wird und es keinen globalen Crawlerbereich gibt, wird der URL der Rückkehrcode 760 zugeordnet, und sie wird aus dem Index entfernt. Die URL wird nicht mehr durchsucht.
3. Wenn eine URL auf Grund einer lokalen Domänenregel aus der Datenbank des Crawlers ausgeschlossen wird, jedoch explizit durch eine Regel im globalen Crawlerbereich zugelassen ist, wird der URL der Rückkehrcode 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 auf Grund 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 Rückkehrcode 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 während der Installation von WebSphere II OmniFind 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`.
5. 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
```

6. Speichern und schließen Sie die Datei.

7. Starten Sie den gestoppten Web-Crawler über die Administrationskonsole für die Unternehmenssuche erneut.

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 die Unternehmensanwendung `ESPACServer.ear` in WebSphere Portal implementieren. Diese Unternehmensanwendung wird bei der Installation von WebSphere Information Integrator OmniFind Edition auf den Suchservern installiert. Führen Sie die folgende Task aus, um diese Unternehmensanwendung zu implementieren:

- „Implementieren der Unternehmensanwendung für den WebSphere Portal-Crawler“ auf Seite 97.

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 herunter, die auf dem angegebenen Portal verfügbar sind.

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.
- 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. Mit Hilfe 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.

Implementieren der Unternehmensanwendung für den WebSphere Portal-Crawler

Bevor Sie einen WebSphere Portal-Crawler erstellen, müssen Sie die Unternehmensanwendung `ESPACServer.ear` in WebSphere Portal implementieren.

Informationen zu dieser Task

Die Datei `ESPACServer.ear` wird bei der Installation von WebSphere II OmniFind Edition im Verzeichnis `ES_INSTALL_ROOT/bin` auf den Suchservern installiert. Die Standardinstallationspfade lauten wie folgt:

UNIX-Systeme:

`/opt/IBM/es/bin/ESPACServer.ear`

Windows-Systeme:

`C:\Programme\IBM\es\bin\ESPACServer.ear`

Vorgehensweise

Gehen Sie wie folgt vor, um die Unternehmensanwendung zu implementieren, die WebSphere Portal-Crawlern das Durchsuchen von WebSphere Portal-Sites ermöglicht:

1. Stoppen Sie die Serverinstanz `WebSphere_Portal`.
2. Starten Sie die WebSphere Application Server-Serverinstanz `server1`.
3. Starten Sie auf dem WebSphere Portal-Server die Administrationskonsole von WebSphere Application Server. Melden Sie sich an, wenn Sie dazu aufgefordert werden.

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.
4. Klicken Sie **Anwendungen** und anschließend **Neue Anwendung installieren** an.
 5. Klicken Sie **Durchsuchen** an, und wählen Sie auf Ihrem System die Datei `ESPACServer.ear` aus.
 6. Klicken Sie **Weiter** zweimal an. Wenn Sie eine Warnung zu den Richtliniendateien erhalten, klicken Sie **Weiter** an.
 7. Klicken Sie **Weiter** an, bis die Seite **Module zu Anwendungsservern zuordnen** angezeigt wird. Nehmen Sie dann die folgenden Einstellungen vor, und klicken Sie **Anwenden** an:
 - a. Wählen Sie **WebSphere:cell=zellenname, node=knotenname, server=WebSphere Portal** im Textfenster **Cluster und Server** aus.

- b. Wählen Sie das Markierungsfeld neben dem Modul **ESPACTServer.ear** aus.
8. Klicken Sie so lange **Weiter** an, bis die Seite **Zusammenfassung** angezeigt wird. Klicken Sie dann **Fertig stellen** an.
9. Klicken Sie den Link **In Master-Konfiguration speichern** und dann den Knopf **Speichern** an, um Ihre Änderungen in der WebSphere Application Server-Konfiguration zu speichern.
10. Starten Sie den WebSphere Portal-Server erneut.

Zugehörige Konzepte

„Integration der Unternehmenssuche in WebSphere Portal“ auf Seite 225
Sie können die Suchfunktionalitäten von IBM WebSphere Portal erweitern, indem Sie die Portlets für die Unternehmenssuche in WebSphere Portal und die WebSphere Portal-Suchzentrale implementieren.

Kopieren der zu durchsuchenden URL-Adresse von WebSphere Portal

Sie können die URL-Adresse der zu durchsuchenden WebSphere Portal-Site kopieren und bei der Konfiguration eines WebSphere Portal-Crawlers in das entsprechende Feld einfügen, um die Eingabe einer falschen URL-Adresse zu vermeiden.

Informationen zu dieser Task

Beim Erstellen eines WebSphere Portal-Crawlers geben Sie die URL-Adresse des Portals 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 vom WebSphere Portal-Crawler durchsucht werden soll:

1. Wenn die Seite **Zu durchsuchender WebSphere Portal-Server** in der Administrationskonsole für die Unternehmenssuche angezeigt wird, stellen Sie sicher, dass der WebSphere Portal-Server gestartet ist, und melden Sie sich bei WebSphere Portal als Administrator an.
2. Wählen Sie **Verwaltung** in der rechten oberen Ecke aus.
3. Wählen Sie **Portaleinstellungen** im Navigationsbereich auf der linken Seite aus, und wählen Sie dann **Verwaltung der Suchfunktion** aus.
4. Wählen Sie **PortalCollection** im Bereich **Suchsammlungen** der Seite **Suchsammlungen verwalten** aus. (Sie können auch eine andere Objektgruppe auswählen, sofern welche zur Verfügung stehen.)
5. Klicken Sie **Inhaltsquelle hinzufügen** im Bereich **Inhaltsquellen in der Objektgruppe** an.
6. Wählen Sie **Portalsite** für die Option **Quellentyp durchlaufen und erfassen** aus. Die URL-Adresse der Site wird im Feld **Über diese URL verknüpfte Dokumente abrufen** angezeigt.
7. 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.)

8. Kehren Sie zur Administrationskonsole für die Unternehmenssuche zurück, und fügen Sie die kopierte URL-Adresse in das Feld **URL der WebSphere Portal-Site** ein.

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 einem UNIX-Computer installieren, können Sie diesen Server nicht zum Durchsuchen von Windows-Dateisystemquellen verwenden (der Crawler für das Windows-Dateisystem wird in der Liste der verfügbaren Crawlertypen nicht angezeigt).

Wenn Sie einen Windows-Dateisystem-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 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 Dokumentenebene. Wenn die Sicherheit bei der Erstellung der Objektgruppe aktiviert war, kann der Crawler Sicherheitsdaten Dokumenten im Index zuordnen. Mit Hilfe 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 Dokumentenebene 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 Dokumentenebene für Dokumente des Windows-Dateisystems“ auf Seite 219

Sie müssen Domänenkontoinformationen auf dem Crawler-Server und auf dem Microsoft Windows-Server konfigurieren, damit aktuelle Berechtigungsnachweise geprüft werden können, wenn ein Benutzer Dokumente durchsucht, die von einem Windows-Dateisystem-Crawler durchsucht wurden.

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.

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.
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 geändert haben, 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 Konzepte

"Data listener" in "Programming Guide and API Reference for Enterprise Search"

Zugehörige Tasks

„Überwachen der Datenempfangsfunktion“ auf Seite 268

Überwachen Sie die Datenempfangsfunktion, um ihren Status und Details zur Aktivität der Datenempfangsfunktions-Clientanwendung anzuzeigen.

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 und um Metadaten Dokumenten in einem Index für die Unternehmenssuche zuzuordnen.

Plug-ins enthalten eine Java-Klasse, die für jedes vom Crawler durchsuchte Dokument aufgerufen wird. Der Java-Klasse werden die Dokumentenkennung (URI) aus dem Index für die Unternehmenssuche, Sicherheitstoken und Metadaten übergeben. Die Klasse gibt einen Satz neuer bzw. geänderter Sicherheitstoken und Metadaten zurück oder gibt an, dass der Crawler ein Dokument ignorieren soll.

Nachdem alle Dokumente im Crawlerbereich einmal durchsucht wurden, wird das Plug-in nur noch bei neuen oder geänderten Dokumenten aufgerufen. Starten Sie eine vollständige Crawlerdurchsuchung aller Dokumente im Crawlerbereich, und reorgani-

sieren Sie den Index, um die Sicherheitstoken und Metadaten für Dokumente zu ändern, die sich im Index für die Unternehmenssuche befinden, aber in der ursprünglichen Datenquelle nicht aktualisiert wurden.

Plug-ins können nicht einem vorhandenen Crawler zugeordnet werden. Bei der Konfiguration der Merkmale eines neuen Crawlers muss der Plug-in-Klassenname und -Klassenpfad angegeben werden.

Verwenden von 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 es Ihnen, zur Ermittlung des Werts der Sicherheitstoken für durchsuchte Dokumente Ihre eigenen Geschäftsregeln anzuwenden. 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.

Verwenden eines Plug-ins zum Hinzufügen von Metadaten zu Dokumenten

Dokumentmetadaten (z. B. das Datum der letzten Änderung eines Dokuments) werden für alle durchsuchten Dokumente erstellt. Das Crawler-Plug-in ermöglicht es Ihnen, zur Ermittlung des Werts der Metadaten, die für jedes Dokument indiziert werden sollen, Ihre eigenen Geschäftsregeln anzuwenden.

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.

Web-Crawler-Plug-in

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 mit einem Token versehen 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.

Zugehörige Konzepte

"Crawler plug-ins" in "Programming Guide and API Reference for Enterprise Search"

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.

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 Repositorys.

persistente_repository-id

Die persistente ID des Repositorys.

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=iio%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=iio%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/REMOTEDB/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

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+icm\lsdb12+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=anlagedateiname*

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-Zielservers.

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%2FADMIN-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
```

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/datenbankreplikat-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.

datenbankreplikat-id

Die Kennung des Datenbankreplikats.

pfad_und_name_der_datenbank

Der Pfad und der Dateiname der NSF-Dokumentdatenbank auf dem QuickPlace-Zielservers.

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%5Csampleplace%5CPageLibrary4925704300021490.nsf/A7986FD2A9CD47090525670800167225/2B02B1DE3A82B2CE49257043001C2498
```

Eine Seitenanlage:

```
quickplace://1twsvr.ibm.com/49257043000214B3/QuickPlace%5Csampleplace%5CPageLibrary4925704300021490.nsf/A7986FD2A9CD47090525670800167225/2B02B1DE3A82B2CE49257043001C2498?AttNo=0&AttName==QPCons3.ppt
```

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`

WebSphere Portal-Crawler

Dokumente, die von einem WebSphere Portal-Crawler durchsucht werden, haben folgendes URI-Format:

`wps://wps-zugriffspfad?portletDefID=portletdefinitions-id&pageID=seiten-id&useSSL=ssl`

Parameter

Die URL-Codierung wird auf alle Felder angewendet.

wps-zugriffspfad

Der Pfad des WebSphere Portal-Servers ohne Protokoll.

portletdefinitions-id

Die Portletdefinitions-ID für den WebSphere Portal-Server.

seiten-id

Die Seiten-ID für den WebSphere Portal-Server.

ssl

`useSSL=true` wird hinzugefügt, wenn das Protokoll des Pfads HTTPS ist. Andernfalls wird `useSSL=false` hinzugefügt.

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&useSSL=false
```

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
```

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

Zugehörige Konzepte

„Crawlerverwaltung für die Unternehmenssuche“ auf Seite 37

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

Zugehörige Tasks

„Konfigurieren von Kategorien“ auf Seite 118

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.

„Konfigurieren von Bereichen“ auf Seite 154

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.

„Entfernen von URIs aus dem Index“ auf Seite 157

Wenn Sie verhindern wollen, dass Benutzer Dokumente einer Objektgruppe durchsuchen, können Sie die URIs der betreffenden Dokumente aus dem Index entfernen.

„Konfigurieren von Quick Links“ auf Seite 168

Beim Erstellen eines Quick Links für eine Objektgruppe für die Unternehmens-

suche ordnen Sie der URI eines Dokuments die Schlüsselwörter zu, die dessen Aufnahme in die Suchergebnisse auslösen.

„Anzeigen von Details zu einer URI“ auf Seite 252

Sie können detaillierte Informationen zu einer URI anzeigen. Sie können aktuelle und protokollierte Informationen dazu anzeigen, wie das Dokument, das durch diese URI dargestellt wird, durchsucht, indexiert und gesucht wird.

„Anzeigen von Berichten über gelöschte Dokumente“ auf Seite 270

Sie können detaillierte Informationen zu Dokumenten anzeigen, die aus einem System für die Unternehmenssuche gelöscht wurden. Diese Informationen sind nur verfügbar, wenn Sie die Dokumentüberwachung für die Objektgruppe aktiviert haben.

Verwaltung von Parsern für die Unternehmenssuche

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, so dass 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 der semantischen Suche" in "Integration der Textanalyse"

"Textanalyse innerhalb der Unternehmenssuche" in "Integration der Textanalyse"

"Unstructured Information Management Architecture (UIMA) - Übersicht" in "Integration der Textanalyse"

"Semantische Suchanwendungen" in "Integration der Textanalyse"

"Begriff der semantischen Suchabfrage" in "Integration der Textanalyse"

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.

Die Unternehmenssuche unterstützt zwei Methoden für das Kategorisieren von Dokumenten:

Regelbasiert

Verwenden Sie diese Methode, wenn Sie Kategorien nach von Ihnen angegebenen Regeln Dokumente zuordnen möchten. Sie können z. B. Dokumente gruppieren, die ein gemeinsames URI-Muster verwenden, oder Dokumente nach Dokumentinhalt gruppieren (z. B. Dokumente, die bestimmte Wörter und Ausdrücke enthalten oder ausschließen).

Modellbasiert

Verwenden Sie diese Methode, wenn Sie IBM WebSphere Portal verwenden und die darin vordefinierten Kategorien zum Durchsuchen von Objektgruppen für die Unternehmenssuche verwenden möchten.

Anwenden von Kategorisierungsänderungen

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 zwischen **Keiner**, **Regelbasiert** und **Modellbasiert** wählen.
- 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 syntaktisch analysieren und den Index reorganisieren.
- 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 syntaktisch analysieren und den Index reorganisieren.

Dokumentinhaltsfelder

Wenn Sie eine Objektgruppe mit modellbasierten Kategorien oder mit regelbasierten Kategorien erstellen, die Regeln für den Dokumentinhalt (im Gegensatz zu Regeln für URI-Muster) verwenden, sollten Sie sicherstellen, dass die zu kategorisierenden Dokumente Inhaltsfelder enthalten.

Modellbasierte Kategorien und Kategorieregeln, die auf Dokumentinhalt basieren, verarbeiten die Inhaltsfelder von Dokumenten und berücksichtigen keine Dokumentmetadaten. Nur Dokumente, die Inhaltsfelder enthalten oder die Felder enthalten, die bei der Konfiguration des Crawlers als Inhaltsfelder konfiguriert werden, können mit diesen Kategorisierungsoptionen kategorisiert werden.

Wenn Sie modellbasierte Kategorien konfigurieren oder Kategorieregeln erstellen, die auf Dokumentinhalt basieren, gelten die Regeln nur für den Inhalt eines Dokuments. Bei den folgenden Crawlertypen gelten die Regeln für Dateien (wie z. B. den Inhalt von UNIX- oder Microsoft Windows-Dateien) oder für den Inhalt von Webseiten:

- Exchange Server
- NNTP
- UNIX-Dateisystem
- Web
- WebSphere Portal
- Windows-Dateisystem

Bei den folgenden Crawlertypen geben Sie bei der Konfiguration von Optionen für einzelne Datenquellen an, welche Felder Dokumentinhaltsfelder sind. Wenn Sie einen Crawler erstellen, ist diese Option auf der letzten Seite des Crawlerassistenten verfügbar. Wenn der Crawler bereits vorhanden ist, bearbeiten Sie den Crawlerbereich, wählen Sie eine Datenquelle aus, für die Optionen bearbeitet werden sollen, und geben Sie an, welche Felder Dokumentinhaltsfelder in dieser Datenquelle sind.

- Content Edition
- DB2 Content Manager
- DB2
- Domino Document Manager
- Notes
- QuickPlace

Regelbasierte Kategorien

Sie können Regeln konfigurieren, um zu steuern, welche Dokumente Kategorien in einer Objektgruppe für die Unternehmenssuche zugeordnet sind.

Sie können Kategorien und Kategorisierungsregeln für Objektgruppen erstellen, die Sie in der Unternehmenssuche erstellen, sowie für regelbasierte Kategorien, die Sie aus IBM WebSphere Portal-Objektgruppen migrieren. Sie müssen beim Erstellen der Objektgruppe bzw. beim Angeben der Syntaxanalyseregeln für die Objekt-

gruppe 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:

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/  
Korrekt URI: file:///c:/program+files/
```

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 es das Wort hr enthält und das Wort benefits nicht enthält:

+hr -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 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.

Regeln für den Dokumentinhalt gelten ausschließlich für den Inhalt eines Dokuments. Bei den folgenden Crawlertypen gelten die Regeln für Dateien (wie z. B. den Inhalt von UNIX- oder Microsoft Windows-Dateien) oder für den Inhalt von Webseiten:

- Exchange Server
- NNTP
- UNIX-Dateisystem
- Web
- WebSphere Portal
- Windows-Dateisystem

Bei den folgenden Crawlertypen geben Sie bei der Konfiguration von Optionen für einzelne Datenquellen an, welche Felder Dokumentinhaltsfelder sind. Wenn Sie einen Crawler erstellen, ist diese Option auf der letzten Seite des Crawlerassistenten verfügbar. Wenn der Crawler bereits vorhanden ist, bearbeiten Sie den Crawlerbereich, wählen Sie eine Datenquelle aus, für die Optionen bearbeitet werden sollen, und geben Sie an, welche Felder Dokumentinhaltsfelder in dieser Datenquelle sind.

- Content Edition
- DB2 Content Manager
- DB2
- Domino Document Manager
- Notes
- QuickPlace

Zugehörige Tasks

„Migration einer Objektgruppe aus WebSphere Portal“ auf Seite 237

Wenn Sie Objektgruppen aus WebSphere Portal in die Unternehmensuche migrieren möchten, bereiten Sie die betreffenden Objektgruppen in WebSphere Portal vor, und rufen Sie anschließend den Migrationsassistenten auf, um sie zu migrieren.

Modellbasierte Kategorien

Wenn Sie in Ihrem IBM WebSphere Portal-System modellbasierte Kategorien verwenden, können Sie diese Kategorien auch in Objektgruppen für die Unternehmenssuche verwenden.

WebSphere Portal enthält eine vordefinierte Taxonomie mit über 2.300 Themen. Diese Themen sind in Kategorien für Hauptgeschäftsbereiche wie Computer, Rechnungswesen und Transport gruppiert. WebSphere Portal-Benutzer können Anwendungen erstellen, die automatisch ermitteln, welche Dokumente mit diesen Themenbereichen übereinstimmen. Sie können zudem die Kategorien Ihren Geschäftsanforderungen anpassen.

Gehen Sie wie folgt vor, wenn Sie die WebSphere Portal-Kategorien für die Unternehmenssuche verwenden wollen:

- Importieren Sie modellbasierte Taxonomiedateien mit Hilfe des Migrationsassistenten in die Unternehmenssuche.
- Geben Sie beim Erstellen einer Objektgruppe oder beim Konfigurieren von Syntaxanalyseregeln für eine Objektgruppe die Verwendung modellbasierter Kategorien an.
- Stellen Sie sicher, dass WebSphere Portal auf dem Indexserver für die Unternehmenssuche installiert ist.
- Verwalten Sie die Kategorien mit den Kategorisierungstools in WebSphere Portal. Modellbasierte Kategorien können nicht über die Administrationskonsole für die Unternehmenssuche verwaltet werden.

Zugehörige Tasks

„Modellbasierte Taxonomien aus WebSphere Portal migrieren“ auf Seite 235
Sie können auswählen, welche modellbasierte Taxonomie Sie für eine Objektgruppe für die Unternehmenssuche verwenden möchten, indem Sie das Portlet zur Taxonomieverwaltung von WebSphere verwenden. Objektgruppen, die Sie bereits in die Unternehmenssuche migriert haben, sind von einer neuen Taxonomiemigration nicht betroffen.

Kategorieebäume

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 Objekt-

gruppe 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.

Zugehörige Tasks

„Migration einer Objektgruppe aus WebSphere Portal“ auf Seite 237

Wenn Sie Objektgruppen aus WebSphere Portal in die Unternehmensuche migrieren möchten, bereiten Sie die betreffenden Objektgruppen in WebSphere Portal vor, und rufen Sie anschließend den Migrationsassistenten auf, um sie zu migrieren.

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 Unternehmensuche 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. Sie können speziell für eine Objektgruppe konfigurierte regelbasierte Kategorien, bereits in Ihrem IBM WebSphere Portal-System vorhandene modellbasierte 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 reorganisieren Sie den Index, nachdem die Dokumente syntaktisch analysiert worden sind, 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 eine Taxonomie mit Kategorisierungsregeln verwenden wollen, die Sie speziell für diese Objektgruppe konfigurieren.

- Wenn Sie eine Objektgruppe konfigurieren, die Sie für die Unternehmenssuche erstellt haben, wählen Sie diese Option aus, um Kategorienamen und Regeln zum Kategorisieren von Dokumenten anzugeben.
- Wenn Sie eine Objektgruppe konfigurieren, die Sie aus WebSphere Portal migriert haben, wählen Sie diese Option aus, wenn Sie die regelbasierten Kategorien, die Sie importiert haben, verwenden oder ändern wollen.

Modellbasiert

Wählen Sie diese Option aus, wenn Sie Dokumente modellbasierten Kategorien zuordnen wollen, die in einem WebSphere Portal-System vorhanden sind. Für diese Option muss WebSphere Portal auf dem Indexserver für die Unternehmenssuche installiert sein. Außerdem müssen Sie die Kategorien mit den Kategorisierungstools in WebSphere Portal verwalten.

3. Klicken Sie **OK** an.

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. Sie müssen alle Dokumente in der Objektgruppe erneut durchsuchen, eine erneute Syntaxisanalyse der Dokumente durchführen und den Index reorganisieren, um die Genauigkeit der Suchergebnisse sicherzustellen.

Vorgehensweise

Gehen Sie wie folgt vor, um eine Kategorie zu konfigurieren:

1. Bearbeiten Sie eine Objektgruppe, wählen Sie die Seite **Syntaxisanalyse** 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 die Unternehmenssuche den URI eines Dokuments verwenden soll, um zu ermitteln, ob das Dokument zu einer bestimmten Kategorie gehört, klicken Sie **URI-Muster** an, und geben Sie anschließend das URI-Muster an.

Ist der von Ihnen angegebene Text im URI vorhanden, wird das Dokument der Kategorie zugeordnet.

Beispiel: `file:///c:/programme/finance`

- Wenn die Unternehmenssuche bei der Ermittlung der Kategorie eines Dokuments die Wörter in den Dokumentinhaltsfeldern analysieren soll, klicken Sie **Dokumentinhalt** an. Wählen Sie die Sprache der Dokumente aus, und geben Sie anschließend die Wörter an, die im Dokumentinhalt vorkommen müssen bzw. nicht vorkommen dürfen. Die Regel formulieren Sie im gleichen Format wie eine Abfrage (zulässig sind nur die Abfrageoperatoren +, - und " ").

Wenn ein Dokument die von Ihnen angegebenen Wörter ein- oder ausschließt, wird das Dokument der Kategorie zugeordnet.

Beispiel: `+Rechnungswesen -Abrechnung +"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 Konzepte

„Migration von WebSphere Portal auf die Unternehmenssuche“ auf Seite 235
Die Unternehmenssuche stellt einen Migrationsassistenten zur Verfügung, mit dem Sie Taxonomien und Objektgruppen aus IBM WebSphere Portal in die Unternehmenssuche migrieren können.

Zugehörige Tasks

„Modellbasierte Taxonomien aus WebSphere Portal migrieren“ auf Seite 235
Sie können auswählen, welche modellbasierte Taxonomie Sie für eine Objektgruppe für die Unternehmenssuche verwenden möchten, indem Sie das Portlet zur Taxonomieverwaltung von WebSphere verwenden. Objektgruppen, die Sie bereits in die Unternehmenssuche migriert haben, sind von einer neuen Taxonomiemigration nicht betroffen.

Zugehörige Verweise

„URI-Formate in einem Index für die Unternehmenssuche“ auf Seite 103
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.

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. Linux

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.

Die Unternehmenssuche kann dem Text in XML-Elementen Suchfeldnamen zuordnen. Beim Konfigurieren von Syntaxanalyseoptionen für eine Objektgruppe geben Sie an, welche XML-Elemente welchen Suchfeldnamen zugeordnet werden sollen. Wenn Sie XML-Elementen Suchfelder zuordnen, können Benutzer diese Elemente durch Angabe der zugeordneten Feldnamen in Abfragen durchsuchen. Abfragen, die bestimmte Felder durchsuchen, liefern präzisere Suchergebnisse als Freiformat-abfragen, die den gesamten Dokumentinhalt durchsuchen.

Wenn z. B. Ihre Objektgruppe XML-Dokumente enthält und Sie angeben, dass die Elemente für den Titel und Autor als Suchfelder im Index markiert werden sollen, können Benutzer diese bestimmten Elemente abfragen. Eine Suche nach `author:Smith` findet XML-Dokumente, die Smith im Element `author` aufweisen.

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 Objektgruppenadministrator der Objektgruppe sein, zu der die XML-Dokumente gehören, um Suchfeldern XML-Elemente zuzuordnen.

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, wenn Sie den Parser stoppen und wieder erneut starten. Die neuen und geänderten Zuordnungen gelten nur für Daten, die nach dem Neustart des Parsers analysiert werden. Die neuen und geänderten Zuordnungen haben keine Auswirkungen auf Daten, die bereits analysiert und indexiert sind.

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, so dass 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.
 - b. Geben Sie `jobrole` in das Feld **XML-Elementname** ein.
 - c. Damit Benutzer das Feld `jobrole` abfragen und die Aufgabenbereiche in den Suchergebnissen anzeigen können, wählen Sie die Markierungsfelder **Feldspezifische Suche** und **Suchergebnisse** aus.
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 `jobroles` in das Feld **Feldname** ein.
 - c. Geben Sie `jobrole` 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 und die Aufgabenbereiche in den Suchergebnissen anzeigen können, wählen Sie die Markierungsfelder **Feldspezifische Suche** und **Suchergebnisse** aus.
7. Ordnen Sie das XML-Element `section` mit dem Attribut `expertise` einem Suchfeld namens `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 **Feldname** ein.
 - d. Geben Sie `id` in das Feld **XML-Attributname** ein.
 - e. Geben Sie `expertise` in das Feld **XML-Attributwert** ein.
 - f. Damit Benutzer das Feld `expertise` abfragen und die zugehörigen Werte in den Suchergebnissen anzeigen können, wählen Sie die Markierungsfelder **Feldspezifische Suche** und **Suchergebnisse** aus.
8. Klicken Sie **OK** an.

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 im Index 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.

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 indexiert werden sollen, können Benutzer diese Elemente abfragen. Bei einer Suche nach `copyright:IBM` werden HTML-Dokumente gefunden, die IBM in den Copyright-Metadaten enthalten.

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 Sie den Parser gestoppt und erneut gestartet haben. Die Änderungen gelten nur für Daten, die nach dem Neustart des Parsers analysiert werden. Wenn Sie Dokumente aktualisieren wollen, die bereits analysiert und indexiert wurden, müssen Sie die Dokumente erneut durchsuchen sowie syntaktisch analysieren und anschließend den Index reorganisieren.

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. Klicken Sie **Feld hinzufügen** auf der Seite **HTML-Feldzuordnungen** an, um der Liste mit Feldzuordnungsregeln eine Leerzeile hinzuzufügen.
3. Geben Sie in das Feld **Feldname** einen Namen ein, den Sie mit dem HTML-Metadatenelement verknüpfen wollen, das Sie gerade zuordnen. Benutzer können diesen Feldnamen bei der Abfrage von HTML-Dokumenten in dieser Objektgruppe angeben.
4. Geben Sie in das Feld **Name des HTML-Metadatenelements** den Namen des HTML-Metadatenelements ein, das Sie zuordnen wollen.
5. Damit Benutzer dieses Feld abfragen und in den Suchergebnissen anzeigen können, wählen Sie die Markierungsfelder **Feldspezifische Suche** und **Suchergebnisse** aus.
6. 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 aktivieren, wählen Sie das Markierungsfeld **Parametrische Suche** aus.
7. Klicken Sie **OK** an.

Beispiele:

Benutzer können die zugeordneten Feldnamen nun abfragen, um nach HTML-Dokumenten mit bestimmten Metadaten zu suchen. Wenn Sie beispielsweise ein HTML-Metadatenelement namens Beschreibung zugeordnet haben, um nach einem Feld namens Kurzdarstellung zu suchen, könnte ein Benutzer eine Abfrage wie die folgende eingeben, wenn er nach Dokumenten zum Thema ThinkPad-Computer sucht:

Kurzdarstellung: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.

WebSphere Information Integrator OmniFind 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 XML-Zuordnungsdateien der Objektgruppe zuordnen und die Ausgabe der XML-Zuordnung einer 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, so dass 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 Datenstrukturen in einer allgemeinen Analysestruktur dem Index für die Unternehmenssuche zu, so dass 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 Begriff "Alex" und einer Telefonnummer).

- Ordnen Sie Datenstrukturen in einer allgemeinen Analysestruktur JDBC-Datenbanktabellen (Java Database Connectivity) zu. Daten können Sie IBM DB2 UDB- (DB2 Universal Database) 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 mit Hilfe von SQL-Abfragen zu durchsuchen.

Zugehörige Konzepte

"Integration der benutzerdefinierten Textanalyse" in "Integration der Textanalyse"

"Unstructured Information Management Architecture (UIMA) - Übersicht" in "Integration der Textanalyse"

"Workflow für die Integration der benutzerdefinierten Analyse" in "Integration der Textanalyse"

"Textanalysealgorithmen" in "Integration der Textanalyse"

"Semantische Suchanwendungen" in "Integration der Textanalyse"

"Begriff der semantischen Suchabfrage" in "Integration der Textanalyse"

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 Verarbeitungseingines (.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 System befinden. Wenn die Datei größer als 8 MB ist, muss sie sich auf dem Indexserver befinden.
7. Klicken Sie **OK** an. Ihre Textanalysesteuerkomponente wird auf der Seite **Textanalysesteuerkomponenten** aufgelistet.

Zugehörige Konzepte

"Workflow für die Integration der benutzerdefinierten Analyse" in "Integration der Textanalyse"

| "Integration der benutzerdefinierten Textanalyse" in "Integration der Text-
| analyse"

| "Unstructured Information Management Architecture (UIMA) - Übersicht" in
| "Integration der Textanalyse"

| "XML-Markup in Analyse und Suche" in "Integration der Textanalyse"

| **Zugehörige Tasks**

| "Erstellen einer Konfigurationsdatei für die Zuordnung von XML- zu UIMA-
| Inhaltstypen" in "Integration der Textanalyse"

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.

4. Klicken Sie **OK** an.

Zugehörige Konzepte

"Workflow für die Integration der benutzerdefinierten Analyse" in "Integration der Textanalyse"

"Integration der benutzerdefinierten Textanalyse" in "Integration der Textanalyse"

"Unstructured Information Management Architecture (UIMA) - Übersicht" in "Integration der Textanalyse"

Zuordnen von XML-Elementen zu einer 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 JDBC-Tabellen über eine semantische Suche zu durchsuchen, können Sie die XML-Elemente einer allgemeinen Analysestruktur zuordnen.

Vorbereitung

Sie müssen ein Administrator für die Unternehmenssuche oder ein Objektgruppenadministrator der Objektgruppe sein, um einer allgemeinen Analysestruktur XML-Elemente zuzuordnen.

Sie können benutzerdefinierte XML-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 XML-Zuordnungsdateien müssen dem UIMA-Framework für die Textanalyse entsprechen.

Wenn Sie einer Objektgruppe, die eine benutzerdefinierte Textanalysesteuerkomponente verwendet, XML-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. Danach können Sie die Analyseergebnisse (in der allgemeinen 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 Benutzer die ursprünglichen XML-Elemente in Abfragen angeben können sollen, müssen Sie keine XML-Zuordnungen definieren. Sie können stattdessen Syntaxanalyseoptionen konfigurieren und die native XML-Suche für die Objektgruppe aktivieren.

Vorgehensweise

Gehen Sie wie folgt vor, um einer 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 allgemeinen Analysestrukturen zuordnen** an.
3. Geben Sie auf der Seite **XML-Elemente einer allgemeinen Analysestruktur zuordnen** einen beschreibenden Anzeigenamen für die XML-Zuordnungsdatei ein.
4. Geben Sie die Speicherposition der Datei an. Wenn die XML-Zuordnungsdatei kleiner als 8 MB ist, können Sie den Pfad eingeben oder nach der Datei suchen. Wenn die XML-Zuordnungsdatei größer als 8 MB ist, stellen Sie sicher, dass sich die Datei auf dem Indexserver befindet, und geben Sie den vollständig qualifizierten Pfad für die Datei an.
5. Klicken Sie **OK** an. Ihre neue XML-Zuordnungsdatei wird der Liste der XML-Zuordnungsdateien auf der Seite **Textverarbeitungsoptionen** hinzugefügt.

Zugehörige Konzepte

"Workflow für die Integration der benutzerdefinierten Analyse" in "Integration der Textanalyse"

"Integration der benutzerdefinierten Textanalyse" in "Integration der Textanalyse"

"Unstructured Information Management Architecture (UIMA) - Übersicht" in "Integration der Textanalyse"

"XML-Markup in Analyse und Suche" in "Integration der Textanalyse"

Zugehörige Tasks

"Erstellen einer Konfigurationsdatei für die Zuordnung von XML- zu UIMA-Inhaltstypen" in "Integration der Textanalyse"

Zuordnen einer allgemeinen Analysestruktur zum Index

Sie können angeben, welche allgemeine Analysestruktur verwendet werden soll, wenn Benutzer eine Objektgruppe mit einer semantischen Suche abfragen.

Vorbereitung

Sie müssen ein Administrator für die Unternehmenssuche oder ein Objektgruppenadministrator der Objektgruppe sein, um dem Index eine allgemeine Analysestruktur zuzuordnen.

Informationen zu dieser Task

Indem Sie dem Index für die Unternehmenssuche eine 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. der Begriff "Alex" und eine Telefonnummer) suchen.

Vorgehensweise

Gehen Sie wie folgt vor, um dem Index eine allgemeine Analysestruktur zuzuordnen:

1. Bearbeiten Sie eine Objektgruppe, wählen Sie die Seite **Syntaxanalyse** aus, und klicken Sie **Textverarbeitungsoptionen konfigurieren** an.
2. Klicken Sie **Allgemeine Analysestruktur auswählen** im Bereich **Eine allgemeine Analysestruktur dem Index zuordnen** an.
3. Wählen Sie auf der Seite **Allgemeine Analysestruktur für diese Objektgruppe auswählen** die Zuordnung aus, die Sie für den Index für die Unternehmenssuche verwenden wollen:
 - Wählen Sie **Standard** aus, wenn Sie die Standardsystemregeln verwenden wollen, um dem Index für die Unternehmenssuche Datenstrukturen in einer allgemeinen Analysestruktur zuzuordnen.
 - Geben Sie die Speicherposition der Datei für die allgemeine Analysestruktur an, um dem Index eine bestimmte allgemeine Analysestruktur zuzuordnen. Wenn die Datei kleiner als 8 MB ist, kann sie sich auf Ihrem System befinden. Wenn die Datei größer als 8 MB ist, muss sie sich auf dem Indexserver befinden.
4. Klicken Sie **OK** an. Die von Ihnen angegebene allgemeine Analysestruktur wird auf der Seite **Textverarbeitungsoptionen** angezeigt.

Zugehörige Konzepte

"Workflow für die Integration der benutzerdefinierten Analyse" in "Integration der Textanalyse"

"Integration der benutzerdefinierten Textanalyse" in "Integration der Textanalyse"

"Unstructured Information Management Architecture (UIMA) - Übersicht" in "Integration der Textanalyse"

"Indexzuordnung für angepasste Analyseergebnisse" in "Integration der Textanalyse"

Zugehörige Tasks

"Erstellen der Konfigurationsdatei für die Indexerstellung" in "Integration der Textanalyse"

Zuordnen einer allgemeinen Analysestruktur zu JDBC-Tabellen

Sie können angeben, welche allgemeine Analysestruktur Sie JDBC-Tabellen zur Verwendung in Datenbankanwendungen zuordnen wollen.

Vorbereitung

Sie müssen ein Administrator für die Unternehmenssuche oder ein Objektgruppenadministrator der Objektgruppe sein, um JDBC-Datenbanktabellen (Java Database Connectivity) eine allgemeine Analysestruktur zuzuordnen.

Informationen zu dieser Task

Durch Zuordnen einer allgemeinen Analysestruktur zu JDBC-Tabellen werden die Daten für Datenbankanwendungen nutzbar. Benutzer können beispielsweise SQL-Abfragen außerhalb der Unternehmenssuche angeben, um die Kommentare zu durchsuchen, die von der allgemeinen Analysestruktur hinzugefügt wurden. Die

Informationen können Sie darüber hinaus zur weiteren Datenverarbeitung verwenden, z. B. zur Verwendung in Data-Mining-Anwendungen.

Vorgehensweise

Gehen Sie wie folgt vor, um JDBC-Tabellen eine allgemeine Analysestruktur 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 **JDBC-Tabellen allgemeine Analysestrukturen zuordnen** an.
3. Geben Sie auf der Seite **Allgemeine Analysestruktur zu JDBC-Tabellen zuordnen** einen beschreibenden Anzeigenamen für die allgemeine Analysestruktur ein, mit der Sie JDBC-Tabellen Informationen zuordnen wollen.
4. Geben Sie die Speicherposition der Datei für die allgemeine Analysestruktur an. Wenn die Datei kleiner als 8 MB ist, kann sie sich auf Ihrem System befinden. Wenn die Datei größer als 8 MB ist, muss sie sich auf dem Indexserver befinden.
5. Klicken Sie **OK** an. Ihre neue Datei wird der Liste der Dateien auf der Seite **Textverarbeitungsoptionen** hinzugefügt. Außerdem werden Datum und Uhrzeit angezeigt, an dem bzw. zu der Sie die Datei dem System hinzugefügt haben.

Zugehörige Konzepte

"Workflow für die Integration der benutzerdefinierten Analyse" in "Integration der Textanalyse"

"Integration der benutzerdefinierten Textanalyse" in "Integration der Textanalyse"

"Unstructured Information Management Architecture (UIMA) - Übersicht" in "Integration der Textanalyse"

"Datenbankzuordnung für ausgewählte Analyseergebnisse" in "Integration der Textanalyse"

Zugehörige Tasks

"Erstellen der Konfigurationsdatei für die XML-Zuordnung" in "Integration der Textanalyse"

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 Objektgruppen-administrator 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, so dass jedes Teilstück als separater Begriff behandelt werden kann.

Vorbereitung

Sie müssen ein Administrator für die Unternehmenssuche oder ein Objektgruppen-administrator 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.

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

| "Linguistische Unterstützung der semantischen Suche" in "Integration der Text-
| analyse"

| "Textanalyse innerhalb der Unternehmenssuche" in "Integration der Text-
| analyse"

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, können genauere Suchergebnisse erzielen, indem die XML-Markup der Dokumente innerhalb der Abfrage genutzt wird. Benutzer können z. B. 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 der semantischen Suche" in "Integration der Text-
analyse"

"Textanalyse innerhalb der Unternehmenssuche" in "Integration der Text-
analyse"

"Semantische Suchanwendungen" in "Integration der Textanalyse"

"Begriff der semantischen Suchabfrage" in "Integration der Textanalyse"

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 keine Leerzeichen zum Begrenzen von Wörtern verwendet werden.

Bei der Unicode-basierten Leerraumsegmentierung 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 keine Leerzeichen als Begrenzer verwenden, können auf Grund der N-Gram-Segmentierung bessere Suchergebnisse zurückgegeben werden als bei der Unicode-basierten Leerraumsegmentierung.

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.

Entfernen von Zeilenvorschubzeichen, die nicht zum ASCII-Zeichenbereich gehören

In Sprachen, in denen Leerraum nicht als Wortbegrenzer verwendet wird, wie z. B. Chinesisch oder Japanisch, können Sie den Parser so konfigurieren, dass bestimmte Leerraumzeichen, die Zeilenumbrüche verursachen, entfernt werden.

Vorbereitung

Sie müssen sich als Administrator für die Unternehmenssuche anmelden, um diese Task auszuführen.

Informationen zu dieser Task

Wenn diese Option für eine Objektgruppe aktiviert ist und ein Dokument in chinesischer oder japanischer Sprache geschrieben wurde, entfernt der Parser jede Folge von Leerzeichen zwischen zwei Zeichen. Die Zeichen müssen aus einem Nicht-ASCII-Unicode-Zeichenbereich stammen. Die folgenden Zeichen werden entfernt:

- 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 syntaktisch analysiert und im Index gespeichert wurden, müssen diese Dokumente erneut durchsucht und syntaktisch analysiert werden. Anschließend muss der Index reorganisiert werden.

Vorgehensweise

Gehen Sie wie folgt vor, um Zeilenvorschubzeichen aus Leerraum zu entfernen:

1. Melden Sie sich am Indexserver als Administrator für die Unternehmenssuche an. Diese Benutzer-ID wurde bei der Installation von WebSphere II OmniFind 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):
`objektgruppen-ID.parserdriver/collection.properties`
3. Ändern Sie den Wert für das Merkmal `removeCjNewlineChars` von `false` in `true`.

Dokumenttypen für Objektgruppenparser und Stellent-Sitzungen

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 in einer Stellent-Sitzung mit Dokumentfilterung 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, RT und Ichitaro).

Da Metadaten irreführend sein können, werden Text- und HTML-Dokumente eventuell irrtümlich an die Stellent-Sitzung gesendet und dann an einen der integrierten Parser zurückgesendet. Dies kann die Leistung beeinträchtigen. 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 zur Stellent-Sitzung 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 die Stellent-Sitzung. 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 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 Dokumente Sie an die Stellent-Sitzung senden möchten, abhängig von der Dateierweiterung bzw. dem Inhaltstyp.
- Vorgehensweise für die syntaktische Analyse von Dokumenten, deren Typ auf Grund fehlender Metadaten unbekannt ist.

Das Dateiformat ist eine Folge von Zeilen, wobei jede Zeile eine der folgenden Regeln darstellt:

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: `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 WebSphere II OmniFind 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 über die 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 die Stellent-Sitzung 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
```

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 jttd stellent
EXTENSION jtcd stellent
EXTENSION jtcdx 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 html
DEFAULT html

```

Zuordnen von Dokumententypen zu einer Stellent-Sitzung

Erstellen Sie eine Konfigurationsdatei, um anzugeben, welche Dokumenttypen von Stellent-Dokumentfiltern analysiert werden sollen. 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

Die Konfigurationsdatei gibt Folgendes an:

- Welche Dateitypen von der Stellent-Sitzung syntaktisch analysiert werden sollen. Ein Dateityp entspricht einem der Dateitypen, die von der Stellent-Bibliothek erkannt werden.
- Welche Dateitypen an den Objektgruppenparser für die Verarbeitung mit einem der integrierten Parser zurückgesendet werden müssen. (Diese Aktion ist erforderlich, weil der Objektgruppenparser auf Grund irreführender Metadaten eventuell irrtümlich ein Dokument an die Stellent-Sitzung sendet.)
- Welche Dateitypen zurückzuweisen sind, 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-Sitzungen 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 der folgenden Regeln darstellt:

```
accept DEFAULT
accept ALL dokumenttyp
accept typ dokumenttyp
native DEFAULT
native typ dokumenttyp
reject 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.

typ Einer der Werte für FI_ in der Stellent-Bibliothek. *dokumenttyp* ist der Wert, der bei Anwendung einer Regel für das Dokumenttyp-Token verwendet werden muss.

DEFAULT

Bedeutet, dass die Liste der akzeptierten oder nativen Typen abhängig vom Regeltyp die Standardliste einschließt. Diese Option ermöglicht Ihnen, die Standardkonfiguration zu erweitern, anstatt 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:

- Wenn es eine Regel `accept` für *typ* gibt, einschließlich der Standardliste, falls `accept DEFAULT` angegeben wurde, wird er akzeptiert.
- Sonst gilt: Falls es eine Regel `reject` für *typ* gibt, wird er nicht akzeptiert.
- Sonst gilt: Wenn `accept ALL` angegeben wurde, wird er akzeptiert.
- Anderenfalls wird er nicht akzeptiert.

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. Wenn der Dokumenttyp nicht akzeptiert wird, tritt folgendes Verhalten ein:

- Wenn es für *typ* eine native Regel gibt (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 bzw. XML.
- Anderenfalls wird das Dokument zurückgewiesen und nicht syntaktisch analysiert.

Vorgehensweise

Gehen Sie wie folgt vor, um der Stellent-Sitzung Dokumenttypen zuzuordnen:

1. Melden Sie sich am Indexserver als Administrator für die Unternehmenssuche an. Diese Benutzer-ID wurde bei der Installation von WebSphere II OmniFind 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, so dass 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 mit Hilfe der Administrationskonsole für die Unternehmenssuche den Parser für die Objektgruppe überwachen und dann den Parser stoppen und erneut starten.

Beispiel

In der folgenden Konfigurationsdatei akzeptiert die Stellent-Sitzung zusätzlich zur Standardliste das 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 unterstützt und können mit dem Dokumenttyp `ps` durchsucht werden; das X Pixmap-Format (XPM) wird an den integrierten Textparser zurückgesendet; das Bildformat PNG wird zurückgewiesen; alle anderen Dateitypen werden akzeptiert und können mit dem Dokumenttyp `other` durchsucht werden.

```
accept DEFAULT
accept FI_POSTSCRIPT ps
native FI_XPIXMAP txt
accept ALL other
reject FI_PNG
```

Standardsyntaxanalyseregeln für Stellent-Sitzungen

Wenn Sie keine Konfigurationsdatei erstellen, um Dokumentfiltern für Stellent-Sitzungen Dateitypen zuzuordnen, werden für die syntaktische Analyse der Dokumente Standardregeln verwendet.

Die Stellent-Sitzung verwendet folgende Standardregeln:

<code>ACCEPT FI_WORD4</code>	<code>doc</code>
<code>ACCEPT FI_WORD5</code>	<code>doc</code>
<code>ACCEPT FI_RTF</code>	<code>rtf</code>
<code>ACCEPT FI_WINWORD1</code>	<code>doc</code>
<code>ACCEPT FI_WINWORD1COMPLEX</code>	<code>doc</code>
<code>ACCEPT FI_WINWORD2</code>	<code>doc</code>
<code>ACCEPT FI_WORD6</code>	<code>doc</code>

ACCEPT FI_WINWORD6	doc
ACCEPT FI_ICHITAR03	jxw
ACCEPT FI_ICHITAR04	jsw
ACCEPT FI_WINWORD1J	doc
ACCEPT FI_WINWORD5J	doc
ACCEPT FI_RTFJ	rtf
ACCEPT FI_WINWORD7	doc
ACCEPT FI_WORDPRO	lwp
ACCEPT FI_WINWORD97	doc
ACCEPT FI_ICHITAR08	jtd
ACCEPT FI_WORDPRO97	lwp
ACCEPT FI_WINWORD2000	doc
ACCEPT FI_WINWORD2002	doc
ACCEPT FI_WINWORD2003	doc
ACCEPT FI_123R1	123
ACCEPT FI_123R2	123
ACCEPT FI_123R3	123
ACCEPT FI_EXCEL	xls
ACCEPT FI_EXCEL3	xls
ACCEPT FI_EXCEL4	xls
ACCEPT FI_123R4	123
ACCEPT FI_EXCEL5	xls
ACCEPT FI_123R6	123
ACCEPT FI_EXCEL97	xls
ACCEPT FI_123R9	123
ACCEPT FI_EXCEL2000	xls
ACCEPT FI_EXCEL2002	xls
ACCEPT FI_EXCEL2003	xls
ACCEPT FI_FREELANCE	prz
ACCEPT FI_POWERPOINT4	ppt
ACCEPT FI_POWERPOINT3	ppt
ACCEPT FI_POWERPOINT7	ppt
ACCEPT FI_FREELANCE3	prz
ACCEPT FI_POWERPOINTMAC3	ppt
ACCEPT FI_POWERPOINTMAC4	ppt
ACCEPT FI_PDF	pdf
ACCEPT FI_EXTPOWERPOINT4	ppt
ACCEPT FI_EXTPOWERPOINTMAC4	ppt
ACCEPT FI_POWERPOINTMACB3	ppt
ACCEPT FI_POWERPOINTMACB4	ppt
ACCEPT FI_POWERPOINT97	ppt
ACCEPT FI_PDFMACBIN	pdf
ACCEPT FI_POWERPOINT9597	ppt
ACCEPT FI_POWERPOINT2000	ppt
ACCEPT FI_POWERPOINT2	ppt
NATIVE FI_HTML	htm
NATIVE FI_HTML_LATIN2	htm
NATIVE FI_HTML_JAPANESESJIS	htm
NATIVE FI_HTML_JAPANESEEUC	htm
NATIVE FI_HTML_CHINESEBIG5	htm
NATIVE FI_HTML_CHINESEEUC	htm
NATIVE FI_HTML_CHINESEGB	htm
NATIVE FI_HTML_KOREANHANGUL	htm
NATIVE FI_HTML_CYRILLIC1251	htm
NATIVE FI_HTML_CYRILLICKO18	htm
NATIVE FI_CYRILLIC1251	txt
NATIVE FI_CYRILLICKO18	txt
NATIVE FI_W2KHTML	htm
NATIVE FI_XL2KHTML	htm
NATIVE FI_PP2KHTML	htm
NATIVE FI_XML	xml
NATIVE FI_WML	xml
NATIVE FI_HTML_JAPANESEJIS	htm
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_JAPANESEJIS	xml
NATIVE FI_WML_JAPANESESJIS	xml
NATIVE FI_WML_JAPANESEEUC	xml
NATIVE FI_WML_KOREANHANGUL	xml
NATIVE FI_WML_LATIN2	xml
NATIVE FI_HTMLUNICODE	htm
NATIVE FI_XML_DOCTYPE_HTML	htm
NATIVE FI_XHTML	htm
NATIVE FI_ASCII	txt
NATIVE FI_ANSI	txt
NATIVE FI_UNICODE	txt
NATIVE FI_ASCII8	txt
NATIVE FI_ANSI8	txt
NATIVE FI_MAC	txt
NATIVE FI_MAC8	txt
NATIVE FI_SHIFTJIS	txt
NATIVE FI_CHINESEGB	txt
NATIVE FI_HANGEUL	txt
NATIVE FI_CHINESEBIG5	txt
NATIVE FI_LATIN2	txt
NATIVE FI_JAPANESE_EUC	txt
NATIVE FI_HEBREW_OLDCODE	txt
NATIVE FI_HEBREW_PC8	txt
NATIVE FI_HEBREW_E0	txt
NATIVE FI_HEBREW_WINDOWS	txt
NATIVE FI_ARABIC_710	txt
NATIVE FI_ARABIC_720	txt
NATIVE FI_ARABIC_WINDOWS	txt
NATIVE FI_7BITTEXT	txt
NATIVE FI_JAPANESE_JIS	txt
NATIVE FI_CENTRALEU_1250	txt
NATIVE FI_UTF8	txt
NATIVE FI_EBCDIC_37	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_500	txt
NATIVE FI_EBCDIC_870	txt
NATIVE FI_EBCDIC_871	txt
NATIVE FI_EBCDIC_1026	txt

Indexverwaltung für die Unternehmenssuche

Um sicherzustellen, dass Benutzer immer Zugriff auf die neuesten Informationen haben, erstellt die Unternehmenssuche für jede Objektgruppe einen Index und pflegt ihn durch regelmäßige Inhaltsaktualisierung und -reorganisation.

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, aktualisiert die Unternehmenssuche den Index durch neuen Inhalt. Der aktualisierte Inhalt muss letztlich in den Basisindex aufgenommen werden. Dieser Aufnahmeprozess wird Reorganisation genannt. Wird der Index aktualisiert bzw. reorganisiert, 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 den Index häufig aktualisieren, können Benutzer die neuesten Daten durchsuchen. Ein kontinuierlich aktualisierter Index muss letztlich reorganisiert werden. Ein aktualisierter Index wird zunehmend größer und belegt damit zunehmend Systemressourcen. Daher sollten Indizes regelmäßig reorganisiert werden, um optimale Leistung zu gewährleisten.

Die Häufigkeit einer Indexreorganisation 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 Indexreorganisation wirksam)
- Häufigkeit eines erzwungenen Crawlerstarts anstatt der Ausführung zu einem terminierten Zeitpunkt
- Häufigkeit des Entfernens bzw. Hinzufügens von URIs durch externe Crawler (diese Crawlertypen interagieren über die API der Datenempfängsfunktion mit der Unternehmenssuche)

Bei Objektgruppen mit mehreren Millionen Dokumenten, die hauptsächlich aus Webdokumenten erzeugt werden, sollte der Index ungefähr einmal täglich reorganisiert und ein- oder zweistündlich aktualisiert werden.

Führen Sie die folgenden Tasks aus, um einen aktuellen, durchsuchbaren Index zu gewährleisten:

- Angabe von Zeitplänen für das Aktualisieren und Reorganisieren des Index
- Ändern des Indexzeitplans
- Aktivieren und Inaktivieren des Indexzeitplans
- Konfigurieren gleichzeitig ablaufender Indexerstellungen

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 265

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.

„Überwachen der Indexierungswarteschlange für die Unternehmenssuche“ auf Seite 266

Sie können den Status aller Indexerstellung in der Indexierungswarteschlange anzeigen, eine Indexerstellung stoppen oder einen Index aus der Warteschlange löschen.

Terminieren von Indexerstellung

Sie können Zeitpläne für das Reorganisieren eines Index und das Aktualisieren eines 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 ein regelmäßiges Reorganisieren und Aktualisieren des Index, um sicherzustellen, dass Benutzer immer auf die aktuellen Informationen in den durchsuchten Quellen zugreifen können. Beim Reorganisieren eines Index wird der gesamte Index erneut erstellt. Die Indexierungsprozesse lesen alle Daten, die von Crawlern erfasst und vom Parser syntaktisch analysiert wurden. Beim Aktualisieren eines Index werden die seit der letzten Indexreorganisation durchsuchten Informationen für die Suche verfügbar gemacht.

Standardmäßig ist die Option zum Terminieren von Indexerstellung ausgewählt. Diese Option weist den Terminierungsprozess an, Tasks zum Aktualisieren und Reorganisieren des Index zu terminieren, wenn das System für die Unternehmenssuche gestartet wird. Sie können die Auswahl des Markierungsfelds **Beim Systemstart aktivieren** jederzeit zurück nehmen, wenn Sie eine terminierte Indexerstellung nicht ausführen wollen. Sie müssen den Zeitplan z. B. inaktivieren, um Probleme zu beheben.

Vorgehensweise

Gehen Sie wie folgt vor, um Indexerstellung zu terminieren:

1. Bearbeiten Sie eine Objektgruppe, wählen Sie die Seite **Index** aus, und klicken Sie **Indexerstellung terminieren** an.
2. Geben Sie auf der Seite **Indexerstellung terminieren** im Bereich **Zeitplan für die Indexaktualisierung angeben** die folgenden Optionen an, um anzugeben, wie oft der Index mit dem neuen Inhalt aktualisiert wird:
 - a. Geben Sie im Bereich **Starten am** in den Feldern **Jahr, Monat, Tag, Stunde** und **Minute** an, wann der Index zum ersten Mal aktualisiert werden soll.

- b. Geben Sie im Bereich **Aktualisierungsintervall** in den Feldern **Tage**, **Stunden** und **Minuten** an, wie oft der Index aktualisiert werden soll.
In der Regel sollte der Index häufig aktualisiert 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 Indexreorganisation angeben** geben Sie an, wie oft der Index komplett 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 Index zum ersten Mal reorganisiert werden soll.
 - b. Geben Sie im Bereich **Aktualisierungsintervall** in den Feldern **Tage**, **Stunden** und **Minuten** an, wie oft der Index reorganisiert werden soll.
Sie sollten den Index regelmäßig, zum Beispiel alle 24 Stunden reorganisieren. 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 das Reorganisieren oder Aktualisieren eines Index 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 häufig der Index mit neuem Inhalt aktualisiert bzw. wie häufig er reorganisiert werden soll.
2. Klicken Sie **Anwenden** an.

Aktivieren und Inaktivieren der Indexzeitpläne

Sie können die Zeitpläne für das Aktualisieren und Reorganisieren des Index 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 anstatt 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 Indexreorganisation zu aktivieren bzw. 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 auf der Seite **Index** ein Index terminiert ist und er nicht zum terminierten Datum und Zeitpunkt erstellt werden soll, klicken Sie **Zeitplan inaktivieren** an. Der Index wird erst erstellt, wenn Sie den Zeitplan wieder aktivieren oder **Starten** 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 Indexerstellung

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 angeben, dass der Index für eine Objektgruppe gleichzeitig aktualisiert und reorganisiert wird, um die Suchqualität verbessern.

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 Reorganisation 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 Indexstellungsanforderungen von verschiedenen Objektgruppen gleichzeitig in der Indexierungswarteschlange befinden. Beim Konfigurieren der Indexierungsoptionen 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 des Index einer Objektgruppe während der Reorganisation des Index verarbeitet werden. Wenn Sie diese Option aktivieren, werden die Suchserver während der länger dauernden Reorganisation des Index mit den neuesten Dokumenten aktualisiert (mittels einer Indexaktualisierung). Die Indexerstellung ist jedoch ein ressourcenintensiver Prozess. Während einer Indexerstellung wird eine große Menge System- und Plattenspeicherplatz belegt. Wenn Sie diese Option aktivieren und nicht genügend Plattenspeicherplatz oder System- und Plattenspeicher zur Verfügung steht, kann sich das negativ auf die Gesamtleistung des Systems auswirken.

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 Indexstellungsanforderungen 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 Indexstellungen für einzelne Objektgruppen vorhanden sind, können Sie die Option auswählen, die Anforderungen für die gleichzeitige Aktualisierung und Reorganisation des Index ermöglicht.
6. Klicken Sie **OK** an.

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 einer 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.

Indexierte Optionen für das Durchsuchen von Dokumenten

Wenn Sie Optionen für das Durchsuchen von mit dem Crawler durchsuchten Daten oder externen Quellen konfigurieren oder XML- und HTML-Metadatenelemente zum Durchsuchen von Feldern zuordnen, geben Sie an, wie die Dokumente durchsucht und in den Suchergebnissen angezeigt werden.

Wenn Sie Optionen für Crawler bearbeiten, 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
- Parametrische Suche
- Suchergebnisse
- Dokumentinhalt

Wenn Sie Optionen für externe Quellen bearbeiten, steht die Option zur Markierung des Felds als Dokumentinhaltsfeld nicht zur Verfügung.

Wenn Sie den Parser konfigurieren und angeben, dass XML-Elemente und HTML-Metadatenelementen Feldern im Index zugeordnet werden sollen, geben Sie an, ob die Felder in einer feldspezifischen Suche, in einer parametrischen Suche oder in den Suchergebnissen verwendet werden können.

Die angegebenen Optionen werden mit den Dokumenten im Index gespeichert. Mit diesen Optionen können Sie einschränken, welche Abfragen ein Benutzer ausführen kann und welche Suchergebnisse angezeigt werden.

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.

Beispiel:

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 die-

sem Wort aus den Suchergebnissen auszuschließen. Umschließen Sie mindestens zwei Wörter durch Anführungszeichen ("), um nach einem genauen Ausdruck zu suchen.

Beispiel:

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 Felder und Metadaten eines Dokuments beschränken. Sie können z. B. angeben, dass bestimmte Wörter im Titel eines Dokuments vorhanden sein müssen.

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.

Beispiel:

Die folgende Abfrage sucht nach Dokumenten, die das Wort ibm und den Ausdruck unternehmenssuche im Titelfeld enthalten müssen:

```
title:ibm title:"unternehmenssuche"
```

Soll ein Feld nach Feldname durchsucht werden, müssen Sie das Feld beim Konfigurieren des Crawlers für die feldspezifische Suche aktivieren.

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.

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
```

Soll ein Feld mit einer parametrischen Suche durchsucht werden, müssen Sie das Feld beim Konfigurieren des Crawlers für die parametrische Suche aktivieren.

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.

Dokumentinhalt

Der Inhalt eines als Dokumentinhaltsfeld markierten Felds kann verwendet werden, um Dokumente einer modellbasierten Kategorie oder Kategorien zuzuordnen, die auf Dokumentinhalt basierte Regeln angeben.

Zugehörige Konzepte

"Query syntax" in "Programming Guide and API Reference for Enterprise Search"

„Suchanwendungen für die Unternehmenssuche“ auf Seite 183

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.

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 kön-

nen alle möglichen Begriffserweiterungen gespeichert werden, oder Suchvorgänge können Begriffe während der Abfrageverarbeitung erweitern.

Alle Änderungen an den Einstellungen für Platzhalterzeichen werden nach der nächsten Indexreorganisation 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*`.

Wenn der Index aktualisiert oder reorganisiert wird, 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-Repositoryys 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.

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 Reorganisation des Index 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. Geben Sie über das Markierungsfeld **Platzhalterzeichen in Abfragen unterstützen, die keine Felder durchsuchen** an, ob Platzhalterzeichen in Abfragen unterstützt werden, die nach Text suchen, der nicht in Feldern enthalten ist. 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 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 bei der Konfiguration von Optionen für Platzhalterzeichen **Hilfe** an, um weitere Informationen zu erhalten.
6. Klicken Sie **OK** an.

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 Index zweimal zu reorganisieren, bevor die Änderungen wirksam werden. Wenn Sie vor der ersten Reorganisation des Index für die Objektgruppe Bereiche konfigurieren, können Benutzer zwar die Objektgruppe durchsuchen, jedoch nicht die Bereichsdaten in den Suchergebnissen anzeigen. Reorganisieren Sie den Index erneut, um sicherzustellen, dass die Suchergebnisse den URI-Bereich des Bereichs wiedergeben.

Wenn der Index mindestens einmal reorganisiert wurde, bevor Sie Bereiche konfigurieren, werden die Änderungen nach der nächsten Indexreorganisation 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 Index zweimal zu reorganisieren, bevor die Änderungen wirksam werden. Wenn Sie vor der ersten Reorganisation des Index für die Objektgruppe Bereiche konfigurieren, können Benutzer zwar die Objektgruppe durchsuchen, jedoch nicht die Bereichsdaten in den Suchergebnissen anzeigen. Reorganisieren Sie den Index erneut, um sicherzustellen, dass die Suchergebnisse den URI-Bereich des Bereichs wiedergeben.

Wenn der Index mindestens einmal reorganisiert wurde, bevor Sie Bereiche konfigurieren, werden die Änderungen nach der nächsten Indexreorganisation 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 103
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.

Ausgeblendete URIs

Die Unternehmenssuche kann die Suchergebnisse so organisieren, dass Dokumente aus Web- und NNTP-Quellen, die das gleiche URI-Präfix aufweisen, als Gruppe angezeigt und 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 wählen.

Benutzer können die Abfragesyntax für die Unternehmenssuche (`samegroupas:uri-prefix`) verwenden, um alle Dokumente zu durchsuchen, die sich in derselben Gruppe wie das 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 jede URI in einer Objektgruppe. Für jede 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.
3. Falls ein Dokument mit keinem URI-Präfix übereinstimmt, gilt Folgendes:
 - Bei Web-URIs verwendet der Indexserver den Hostnamen der URL als URI-Präfix.
 - Bei NNTP-URIs 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 Web- und NNTP-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 der Suchergebnisse werden erst wirksam, wenn der Index das nächste Mal reorganisiert wird.

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/

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. Mit Hilfe 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. Mit Hilfe 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 eine vollständig qualifizierte URI angeben, wird Benutzern die 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 die entfernte URI weiterhin in den Suchergebnissen zurückgegeben. Der Suchcache wird nicht aktualisiert und die URI wird erst aus dem Index entfernt, wenn dieser das nächste Mal aktualisiert oder reorganisiert wird.

Wenn Sie ein Muster zum Entfernen mehrerer URIs angeben, werden den Benutzern weiterhin die URIs, die mit dem Muster übereinstimmen, in den Suchergebnissen angezeigt, bis der Index das nächste Mal aktualisiert oder reorganisiert wird.

Wenn Sie eine URI aus dem Index entfernen, entfernen Sie sie nicht aus dem Crawlerbereich. Wenn der Crawler das Dokument das nächste Mal durchsucht, wird die URI in den Index eingetragen und steht erneut für die Suche zur Verfügung. Sie müssen die Crawlerregeln aktualisieren, um das Dokument auszuschließen, und anschließend den Crawler stoppen und erneut starten, um eine 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*
```

3. Klicken Sie **OK** an.

Zugehörige Verweise

„URI-Formate in einem Index für die Unternehmenssuche“ auf Seite 103
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.

Verwaltung des Suchservers für die Unternehmenssuche

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 mit einem Token versehenen 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 der Abfragen beeinflussen. Dazu gehören auch Optionen, die sich auf die Abfrageleistung auswirken können:

Konfigurieren eines Suchcaches

Zur Optimierung der Abfrageleistung können Sie angeben, dass Ergebnisdokumente in den Cache gestellt werden sollen, und Sie können die Cachespeichergröße angeben, die für diese Dokumente reserviert 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 die 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 bei der Unternehmenssuche“ auf Seite 171

Wenn ein Benutzer eine Objektgruppe durchsucht, geben die Suchvorgänge die relevantesten Ergebnisse für die Abfragebedingungen zurück.

„Benutzerdefinierte Verzeichnisse von Boostwörtern“ auf Seite 172
Zur Verbesserung der Qualität der Suchergebnisse können Sie die Rangfolge von Dokumenten in den Suchergebnissen beeinflussen, indem Sie ein benutzerdefiniertes Verzeichnis von Boostwörtern erstellen.

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 Suchanfrage im Cache vorhanden sind. Wenn die Suchserver die geeigneten Ergebnisdokumente finden, können sie die Suchergebnisse rasch an den Benutzer zurückgeben. Wenn die Suchserver die entsprechenden Ergebnisdokumente nicht finden, durchsuchen sie den Index.

Die ältesten Ergebnisdokumente und Ergebnisdokumente für seltene Abfragen 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 Suchanfragen, deren Ergebnisse gleichzeitig zwischengespeichert werden können).

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 Suchergebnissätze 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 eine Abfrage, die ein Akronym wie z. B. ACL verwendet, Dokumente als Suchergebnis zurückgibt, 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.

Für die Erstellung 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 verwendet werden, das mit WebSphere II OmniFind Edition geliefert wird.

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.

Zugehörige Konzepte

"Synonymunterstützung in Suchanwendungen" in "Integration der Textanalyse"

Zugehörige Tasks

"Erstellen einer XML-Datei für Synonyme" in "Integration der Textanalyse"

"Erstellen eines Synonymverzeichnisses" in "Integration der Textanalyse"

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.

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 die Datei kleiner als 8 MB ist, kann sie sich auf Ihrem System befinden. Wenn die Datei größer als 8 MB ist, muss sie sich auf dem Indexserver befinden.
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 benutzer-

definierten Wörterverzeichnissen enthalten sind, sowie Wörter, die für die Unternehmenssuche vordefiniert sind (z. B. häufige Präpositionen und Artikel).

WebSphere II OmniFind Edition führt standardmäßig eine sprachspezifische Erkennung von Stoppwörtern aus. 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.

Für die Erstellung 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 verwendet werden, das mit WebSphere II OmniFind Edition geliefert wird.

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.

Zugehörige Konzepte

"Benutzerdefinierte Verzeichnisse von Stoppwörtern" in "Integration der Textanalyse"

Zugehörige Tasks

"Erstellen einer XML-Datei für Stoppwörter" in "Integration der Textanalyse"

"Erstellen eines Verzeichnisses von Stoppwörtern" in "Integration der Textanalyse"

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 zuordnen. 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.

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 die Datei kleiner als 8 MB ist, kann sie sich auf Ihrem System befinden. Wenn die Datei größer als 8 MB ist, muss sie sich auf dem Indexserver befinden.
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.

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, die ein Beschreibungssatz enthalten kann. Der Standardwert ist 4.

MaxWordsPerSentence

Die maximale Anzahl Wörter, die ein Beschreibungssatz enthalten kann. Der Standardwert ist 20.

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 WebSphere II OmniFind 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`

Tipp: 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 eine 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 die URI angeben. Dadurch helfen Sie Benutzern, das Dokument einfach zu erkennen und schnell zu ermitteln, ob es abgerufen werden soll.

Beispielsweise könnten Sie für die 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 der 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, die 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 103

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.

Dokumentrangfolge bei der Unternehmenssuche

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 177

Durch die Zuordnung von Feldern zu Boostklassen können Sie beeinflussen, wie Dokumente in den Suchergebnissen eingestuft werden.

„Auf URI-Mustern basierende Dokumentrangfolge“ auf Seite 175

Sie können den Rang von Dokumenten erhöhen oder herabsetzen, indem Sie URI-Mustern Boostfaktoren zuordnen.

„Benutzerdefinierte Verzeichnisse von Boostwörtern“ auf Seite 172

Zur Verbesserung der Qualität der Suchergebnisse können Sie die Rangfolge von Dokumenten in den Suchergebnissen beeinflussen, indem Sie ein benutzerdefiniertes Verzeichnis von Boostwörtern erstellen.

Textbasierte Bewertung

Die Unternehmenssuche berechnet dynamisch eine Bewertung für jedes Dokument, das mit den Begriffen in einer Abfrage übereinstimmt.

Beim Berechnen der Textbewertung für jedes Dokument, das mit einer Abfrage übereinstimmt, zieht die Unternehmenssuche viele Faktoren in Betracht, wie z. B.:

- Die Häufigkeit der einzelnen Abfragebegriffe in der gesamten Objektgruppe. Im Allgemeinen tragen Abfragebegriffe, die in den meisten Dokumenten vorkommen, weniger zur Bewertung eines Dokuments bei als Abfragebegriffe, die in selektiveren Dokumentgruppen vorkommen.
- Die Anzahl der Vorkommen jeden Abfragebegriffs im übereinstimmenden Dokument. Im Allgemeinen ist die Bewertung um so höher, je häufiger die Abfragebegriffe in einem Dokument vorkommen.
- Die räumliche Nähe, mit der Abfragebegriffe in jedem übereinstimmenden Dokument vorkommen. Im Allgemeinen tragen Abfragebegriffe, die in einem Dokument nah zueinander angeordnet sind, mehr zur Dokumentbewertung bei als Abfragebegriffe mit entfernteren Vorkommen.
- Der Kontext, in dem Abfragebegriffe in jedem übereinstimmenden Dokument vorkommen. Beispielsweise trägt ein im Titel eines Dokuments angeordneter

Abfragebegriff mehr zur Bewertung des Dokuments bei als der gleiche Begriff, wenn er im Textkörper des Dokuments vorkommt.

Die Länge eines Dokuments und der Reichtum seines Vokabulars sind ebenfalls Faktoren bei der Ermittlung seiner Bewertung.

Statische Rangfolge

Bei bestimmten Dokumenttypen können Sie einen statischen Rangfolgefaktor zuordnen, der den Rang dieser Dokumente in den Suchergebnissen erhöht.

Beim Erstellen einer Objektgruppe geben Sie an, ob Sie den Dokumenten in der Objektgruppe einen statischen Rangfolgefaktor zuordnen wollen. Bei Webinhalten können die Anzahl Links zu einem Dokument aus anderen Dokumenten und die Ursprünge dieser Links die Relevanz dieses Dokuments in den Suchergebnissen steigern.

Bei Dokumenten mit Datumsfeldern oder Datumsmetadaten können Sie das Datum des Dokuments verwenden, um seine Relevanz zu erhöhen. Beispielsweise sind neuere Artikel in NNTP-Newsgruppen wahrscheinlich relevanter als ältere Artikel. Wenn eine Datenquelle mehrere Datumswerte enthält, können Sie auswählen, welches das wichtigste für das Ermitteln der Relevanz von Dokumenten in der Datenquelle 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 Dokument als den statischen Rangfolgefaktor verwenden wollen, stellen Sie sicher, dass die Objektgruppe nur Webdokumente enthält. Die Suchqualität wird vermindert, wenn Quellen mit verschiedenen Rangfolgemodellen in der gleichen Objektgruppe kombiniert werden.

Sie müssen zudem sicherstellen, dass die Dokumente Felder und Werte enthalten, auf die eine statische Rangfolge angewendet werden kann. Wenn Sie z. B. angeben, dass Sie das Dokumentdatum als einen Faktor für die Rangfolge von Dokumenten in der Objektgruppe verwenden wollen, und die Dokumente keine Datumsfelder oder -attribute enthalten, wird die Suchqualität vermindert.

Zugehörige Tasks

„Migration einer Objektgruppe aus WebSphere Portal“ auf Seite 237

Wenn Sie Objektgruppen 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.

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.

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. (Nur Abfragebegriffe, auf die ein Boostfaktor angewendet wird, beeinflussen den Rang eines Dokuments in den Suchergebnissen, und 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 Boostklassen Feldnamen zuordnen.)

Für die Erstellung 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 verwendet werden, das mit WebSphere II OmniFind Edition geliefert wird.

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.

Zugehörige Konzepte

„Dokumentrangfolge bei der Unternehmenssuche“ auf Seite 171

Wenn ein Benutzer eine Objektgruppe durchsucht, geben die Suchvorgänge die relevantesten Ergebnisse für die Abfragebedingungen zurück.

„Benutzerdefinierte Verzeichnisse von Boostwörtern“ in „Integration der Textanalyse“

Zugehörige Tasks

„Erstellen einer XML-Datei für Boostwörter“ in „Integration der Textanalyse“

„Erstellen eines Verzeichnisses von Boostwörtern“ in „Integration der Textanalyse“

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.

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 die Datei kleiner als 8 MB ist, kann sie sich auf Ihrem System befinden. Wenn die Datei größer als 8 MB ist, muss sie sich auf dem Indexserver befinden.
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 jede 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.

Zugehörige Konzepte

„Dokumentrangfolge bei der Unternehmenssuche“ auf Seite 171

Wenn ein Benutzer eine Objektgruppe durchsucht, geben die Suchvorgänge die relevantesten Ergebnisse für die Abfragebedingungen zurück.

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/*` über-

ein. 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. Reorganisieren Sie den Index, um die Boostfaktoren auf bereits indexierte Dokumente anzuwenden.

Zugehörige Konzepte

„Dokumentrangfolge bei der Unternehmenssuche“ auf Seite 171

Wenn ein Benutzer eine Objektgruppe durchsucht, geben die Suchvorgänge die relevantesten Ergebnisse für die Abfragebedingungen zurück.

Zugehörige Verweise

„URI-Formate in einem Index für die Unternehmenssuche“ auf Seite 103

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.

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

Bei der Zuordnung eines Felds zu einer Boostklasse müssen Sie angeben, ob das Feld zur Erkennung von doppelten Dokumenten verwendet wird:

- Wenn ein Feld zur Erkennung von doppelten Dokumenten verwendet wird, wird das Feld als Inhaltsfeld betrachtet, und nur die für Inhaltsfelder konzipierten Boostklassen können ausgewählt 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.

Werte für viele oder wenige Ergebnisse

Bei der Bewertung einer Abfrage nehmen die Suchvorgänge eine Schätzung vor, wie viele Ergebnisdokumente zurückgegeben werden. Schwellenwerte legen fest, ob eine Abfrage wenige oder viele Ergebnisse zurückgibt:

Wenig 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.

Zwischenwert

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 bei der Unternehmenssuche“ auf Seite 171

Wenn ein Benutzer eine Objektgruppe durchsucht, geben die Suchvorgänge die relevantesten Ergebnisse für die Abfragebedingungen zurück.

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 Felder, um die Bewertungen für Felder und Text berechnen zu können, die bzw. der keine weiteren definierenden Merkmale enthalten bzw. enthält (z. B. ein Feld, das kein Titelfeld ist, oder nicht hervorgehobener Text). 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 Sie das Markierungsfeld inaktivieren, enthält die Liste mit verfügbaren Boostklassen Klassen, die auf Metadatenfelder angewendet werden.
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. Zur Anwendung von Änderungen auf bereits indexierte Dokumente müssen diese Dokumente erneut durchsucht, syntaktisch analysiert und indexiert werden.

Zugehörige Konzepte

„Dokumentrangfolge bei der Unternehmenssuche“ auf Seite 171

Wenn ein Benutzer eine Objektgruppe durchsucht, geben die Suchvorgänge die relevantesten Ergebnisse für die Abfragebedingungen zurück.

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 bei der Unternehmenssuche“ auf Seite 171

Wenn ein Benutzer eine Objektgruppe durchsucht, geben die Suchvorgänge die relevantesten Ergebnisse für die Abfragebedingungen zurück.

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
```

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 3. Standardwerte für Boostklassen

Hohe und niedrige Standardboostfaktoren				
Boostklassenname	Keine statische Rangfolge	Dokumentverknüpfung	Dokumentdatum	Vordefinierte Feldzuordnungen
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

Tabelle 3. Standardwerte für Boostklassen (Forts.)

Hohe und niedrige Standardboostfaktoren				
Boostklassenname	Keine statische Rangfolge	Dokumentverknüpfung	Dokumentdatum	Vordefinierte Feldzuordnungen
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	
Metadatenklasse A	Niedrig: 1 Hoch: 1	Niedrig: 1 Hoch: 1	Niedrig: 1 Hoch: 1	es_special_field.default_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
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	description

Zugehörige Konzepte

„Dokumentrangfolge bei der Unternehmenssuche“ auf Seite 171
 Wenn ein Benutzer eine Objektgruppe durchsucht, geben die Suchvorgänge die relevantesten Ergebnisse für die Abfragebedingungen zurück.

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.

Zugehörige Konzepte

„Indexierte Optionen für das Durchsuchen von Dokumenten“ auf Seite 148
Wenn Sie Optionen für das Durchsuchen von mit dem Crawler durchsuchten Daten oder externen Quellen konfigurieren oder XML- und HTML-Metadaten-elemente zum Durchsuchen von Feldern zuordnen, geben Sie an, wie die Dokumente durchsucht und in den Suchergebnissen angezeigt werden.

„Sicherheit mit Suchanwendungs-IDs“ auf Seite 214

Wenn Sie Sicherheit auf Objektgruppenebene bereitstellen möchten, geben Sie an, welche Suchanwendungen die jeweiligen Objektgruppen und externen Quellen durchsuchen dürfen.

„Sicherheit auf Dokumentebene“ auf Seite 214

Wenn die Sicherheit bei der Erstellung einer Objektgruppe aktiviert ist, können Sie Sicherheitsmaßnahmen auf Dokumentebene konfigurieren. Die Sicherheit auf Dokumentebene sorgt dafür, dass Benutzer, die Objektgruppen durchsuchen, nur auf die Dokumente zugreifen können, für deren Anzeige sie die Berechtigung haben.

„Sicherheit auf Dokumentenebene mit Portal Search Engine“ auf Seite 227
Mit IBM WebSphere Portal Search Engine können Sie die Sicherheit auf Dokumentenebene verbessern, wenn der Benutzer Objektgruppen für die Unternehmenssuche durchsucht.

„Search and index API overview“ in „Programming Guide and API Reference for Enterprise Search“

„Query syntax“ in „Programming Guide and API Reference for Enterprise Search“

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 eine, mehrere oder 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. nach bestimmten Feldern in einem Dokument suchen oder mit 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 der Dokumente nach Relevanz und nach Dokumentdatum (in aufsteigender Reihenfolge oder absteigender Reihenfolge).
- 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 der Dokument-URI. Wenn Clientanwendungen verfügbar sind, können Sie ein Ergebnisdokument möglicherweise auch mit einer nativen Anzeigefunktion aufrufen. Wenn zum Beispiel eine Notes-Clientanwendung installiert ist, können Sie mit dieser Anwendung Dokumente aus einer Lotus Notes-Datenbank anzeigen.

Wenn für einen Crawler die Sicherheit auf Dokumentenebene 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.

Merkmale von Mustersuchanwendungen

Sie können die Datei `config.properties` für die Mustersuchanwendung 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.

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.

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. Ihr Computer-Hostname "MeineMaschine" und als Hostname für den Web-Server "www.ibm.com" konfiguriert ist, geben Sie `www.ibm.com` an.

port Gibt die Portnummer des Web-Servers an, der zur Unterstützung Ihrer Instanz von WebSphere Application Server konfiguriert ist. Der Standardwert ist 80. Ändern Sie den Standardwert nur, wenn Sie den Port für Ihre Web-Server-Instanz ändern.

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.

username

Gibt einen Benutzernamen an, mit dem die Unternehmenssuche die Benutzerauthentifizierung umsetzen und Benutzer für WebSphere Application Server authentifizieren kann, wenn diese Suchanforderungen übergeben. Dieses Feld wird nur dann verwendet, wenn Sie die globale Sicherheit in WebSphere Application Server aktiviert haben.

password

Gibt das Kennwort für den angegebenen Benutzernamen 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.

INFO Es werden Informationsnachrichten protokolliert.

FINE Es werden Fehlernachrichten protokolliert.

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 deren Hilfe die bei der Installation von WebSphere Information Integrator OmniFind Edition unterstützten Crawler und externen Quellen angegeben werden, sind in der Datei `config.properties` vordefiniert.



documentSource.vbr.icon

Gibt den Pfad und 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 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 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 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 Namen einer Bilddatei an, die darauf hinweist, dass das

Dokument von einem Exchange Server-Crawler durchsucht wurde. Das Standardsymbol ist /images/sourceExchange.gif.



documentSource.nntp.icon

Gibt den Pfad und 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 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 Namen einer Bilddatei an, die darauf hinweist, dass das Dokument von einem QuickPlace-Crawler durchsucht wurde. Das Standardsymbol ist /images/sourceWorkplace.gif.



documentSource.unixfs.icon

Gibt den Pfad und Namen einer Bilddatei an, die darauf hinweist, dass das Dokument von einem Crawler für das UNIX-Dateisystem durchsucht wurde. Das Standardsymbol ist /images/sourceUnixFS.gif.



documentSource.web.icon

Gibt den Pfad und Namen einer Bilddatei an, die darauf hinweist, dass das Dokument von einem Web-Crawler durchsucht wurde. Das Standardsymbol ist /images/sourceWeb.gif.



documentSource.wps.icon

Gibt den Pfad und 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 Namen einer Bilddatei an, die darauf hinweist, dass das Dokument von einem Crawler für das Windows-Dateisystem durchsucht wurde. Das Standardsymbol ist /images/sourceWindowsFS.gif.



documentSource.ldap.icon

Gibt den Pfad und 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 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. Wenn eine für den Ergebnisdokumententyp native Clientanwendung vorhanden ist, kann das Dokument auch mit der Clientanzeigefunktion angezeigt werden. Sie können die

Bilder anpassen, die den Typ der Clientanzeigefunktion darstellen, mit der das Dokument angezeigt wird. Im folgenden Beispiel weist das Lotus Notes-Symbol darauf hin, dass das Dokument mit einer Lotus Notes-Clientanwendung angezeigt werden kann:

```
client.notes.icon=/images/notes.gif
```

In den Suchergebnissen werden das Symbol und der Link für die Clientanwendung 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\://=servername datenbankname databanktitel dateiname  
erweiterung \ erstellungsdatum Änderungsdatum  
fields.cm\://=servername elementtypname erstellungsdatum Änderungsdatum mime-typ  
fields.file\://=verzeichnis dateiname erweiterung Änderungsdatum dateigröße titel
```

Feldsymbole

Für Datenquellentypen und Dokumente mit Feldern können Sie die Bilder anpassen, die zur Darstellung der Felder verwendet werden. Die folgenden Feldsymbole sind in der Datei config.properties vordefiniert.



field.icon.databasetitle

Gibt den Pfad und Namen einer Bilddatei an, die darauf hinweist, dass das Feld den Dokumenttitel enthält. Das Standardsymbol ist /images/notesdb.gif.



field.icon.databasesname

Gibt den Pfad und 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 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 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.

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 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 Text 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, Präsentationsvorlage der IBM Software Group, \
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 die mit einer URI-Präfixregel übereinstimmenden URIs in den Suchergebnissen ausgeblendet werden, wenn Regeln für das Ausblenden von Sites in der Administrationskonsole für die Unternehmenssuche konfiguriert wurden.

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.

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

Gibt an, ob der Knopf **Aktualisieren** angezeigt wird, und steuert die Funktionalität zur Aktualisierung der Suchanwendung. Wenn Sie diese Option auf "true" setzen, können die Benutzer die Liste der Objektgruppen und externen Quellen aktualisieren, die für die Suche zur Verfügung stehen (z. B. wenn die Suchanwendung in der Administrationskonsole für die Unternehmenssuche zusätzlichen Objektgruppen oder externen Quellen zugeordnet wurde).

Möglicherweise wollen Sie den Knopf **Aktualisieren** anzeigen, wenn Sie an der Datei `config.properties` 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 Unternehmensanwendung "ESSearchApplication" 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 Unternehmensanwendung "ESSearchServer" in WebSphere Application Server nicht gestartet wurde), wird der Knopf **Aktualisieren** automatisch angezeigt, um Sie bei der Fehlerbehebung zu unterstützen.

extraMessages.show=false

Steuert die Anzeige eines Bereichs unten in den Suchergebnissen, in dem Warnungen und Informationsnachrichten angezeigt werden. Fehlermeldungen werden stets oben auf der Seite angezeigt. Setzen Sie diese Option auf "true", wenn Sie zusätzliche Nachrichten anzeigen möchten.

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.

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.

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.

Benutzerdefiniertes Banner und Logo

Sie können die Bilder im Bannerbereich oben in der Suchanwendung anpassen. Sie könnten z. B. die Standardbilder für WebSphere II OmniFind Edition durch Bilder ersetzen, die das Branding Ihres Unternehmens widerspiegeln. Wenn Sie keinen 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_OFEdition.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 WebSphere II OmniFind Edition durch Bilder ersetzen, die das Branding 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 für die Suche der Datei `config.properties` geben die Namen der JSPs (Java Server Pages) an, die für Seiten mit Indexzungen in der Sicht **Suchen** der Suchanwendung (**Suchen**, **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 Indexungen 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 Indexungen 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 für die Suche 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 SIAPI (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
```

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 benutzerdefinierten 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.

Mehrere benutzerdefinierte Filter sind auf Kommentar gesetzt und stehen als Beispiele in der Datei `config.properties` zur Verfügung.

Bearbeiten der Merkmale der Mustersuchanwendung

Die Mustersuchanwendung für die Unternehmenssuche kann alle aktiven Objektgruppen und externen Quellen in Ihrem System durchsuchen. Sie können eine Merkmaldatei bearbeiten, um Optionen für Ihre Web-Serverumgebung anzugeben, eine andere Suchanwendung als Standardanwendung zu verwenden oder 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 Merkmaldatei `config.properties`.

Stoppen Sie die Unternehmensanwendung "ESSearchApplication" 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. 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 WebSphere II OmniFind Edition-Installationsverzeichnis auf dem Suchserver:

```
ES_INSTALL_ROOT/installedApps/ESSearchApplication.ear/  
ESSearchApplication.war/WEB-INF/config.properties
```

2. Bearbeiten Sie die Merkmale, und geben Sie Informationen zu Ihrer Web-Serverumgebung und den Suchvorgaben an. Speichern und schließen Sie dann die Datei. (Das Nummernzeichen (#) in der Datei gibt eine Kommentarzeile an.)
3. Stoppen Sie die Anwendung "ESSearchApplication", und starten Sie sie erneut:

- a. Starten Sie die Administrationskonsole von WebSphere Application Server auf dem Suchserver.
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.
- b. Wenn Sie zur Eingabe von Benutzer-ID und Kennwort aufgefordert werden, geben Sie eine Benutzer-ID und ein Kennwort ein, die in der WebSphere Application Server-Benutzerregistry registriert sind. Wenn Sie das lokale Betriebssystem als Benutzerregistry verwenden, können Sie ID und Kennwort des Administrators für die Unternehmenssuche angeben.
- c. Nachdem Sie sich an der Administrationskonsole angemeldet haben, klicken Sie **Anwendungen** und anschließend **Enterprise-Anwendungen** an.
- d. Wählen Sie das Markierungsfeld für ESSearchApplication aus, und klicken Sie **Stoppen** an.
- e. Wählen Sie das Markierungsfeld für die Anwendung erneut aus, nachdem die Anwendung gestoppt wurde, und klicken Sie **Starten** an.

Zugreifen auf die Mustersuchanwendung

Sie greifen auf die Mustersuchanwendung zu, indem Sie eine URL-Adresse in einem Web-Browser angeben.

Vorbereitung

Sie müssen die Mustersuchanwendung für Ihre Web-Serverumgebung 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 die Mustersuchanwendung zu starten:

1. Geben Sie die URL-Adresse für die Suchanwendung in einem Web-Browser ein.
Beispiel:
`http://suchserver.com/ESSearchApplication/`

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/`

2. Wenn die Sicherheit in WebSphere Application Server aktiviert ist, melden Sie sich an 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 anmelden zu müssen.

3. Wählen Sie auf der Seite **Suche** die zu durchsuchenden Objektgruppen und externen Quellen aus, und übergeben Sie eine Abfrage.

Aktivieren der Sicherheit für die Mustersuchanwendung

Wenn Sie globale Sicherheit in WebSphere Application Server aktivieren und die Mustersuchanwendung für das Durchsuchen sicherer Objektgruppen verwenden möchten, müssen Sie die Konfigurationseinstellungen in der Musteranwendung sowie in WebSphere Application Server ändern.

Vorbereitung

- Sie müssen die Rolle eines Administrators für die Unternehmenssuche haben.
- Sie müssen die globale Sicherheit in WebSphere Application Server aktivieren. Wenn Sie WebSphere II OmniFind Edition für eine Konfiguration mit mehreren Servern installiert haben, aktivieren Sie die globale Sicherheit auf den Suchservern. Anweisungen zur Aktivierung der globalen Sicherheit finden Sie in der Dokumentation von WebSphere Application Server.
- Wenn Sie für die Benutzerregistry in WebSphere Application Server Lightweight Directory Access Protocol (LDAP) wählen, müssen Sie ID und Kennwort des Administrators für die Unternehmenssuche Ihrer LDAP-Registry hinzufügen (diese ID und das Kennwort wurden bei der Installation von WebSphere II OmniFind Edition angegeben).

Angenommen, die ID des Administrators für die Unternehmenssuche ist `adminUser`, dann könnte der Benutzereintrag in der LDAP-Registry lauten:
`uid=adminUser,ou=Employees,o=IBM,c=US`. Anweisungen hierzu finden Sie in der Dokumentation für LDAP-Server.

Vorgehensweise

Gehen Sie wie folgt vor, um die Sicherheit für die Musteranwendung für die Unternehmenssuche zu aktivieren:

1. Aktualisieren Sie die Merkmale der Suchanwendung in der Administrationskonsole:
 - a. Melden Sie sich an der Administrationskonsole für die Unternehmenssuche als Benutzer mit den Zugriffsrechten eines Administrators für die Unternehmenssuche an.
 - b. Klicken Sie in der Funktionsleiste **Sicherheit** an.
 - c. Klicken Sie **Suchanwendungen konfigurieren** auf der Seite **Suchanwendungen** an.

- d. Klicken Sie **Suchanwendung hinzufügen** an, und geben Sie in das Feld **Name der Suchanwendung** die ID des Administrators für die Unternehmenssuche ein, die bei der Installation von WebSphere II OmniFind Edition angegeben wurde.
 - e. Stellen Sie sicher, dass **Alle Objektgruppen und externen Quellen** ausgewählt ist, und klicken Sie **OK** an.
2. Bearbeiten Sie die Datei config.properties:
 - a. Wenn Sie unter UNIX arbeiten, öffnen Sie das Konsolfenster. Wenn Sie Microsoft Windows verwenden, öffnen Sie ein Fenster mit Eingabeaufforderung.
 - b. Wechseln Sie in das Verzeichnis WEB-INF für die Mustersuchanwendung. Die folgenden Beispiele werden der Lesbarkeit halber auf zwei Zeilen angezeigt; geben Sie den Befehl jedoch in einer einzigen Zeile an:

UNIX:

```
cd $ES_INSTALL_ROOT/installedApps/ESSearchApplication.ear/ESSearchApplication.war/WEB-INF
```

Windows:

```
cd %ES_INSTALL_ROOT%\installedApps\ESSearchApplication.ear\ESSearchApplication.war\WEB-INF
```
 - c. Verwenden Sie für das Bearbeiten der Datei config.properties einen Texteditor.
 - d. Ändern Sie das Benutzernamenmerkmal in den Namen eines gültigen WebSphere Application Server-Benutzers.
 - e. Ersetzen Sie das Kennwortmerkmal durch das Kennwort für den angegebenen Benutzer.
 - f. Speichern und schließen Sie die Datei.
 3. Starten Sie die Anwendung ESSearchApplication in WebSphere Application Server erneut:
 - a. 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.
 - b. 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.
 - c. Nachdem Sie sich an der Administrationskonsole angemeldet haben, klicken Sie **Anwendungen** und anschließend **Enterprise-Anwendungen** an.

- d. Wählen Sie das Markierungsfeld für ESSearchApplication aus, und klicken Sie **Stoppen** an.
- e. Wählen Sie das Markierungsfeld für ESSearchApplication erneut aus, nachdem die Anwendung gestoppt wurde, und klicken Sie **Starten** an.

Externe Quellen für die Unternehmenssuche

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:

- JDBC-Datenbanktabellen (Java Database Connectivity) (nur IBM DB2 Universal Database-Datenbanken (DB2 UDB) und Oracle-Datenbanken). Für jede Tabelle in einer JDBC-Datenbank wird eine separate externe Quelle erstellt.
- 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

"Search and index API federators" in "Programming Guide and API Reference for Enterprise Search"

Hinzufügen von externen Quellen zum System

Wenn Sie dem System für die Unternehmenssuche eine externe Quelle hinzufügen, müssen Sie den gewünschten Quellentyp angeben. Mit Hilfe 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.

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 mit Hilfe 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 mit Hilfe 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.

Vorgehensweise

Gehen Sie wie folgt vor, um dem System eine externe Quelle hinzuzufügen:

1. Klicken Sie **Externe Quellen** an, um die Sicht **Externe Quellen** zu öffnen.
2. Klicken Sie **Externe Quelle hinzufügen** an.
3. Wählen Sie den Typ der externen Quelle aus, den Sie hinzufügen wollen (LDAP-Server oder JDBC-Datenbank).
4. 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 Universal Database (DB2 UDB) und Oracle-Datenbanken behilflich sein:

DB2: Traditioneller JDBC-Treiber

Treibername: `COM.ibm.db2.jdbc.app.DB2Driver`

Speicherposition: `db2-installationsroot/java/db2java.zip`

DB2: Universeller JDBC-Treiber

Treibername: `com.ibm.db2.jcc.DB2Driver`

Speicherposition:

`db2-installationsroot/java/db2jcc.jar`

`db2-installationsroot/java/db2jcc_license_cu.jar`

Oracle Treibername: `oracle.jdbc.driver.OracleDriver`

Speicherposition:

`oracle-ausgangsverzeichnis/jdbc/lib/classes12.zip`

`oracle-ausgangsverzeichnis/jdbc/lib/nls_charset12.zip`

5. 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

"Search and index API federators" in "Programming Guide and API Reference for Enterprise Search"

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

"Search and index API federators" in "Programming Guide and API Reference for Enterprise Search"

Sicherheit für die Unternehmenssuche

Über die Sicherheitsmechanismen für die Unternehmenssuche können Sie Ihre Quellen gegen unberechtigtes Durchsuchen schützen und Verwaltungsfunktionen auf bestimmte Benutzer beschränken.

Mit Hilfe 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.

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 versuchen, das System zu 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 ebenfalls die Unterstützung für die Authentifizierung in WebSphere Application Server verwenden, um Benutzer zu authentifizieren, die mit der Suchanwendung 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.

Wenn die Sicherheit auf Objektgruppenebene aktiviert ist, wenden die globalen Analyseprozesse die folgenden speziellen Regeln an:

- Dokumente mit identischem oder fast identischem Inhalt werden unabhängig voneinander indexiert, anstatt den Inhalt gemeinsam in einer kanonischen Darstellung zu indexieren, um sicherzustellen, dass die Sicherheitsmaßnahmen für jedes Dokument bewertet werden.
- 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. Dadurch kann das Zieldokument von einer Abfrage abgerufen werden, in der Text angegeben ist, der in dem Quelldokument erscheint. Dieser Typ der Ankertextverarbeitung stellt ein Sicherheitsrisiko dar, falls es dem Benutzer erlaubt ist, die Zieldokumente anzuzeigen, aber nicht die Quelldokumente. Wenn die Objektgruppensicherheit aktiviert ist, wird der Ankertext in den Links zu verbotenen Dokumenten vom Index ausgeschlossen. Ein Dokument wird nur als Suchergebnis zurückgegeben, wenn sein Inhalt oder seine Metadaten mit der Abfrage übereinstimmen.

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 Objektgruppenebene ist für Ihre Suchanwendungen auch über eine Anwendungs-ID verfügbar. Für das Durchsuchen von Objektgruppen muss ein Administrator für die Unternehmenssuche die Suchanwendung den jeweiligen Objektgruppen zuordnen, die Ihre Anwendung durchsuchen darf. Sie können dann die Standardzugriffssteuerungsmaßnahmen verwenden, um Benutzern den Zugriff auf Suchanwendungen zu erlauben oder zu verweigern.

Sicherheit auf Dokumentebene

Wenn Sie Crawler für eine Objektgruppe konfigurieren, können Sie die Sicherheit auf Dokumentebene aktivieren. Wenn Sie sich für diese Option entscheiden, kann der Crawler jedem Dokument, das er durchsucht, Sicherheitstoken zuordnen. Die Sicherheitstoken werden für die Dokumente im Index gespeichert. Für einige Crawler können Sie außerdem angeben, ob die Benutzerberechtigungen bei der Abfrageverarbeitung mit aktuellen Zugriffsteuerungsdaten (wie in der nativen Datenquelle konfiguriert) geprüft werden sollen.

Ihre Suchanwendungen können die Sicherheitstoken und Benutzerberechtigungen für die Umsetzung der Zugriffssteuerung verwenden. Eine Suchanwendung kann Berechtigungsnachweise des angemeldeten Benutzers in die Abfragen einbeziehen, die sie an die Suchserver weitergibt, um sicherzustellen, dass der Benutzer lediglich die Dokumente durchsucht und abrufen, für die er eine Zugriffsberechtigung hat.

Die Sicherheit für Ihre Objektgruppe umfasst die Authentifizierung und Zugriffssteuerungsmechanismen, die die Unternehmenssuche verwenden kann, um indextierten 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 Ersatzports 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.

Verwaltungsrollen

Bei der Unternehmenssuche wird das Konzept der Rollen verwendet, um den Zugriff auf verschiedene Funktionen in der Administrationskonsole zu steuern.

Während der Installation von WebSphere Information Integrator OmniFind Edition (WebSphere II OmniFind Edition) 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 WebSphere II OmniFind 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' andere Benutzer zuordnen.

Objektgruppenadministrator

Dieser Benutzer kann die Handhabung bestimmter oder aller Objektgruppen bearbeiten, überwachen und steuern. 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

Diese Benutzer können die Handhabung bestimmter Objektgruppen oder aller Objektgruppen überwachen und steuern. 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

Diese Benutzer können bestimmte Objektgruppen oder alle Objektgruppen überwachen. 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.

Zugehörige Tasks

„Anmelden an der Administrationskonsole“ auf Seite 19

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

„Starten der Server für die Unternehmenssuche“ auf Seite 243

Damit Benutzer eine Objektgruppe durchsuchen können, müssen Sie die Systemprozesse und anschließend die Server für die Crawlersuche, die Syntaxanalyse, die Indexierung und die Suche für die Objektgruppe starten.

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 Funktionsleiste **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.

Authentifizierung und Zugriffssteuerung

Die Unternehmenssuche unterstützt die Benutzerauthentifizierung und die Zugriffssteuerung, um Ihre Inhalte vor nicht berechtigten Benutzern zu schützen, und um den Zugriff auf Verwaltungsfunktionen zu steuern.

Authentifizierung

Die Authentifizierung ist ein Prozess, mit dem ein System prüft, ob ein Benutzer die Person ist, die er vorgibt zu sein. Da der Zugriff normalerweise auf der Identi-

tät des Benutzers beruht, der die Ressource angefordert hat, ist die Authentifizierung entscheidend für effektive Sicherheit.

Die Unternehmenssuche nutzt für die Authentifizierung von Benutzern, die versuchen, auf die Administrationskonsole zuzugreifen, die Authentifizierungsunterstützung, die in WebSphere Application Server bereitgestellt wird.

Ihre Suchanwendungen können die Sicherheit in WebSphere Application Server nutzen, um Benutzer zu authentifizieren, die Objektgruppen für die Unternehmenssuche durchsuchen, und um die von Ihnen bevorzugten Methoden für die Authentifizierung von Benutzerberechtigungen zu implementieren. Im Normalfall besteht die Benutzerberechtigung aus einer Benutzer-ID und einem Kennwort, die an die Suchanwendung übergeben werden, wenn ein Benutzer sich anmeldet oder versucht, auf die Suchanwendung zuzugreifen.

Die Benutzerauthentifizierung kann auch auf andere Art implementiert werden, je nach den Ressourcen und Protokollen, die in Ihren Unternehmen verfügbar sind. Sie könnten z. B. Benutzer dazu zwingen, sich durch eine Smart Card, durch das Verwalten von digitalen Zertifikaten und einer PKI-Infrastruktur (public key), oder durch das Zuweisen von Tickets zu identifizieren, wenn der Benutzer sich anmeldet, um seinen Authentifizierungsstatus zu verfolgen.

Zugriffssteuerung

Die Zugriffssteuerung grenzt ein, was ein Benutzer tun darf, nachdem er sich identifiziert hat und authentifiziert wurde. Eine Zugriffssteuerungsliste (ACL - Access Control List) ist die allgemein üblichste Methode, mit der der Zugriff auf Ressourcen eingeschränkt werden kann. Eine ACL ist eine Liste von Benutzer-IDs (Benutzernamen, Gruppennamen, Benutzerrollen usw.). Jede Benutzer-ID ist einer Gruppe von Berechtigungen zugeordnet, die die Berechtigungen und Zugriffsrechte des Benutzers festlegen.

So können z. B. Zugriffssteuerungen den Zugriff auf Dateien auf einem Dateiserver zulassen oder verweigern und steuern, ob ein Benutzer, dem der Zugriff erlaubt ist, Dateien auf diesem Server lesen, erstellen, bearbeiten oder löschen darf.

In der Unternehmenssuche hängt die Zugriffssteuerung davon ab, ob ein Benutzer die Berechtigung hat, Daten im Index zu lesen. Sie können Folgendes tun, je nachdem, wie Sie die Zugriffssteuerungen in Ihren Suchanwendungen aktivieren, und je nach den Regeln, die Sie für die Objektgruppe und für die Crawler festlegen, wenn Sie die Unternehmenssuche verwalten:

- Allen Benutzern erlauben, alle Dokumente in einer Objektgruppe zu durchsuchen.
- Allen Benutzern erlauben, alle Dokumente zu durchsuchen, die von bestimmten Crawlern durchsucht wurden, und den Zugriff auf Dokumente begrenzen, die von anderen Crawlern durchsucht wurden.
- Bestimmten Benutzern das Durchsuchen bestimmter Dokumente erlauben. Beispiel: Wenn Sie die Datenbanken angeben, die Sie mit einem Notes-Crawler durchsuchen möchten, können Sie Optionen angeben, die es bestimmten Benutzern ermöglichen, Dokumente in bestimmten Sichten und Ordnern zu durchsuchen, und andere Benutzer daran hindern, diese Dokumente zu durchsuchen.

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, welche WebSphere II OmniFind Edition-Aktivitäten eine Benutzerauthentifizierung erfordern.

Informationen zu dieser Task

Das Installationsprogramm von WebSphere II OmniFind 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 WebSphere II OmniFind Edition-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 WebSphere II OmniFind Edition-Unternehmensanwendungen aktivieren bzw. inaktivieren. Beispiel: Sie möchten alle Benutzer authentifizieren, die auf die Administrationskonsole für die Unternehmenssuche zugreifen, aber nicht die Benutzer, die die WebSphere II OmniFind Edition-SI-API-Schnittstellen oder die Mustersuchanwendung verwenden.

Vorgehensweise

Gehen Sie wie folgt vor, um die Sicherheit für eine bestimmte 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. Klicken Sie wieder **Anwendungen** und dann **Enterprise-Anwendungen** an.
12. Wählen Sie das Markierungsfeld für die Anwendung aus, die Sie gerade geändert haben, und klicken Sie **Stoppen** an.
13. Wählen Sie nach dem Stoppen der Anwendung das Markierungsfeld für die Anwendung erneut aus, und klicken Sie **Starten** an.

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 und Ankertext in Dokumenten an.

Nachdem Sie eine Suchanwendung erstellt haben, können Sie mit Hilfe einer Suchanwendungs-ID festlegen, welche Objektgruppen die Suchanwendung durchsuchen darf, und welche Benutzer auf die Suchanwendung zugreifen dürfen.

Analyse auf doppelte Dokumente

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. Dies stellt sicher, dass der Benutzer nur jene Dokumente durchsucht, deren Sicherheitstoken mit seinem Berechtigungsnachweis übereinstimmen. Zwei

Dokumente können z. B. nahezu identischen Inhalt haben, aber verschiedene Zugriffssteuerungslisten für die Umsetzung der Sicherheit verwenden.

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.

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 indexiert 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` auf einer Website eine Verbindung herstellt:

```
<a href=" ../doc/syntax.htm">Abfragesyntax</a>
```

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 stehenden Beispiel ist der Ankertext `Abfragesyntax` der Zielseite `syntax.htm` zugeordnet, sowie der Seite, die das Ankerkonstrukt enthält.

Wenn Sie beim Erstellen einer Objektgruppe die Objektgruppensicherheit aktivieren, wird die Ankertextverarbeitung inaktiviert. Der Ankertext wird nicht mehr mit einem Dokument indexiert, 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, da der Ankertext zu verbotenen Dokumenten nicht den Dokumenten zugeordnet wird, auf die die Benutzer zugreifen dürfen.

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 der einzelnen Web-Crawler eine Option auswählen, mit der der Ankertext in Links zu verbotenen Dokumenten indexiert werden soll. Wenn Sie angeben, dass der Ankertext indexiert werden soll, indexieren die Analyse- und Indexierungsprozesse den Ankertext für alle Seiten, die vom Web-Crawler abgerufen wurden. Wenn Sie angeben, dass der Ankertext nicht indexiert werden soll, wird der Ankertext in Links zu verbotenen Dokumenten aus dem Index ausgeschlossen.

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, wählen Sie auf der Seite **Crawler** den Web-Crawler aus, 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 Sie die Dokumente erneut vom Crawler durchsuchen lassen, so dass 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.

Sicherheit mit Suchanwendungs-IDs

Wenn Sie Sicherheit auf Objektgruppenebene bereitstellen möchten, geben Sie an, welche Suchanwendungen die jeweiligen Objektgruppen und externen Quellen durchsuchen dürfen.

Alle Suchanwendungen werden aufgefordert, eine Anwendungs-ID an die APIs für die Unternehmenssuche zu übergeben. Ein Administrator für die Unternehmenssuche und Ihre Suchanwendungen können diese ID verwenden, um die Sicherheit für Objektgruppen und externe Quellen umzusetzen.

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.

Für die Umsetzung der Zugriffssteuerung können Sie Ihrer Suchanwendung Sicherheitstoken (z. B. Benutzer-IDs, Gruppen-IDs oder Benutzerrollen) zuordnen und nur diesen Benutzern den Zugriff auf die Anwendung erlauben. 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

„Suchanwendungen für die Unternehmenssuche“ auf Seite 183

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.

"Search and index API overview" in "Programming Guide and API Reference for Enterprise Search"

"Search and index API security" in "Programming Guide and API Reference for Enterprise Search"

Sicherheit auf Dokumentenebene

Wenn die Sicherheit bei der Erstellung einer Objektgruppe aktiviert ist, können Sie Sicherheitsmaßnahmen auf Dokumentenebene konfigurieren. Die Sicherheit auf Dokumentenebene sorgt dafür, dass Benutzer, die Objektgruppen durchsuchen, nur auf die Dokumente zugreifen können, für deren Anzeige sie die Berechtigung haben.

Crawler können Sicherheitseinstellungen für den Index übernehmen, um den Zugriff auf Dokumente in einer Objektgruppe zu steuern. Für manche Datenquellentypen können Sie auch aktuelle Berechtigungsnachweise des Benutzers prüfen, wenn eine Abfrage übergeben wird. Zur Prüfung des aktuellen Berechtigungsnachweises des Benutzers können Sie in Ihre benutzerdefinierten Suchanwendungen Unterstützung für Benutzerprofile integrieren. Durch die Speicherung der Benutzerprofile können Benutzer auf Dokumente zugreifen, ohne dass sie mehrfach aufgefordert werden, ihre Berechtigungsnachweise anzugeben.

Zugehörige Konzepte

„Suchanwendungen für die Unternehmenssuche“ auf Seite 183

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.

„Sicherheit auf Dokumentebene mit Portal Search Engine“ auf Seite 227

Mit IBM WebSphere Portal Search Engine können Sie die Sicherheit auf Dokumentebene verbessern, wenn der Benutzer Objektgruppen für die Unternehmenssuche durchsucht.

“Search and index API security” in “Programming Guide and API Reference for Enterprise Search”

Prüfung durch gespeicherte Sicherheitstoken

Wenn die Sicherheit für eine Objektgruppe bei ihrer Erstellung aktiviert ist, können Sie Sicherheitsmaßnahmen auf Dokumentebene konfigurieren, indem Sie im Index Sicherheitsdaten speichern.

Jedes Dokument wird standardmäßig als allgemein zugängliches Dokument betrachtet, d. h., es kann von allen Benutzern durchsucht werden. Bei den meisten Dokumenttypen können Sie die Sicherheit auf Dokumentebene erzielen, indem mindestens ein Sicherheitstoken den Dokumenten zugeordnet wird und diese Token mit den Dokumenten im Index gespeichert werden. 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 Datenquellentyp Felder einschließt, können Sie festlegen, dass Sie den Wert in einem dieser Felder für die Umsetzung der Zugriffssteuerung verwenden möchten. Wenn die Datenquelle keine Felder aufweist, wenn Sie keinen Feldwert zu Sicherheitszwecken verwenden möchten, oder wenn von Ihnen angegebene Feld keinen Wert enthält, der die umzusetzenden Zugriffssteuerungen aktiviert, können Sie Sicherheitstoken für den Crawler definieren, die Dokumenten zugeordnet werden.

Der Administrator der jeweiligen Objektgruppe entscheidet über den Sicherheitstoken, den der Crawler Dokumenten zuordnen soll. Ein Sicherheitstoken kann z. B. durch eine Benutzer-ID, eine Gruppen-ID, eine Benutzerrolle, oder einen beliebigen anderen Wert dargestellt werden, den sie als gültig für die Datenquelle festlegen. Wenn ein Datenquellenadministrator die native Zugriffssteuerungsliste aktualisiert, sind die aktualisierten Sicherheitsmaßnahmen beim nächsten Aktualisieren oder Reorganisieren des Index verfügbar.

Sicherheitstoken begleiten Dokumente während der Arbeitsgänge des Parsing, der Analyse und der Indexierung. Wenn Ihre Suchanwendungen die Sicherheit aktivieren, können Sie mit den Sicherheitstoken den Dokumentzugriff steuern. Benutzer, die die Objektgruppe durchsuchen, können nur die Dokumente durchsuchen, für die sie auf Grund ihrer Berechtigungsnachweise berechtigt sind. Wenn der Berechtigungsnachweis eines Benutzers nicht den Sicherheitsregeln entspricht, kann der Benutzer die durch die Sicherheitstoken geschützten Dokumente nicht durchsuchen.

Sie können benutzerdefinierte Geschäftsregeln zur Bestimmung des Werts für die Sicherheitstoken anwenden, 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.

Prüfung aktueller Berechtigungsnachweise während der Abfrageverarbeitung

Wenn die Sicherheit für eine Objektgruppe bei ihrer Erstellung aktiviert wurde, können Sie mit bestimmten Domämentypen die aktuellen Berechtigungsnachweise des Benutzers prüfen, wenn der Benutzer eine Abfrage übergibt.

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:

- Content Edition-Crawler (nur Repositorytypen für Documentum, FileNet Pangen Content Services und Portal Document Manager)
- Domino Document Manager-Crawler
- Notes-Crawler
- QuickPlace-Crawler
- WebSphere Portal-Crawler
- Windows-Dateisystem-Crawler

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.

Diese Methode zur Umsetzung der Sicherheit auf Dokumentenebene bietet ein hohes Maß an Sicherheit, da die Berechtigungsnachweise des Benutzers mit den aktuellen Sicherheitsdaten verglichen werden und nicht mit denen, die im Index gespeichert sind. Hierdurch wird auch sichergestellt, dass der Zugriff von den Sicherheitsmechanismen des nativen Repositories ohne Rücksicht darauf kontrolliert wird, wie komplex jene Mechanismen sein könnten. Da das Filtern der Dokumente in Echtzeit erfolgt, spiegeln die Suchergebnisse die neuesten Zugriffssteuerungseinstellungen für jedes Dokument wider, das die Suchbedingungen erfüllt.

Ein weiterer Vorteil dieser Methode liegt darin, dass sie sich nicht auf die Indexgröße auswirkt (für die Indexierung der Sicherheitstoken ist kein zusätzlicher Speicherplatz erforderlich). Da jedoch für die Prüfung eine Verbindung zu den nativen Repositories erforderlich ist, kann diese Methode Auswirkungen auf die Abfrageleistung haben.

Maximale Sicherheit und minimale Auswirkungen auf die Abfrageleistung erreichen Sie, wenn Sie die Option für das Speichern der Sicherheitstoken im Index mit der Option für die Prüfung der aktuellen Zugriffssteuerungen kombinieren. Wenn ein Benutzer eine Abfrage übergibt, wird die Prüfung in zwei Schritten ausgeführt:

- Zuerst bestimmen die Suchserver mit Hilfe der indexierten Sicherheitsdaten schnell, ob der Benutzer eine Zugriffsberechtigung für den Server und die Datenbank hat, in der das Dokument durchsucht wurde (der Index ist für eine höchstmögliche Verarbeitungsgeschwindigkeit optimiert und gibt Antworten in Sekundenbruchteilen aus).
- Als Nächstes erstellen die Suchserver eine vorläufige Liste mit nur den Dokumenten, die sich in Domänen auf den Servern befinden, für die der Benutzer

eine Zugriffsberechtigung hat. Mit Hilfe dieser Liste stellen die Suchserver dann eine Verbindung zu den nativen Repositories her und legen fest, ob der Benutzer das angeforderte Dokument anzeigen darf.

Wenn ein Benutzer auf einen Server und eine Domäne zugreifen darf, darf er auch mit hoher Wahrscheinlichkeit auf die Dokumente zugreifen. Durch diesen letzten Filterschritt wird jedoch sichergestellt, dass nur Dokumente in den Suchergebnissen zurückgegeben werden, die mit den aktuellen Berechtigungseinstellungen des Benutzers übereinstimmen.

Zugehörige Konzepte

„Umsetzung der Sicherheit auf Dokumentenebene für Dokumente des Windows-Dateisystems“ auf Seite 219

Sie müssen Domänenkontoinformationen auf dem Crawler-Server und auf dem Microsoft Windows-Server konfigurieren, damit aktuelle Berechtigungsnachweise geprüft werden können, wenn ein Benutzer Dokumente durchsucht, die von einem Windows-Dateisystem-Crawler durchsucht wurden.

„Umsetzung der Sicherheit auf Dokumentenebene für Lotus Domino-Dokumente“ auf Seite 221

Wenn der Lotus Notes-Server, der durchsucht werden soll, das Protokoll NRPC (Notes Remote Procedure Call) verwendet, müssen Sie den Crawler-Server entsprechend konfigurieren, damit die Zugriffssteuerungen auf Dokumentenebene umgesetzt werden können.

Zugehörige Tasks

„Konfigurieren der Lotus Domino-Trusted Server, um Benutzerberechtigungen zu prüfen“ auf Seite 222

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 Protokoll NRPC (Notes Remote Procedure Call) verwendet.

Benutzerprofile und Identitätsmanagement

Benutzer können Berechtigungsnachweise speichern, mit denen Sie sichere Domänen durchsuchen können, indem Sie Benutzerprofile für die Unternehmenssuche erstellen.

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* von WebSphere Information Integrator OmniFind Edition können Benutzer Berechtigungsnachweise für eine beliebige Anzahl Domänen in einem Benutzerprofil speichern. Die Berechtigungsnachweise werden verschlüsselt und sicher im System für die Unternehmenssuche gespeichert.

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).

Im Benutzerprofil werden die verschiedenen Berechtigungsnachweise gespeichert, die der Benutzer zur Anmeldung an den Domänen angeben muss, die durchsucht werden sollen. Benutzer können ein Benutzerprofil erstellen, wenn alle folgenden Bedingungen wahr sind:

- Die globale Sicherheit ist in WebSphere Application Server aktiviert.

- Das Identitätsmanagement von WebSphere II OmniFind Edition ist in der Administrationskonsole für die Unternehmenssuche aktiviert.
- Die Sicherheit ist in mindestens einer der Objektgruppen aktiviert, die die Suchanwendung durchsuchen kann.
- Zumindest eine sichere Objektgruppe schließt Dokumente ein, die von einem Crawler durchsucht wurden, der die Zugriffssteuerung umsetzt, indem er das Prüfen der aktuellen Berechtigungsnachweise des Benutzers beim Übergeben einer Abfrage erfordert.
- Die Option für das Prüfen der aktuellen Berechtigungsnachweise während der Abfrageverarbeitung wurde ausgewählt, wenn die Sicherheit auf Dokumentenebene für mindestens einen der folgenden Crawlertypen ausgewählt wurde:
 - Content Edition (für manche Repositorytypen)
 - Domino Document Manager
 - Notes
 - QuickPlace
 - WebSphere Portal
 - Windows-Dateisystem

Das Profil listet alle Domänen auf, die für die Suchanwendung verfügbar sind und für die die Benutzerberechtigungen während der Abfrageverarbeitung geprüft werden müssen. Benutzer können wählen, für welche Domänen sie Berechtigungsnachweise speichern wollen. Wenn der Benutzer für eine Domäne keine Berechtigungsnachweise angibt, werden Dokumente aus den Datenquellen in dieser Domäne aus den Suchergebnissen ausgeschlossen.

Wenn Sie das Identitätsmanagement von WebSphere II OmniFind Edition nicht verwenden, muss die Suchanwendung die Zeichenfolge für den Sicherheitskontext des Benutzers (User Security Context, USC) bereitstellen, wenn die Benutzer Domänen abfragen, für die aktuelle Berechtigungsnachweise geprüft werden müssen.

Konfigurieren des Identitätsmanagements

Sie können das Identitätsmanagement von WebSphere II OmniFind Edition verwenden, um Benutzerprofile zu speichern. Mit den Profilen können Benutzer Domänen durchsuchen, für die die Benutzerberechtigung während der Abfrageverarbeitung geprüft werden muss.

Vorbereitung

Sie müssen ein Administrator für die Unternehmenssuche sein, um Identitätsmanagementoptionen zu konfigurieren.

Informationen zu dieser Task

Wenn Sie angeben, dass Benutzerberechtigungen während der Abfrageverarbeitung geprüft werden sollen, sobald Sie Sicherheitsoptionen auf Dokumentenebene für einen Crawler konfigurieren, müssen die Benutzer ihre Berechtigungsnachweise angeben, wenn sie eine Domäne durchsuchen, die eine Prüfung erforderlich macht. Mit dem Identitätsmanagement von WebSphere II OmniFind Edition können die Benutzer ein Benutzerprofil erstellen und die Berechtigungsnachweise registrieren, die sie zum Anmelden an den sicheren Domänen verwenden. Die Berechtigungsnachweise sind in einer sicheren Datenbank verschlüsselt, die von WebSphere II OmniFind Edition verwaltet wird.

Die Suchserver verwenden die gespeicherten Berechtigungsnachweise zum Authentifizieren des Benutzers, wenn dieser eine sichere Domäne durchsucht. Wenn die Berechtigungsnachweise nicht gültig sind, werden Dokumente aus der sicheren Domäne aus den Suchergebnissen ausgeschlossen.

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 aus, das es WebSphere II OmniFind Edition ermöglicht, Benutzerberechtigungen in Benutzerprofilen zu verwalten.
4. Klicken Sie **OK** an.

Umsetzung der Sicherheit auf Dokumentenebene für Dokumente des Windows-Dateisystems

Sie müssen Domänenkontoinformationen auf dem Crawler-Server und auf dem Microsoft Windows-Server konfigurieren, damit aktuelle Berechtigungsnachweise geprüft werden können, wenn ein Benutzer Dokumente durchsucht, die von einem Windows-Dateisystem-Crawler durchsucht wurden.

Beim Konfigurieren eines Windows-Dateisystem-Crawlers 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.

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. (Verwenden Sie zum Zuordnen von Benutzerberechtigungen **Verwaltung** unter Windows: **Verwaltung** → **Lokale Sicherheitsrichtlinie** → **Lokale Richtlinien** → **Zuweisen von Benutzerrechten**.)

- 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 WebSphere Information Integrator OmniFind 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. (Verwenden Sie zum Zuordnen von Benutzerberechtigungen **Verwaltung** unter Windows: **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 WebSphere Information Integrator OmniFind Edition installiert wird.)
- Die Benutzer müssen über die Benutzerberechtigung **Anmelden als Stapelverarbeitungsauftrag** verfügen.

Zugehörige Konzepte

„Prüfung aktueller Berechtigungsnachweise während der Abfrageverarbeitung“ auf Seite 216

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.

„Crawler für das Windows-Dateisystem“ auf Seite 99

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.

Umsetzung der Sicherheit auf Dokumentenebene für Lotus Domino-Dokumente

Wenn der Lotus Notes-Server, der durchsucht werden soll, das Protokoll NRPC (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 mit Dokumenten umzusetzen, die auf einem Lotus Notes-Server durchsucht wurden, der das Protokoll NRPC 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 Lotus Notes-Dokumenten verfügt, die mit den Suchkriterien übereinstimmen. Dokumente, für die der Benutzer keine Berechtigung zum Anzeigen hat, werden aus den Suchergebnissen entfernt, bevor die Ergebnisse dem Benutzer zurückgegeben werden.

- „Konfigurieren der Lotus Domino-Trusted Server, um Benutzerberechtigungen zu prüfen“ auf Seite 222.
- Aktivieren der globalen Sicherheit in WebSphere Application Server und Konfigurieren der Suchanwendung für die Verwendung der Sicherheit. Durch diesen Schritt wird sichergestellt, dass Benutzer zur Angabe von Berechtigungsnachweisen aufgefordert werden, wenn sie versuchen, die Suchanwendung zu verwenden. Die Suchserver können anschließend diese Berechtigungsnachweise verwenden, um den Zugriff jedes Benutzers auf Lotus Notes-Dokumente zu prüfen.

Zugehörige Konzepte

„Prüfung aktueller Berechtigungsnachweise während der Abfrageverarbeitung“ auf Seite 216

Wenn die Sicherheit für eine Objektgruppe bei ihrer Erstellung aktiviert wurde,

können Sie mit bestimmten Domämentypen die aktuellen Berechtigungsnachweise des Benutzers prüfen, wenn der Benutzer eine Abfrage übergibt.

„Notes-Crawler“ auf Seite 63

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

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 Protokoll NRPC (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. Wählen Sie **Datei** → **Server öffnen** (File > 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 Knopf mit dem 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-Prozeduren für die gegenseitige Zertifizierung der Domino-Server-ID-Datei von WebSphere II OmniFind Edition mit der anderen Domino-Domäne befolgen. Informationen zu diesen Prozeduren 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 216

Wenn die Sicherheit für eine Objektgruppe bei ihrer Erstellung aktiviert wurde, können Sie mit bestimmten Domämentypen die aktuellen Berechtigungsnachweise des Benutzers prüfen, wenn der Benutzer eine Abfrage übergibt.

„Notes-Crawler“ auf Seite 63

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

Inaktivieren der Sicherheit auf Dokumentebene

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 Objektgruppen-administrator der Objektgruppe sein, um Sicherheitsoptionen auf Dokumentebene zu konfigurieren.

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

Beim Konfigurieren eines Crawlers können Sie den zu durchsuchenden Dokumenten Sicherheitstoken zuordnen. 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 ignorieren 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 konfigurieren** an.
2. Wählen Sie auf der Seite **Sicherheit auf Dokumentenebene** das Markierungsfeld **Zugriffssteuerungen auf Dokumentenebene im Index ignorieren** aus, wenn Sie die den Dokumenten von den Crawlern zugeordneten Sicherheitstoken nicht verwenden wollen, wenn die Objektgruppe von Benutzern abgefragt wird.

Die Crawler fügen den Dokumenten zwar weiterhin Sicherheitstoken hinzu, diese werden jedoch von den Suchservern ignoriert, so dass 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 Repositories 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 angegeben haben, gelten jene Sicherheitsmaßnahmen weiterhin, solange Sie nicht auch das Markierungsfeld **Zugriffssteuerungen auf Dokumentenebene im Index ignorieren** ausgewählt haben.

4. Klicken Sie **OK** an.

Integration der Unternehmenssuche in WebSphere Portal

Sie können die Suchfunktionalitäten von IBM WebSphere Portal erweitern, indem Sie die Portlets für die Unternehmenssuche in WebSphere Portal und die WebSphere Portal-Suchzentrale implementieren.

Integrationspunkte

Die Portlets für die Unternehmenssuche werden in mehrerer Hinsicht in WebSphere Portal integriert:

WebSphere Portal

WebSphere Portal bietet dem Benutzer einen zentralen Zugang für die Interaktion mit Anwendungen, Inhalt, Prozessen und Personen. Das WebSphere Portal-Framework aktiviert neue Anwendungen, genannt Portlets, die integriert und implementiert werden sollen, ohne andere Anwendungen im Portal zu beeinträchtigen.

Wenn Sie das Portlet für die Unternehmenssuche in WebSphere Portal implementieren, können Sie über die Schnittstelle von WebSphere Portal Objektgruppen für die Unternehmenssuche durchsuchen. 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.

Portal Search Engine

WebSphere Portal Search Engine durchsucht Websites. Administrationsportlets ermöglichen es Administratoren, indexierte Objektgruppen zu erstellen, und Suchportlets ermöglichen es dem Benutzer, diese Objektgruppen zu durchsuchen.

Wenn Sie WebSphere Portal Version 5.0.2 oder aktueller verwenden, können Sie die Konfigurationsdaten für Portal Search Engine-Objektgruppen und Taxonomien in die Unternehmenssuche migrieren.

Wenn Sie WebSphere Portal Version 5.1 oder höher verwenden, können Sie das Portlet für die Dokumentsuche von Portal Search Engine verwenden, um Objektgruppen für die Unternehmenssuche zu durchsuchen. Die Konfigurationsmerkmale ermöglichen es dem Benutzer, ohne großen Aufwand und nach Bedarf zwischen den beiden Suchfunktionalitäten hin- und herzuwechseln.

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 es Ihnen, WebSphere Portal-Inhalt zu durchsuchen, sowie beliebige andere Objektgruppen, die ein Administrator in der Suchzentrale registriert.

Die Suchzentrale verfügt über eine paginierte Schnittstelle. Sie können alle verfügbaren Objektgruppen über eine einheitliche Seite durchsuchen, oder Sie können eine Seite auswählen, um eine individuelle Objektgruppe zu durchsuchen. Es gibt z. B. eine Seite für Portal Search Engine-Indizes und eine Seite für Bibliotheken für die Portal-Dokumentverwaltung.

WebSphere Information Integrator OmniFind Edition (WebSphere II OmniFind Edition) enthält einen Adapter und ein Registrierungsportlet, damit Objektgruppen für die Unternehmenssuche über die Suchzentrale durchsucht werden können. Der Adapter fügt der Schnittstelle der Suchzentrale eine Indexzunge für die Unternehmenssuche hinzu, und das Registrierungsportlet registriert das Portlet für die Unternehmenssuche in der Suchzentrale.

Bevor Sie die Suchzentrale zum Durchsuchen von Objektgruppen für die Unternehmenssuche verwenden können, müssen Sie den Adapter und das Registrierungsportlet installieren.

WebSphere Portal-Crawler

Bei der Installation von WebSphere II OmniFind Edition wird die Unternehmensanwendung ESPACServer.ear auf den Suchservern installiert. Nach der Implementierung dieser Unternehmensanwendung in WebSphere Portal können Sie die Administrationskonsole für die Unternehmenssuche zum Konfigurieren eines WebSphere Portal-Crawlers verwenden und Ihren Objektgruppen für die Unternehmenssuche WebSphere Portal-Sites hinzufügen.

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 sonstigen Datenquellentypen durchsuchen, die von WebSphere II OmniFind Edition 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 von Unternehmenskapazitäten verteilt die Unternehmenssuche die Auslastung auf vier Server (einer für das Durchsuchen von Daten, einer für das Parsing und die Indexierung von Daten, weitere zwei für die Unterstützung von Such- und Abrufprozessen).

Überblick über die Portletimplementierung

Die Portlets, die Sie für die Unternehmenssuche verwenden können, hängen von der WebSphere Portal-Version ab, die Sie verwenden:

WebSphere Portal Version 5.0.2

Sie können das Portlet für die Unternehmenssuche implementieren und es für das Durchsuchen von Objektgruppen für die Unternehmenssuche verwenden. Dieses Portlet kann mit den Portlets von Portal Search Engine koexistieren.

WebSphere Portal Version 5.1 oder höher

- Sie können das Portlet für die Unternehmenssuche implementieren und es für das Durchsuchen von Objektgruppen für die Unternehmenssuche verwenden. Dieses Portlet kann mit den Portlets von Portal Search Engine koexistieren.
- Nach der Implementierung des Portlets für die Unternehmenssuche können Sie das Portlet für die Dokumentsuche in Portal Search Engine so konfigurieren, dass Objektgruppen für die Unternehmenssuche durchsucht werden.

- Nach der Registrierung des Portlets für die Unternehmenssuche in der WebSphere Portal-Suchzentrale können Sie die Suchzentrale einsetzen, um Objektgruppen für die Unternehmenssuche zu durchsuchen. Sie können eine 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 alle anderen Objektgruppen durchsucht, die in der Suchzentrale verfügbar sind.

Zugehörige Konzepte

„Migration von WebSphere Portal auf die Unternehmenssuche“ auf Seite 235
Die Unternehmenssuche stellt einen Migrationsassistenten zur Verfügung, mit dem Sie Taxonomien und Objektgruppen aus IBM WebSphere Portal in die Unternehmenssuche migrieren können.

Zugehörige Tasks

„Implementieren der Unternehmensanwendung für den WebSphere Portal-Crawler“ auf Seite 97

Bevor Sie einen WebSphere Portal-Crawler erstellen, müssen Sie die Unternehmensanwendung ESPACServer.ear in WebSphere Portal implementieren.

Sicherheit auf Dokumentebene mit Portal Search Engine

Mit IBM WebSphere Portal Search Engine können Sie die Sicherheit auf Dokumentebene verbessern, wenn der Benutzer Objektgruppen für die Unternehmenssuche durchsucht.

Wenn ein Crawler für die Unternehmenssuche den Sicherheitstoken "Gruppen-ID" den Dokumenten zuordnet, die er durchsucht, und wenn Sie das Portlet für die Dokumentsuche für Portal Search Engine so konfigurieren, dass Objektgruppen für die Unternehmenssuche durchsucht werden, kann Portal Search Engine die Gruppen-ID des angemeldeten Benutzers daraus ableiten und den Sicherheitstoken für diese Gruppen-ID mit der Suchanfrage an die Unternehmenssuche übergeben. Der Sicherheitstoken stellt sicher, dass nur die Dokumente, für die der Benutzer die Berechtigung hat, als Suchergebnisse zurückgegeben werden.

Wenn der Crawler den Dokumenten einen anderen Sicherheitstokentyp zugeordnet hat, wie z. B. eine Benutzer-ID oder eine Benutzerrolle, und Sie die Sicherheit auf Dokumentebene umsetzen möchten, wenn Objektgruppen für die Unternehmenssuche durchsucht werden, müssen Sie ein benutzerdefiniertes Suchportlet erstellen. Portal Search Engine leitet die Sicherheitstoken nur für Gruppen-IDs ab.

Zugehörige Konzepte

„Suchanwendungen für die Unternehmenssuche“ auf Seite 183

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.

„Sicherheit auf Dokumentebene“ auf Seite 214

Wenn die Sicherheit bei der Erstellung einer Objektgruppe aktiviert ist, können Sie Sicherheitsmaßnahmen auf Dokumentebene konfigurieren. Die Sicherheit auf Dokumentebene sorgt dafür, dass Benutzer, die Objektgruppen durchsuchen, nur auf die Dokumente zugreifen können, für deren Anzeige sie die Berechtigung haben.

Implementieren des Suchportlets

Das Suchportlet `ESSearchPortlet.war` ermöglicht die Verwendung von WebSphere Portal Search Engine, mit der Sie Objektgruppen für die Unternehmenssuche durchsuchen können.

Vorbereitung

Das Suchportlet für die Unternehmenssuche muss in WebSphere Portal implementiert werden, bevor Sie Portal Search Engine für die Verwendung des Portlets konfigurieren können.

Informationen zu dieser Task

Nach der Implementierung des Suchportlets können Sie Portal Search Engine weiter verwenden, um indizierte Daten in WebSphere Portal zu durchsuchen. Ein WebSphere Portal-Administrator bearbeitet die Konfigurationsmerkmale von Portal Search Engine so, dass der Benutzer Objektgruppen für die Unternehmenssuche durchsuchen kann. Sie können zwischen diesen beiden Suchfunktionalitäten wechseln und die für Ihre Bedürfnisse am besten geeignete Suchlösung verwenden.

Die Datei `ESSearchPortlet.war` wird bei der Installation von WebSphere II OmniFind Edition im Verzeichnis `ES_INSTALL_ROOT/bin` auf den Suchservern installiert. Die Standardinstallationspfade lauten wie folgt:

- UNIX-Systeme: `/opt/IBM/es/bin/ESSearchPortlet.war`
- Windows-Systeme: `C:\Programme\IBM\es\bin\ESSearchPortlet.war`

Vorgehensweise

Gehen Sie wie folgt vor, um das Suchportlet in WebSphere Portal zu implementieren:

1. Installieren Sie das Portlet:
 - a. Melden Sie sich mit der Administrator-ID des WebSphere Portal-Servers am WebSphere Portal-Server an.
 - b. Wählen Sie in der Funktionsleiste **Verwaltung** aus.
 - c. Wählen Sie im Navigationsbereich auf der linken Seite **Portlets** aus, und wählen Sie dann **Installieren** im Menü **Portlets** aus.
 - d. Klicken Sie **Durchsuchen** an, wählen Sie die Datei `ESSearchPortlet.war` auf Ihrem System aus, und klicken Sie **Weiter** an.
 - e. Klicken Sie auf der nächsten Seite **Installieren** an. Nach der Installation des Portlets wird die folgende Nachricht angezeigt: `APIN0005I: Portlets wurden erfolgreich installiert.`
2. Modifizieren Sie die Portletparameter:
 - a. Wählen Sie im Navigationsbereich auf der linken Seite **Verwalten von Portlets** aus, und wählen Sie dann **IBM Enterprise Search** aus.
 - b. Wenn rechts neben dem Auswahlfenster neue Symbole angezeigt werden, wählen Sie **Parameter ändern** aus.
 - c. Modifizieren Sie in der Liste mit den Portletparametern die folgenden Parameter:

port Setzen Sie diesen Wert auf die Portnummer des Web-Servers für den WebSphere II OmniFind Edition-Suchserver. Der Standardwert ist 80.

applicationName

Setzen Sie diesen Wert auf den Namen einer gültigen Suchanwendung für die Unternehmenssuche. (Die Namen der verfügbaren Suchanwendungen sind auf der Seite **Suchanwendungen** in der Administrationskonsole für die Unternehmenssuche aufgelistet.) Der Standardwert ist Default. Dies ist die Suchanwendung, die auf den Suchservern immer konfiguriert ist.

hostname

Setzen Sie diesen Wert auf den vollständig qualifizierten Hostnamen eines WebSphere II OmniFind Edition-Suchservers.

- d. Klicken Sie **Speichern** an, um Ihre Änderungen zu speichern. Klicken Sie dann **Abbrechen** an, um die Seite **Parameter ändern** zu verlassen.
3. Erstellen Sie eine Portalseite für das Portlet:
 - a. Wählen Sie im Navigationsbereich auf der linken Seite **Benutzeroberfläche des Portals** aus, und wählen Sie dann **Verwalten von Seiten** aus.
 - b. Wählen Sie den Link **Mein Portal** aus.
 - c. Klicken Sie **Neue Seite** an, geben Sie einen Seitentitel an, und klicken Sie **OK** an. Nach Erstellung der Seite wird die folgende Nachricht angezeigt:
APPR0010I: *seitentitel* wurde erfolgreich erstellt.
 - d. Klicken Sie **OK** an. Die erstellte Seite wird in der Liste mit verfügbaren Seiten angezeigt.
4. Fügen Sie das Suchportlet der Seite hinzu:
 - a. Wählen Sie das Bearbeitungssymbol (Stift) für die erstellte Seite aus.
 - b. Klicken Sie **Portlets hinzufügen** an.
 - c. Geben Sie Enterprise in das Textfeld **Suchen nach** ein, und klicken Sie **Suchen** an.
 - d. Wählen Sie das Markierungsfeld neben **IBM Enterprise Search** aus, und klicken Sie **OK** an.
 - e. Klicken Sie **Fertig** an. Dieser Schritt gibt an, dass Sie das Suchportlet erfolgreich installiert und einer WebSphere Portal-Seite hinzugefügt haben.
5. Rufen Sie das Portlet auf:
 - a. Wählen Sie in der Navigationsleiste oben auf der Seite **Mein Portal** aus.
 - b. Wählen Sie die Seite aus, die mit dem Seitentitel übereinstimmt, den Sie beim Hinzufügen des Portlets zum System angegeben haben.
Wenn Sie das erste Mal auf die Portletseite zugreifen, wird sie möglicherweise nur langsam aufgebaut, weil das System die Java Server Pages (JSP-Dateien) für das Portlet kompilieren muss.

Konfigurieren des Suchportlets von WebSphere Portal für die Unternehmenssuche

WebSphere Portal stellt ein Portlet zum Durchsuchen nativer WebSphere Portal-Objektgruppen zur Verfügung. Dieses Portlet kann zum Durchsuchen von Objektgruppen für die Unternehmenssuche konfiguriert werden.

Vorgehensweise

Gehen Sie wie folgt vor, um das Suchportlet zum Durchsuchen von Objektgruppen für die Unternehmenssuche zu konfigurieren:

1. Stoppen Sie die Serverinstanz `WebSphere_Portal`.

2. Kopieren Sie die folgenden Dateien aus dem WebSphere II OmniFind Edition-Verzeichnis ES_INSTALL_ROOT/lib in das Verzeichnis /WebSphere/PortalServer/shared/app auf dem WebSphere Portal-Server:
 - esapi.jar
 - siapi.jar
3. Starten Sie die Serverinstanz WebSphere_Portal, und melden Sie sich mit der Administrator-ID für WebSphere Portal am WebSphere Portal-Server an.
4. Kopieren Sie das Portlet:
 - a. Klicken Sie **Verwaltung** in der rechten oberen Ecke an, erweitern Sie das Objekt **Portletverwaltung**, und wählen Sie **Portlets** aus.
 - b. Suchen Sie nach dem Wort search.
 - c. Klicken Sie für das Suchportlet **Portlet kopieren** an, geben Sie einen eindeutigen Namen für das Portlet an, und klicken Sie **OK** an.
5. Konfigurieren Sie das Portlet:
 - a. Klicken Sie das Symbol **Portlet konfigurieren** neben dem Portletnamen an, den Sie angegeben haben.
 - b. Klicken Sie den Rechtspfeil an, um auf die zweite Seite mit Parametern zu wechseln.
 - c. Geben Sie ApplicationInfoId unter **Neuer Parameter** ein, geben Sie im Feld **Neuer Wert** den Namen einer gültigen WebSphere II OmniFind Edition-Suchanwendung an, und klicken Sie dann **Hinzufügen** an, um den neuen Parameter hinzuzufügen.
 Die Namen der verfügbaren Suchanwendungen werden auf der Seite **Suchanwendungen** in der Administrationskonsole für die Unternehmenssuche aufgelistet. Der Standardwert ist Default. Dies ist die Suchanwendung, die auf den Suchservern immer konfiguriert ist.
 - d. Löschen Sie die folgenden Parameter:
 - ApplicationInfoId_EXAMPLE
 - QueryFactoryImp
 - BrowseFactoryImp
 - SearchFactoryImp
 - IndexName
 - EJB_Example
 - IIOP_URL_Example
 - SOAP_URL_Example
 - EJB
 - SOAP_URL
 - IIOP_URL
 - e. Fügen Sie die folgenden Parameter hinzu, und geben Sie die unten aufgeführten Werte an. Klicken Sie **Hinzufügen** an, um jeden Parameternamen und -wert hinzuzufügen. Bei Parameternamen muss die Groß-/Kleinschreibung beachtet werden.
 - IMPLEMENT = ES
 - ApplicationInfoId = *name_der_suchanwendung* (z. B. 'Default')
 - SearchFactoryImp = com.ibm.es.api.search.RemoteSearchFactory
 - BrowseFactoryImp = com.ibm.es.api.browse.RemoteBrowseFactory
 - hostname = *hostname_des_suchservers* (z. B. omnifind.ibm.com)
 - port = *portnummer_des_suchservers* (z. B. 80)
6. Ordnen Sie die neue Kopie des Suchportlets einer Portalseite zu. Weitere Informationen finden Sie in der Verwaltungsdokumentation für WebSphere Portal.
7. Rufen Sie das Portlet auf:
 - a. Wählen Sie in der Navigationsleiste oben auf der Portalseite **Mein Portal** aus.

- b. Wählen Sie die Seite aus, die mit dem Namen übereinstimmt, den Sie beim Kopieren des Portlets in Schritt 4 auf Seite 230 angegeben haben.

Installieren des Adapters für die Unternehmenssuche für die Suchzentrale

Bevor Sie die WebSphere Portal-Suchzentrale zum Durchsuchen von Objektgruppen für die Unternehmenssuche verwenden können, müssen Sie einen Adapter für die Unternehmenssuche installieren.

Informationen zu dieser Task

Der Adapter `ESSearchAdapter.ear` fügt der Schnittstelle der Suchzentrale eine Seite für die Unternehmenssuche hinzu. Nachdem Sie diese Anwendung und das Registrierungsportlet installiert haben, können Sie der Suchzentrale Objektgruppen für die Unternehmenssuche hinzufügen und diese Objektgruppen mit dem universalen Suchportlet durchsuchen.

Die Datei `ESSearchAdapter.ear` wird bei der Installation von WebSphere II OmniFind Edition im Verzeichnis `ES_INSTALL_ROOT/bin` auf den Suchservern installiert. Die Standardinstallationspfade lauten wie folgt:

- UNIX-Systeme: `/opt/IBM/es/bin/ESSearchAdapter.ear`
- Windows-Systeme: `C:\Programme\IBM\es\bin\ESSearchAdapter.ear`

Vorgehensweise

Gehen Sie wie folgt vor, um den Adapter für die Unternehmenssuche in der Suchzentrale zu installieren:

1. Stoppen Sie die Serverinstanz `WebSphere_Portal`.
2. Starten Sie die WebSphere Application Server-Serverinstanz `server1`.
3. Kopieren Sie die folgenden Dateien aus dem WebSphere II OmniFind Edition-Verzeichnis `ES_INSTALL_ROOT/lib` in das Verzeichnis `/WebSphere/PortalServer/shared/app` auf dem WebSphere Portal-Server:
`esapi.jar`
`siapi.jar`

Wenn Sie aufgefordert werden, die Datei `siapi.jar` zu überschreiben, geben Sie **Ja** an.

4. Starten Sie auf dem WebSphere Portal-Server die Administrationskonsole von WebSphere Application Server. Melden Sie sich an, wenn Sie dazu aufgefordert werden.

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.

5. Klicken Sie **Anwendungen** an, und klicken Sie dann **Neue Anwendung installieren** an.
6. Klicken Sie **Durchsuchen** an, wählen Sie die Datei `ESSearchAdapter.ear` auf Ihrem System aus, und klicken Sie zweimal **Weiter** an. Wenn Sie eine Warnung zu Richtliniendateien erhalten, klicken Sie **Weiter** an.
7. Klicken Sie so lange **Weiter** an, bis die Seite **Module zu Anwendungsservern zuordnen** angezeigt wird. Treffen Sie die folgende Auswahl, und klicken Sie **Anwenden** an:
 - Wählen Sie **WebSphere:cell=zellenname, node=knotenname, server=server1** im Feld **Cluster und Server** aus.
 - Wählen Sie das Markierungsfeld neben dem Modul `ESSearchAdapterEJB` aus.
8. Klicken Sie so lange **Weiter** an, bis die Zusammenfassungsseite angezeigt wird. Klicken Sie dann **Fertig stellen** an.
9. Klicken Sie den Link **In Master-Konfiguration speichern** an, und klicken Sie dann den Knopf **Speichern** an, um Ihre Änderungen in der WebSphere Application Server-Konfiguration zu speichern.
10. Starten Sie den WebSphere Portal-Server erneut.

Installieren des Registrierungsportlets für die Unternehmenssuche für die Suchzentrale

Bevor Sie die WebSphere Portal-Suchzentrale zum Durchsuchen von Objektgruppen für die Unternehmenssuche verwenden können, müssen Sie ein Registrierungsportlets für die Unternehmenssuche installieren.

Informationen zu dieser Task

Das Registrierungsportlet `ESSearchAdapterPortlet.war` registriert das Portlet für die Unternehmenssuche bei der Suchzentrale. Nachdem Sie dieses Portlet (und die Adapteranwendung für die Unternehmenssuche) installiert haben, können Sie der Suchzentrale Objektgruppen für die Unternehmenssuche hinzufügen und diese Objektgruppen mit dem universalen Suchportlet durchsuchen.

Die Datei `ESSearchAdapterPortlet.war` wird bei der Installation von WebSphere II OmniFind Edition im Verzeichnis `ES_INSTALL_ROOT/bin` auf den Suchservern installiert. Die Standardinstallationspfade lauten wie folgt:

- UNIX-Systeme: `/opt/IBM/es/bin/ESSearchAdapterPortlet.war`
- Windows-Systeme: `C:\Programme\IBM\es\bin\ESSearchAdapterPortlet.war`

Vorgehensweise

Gehen Sie wie folgt vor, um das Registrierungsportlet für die Unternehmenssuche zu installieren:

1. Melden Sie sich mit der Administrator-ID für WebSphere Portal am WebSphere Portal-Server an.
2. Wählen Sie in der Funktionsleiste **Verwaltung** aus.
3. Wählen Sie im Navigationsbereich auf der linken Seite **Portlets** aus, und wählen Sie dann **Installieren** im Menü **Portlets** aus.
4. Klicken Sie **Durchsuchen** an, wählen Sie die Datei `ESSearchAdapterPortlet.war` auf Ihrem System aus, und klicken Sie **Weiter** an.

5. Klicken Sie auf der nächsten Seite **Installieren** an. Nach der Installation des Portlets wird die folgende Nachricht angezeigt: APIN0005I: Portlets wurden erfolgreich installiert.
6. Ordnen Sie das Portlet (mit der Bezeichnung IBM WebSphere II OmniFind Edition-Registrierungsportlet für die Unternehmenssuche) einer Portalseite zu. Weitere Informationen finden Sie in der Verwaltungsdokumentation für WebSphere Portal.

Migration von WebSphere Portal auf die Unternehmenssuche

Die Unternehmenssuche stellt einen Migrationsassistenten zur Verfügung, mit dem Sie Taxonomien und Objektgruppen aus IBM WebSphere Portal in die Unternehmenssuche migrieren können.

Wenn Sie Taxonomien und Objektgruppen migrieren möchten, führen Sie den Migrationsassistenten auf dem Indexserver für die Unternehmenssuche aus. Nach der Migration einer Taxonomie können Sie sie für Objektgruppen für die Unternehmenssuche verwenden. Sie können die Unternehmenssuche auch verwenden, um Objektgruppen zu verwalten und zu durchsuchen, die Sie aus WebSphere Portal migrieren.

In der Unternehmenssuche wird eine Taxonomie als "Kategoriebaum" bezeichnet. Nach der Migration einer regelbasierten Taxonomie können Sie die Administrationskonsole für die Unternehmenssuche verwenden, um den Kategoriebaum zu verwalten. Für die Verwendung modellbasierter Taxonomien für die Unternehmenssuche muss WebSphere Portal auf dem Indexserver installiert sein.

Wenn Sie Taxonomien und Objektgruppen migrieren möchten, sollten Sie immer zuerst die modellbasierten Taxonomiedateien migrieren, bevor Sie Objektgruppen migrieren. Wenn Sie dies versäumen, wird die modellbasierte Kategorisierung nicht für die Objektgruppen funktionieren, die Sie aus WebSphere Portal migrieren.

Zugehörige Konzepte

„Integration der Unternehmenssuche in WebSphere Portal“ auf Seite 225
Sie können die Suchfunktionalitäten von IBM WebSphere Portal erweitern, indem Sie die Portlets für die Unternehmenssuche in WebSphere Portal und die WebSphere Portal-Suchzentrale implementieren.

Zugehörige Tasks

„Konfigurieren von Kategorien“ auf Seite 118
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.

Modellbasierte Taxonomien aus WebSphere Portal migrieren

Sie können auswählen, welche modellbasierte Taxonomie Sie für eine Objektgruppe für die Unternehmenssuche verwenden möchten, indem Sie das Portlet zur Taxonomieverwaltung von WebSphere verwenden. Objektgruppen, die Sie bereits in die Unternehmenssuche migriert haben, sind von einer neuen Taxonomie-migration nicht betroffen.

Vorbereitung

Bevor Sie den Migrationsassistenten zum ersten Mal ausführen, stoppen Sie das System für die Unternehmenssuche, damit Änderungen an den Konfigurationsdateien vorgenommen werden können. Beim erstmaligen Ausführen des Migrationsassistenten müssen Sie den Pfad angeben, in dem WebSphere Application Server und WebSphere Portal installiert sind. Wenn Sie den Migrationsassistenten nach dem ersten Mal ausführen, kann das System für die Unternehmenssuche aktiv sein.

Informationen zu dieser Task

Wenn Sie eine modellbasierte Taxonomie migrieren möchten, müssen Sie die betreffende Taxonomie in WebSphere Portal auswählen und exportieren. Verwenden Sie anschließend den Migrationsassistenten für die Unternehmenssuche, um die Taxonomie in die Unternehmenssuche zu migrieren.

Vorgehensweise

Gehen Sie wie folgt vor, um eine modellbasierte Taxonomie aus WebSphere Portal in die Unternehmenssuche zu migrieren:

1. Exportieren Sie Ihre aktuelle modellbasierte Taxonomie aus dem Portlet für die Taxonomieverwaltung von WebSphere Portal. Zu der Taxonomie gehören die folgenden XML-Dateien:

```
synonyms.xml  
titles.xml  
treenodes.xml
```

2. Kopieren Sie diese Dateien auf den Indexserver für die Unternehmenssuche.
3. Melden Sie sich am Indexserver für die Unternehmenssuche als Administrator für die Unternehmenssuche an. Diese Benutzer-ID wurde bei der Installation von WebSphere II OmniFind Edition angegeben.
4. Wechseln Sie in das Installationsverzeichnis der Unternehmenssuche:

```
UNIX: cd $ES_INSTALL_ROOT/bin  
Windows: cd %ES_INSTALL_ROOT%\bin
```

5. Wenn Sie den Migrationsassistenten das erste Mal starten, geben Sie den folgenden Befehl ein, um das System für die Unternehmenssuche zu stoppen.
esadmin stop
6. Geben Sie den folgenden Befehl ein, um den Migrationsassistenten zu starten, und klicken Sie **Weiter** an.

```
UNIX: ./eswpsmigrate.sh  
Windows: eswpsmigrate.bat
```

Wenn Sie den Migrationsassistenten nicht zum ersten Mal ausführen und das System für die Unternehmenssuche aktiv ist, stellen Sie sicher, dass Sie keine Objektgruppe für die Unternehmenssuche erstellen, während der Migrationsassistent ausgeführt wird:

7. Wählen Sie **Modellbasierte Taxonomiedateien aus WebSphere Portal importieren** aus, und klicken Sie **Weiter** an.
8. Wenn Sie den Migrationsassistenten das erste Mal starten, geben Sie die Pfade an, in denen WebSphere Application Server und WebSphere Portal installiert sind. Klicken Sie **Weiter** an.
9. Suchen Sie nach dem Verzeichnis, das die modellbasierten Taxonomiedateien enthält, wählen Sie die XML-Dateien aus, die Sie migrieren müssen, um die modellbasierte Taxonomie verwenden zu können (synonyms.xml, titles.xml und treenodes.xml), und klicken Sie **Weiter** an.
10. Wenn Sie den Migrationsassistenten zum ersten Mal gestartet haben, geben Sie den folgenden Befehl ein, nachdem die Migration beendet ist, um das System für die Unternehmenssuche erneut zu starten:
esadmin start

Wenn Fehler auftreten, sehen Sie sich die Datei MigrationWizard.log an, die sich in dem Verzeichnis befindet, in dem der Migrationsassistent installiert ist.

Zugehörige Konzepte

„Modellbasierte Kategorien“ auf Seite 116

Wenn Sie in Ihrem IBM WebSphere Portal-System modellbasierte Kategorien verwenden, können Sie diese Kategorien auch in Objektgruppen für die Unternehmenssuche verwenden.

Zugehörige Tasks

„Konfigurieren von Kategorien“ auf Seite 118

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.

Migration einer Objektgruppe aus WebSphere Portal

Wenn Sie Objektgruppen 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 modellbasierte Taxonomien und Objektgruppen migrieren wollen, müssen Sie zuerst die modellbasierten Taxonomie-dateien migrieren, bevor Sie diese Prozedur zum Migrieren von Objektgruppen verwenden. Wenn Sie dies versäumen, wird die modellbasierte Kategorisierung nicht für die Objektgruppen funktionieren, die Sie aus WebSphere Portal migrieren.

Vorgehensweise

Gehen Sie wie folgt vor, um eine Objektgruppe 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 WebSphere II OmniFind Edition angegeben.
5. Wechseln Sie in das Installationsverzeichnis der Unternehmenssuche:

UNIX: `cd $ES_INSTALL_ROOT/bin`

Windows: `cd %ES_INSTALL_ROOT%\bin`

6. Geben Sie den folgenden Befehl ein, um den Migrationsassistenten zu starten, und klicken Sie **Weiter** an.

UNIX: `./eswpsmigrate.sh`

Windows: `eswpsmigrate.bat`

7. Wählen Sie **Sucheinstellungen von Portal Search Engine nach WebSphere Portal migrieren** aus, und klicken Sie **Weiter** an.
8. 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.
9. 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 "keiner" sein, auf Dokumentdaten basieren, oder auf Links zu Webdokumenten von anderen Webdokumenten basieren.
 - Der Typ der Kategorien, die Sie für diese Objektgruppe verwenden möchten. Sie können entweder "keiner", regelbasierte Kategorien oder modellbasierte Kategorien auswählen.
 Wenn Sie regelbasierte Kategorien auswählen, werden die Taxonomie und die Regeln der WebSphere Portal-Objektgruppe in die Unternehmenssuche migriert.
 Sie können modellbasierte Kategorien nur dann auswählen, wenn Sie zuvor die modellbasierte Taxonomie aus WebSphere Portal in die Unternehmenssuche migriert haben.

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).

10. Starten Sie für die migrierten Objektgruppen die Prozesse für die Crawler-suche, das Parsing und die Indexierung über die Administrationskonsole für die Unternehmenssuche.
11. Nachdem Sie festgelegt haben, dass die migrierte Objektgruppe in der Unternehmenssuche durchsucht werden kann, können Sie die ursprüngliche Objektgruppe in Portal Search Engine löschen.
12. 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.

Zugehörige Konzepte

„Objektgruppen für die Unternehmenssuche“ auf Seite 29

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.

„Regelbasierte Kategorien“ auf Seite 113

Sie können Regeln konfigurieren, um zu steuern, welche Dokumente Kategorien in einer Objektgruppe für die Unternehmenssuche zugeordnet sind.

„Modellbasierte Kategorien“ auf Seite 116

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.

„Statische Rangfolge“ auf Seite 172

Bei bestimmten Dokumenttypen können Sie einen statischen Rangfolgefaktor zuordnen, der den Rang dieser Dokumente in den Suchergebnissen erhöht.

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-URL
- Die Anzahl paralleler Crawlersuchprozesse
- Die Crawlersuchtiefe

- Das Zeitlimit (in Sekunden) für das Abrufen eines Dokuments
- Den Standardzeichensatz
- Einschließungs- und Ausschlussregeln für die Crawlersuche

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, wie z. B. regelbasiert, oder "keiner"
- Ob der Suchcache verwendet werden soll, und wie viele Abfragen mit Suchergebnissen 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 aktualisierte Anzeige des Index
- Zeitplan für die Reorganisation des Index
- 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.

Starten und Stoppen der Server für die Unternehmenssuche

Nach dem Erstellen einer Objektgruppe müssen Sie die Server für Crawlersuche, Syntaxanalyse, Indexieren und Suche von Daten starten. 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 das Reorganisieren oder Aktualisieren des Index angeben. Nachdem Sie die Server für die Datensyntaxanalyse und das Durchsuchen des Index gestartet haben, brauchen Sie sie in der Regel nur dann zu stoppen und erneut zu starten, wenn Sie die Konfigurationseinstellungen ändern (wie 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 gewöhnlich stoppen und erneut starten, damit die Änderungen wirksam werden. Wenn Sie die Crawlersuchregeln nicht ändern, werden die Crawler (im Fall von Web- und NNTP-Crawlern) kontinuierlich oder nach den angegebenen Zeitplänen ausgeführt.

Starten der Server für die Unternehmenssuche

Damit Benutzer eine Objektgruppe durchsuchen können, müssen Sie die Systemprozesse und anschließend die Server für die Crawlersuche, die Syntaxanalyse, die Indexierung und die Suche für die Objektgruppe starten.

Vorbereitung

Konfigurieren Sie die Datenquellen, die Sie durchsuchen wollen, und geben Sie Optionen an, wie die Daten syntaktisch analysiert, indexiert 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 Dokumente syntaktisch analysieren und indexieren können.

Vorgehensweise

Gehen Sie wie folgt vor, um die Server für die Unternehmenssuche zu starten:

1. Gehen Sie wie folgt vor, um die Administrationskonsole für die Unternehmenssuche oder Suchanwendung zu verwenden:
 - a. Starten Sie den IBM HTTP-Server, wenn dieser noch nicht aktiv ist.

- b. Verwenden Sie die WebSphere Application Server-Administrationskonsole, um die Unternehmensanwendungen ESAdmin und ESSearchApplication zu starten, falls diese noch nicht aktiv sind.
2. Starten Sie die allgemeine Übertragungsschicht (CCL = Common Communication Layer) für die Unternehmenssuche, falls diese noch nicht aktiv ist:
 - a. Melden Sie sich auf dem Indexserver mit der Benutzer-ID des Administrators für die Unternehmenssuche an, die bei der Installation von WebSphere II OmniFind Edition angegeben wurde.

UNIX-Systeme



Geben Sie den folgenden Befehl ein:


```
startccl.sh -bg
```

Windows-Systeme

Wählen Sie **Start** → **Programme** → **Verwaltung** → **Dienste** aus, und starten Sie anschließend den Service "IBM WebSphere Information Integrator OmniFind Edition" erneut.

3. Starten Sie das System für die Unternehmenssuche, wenn es noch nicht aktiv ist:
 - a. Melden Sie sich auf dem Indexserver mit der Benutzer-ID des Administrators für die Unternehmenssuche an, die bei der Installation von WebSphere II OmniFind Edition angegeben wurde.
 - b. Geben Sie den folgenden Befehl ein:


```
esadmin start
```
4. 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.
5. Suchen Sie die Objektgruppe in der Sicht **Objektgruppen**, die Sie verwalten wollen, und klicken Sie  **Überwachen** an.
6. Klicken Sie auf der Seite **Crawler** für jeden Crawler, den Sie starten wollen,
 -  **Starten** an.
 - Wenn Sie einen Web- oder NNTP-Crawler starten, beginnt der Crawler unverzüglich, Daten zu durchsuchen. Diese Crawlertypen führen die Crawler suche kontinuierlich aus, wobei sie Dokumente von Websites und NNTP- Newsgroups durchsuchen und erneut durchsuchen.
 - Wenn Sie einen anderen Crawlertyp starten, wird die Crawler sitzung gestartet. Der Crawler beginnt zum terminierten Datum und zur terminierten Zeit mit der Crawler suche. 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.
7. 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).

8. Optional: Wenn Sie den Start der Indexierungsprozesse erzwingen wollen, anstatt 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 aktualisiert und reorganisiert.

9. Klicken Sie **Starten** auf der Seite **Suche** an.

Sie können die Suchserver kontinuierlich ausführen. Normalerweise ist es erst erforderlich, die Suchserver zu stoppen, wenn Sie Änderungen an den Einstellungen des Suchcaches oder der Dokumentzusammenfassung vornehmen.

Zugehörige Konzepte

„Übersicht über die Verwaltung der Unternehmenssuche“ auf Seite 15

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.

„Verwaltungsrollen“ auf Seite 206

Bei der Unternehmenssuche wird das Konzept der Rollen verwendet, um den Zugriff auf verschiedene Funktionen in der Administrationskonsole zu steuern.

Zugehörige Tasks

„Anmelden an der Administrationskonsole“ auf Seite 19

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

Zugehörige Verweise

„Befehle, Rückkehrcodes und Sitzungs-IDs der Unternehmenssuche“ auf Seite 287

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.

Stoppen der Server 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 den Parser zu stoppen und erneut zu starten.


Wenn Sie statt einzelner Server das System für die Unternehmenssuche stoppen wollen, müssen Sie sich auf dem Indexserver mit der ID des Administrators für die

Unternehmenssuche anmelden (das ist die ID, die bei der Installation von WebSphere II OmniFind Edition angegeben wurde). Geben Sie danach den folgenden Befehl ein:

```
esadmin stop
```

Vorgehensweise

Gehen Sie wie folgt vor, um Server für die Unternehmenssuche zu stoppen:


1. Suchen Sie die Objektgruppe in der Sicht **Objektgruppen**, 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 in die Objektgruppe übernommene 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. Wenn Sie die Änderungen auf Dokumente anwenden wollen, 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 indexiert werden.

4. Klicken Sie  **Stoppen** auf der Seite **Index** an, um einen Index zu stoppen, der gerade aktualisiert oder reorganisiert 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.

Zugehörige Konzepte

„Übersicht über die Verwaltung der Unternehmenssuche“ auf Seite 15

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.

Zugehörige Verweise

„Befehle, Rückkehrcodes und Sitzungs-IDs der Unternehmenssuche“ auf Seite 287

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.

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 die Anzahl der bei einer Crawler- bzw. Indexerstellungssitzung 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 Konzepte

„Übersicht über die Verwaltung der Unternehmenssuche“ auf Seite 15

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.

Zugehörige Tasks

„Starten der Server für die Unternehmenssuche“ auf Seite 243

Damit Benutzer eine Objektgruppe durchsuchen können, müssen Sie die Systemprozesse und anschließend die Server für die Crawlersuche, die Syntaxanalyse, die Indexierung und die Suche für die Objektgruppe starten.

„Stoppen der Server für die Unternehmenssuche“ auf Seite 245

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.

Zugehörige Verweise

„Befehle, Rückkehrcodes und Sitzungs-IDs der Unternehmenssuche“ auf Seite 287

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.

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. Der Ressourcenmanager schätzt mit Hilfe 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. Der Ressourcenmanager gibt Warnungen aus, wenn der von Ihnen geschätzte Speicher- und Datenträgerressourcenbedarf größer ist als im Moment auf dem System verfügbar ist. Mit Hilfe dieser Warnungen können Sie in Zukunft Probleme, die wegen ungenügender Ressourcen entstehen, vermeiden.

Falls Sie Alerts für die Objektgruppe konfigurieren und die Option aktivieren, 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 Überwachungsprogramm verwendet diese Zahl zusammen mit dem von Ihnen angegebenen prozentualen Schwellenwert für Alerts 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.

Überprüfen der Verfügbarkeit von Systemressourcen

Wenn Sie die Systemressourcen nach dem Erstellen einer Objektgruppe oder eines Crawlers und noch vor dem Durchsuchen von Dokumenten überprüfen, können Sie feststellen, ob die verfügbaren Systemressourcen ausreichen, um WebSphere II OmniFind Edition mit dem vollen Nutzungsvolumen unter Ihren aktuellen Konfigurationseinstellungen auszuführen.

Vorbereitung

Sie müssen ein Administrator für die Unternehmenssuche sein, um Systemressourcen zu überprüfen.

Informationen zu dieser Task

Beim Erstellen einer Objektgruppe oder eines Crawlers überprüft das System automatisch die Verfügbarkeit von Ressourcen. Sie können auch eine Option auswählen, um die Verfügbarkeit von Systemressourcen zu einem beliebigen Zeitpunkt zu überprüfen.

Das System vergleicht den von Ihnen beim Konfigurieren von Objektgruppen- und Crawlermerkmalen angegebenen geschätzten verwendeten Speicherplatz mit dem verfügbaren Systemspeicherbereich und zeigt anschließend eine Nachricht mit Angaben zur Ressourcenverfügbarkeit an. Die Nachricht gibt an, welcher Server möglicherweise unzureichende Ressourcen aufweist, wie viel Platz laut Ihrer Schätzung der Objektgruppengröße erforderlich ist und wie viel freier Speicherbereich verfügbar ist.

Wenn nicht genügend Ressourcen verfügbar sind, versuchen Sie, eine der folgenden Fehlerberichtigungen durchzuführen:

- Vergrößern Sie das im Nachrichtentext angegebene Dateisystem.
- Bearbeiten Sie die allgemeinen Optionen für die Objektgruppe, und geben Sie für die geschätzte Anzahl Dokumente einen niedrigeren Wert an.
- Bearbeiten Sie die Crawlermerkmale, und geben Sie für die maximale Anzahl der zu durchsuchenden Dokumente, die maximale Seitengröße und die maximale Anzahl Threads einen niedrigeren Wert an.
- Bearbeiten Sie für einen Web-Crawler die Crawlermerkmale, und geben Sie für die maximale Anzahl der aktiven Hosts, der neuen Dokumente und der Dokumente, die sich im temporären Speicher befinden, einen niedrigeren Wert an.
- Bearbeiten Sie die Indexmerkmale für das System, und geben Sie für die Anzahl der Indexstellungen, die gleichzeitig ausgeführt werden dürfen, einen niedrigeren Wert an.
- Vermeiden Sie zu viele Prozesse, die in mehreren Objektgruppen gleichzeitig aktiv sind. Zum Beispiel kann sich die gleichzeitige Ausführung von Crawlern und Parsern mehrerer Objektgruppen negativ auf die Ressourcen auswirken.
- Löschen Sie Crawler aus dem System.
- Löschen Sie Objektgruppen aus dem System.

Wenn keine Warnungen auf Grund eines Mangels an Ressourcen angezeigt werden, sind wahrscheinlich genügend Ressourcen vorhanden, um die Daten zu durchsuchen und zu indexieren.

Vorgehensweise

Gehen Sie wie folgt vor, um festzustellen, ob ausreichende Ressourcen für die aktuelle Konfiguration Ihres Systems für die Unternehmenssuche vorhanden sind:

1. Klicken Sie **System** an, um die Sicht **System** zu öffnen.
2. Klicken Sie **Systemressourcen überprüfen** auf der Seite **Allgemein** an.

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.

Tip: 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 einer URI anzeigen wollen, klicken Sie  **URI-Details** an.

Sie wollen z. B. prüfen, ob sich eine bestimmte URI im Index befindet oder ob der Index, in dem sich die 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.

Zugehörige Konzepte

„Objektgruppen für die Unternehmenssuche“ auf Seite 29

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.

Anzeigen von Details zu einer URI

Sie können detaillierte Informationen zu einer URI anzeigen. Sie können aktuelle und protokollierte Informationen dazu anzeigen, wie das Dokument, das durch diese 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 einer 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 eine 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 die URI eines von einem Web-Crawler durchsuchten Dokuments angeben.

Vorgehensweise

Gehen Sie wie folgt vor, um Details zu einer 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** die URI ein, für die 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 indiziert 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 indiziert 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 269

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.

Zugehörige Tasks

„Anzeigen von Berichten über gelöschte Dokumente“ auf Seite 270

Sie können detaillierte Informationen zu Dokumenten anzeigen, die aus einem System für die Unternehmenssuche gelöscht wurden. Diese Informationen sind nur verfügbar, wenn Sie die Dokumentüberwachung für die Objektgruppe aktiviert haben.

Zugehörige Verweise

„URI-Formate in einem Index für die Unternehmenssuche“ auf Seite 103

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.

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 Überwachung von Crawlern 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 Crawler suche aktivieren oder inaktivieren.

4. Wenn ein Crawler gestoppt 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 Crawler suche und durchsucht den gesamten Crawlerbereich. Wurde der Crawler angehalten, nimmt er die Crawler suche 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 NNTP-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.

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.

Zugehörige Konzepte

„Crawlerverwaltung für die Unternehmenssuche“ auf Seite 37

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

Zugehörige Verweise

„Befehle, Rückkehrcodes und Sitzungs-IDs der Unternehmenssuche“ auf Seite 287

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.

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.
 - a. Klicken Sie **Threaddetails** an, um anzuzeigen, wie viele Threads aktiv die Websites durchsuchen, und wie viele sich im inaktiven Status befinden.
 - b. Klicken Sie **Aktive Sites** an, um Informationen über die Websites anzuzeigen, die der Crawler aktiv durchsucht.
 - c. 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.
 - d. Klicken Sie **Crawlerverlauf** an, um Berichte über vergangene Crawleraktivitäten anzuzeigen.
 - e. Geben Sie im Bereich **URL-Status** die URL einer Website an, für die Sie Informationen anzeigen möchten, und klicken Sie anschließend **Anzeigen** an.

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.

Nach dem Anzeigen der Details zur URL können Sie **Siteverlauf** anklicken, um zusätzliche statistische Informationen über die Crawleraktivität in dieser URL anzuzeigen.

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 über die 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 Ide-

allfall ist die Anzahl aktivierter URLs dicht an dem Wert, der für das Feld **Maximale Anzahl aktiver Hosts** in den Crawlermerkmalen 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, bearbeiten Sie die Crawlermerkmale und passen die Zeitlimitwerte im Feld **Maximale Zeit für URL-Adressen im Speicher** 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 Eigenschaften, 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 über die 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. Der Sitebericht und der HTTP-Rückkehrcodebericht benötigen eine gewisse Zeit, um erstellt zu werden. Wenn Sie diese Berichtstypen erstellen, können Sie eine E-Mail-Adresse für den Empfang des Berichts angeben, anstatt 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**.

Tip: 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 oder genaue Site an, für die Sie einen Bericht erstellen möchten, klicken Sie **Anzeigen** und anschließend **Siteverlauf** an.
5. Wählen Sie für die Crawlerverlaufs- und Siteverlaufsberichte die Markierungsfelder aller Statistikdaten aus, die Sie in einem Bericht sehen möchten, und klicken Sie anschließend **Bericht anzeigen** an.

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, geben Sie die Optionen für das Erstellen eines Siteberichts an, und klicken Sie anschließend **Bericht ausführen** an.

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, geben Sie die Optionen für das Erstellen eines HTTP-Rückkehrcodeberichts an, und klicken Sie anschließend **Bericht ausführen** an.

Dieser Bericht enthält Informationen über die Anzahl pro Site ausgegebener HTTP-Rückkehrcodes. 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 an 4er-Rückkehrcodes (die auf nicht gefundene Seiten hinweisen), 5er-Rückkehrcodes (die auf Serverprobleme hinweisen), 6er-Rückkehrcodes (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 großen Anzahlen an HTTP-Rückkehrcodes nicht zum Index beitragen, können Sie die Leistung des Crawlers verbessern, indem Sie die Sites aus dem Crawlerbereich entfernen.

Web-Crawler-HTTP-Rückkehrcodes

Wenn Sie einen Web-Crawler überwachen, können Sie Informationen zu den HTTP-Rückkehrcodes anzeigen, die der Crawler von den Seiten empfängt, die er versucht, zu durchsuchen.

Tabellenzusammenfassung

Wenn Sie den Web-Crawler-Verlauf oder den Status einer bestimmten URL-Adresse überwachen, können Sie die Informationen zu den HTTP-Rückkehrcodes 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 von HTTP-Rückkehrcodes für eine URL empfängt und der Rückkehrcode 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-Rückkehrcodes und ihre Interpretation durch den Web-Crawler aufgelistet. Die Werte von 100 bis 505 sind Standard-HTTP-Rückkehrcodes (weitere Informationen siehe <http://www.w3.org/Protocols/rfc2616/rfc2616.html>). Die verbleibenden HTTP-Rückkehrcodes gehören zur Unternehmenssuche und zum Web-Crawler.

Tabelle 4. HTTP-Rückkehrcodes vom Web-Crawler

Code	Beschreibung	Code	Beschreibung	Code	Beschreibung	Code	Beschreibung
NULL	Nicht durchsucht	401	Nicht berechtigt	500	Interner Serverfehler	700	Parserfehler (Kein Headerende)
100	Weiter	402	Zahlung erforderlich	501	Nicht implementiert	710	Parserfehler (Header)
200	Erfolgreich	403	Nicht erlaubt	502	Fehlerhaftes Gateway	720	Parserfehler (kein HTTP-Code)
201	Erstellt	404	Nicht gefunden	503	Service nicht verfügbar	730	Parserfehler (Hauptteil)
202	Akzeptiert	405	Methode nicht zulässig	504	Gateway-Zeitlimitüberschreitung	740	Ausgeschlossen durch die Datei robots.txt
203	Nicht berechtigte Informationen	406	Nicht zulässig	505	HTTP-Version nicht unterstützt	741	Robots vorübergehend nicht verfügbar
204	Kein Inhalt	407	Proxy-Authentifizierung erforderlich	611	Lesefehler	760	Ausgeschlossen durch Crawlerbereichsdefinition
205	Inhalt zurücksetzen	408	Zeitlimitüberschreitung der Anforderung	612	Verbindungsfehler	770	Fehlerhaftes Protokoll oder vom Standard abweichender Systemport
206	Teilinhalte	409	Konflikt	613	Lesezeitüberschreitung	780	Ausgeschlossen durch Dateityp-ausschlüsse
300	Multiple Choice	410	Fehlend	614	SSL-Handshake fehlgeschlagen	2004	Kein Index-META-Tag
301	Die Site ist umgezogen	411	Länge erforderlich	615	Anderer Lesefehler	3020	Weiche Umleitung
302	Gefunden	412	Bedingung fehlgeschlagen	616	FBA-Unregelmäßigkeit	4044	Ausgeschlossen durch die Datei robots.txt
303	Siehe Sonstiges	413	Anforderungseinheit zu groß	617	Verschlüsselungsfehler		
304	Nicht geändert	414	Anforderungs-URI zu groß	680	Störung der DNS-Suchfunktion		
305	Proxy verwenden	415	Datenträgertyp wird nicht unterstützt				
306	(Wird nicht verwendet)	417	Erwartung ist fehlgeschlagen				
307	Temporäre Umleitung						

Anmerkungen zur Tabelle

4er-Rückkehrcodes

Sie werden nur selten den Code 400 (fehlerhafte Anforderung) erhalten. Gemäß den HTTP-Rückkehrcodestandards sollen die 4er-Codes angegeben, 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-Rückkehrcode 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 Rückkehrcodes dieses Typs kann darauf hindeuten, dass der Crawler einen Fehler hat.

6er-Rückkehrcodes

Mit Ausnahme des Codes 615 weisen 6er-Rückkehrcodes auf zu erwartende Probleme bei der Crawlersuche hin, wie z. B. Zeitlimitüberschreitungen. Die folgenden Rückkehrcodes erfordern möglicherweise Korrekturmaßnahmen:

611, 612 und 613

Weisen auf langsame Sites oder eine schlechte Netzleistung hin.

- 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 Domänennamen der Seite übereinstimmen, die angefordert wird.

- 616** Gibt an, dass das Anmeldeformular auch nach erneuter Authentifizierung immer noch im Download angezeigt wird.

- 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.
- 680 Gibt an, dass der Crawler die IP-Adressen für Hosts im Crawlerbereich nicht beschaffen konnte, möglicherweise auf Grund 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 von Rückkehrcodes dieses Typs verringert den Durchsatz erheblich.

7er-Rückkehrcodes

Die 7er-Rückkehrcodes entstehen hauptsächlich auf Grund der Regeln im Crawlerbereich:

710 - 730

Geben an, dass der Crawler auf Grund von Problemen kein vollständiges Download ausführen konnte, oder dass der Crawler in einer Site auf ungültige HTML-Daten gestoßen ist. Wenn Sie viele Rückkehrcodes 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 von Rückkehrcodes dieses Typs sollten Sie prüfen, ob die Zielseite vorübergehend oder dauerhaft nicht verfügbar ist. Wenn die Zielseite nicht mehr verfügbar ist, entfernen Sie sie aus dem Crawlerbereich.

Die anderen 7er-Rückkehrcodes treten hauptsächlich dann auf, wenn Sie Änderungen am Crawlerbereich durchführen, nachdem der Crawler bereits einige Zeit aktiv war. Solche Rückkehrcodes deuten normalerweise nicht auf Probleme hin, die Sie beheben müssten.

3020 Gibt an, dass ein Dokument mit dem Rückkehrcode 200 einen Positionsholder 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 automatischer 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 unter Umständen angehalten sein (beispielsweise kann der Parser so lange angehalten sein, bis eine Indexreorganisation 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 wurde 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 Indexstellungsprozess 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 wurde 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 Indexerstellung zu stoppen, z. B. um die Reorganisation eines Index zu erzwingen, nachdem Sie den in der Objektgruppe verwendeten Kategorisierungstyp geändert haben.

7. Wenn Fehler während einer Indexerstellung 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.

Zugehörige Konzepte

„Indexverwaltung für die Unternehmenssuche“ auf Seite 143

Um sicherzustellen, dass Benutzer immer Zugriff auf die neuesten Informationen haben, erstellt die Unternehmenssuche für jede Objektgruppe einen Index und pflegt ihn durch regelmäßige Inhaltsaktualisierung und -reorganisation.

Überwachen der Indexierungswarteschlange für die Unternehmenssuche

Sie können den Status aller Indexerstellung 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 Indexierungswarteschlange 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 Indexierungswarteschlange befinden. Für jeden Index wird der zu erstellende Indextyp (Aktualisierung oder Reorganisation), der Zeitpunkt, zu der der Index in die Indexierungswarteschlange 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 unter Umständen die Indexaktualisierung stoppen, so dass Sie stattdessen den Start einer Indexreorganisation 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 den Index zu aktualisieren oder zu reorganisieren.

5. Wenn Sie einen Index aus der Indexierungwarteschlange entfernen wollen, klicken Sie  **Entfernen** an.

Zugehörige Konzepte

„Indexverwaltung für die Unternehmenssuche“ auf Seite 143

Um sicherzustellen, dass Benutzer immer Zugriff auf die neuesten Informationen haben, erstellt die Unternehmenssuche für jede Objektgruppe einen Index und pflegt ihn durch regelmäßige Inhaltsaktualisierung und -reorganisation.

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 wurde 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. Mit dem Erstellen von Quick Links können Sie die Suchqualität für viele Benutzer positiv beeinflussen. 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).

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.

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 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 ihre Portnummer ändern.

Zugehörige Tasks

„Konfigurieren der Unterstützung für Anwendungen der Datenempfangsfunktion“ auf Seite 100

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.

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 die URI aus dem Index gelöscht hat.

Zugehörige Tasks

„Anzeigen von Details zu einer URI“ auf Seite 252

Sie können detaillierte Informationen zu einer URI anzeigen. Sie können aktu-

elle und protokollierte Informationen dazu anzeigen, wie das Dokument, das durch diese URI dargestellt wird, durchsucht, indexiert und gesucht wird.

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, so dass 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.
4. Klicken Sie **OK** an.

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 Sie die Dokumentüberwachung für die Objektgruppe aktiviert haben.

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** die URI ein, für die 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 indexiert 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 einer URI“ auf Seite 252

Sie können detaillierte Informationen zu einer URI anzeigen. Sie können aktuelle und protokollierte Informationen dazu anzeigen, wie das Dokument, das durch diese URI dargestellt wird, durchsucht, indexiert und gesucht wird.

| **Zugehörige Verweise**

| „URI-Formate in einem Index für die Unternehmenssuche“ auf Seite 103
| 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.

Protokolldateien und Alerts für die Unternehmenssuche

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, so dass 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" in "Fehlernachrichten"

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, und eine weitere Nachricht wird protokolliert, wenn der Notes-Crawler 900 000 Dokumente durchsucht.

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

„Empfangen von E-Mail zu protokollierten Nachrichten“ auf Seite 279
Sie können Optionen angeben, um E-Mail zu empfangen, wenn bestimmte Nachrichten oder Nachrichtentypen protokolliert werden.

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 279
Sie können Optionen angeben, um E-Mail zu empfangen, wenn bestimmte Nachrichten oder Nachrichtentypen protokolliert werden.

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, so dass 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.

4. Geben Sie in das Feld **Maximale Größe jeder Protokolldatei** die maximale Anzahl Megabyte für jede Protokolldatei ein. Der Standardwert ist 5 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 10.
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.

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 einer 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" in "Fehlernachrichten"

Zugehörige Tasks

„Konfigurieren von Alerts auf Objektgruppenebene“ auf Seite 274

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.

„Konfigurieren von Alerts auf Systemebene“ auf Seite 276

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.


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 Objektgruppe, die Sie anzeigen wollen, in der Objektgruppenliste, 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 eine System- oder Objektgruppen-ID), das Datum der Dateierstellung und ein numerisches Suffix ein, das die Reihenfolge angibt, in der die Datei an dem Datum erstellt wurde. Beispiel:

```
protokolldateityp_2005-05-26_1.log  
protokolldateityp_2005-05-26_2.log  
protokolldateityp_2005-05-25_1.log  
protokolldateityp_2005-05-25_2.log  
protokolldateityp_2005-05-25_3.log
```

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.

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.

Wenn das System auf Grund eines nicht behebbaren Fehlers fehlschlägt, müssen Sie WebSphere Information Integrator OmniFind Edition erneut installieren und anschließend das Restore-Script ausführen. Sie können diese Scripts auch verwenden, um wichtige Systemdateien auf mindestens einem neuen Server wiederherzustellen.

Die Scripts können die folgenden Dateien sichern und wiederherstellen:

- Konfigurationsdateien aus dem Verzeichnis `ES_NODE_ROOT/master_config`.
- Datenbankdateien für die Crawler.
- Indexdateien. Wenn sich die Indizes für die Unternehmenssuche nicht im Verzeichnis `ES_NODE_ROOT/data` befinden, können Sie die Scripts für die Unternehmenssuche zum Sichern und Wiederherstellen der Indexdateien nicht verwenden.

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

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. Wenn Sie ein Backup starten, sind die meisten Systemsitzungen vorübergehend nicht verfügbar. Der Suchvorgang wird weiterhin ausgeführt. Sie sollten das Backup starten, nachdem Sie den Index reorganisiert haben, um sicherzustellen, dass Sie über den neuesten Index verfügen.

Führen Sie in einer Installation mit mehreren Servern das Backup und den Restore des Systems über den Indexserver für die Unternehmenssuche aus. Da der Indexserver einen Datenbankkatalog erstellt, kann der Indexserver auf die Produktdatenbanktabellen auf dem Crawler-Server zugreifen und sie sichern.

Zugehörige Verweise

„Befehle, Rückkehrcodes und Sitzungs-IDs der Unternehmenssuche“ auf Seite 287

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.

Sichern des Systems für die Unternehmenssuche

Sie können das System für die Unternehmenssuche sichern, indem Sie unter UNIX das Script `esbackup.sh` oder unter Microsoft Windows das Script `esbackup.bat` verwenden.

Einschränkungen

Alle Systemsitzungen werden gestoppt, während die Backup- und Restore-Skripts ausgeführt werden. Damit keine fehlerhaften oder inkonsistenten Systeminformationen angezeigt werden, verwenden Sie die Administrationskonsole für die Unternehmenssuche nicht, während die Skripts ausgeführt werden.

Die ID des Administrators für die Unternehmenssuche muss über eine Schreibberechtigung für das Verzeichnis verfügen, das Sie angeben, wenn Sie das Backup-Skript ausführen.

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 WebSphere II OmniFind Edition angegeben.
2. Geben Sie den folgenden Befehl ein, wobei *backupverzeichnis* das Verzeichnis ist, in dem die Dateien gesichert werden sollen:

UNIX: `esbackup.sh -c -d -i backupverzeichnis`

Windows: `esbackup.bat -c -d -i backupverzeichnis`

Sie können die folgenden Optionen angeben:

- c Sichert die Konfigurationsdateien. Dies ist die Standardoption.
- d Sichert durchsuchte Dokumente im Datenspeicher des Crawlers (diese Dokumente sind noch nicht syntaktisch analysiert oder indexiert).
- i Sichert die Indexdateien. Wenn sich die Indexdateien nicht im Verzeichnis `ES_NODE_ROOT/data` befinden, können Sie das Skript `esbackup` zum Sichern der Indexdateien nicht verwenden.

Zugehörige Verweise

„Befehle, Rückkehrcodes und Sitzungs-IDs der Unternehmenssuche“ auf Seite 287

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.

Wiederherstellung des Systems für die Unternehmenssuche

Sie können Systemkonfigurationsdateien wiederherstellen, nachdem Sie WebSphere II OmniFind Edition erneut installiert haben, indem Sie das Skript `esrestore.sh` unter UNIX oder das Skript `esrestore.bat` unter Microsoft Windows verwenden.

Einschränkungen

Alle Systemsitzungen werden gestoppt, während die Backup- und Restore-Skripts ausgeführt werden. Damit keine fehlerhaften oder inkonsistenten System-

informationen angezeigt werden, verwenden Sie die Administrationskonsole für die Unternehmenssuche nicht, während die Scripts ausgeführt werden.

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 WebSphere II OmniFind Edition angegeben.
2. Starten Sie die allgemeine Übertragungsschicht (CCL = Common Communications Layer), falls diese noch nicht gestartet ist:

UNIX: `startccl.sh -bg`

Windows: `startccl.bat`

Wenn Sie die allgemeine Übertragungsschicht im Hintergrund auf einem Windows-System starten möchten, klicken Sie **Start** → **Programme** → **Verwaltung** → **Dienste** an, und starten Sie den Dienst "WebSphere Information Integrator OmniFind Edition" erneut.

3. Stoppen Sie den Controller, falls er nicht bereits gestoppt wurde:

`esadmin stop`

4. Geben Sie den folgenden Befehl ein, wobei *backupverzeichnis* das Verzeichnis ist, in dem Sie die Dateien gesichert haben:

UNIX: `esrestore.sh -c -d -i backupverzeichnis`

Windows: `esrestore.bat -c -d -i backupverzeichnis`

Sie können die folgenden Optionen angeben:

-c Stellt die Konfigurationsdateien wieder her. Dies ist die Standardoption.

-d Stellt durchsuchte Dokumente im Datenspeicher des Crawlers wieder her. (Diese Dokumente wurden noch nicht syntaktisch analysiert und indiziert.)

-i Stellt die Indexdateien wieder her. Wenn sich die Indexdateien nicht im Verzeichnis `ES_NODE_ROOT/data` befinden, können Sie das Script `esrestore` zum Wiederherstellen der Indexdateien nicht verwenden.

Zugehörige Verweise

„Befehle, Rückkehrcodes und Sitzungs-IDs der Unternehmenssuche“ auf Seite 287

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.

Wiederherstellen von Systemdateien für die Unternehmenssuche auf neuen Servern

Sie können Systemdateien aus einem Server für die Unternehmenssuche sichern und die Systemdateien auf mindestens einem neuen Server wiederherstellen, anstatt sie auf den aktuellen Servern wiederherzustellen.

Vorbereitung

Sie müssen WebSphere II OmniFind Edition auf den neuen Servern installieren, bevor Sie das Script für den Restore ausführen.

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.

Die ID des Administrators für die Unternehmenssuche muss über eine Schreibberechtigung für das Verzeichnis verfügen, das Sie angeben, wenn Sie das Backup-Script ausführen.

Informationen zu dieser Task

Die Serverinformationen, die in der Datei `ES_NODE_ROOT/master_config/nodes.ini` gespeichert sind, sind nicht in den Backup-Dateien enthalten.

Vorgehensweise

Gehen Sie wie folgt vor, um die Systemdateien für die Unternehmenssuche auf mindestens einem neuen Server wiederherzustellen:

1. Melden Sie sich am Indexserver als Administrator für die Unternehmenssuche an. Diese Benutzer-ID wurde bei der Installation von WebSphere II OmniFind Edition angegeben.
2. Führen Sie das Backup-Script auf dem derzeitigen (alten) Indexserver aus, wobei *backupverzeichnis* das Verzeichnis ist, in dem die Dateien gesichert werden sollen:

UNIX: `esbackup.sh backupverzeichnis`

Windows: `esbackup.bat backupverzeichnis`

3. Verwenden Sie ein FTP-Programm, um alle Dateien in *backupverzeichnis* an den neuen Indexserver zu senden.
4. Führen Sie das Restore-Script auf dem neuen Indexserver aus:

UNIX: `esrestore.sh backupverzeichnis`

Windows: `esrestore.bat backupverzeichnis`

Zugehörige Verweise

„Befehle, Rückkehrcodes und Sitzungs-IDs der Unternehmenssuche“ auf Seite 287

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.

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 5. *esadmin*-Befehle für die Unternehmenssuche

Befehl	Beschreibung
<code>esadmin <i>crawler-sitzungs-id</i> start</code>	Startet eine Crawlersitzung. Dieser Befehl startet keine Crawleraktivität. Beispielbefehl: <code>esadmin col1.WEB1.esadmin start</code> Beispiele für Nachrichten und Rückkehrcodes: FFQC5310I WEBCrawler1 (sid: col1.WEB1.esadmin) ist nicht aktiv. FFQC5314I Ergebnis: 0
<code>esadmin <i>crawler-sitzungs-id</i> startCrawl</code>	Startet die Crawler suche. Beispielbefehl: <code>esadmin col3.DB21.esadmin startCrawl</code> Beispiele für Nachrichten und Rückkehrcodes: FFQC5303I DB2Crawler1 (sid: col3.DB21.esadmin) ist bereits aktiv. PID: 23650 FFQC5314I Ergebnis: 0

| *Tabelle 5. esadmin-Befehle für die Unternehmenssuche (Forts.)*

Befehl	Beschreibung
esadmin <i>crawler-sitzungs-id</i> pause	Hält die Crawlersuche an.
	Beispielbefehl:
	esadmin col3.DB21.esadmin pause
	Beispiele für Nachrichten und Rückkehrcodes:
	FFQC5303I DB2Crawler1 (sid: col3.DB21.esadmin)
	ist bereits aktiv. PID: 23650
	FFQC5314I Ergebnis: 0
esadmin <i>crawler-sitzungs-id</i> resume	Setzt die Crawlersuche fort.
	Beispielbefehl:
	esadmin col3.DB21.esadmin resume
	Beispiele für Nachrichten und Rückkehrcodes:
	FFQC5303I DB2Crawler1 (sid: col3.DB21.esadmin)
	ist bereits aktiv. PID: 23650
	FFQC5314I Ergebnis: 0
esadmin <i>crawler-sitzungs-id</i> stopCrawl	Stoppt die Crawlersuche.
	Beispielbefehl:
	esadmin col3.DB21.esadmin stopCrawl
	Beispiele für Nachrichten und Rückkehrcodes:
	FFQC5303I DB2Crawler1 (sid: col3.DB21.esadmin)
	ist bereits aktiv. PID: 23650
	FFQC5314I Ergebnis: 0
esadmin <i>crawler-sitzungs-id</i> stop	Stoppt eine Crawlersitzung.
	Beispielbefehl:
	esadmin col3.DB21.esadmin stop
	Beispiele für Nachrichten und Rückkehrcodes:
	FFQC5303I DB2Crawler1 (sid: col3.DB21.esadmin)
	ist bereits aktiv. PID: 23650
	FFQC5314I Ergebnis: 0

Tabelle 5. esadmin-Befehle für die Unternehmenssuche (Forts.)

Befehl	Beschreibung
esadmin <i>crawler-sitzungs-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>Beispiel für einen Web-Crawler:</p> <pre>esadmin col1.WEB1.esadmin getCrawlerStatus</pre> <p>Mögliche Rückkehrcodes und Nachrichten für einen Web-Crawler:</p> <pre>FFQC5303I WebCrawler1 (sid: col1.WEB1.esadmin) ist bereits aktiv. PID: 23650</pre> <p>Beispiel für einen Nicht-Web-Crawler:</p> <pre>esadmin col3.DB21.esadmin getCrawlerStatus</pre> <p>Mögliche Rückkehrcodes und Nachrichten für einen Nicht-Web-Crawler:</p> <pre>FFQC5303I db2crawler (sid: db2col.DB2_96945) ist bereits aktiv. PID: 5936</pre> <p>Weitere Informationen zu zurückgegebenen Statusnachrichten finden Sie in „Detaillierte Informationen zu Statusbefehlen“ auf Seite 293.</p>
esadmin <i>dscrawler-sitzungs-id</i> getCrawlSpaceStatus	<p>Ruft den allgemeinen Status des Crawlerbereichs für jeden anderen Crawler als den Web-Crawler ab.</p>
esadmin <i>web-crawler-sitzungs-id</i> getCrawlStatus -selections <i>wert</i>	<p>Beispielbefehl:</p> <pre>esadmin col3.DB21.esadmin getCrawlSpaceStatus</pre> <p>Beispiele für Nachrichten und Rückkehrcodes:</p> <pre>FFQC5303I DB2Crawler1 (sid: col3.DB21.esadmin) ist bereits aktiv. PID: 23650</pre> <p>Ruft den allgemeinen Status des Crawlerbereichs für den Web-Crawler ab.</p> <p>Beispielbefehl:</p> <pre>esadmin col1.WEB1.esadmin getCrawlStatus</pre> <p>Weitere Informationen zu zurückgegebenen Statusnachrichten finden Sie in „Detaillierte Informationen zu Statusbefehlen“ auf Seite 293.</p>

Tabelle 5. esadmin-Befehle für die Unternehmenssuche (Forts.)

Befehl	Beschreibung
<pre>esadmin dscrawler-sitzungs-id getCrawlSpaceStatusDetail -ts zielserver-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>Beispielbefehl:</p> <pre>esadmin col3.DB21.esadmin getCrawlSpaceStatusDetail -ts FOUNTAIN</pre> <p>Beispiele für Nachrichten und Rückkehrcodes:</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>Beispielbefehl:</p> <pre>esadmin col1.WEB1.esadmin getCrawlDetailsPerSite</pre> <p>Weitere Informationen zu zurückgegebenen Statusnachrichten finden Sie in „Detaillierte Informationen zu Statusbefehlen“ auf Seite 293.</p>
<pre>esadmin startParser -cid objektgruppen-id</pre>	<p>Startet den Parser.</p> <p>Beispielbefehl:</p> <pre>esadmin startParser -cid col1</pre> <p>Beispiele für Nachrichten und Rückkehrcodes:</p> <pre>FFQC5303I Controller (server1) (sid: controller) ist bereits aktiv. PID: 25917 FFQC5314I Ergebnis: 0</pre>
<pre>esadmin stopParser objektgruppen-id</pre>	<p>Stoppt den Parser.</p> <p>Beispielbefehl:</p> <pre>esadmin stopParser -cid col1</pre> <p>Beispiele für Nachrichten und Rückkehrcodes:</p> <pre>FFQC5303I Controller (knoten1) (sid: controller) ist bereits aktiv. PID: 15292 FFQC5312E Fehler beim Ausführen der Aktion: [stopParser] FFQC4823E Parsertreiber der Sitzung - Objektgruppe obj-gr1 (knoten1) [col1.parserdriver] ist nicht aktiv.</pre>

Tabelle 5. esadmin-Befehle für die Unternehmenssuche (Forts.)

Befehl	Beschreibung
esadmin monitor getCollectionParserMonitorStatus -cid <i>objektgruppen-id</i>	Ruft den Parserstatus ab. Beispielbefehl: esadmin monitor getCollectionParserMonitorStatus -cid coll Beispiele für Nachrichten und Rückkehrcodes: FFQC5303I Monitor (knoten1) (sid: monitor) ist bereits aktiv. PID: 12543 Weitere Informationen zu zurückgegebenen Statusnachrichten finden Sie in „Detaillierte Informationen zu Statusbefehlen“ auf Seite 293.
esadmin startMain -cid <i>objektgruppen-id</i>	Startet das Reorganisieren eines Index. Beispielbefehl: esadmin startMain -cid coll Beispiele für Nachrichten und Rückkehrcodes: FFQC5303I Controller (knoten1) (sid: controller) ist bereits aktiv. PID: 25917 FFQC5314I Ergebnis: 1117671147056
esadmin startDelta -cid <i>objektgruppen-id</i>	Startet das Aktualisieren eines Index. Beispielbefehl: esadmin startDelta -cid coll Beispiele für Nachrichten und Rückkehrcodes: FFQC5303I Controller (knoten1) (sid: controller) ist bereits aktiv. PID: 4548 FFQC5314I Ergebnis: 1117670603408
esadmin monitor getCollectionIndexMonitorStatus -cid <i>objektgruppen-id</i> -buildType [main delta] -numrecords <i>letzte-n-datensätze</i>	Ruft den Status einer Indexaktualisierung oder -reorganisation ab. (Der Typ main verweist auf einen reorganisierten Index; der Typ delta verweist auf einen aktualisierten Index). Die Option numrecords zeigt die letzten N Statusdatensätze der Indexer- stellung. Wenn numrecords nicht angegeben wird, wird der Sta- tus für die letzten 20 Indexerstellungen zurückgegeben. Beispielbefehl: esadmin monitor getCollectionIndexMonitorStatus -cid coll -buildType main -numrecords 4 Beispiele für Nachrichten und Rückkehrcodes: FFQC5303I Monitor (knoten1) (sid: monitor) ist bereits aktiv. PID: 12649 Weitere Informationen zu zurückgegebenen Statusnachrichten finden Sie in „Detaillierte Informationen zu Statusbefehlen“ auf Seite 293.

Tabelle 5. esadmin-Befehle für die Unternehmenssuche (Forts.)

Befehl	Beschreibung
esadmin startSearch -cid <i>objektgruppen-id</i>	<p>Startet die Suchserverprozesse.</p> <p>Beispielbefehl:</p> <pre>esadmin startSearch -cid coll</pre> <p>Beispiele für Nachrichten und Rückkehrcodes:</p> <pre>FFQC5303I Controller (knoten1) (sid: controller) ist bereits aktiv. PID: 25917 FFQC5314I Ergebnis: 0</pre>
esadmin stopSearch -cid <i>objektgruppen-id</i>	<p>Stoppt die Suchserverprozesse.</p> <p>Beispielbefehl:</p> <pre>esadmin stopSearch -cid coll</pre> <p>Beispiele für Nachrichten und Rückkehrcodes:</p> <pre>FFQC5303I Controller (knoten1) (sid: controller) ist bereits aktiv. PID: 15292 FFQC5314I Ergebnis: 0</pre>
esadmin monitor getCollectionSearchMonitorStatus -cid <i>objektgruppen-id</i> esadmin <i>suchmanager-sitzungs-id</i> getStatus -cid <i>objektgruppen-id</i>	<p>Ruft den Status des Suchservers ab.</p> <p>Beispielbefehl:</p> <pre>esadmin monitor getCollectionSearchMonitorStatus -cid coll</pre> <p>Beispiele für Nachrichten und Rückkehrcodes:</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>Beispielbefehl:</p> <pre>esadmin searchmanager.node1 getStatus -cid coll</pre> <p>Beispiele für Nachrichten und Rückkehrcodes:</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“ auf Seite 293.</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 287 sind mögliche zurückgegebene Informationen der jeweiligen **esadmin**-Befehle 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 6. Informationen zum Web-Crawler-Status

Element	Attribute	Beschreibung
CrawlerStatus	<ul style="list-style-type: none"> • CrawlerThreadStateDist • ActiveBucketList • CrawlRate • RecentlyCrawledURLList • NumURLsThisSession 	Crawlerstatus.
CrawlerRunLevel Value	<ul style="list-style-type: none"> • Zeichenfolge (Englisch) „Not started“: Die Crawler Sitzung ist vorhanden, aber sie hat noch keine Startnachricht zum Verarbeiten von Dokumenten empfangen. • „Started“: Der Crawler wird gestartet. • „Running“: Der Crawler hat die Initialisierung und den Start beendet und führt gerade eine Crawler suche aktiv aus. • „Paused“: Der Crawler wurde aufgefordert, die aktive Crawler suche auszusetzen, aber nicht zu beenden. • „Stopping“: Der Crawler hat das Stoppsignal empfangen und wird gestoppt. • „Error“: Der Crawler befindet sich in einem Status, in dem keine Recovery durchgeführt werden kann. Er muss gestoppt und erneut gestartet werden, um die Crawler suche fortzusetzen. 	Informationen zur Aktivität des Crawlers.
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"> • URL: Zeichenfolge (URL-Angabe) Das Protokoll, der Host und der Port, dessen URL-Adressen gerade durchsucht werden. • NumActURLs: Ganze Zahl (positiv) Die Anzahl URL-Adressen im Bucket, als er für die Crawler suche zur Verfügung gestellt wurde (aktiviert). • NumProcURLs: Ganze Zahl (nicht negativ) Die Anzahl URL-Adressen aus dem Bucket, die bisher verarbeitet, d. h. durchsucht oder zurückgewiesen wurden. • TimeRem: Ganze Zahl Die Anzahl Sekunden bis zur Zeitlimit-überschreitung des Bucket. • Duration: Ganze Zahl (nicht negativ) Die Anzahl Sekunden, seitdem der Bucket aktiviert wurde. 	Die aktuelle Aktivität einer angegebenen Website.
CrawlRate	Value: Ganze Zahl (nicht negativ) Seiten, die pro Sekunde durchsucht werden (alle Buckets kombiniert).	Der Messwert des Crawler-Durchsatzes.

Tabelle 6. Informationen zum Web-Crawler-Status (Forts.)

Element	Attribute	Beschreibung
RecentlyCrawledURL	URL: Zeichenfolge (URL-Angabe) Zeichenfolge, die ein Protokoll, einen Host, einen Port und eine Datei angibt, die durchsucht wurde.	Eine Seite, die kürzlich durchsucht wurde.
NumURLsThisSession	Value: Ganze Zahl (nicht negativ)	Die Anzahl URL-Adressen, die durchsucht wurden, seitdem diese Instanz des Crawlers (Prozess) die Crawler suche 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 Statusbefehl für den 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 der folgenden Tabelle 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 7. Elemente und Attribute für den Statusbefehl des Crawlers für den NNTP-Crawler, den DB2-Crawler und den Notes-Crawler

Element- und Attributname	NNTP-Crawler	DB2-Crawler	Notes-Crawler
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 8. Elemente und Attribute für den Statusbefehl des Crawlers für den Exchange Server-Crawler, den DB2 Content Manager-Crawler und den Content Edition-Crawler

Element- und Attributname	Exchange Server-Crawler	DB2 Content Manager-Crawler	Content Edition-Crawler
Status	Status (0, 1, 2, -1)	Status (0, 1, 2, -1)	Status (0, 1, 2, -1)
StatusMessage	Status: 0 - 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 9. Elemente und Attribute für den Statusbefehl des Crawlers für den QuickPlace-Crawler, den Domino Document Manager-Crawler sowie für den Crawler für das UNIX- und Windows-Dateisystem.

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.

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 10. 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-Rückkehrcodes

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 11. Informationen, die durch den Statusbefehl des Crawlerbereichs für den Web-Crawler zurückgegeben wurden

Element	Attribute	Beschreibung
CrawlerStatus	<ul style="list-style-type: none"> • NumPagesInRDS • NumSitesDiscovered • NumSitesWithDNS • NumSitesWithoutDNS • NumURLsDiscovered • NumUniquePagesSaved • NumURLsCrawled • NumURLsUncrawled • NumURLsOverdue • HTTPCodeDist 	Informationen zum kumulativen Status der Crawlersuche (alle Sitzungen), die schnell abgerufen werden können.
NumPagesInRDS	<p>Value: Nicht negative ganze Zahl</p> <p>Anzahl Seiten, die sich gegenwärtig im Zwischenspeicherungsbereich des Rohdatenspeichers (RDS) befinden (nur von diesem Crawler)</p>	Auslastung des Rohdatenspeichers (RDS) (nur durch Daten von diesem Crawler).
NumSitesDiscovered	<p>Value: Nicht negative ganze Zahl</p> <p>Anzahl Hosts, die durch die Crawlersuche (oder anhand von Einstiegspunkten) aufgespürt wurden.</p>	Eine Kennzahl für den Geltungsbereich der Domäne des Crawlers, die durchsucht werden soll (Hostanzahl).
NumSitesWithDNS	<p>Value: Nicht negative ganze Zahl</p> <p>Anzahl Hosts mit zugeordneten IP-Adressen (vom Crawler im Hintergrund aufgelöst)</p>	Eine Kennzahl, wie effektiv der Crawler IP-Adressen für Hosts abrufen kann, die durch DNS-Namen in den URL-Adressen aufgespürt werden.
NumSitesWithoutDNS	<p>Value: Nicht negative ganze Zahl</p> <p>Anzahl Hosts ohne zugeordnete IP-Adressen (vom Crawler im Hintergrund aufgelöst)</p>	Eine Kennzahl, wie effektiv der Crawler IP-Adressen für Hosts abrufen kann, die durch DNS-Namen in den URL-Adressen aufgespürt werden.
NumURLsDiscovered	<p>Value: Nicht negative ganze Zahl</p> <p>Anzahl eindeutiger URL-Adressen, die durch den Crawler aufgerufen wurden.</p>	Eine Kennzahl für den Geltungsbereich der Domäne des Crawlers, die durchsucht werden soll (URL-Anzahl).
NumUniquePagesSaved	<p>Value: Nicht negative ganze Zahl</p> <p>Anzahl eindeutiger Seiten, die zur weiteren Verarbeitung durch andere Komponenten für die Unternehmenssuche in den Rohdatenspeicher (RDS) geschrieben wurden.</p>	Der Beitrag dieses Crawlers zur Größe des Index.
NumURLsCrawled	<p>Value: Nicht negative ganze Zahl</p> <p>Anzahl eindeutiger URL-Adressen, die vom Crawler durchsucht wurden.</p>	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	<p>Value: Nicht negative ganze Zahl</p> <p>Anzahl eindeutiger URL-Adressen, die erneut durchsucht werden können.</p>	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 Statusbefehl für den 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 der folgenden Tabelle werden die XML-Elemente und -Attribute für jeden Crawler der 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 12. Elemente und Attribute für den Statusbefehl des Crawlerbereichs für den NNTP-Crawler, den DB2-Crawler und den Notes-Crawler

Element- und Attributname	NNTP-Crawler	DB2-Crawler	Notes-Crawler
Server@Name	Name des News-Servers	Datenbankname	Datenbankname oder Verzeichnisname
Server/Status	Status: (0, 1, 2, 3, 4, -1) <ul style="list-style-type: none"> 0: Keine Crawlersuche ausgeführt 1: Crawlersuche wird ausgeführt 2: Abgeschlossen (nicht terminiert) 3: Wartestatus 4: Angehalten -1: Fehler 	Status (0, 1, 2, 3, 4, 5, -1) <ul style="list-style-type: none"> 0: Keine Crawlersuche ausgeführt 1: Crawlersuche wird ausgeführt 2: Abgeschlossen (nicht terminiert) 3: Wartestatus 4: Angehalten 5: Terminiert -1: Fehler 	Status (0, 1, 2, 3, 4, 5, -1) <ul style="list-style-type: none"> 0: Keine Crawlersuche ausgeführt 1: Crawlersuche wird ausgeführt 2: Abgeschlossen (nicht terminiert) 3: Wartestatus 4: Angehalten 5: Terminiert -1: Fehler
Server/StatusMessage	<ul style="list-style-type: none"> 0: Keine Crawlersuche ausgeführt 1: Crawlersuche wird ausgeführt 2: Abgeschlossen (nicht terminiert) 3: Wartestatus 4: Angehalten 5: Terminiert -1: Fehler 	<ul style="list-style-type: none"> 0: Keine Crawlersuche ausgeführt 1: Crawlersuche wird ausgeführt 2: Abgeschlossen (nicht terminiert) 3: Wartestatus 4: Angehalten 5: Terminiert -1: Fehler 	<ul style="list-style-type: none"> 0: Keine Crawlersuche ausgeführt 1: Crawlersuche wird ausgeführt 2: Abgeschlossen (nicht terminiert) 3: Wartestatus 4: Angehalten 5: Terminiert -1: Fehler

Table 12. Elements and Attributes for the Status Command of the Crawler Area for the NNTP-Crawler, the DB2-Crawler and the Notes-Crawler (Forts.)

Element- und Attributname	NNTP-Crawler	DB2-Crawler	Notes-Crawler
Server/NumberOfTargets	Die Anzahl Newsgroups im Crawlerbereich.	Die Anzahl Datenbanken im Crawlerbereich.	Die Anzahl Sichten und Ordner im Crawlerbereich.
Server/NumberOfCompletedTargets	Die Anzahl durchsuchter Newsgroups.	Die Anzahl durchsuchter Tabellen.	Die Anzahl durchsuchter Sichten und Ordner.
Server/NumberOfErrors	Nicht zutreffend.	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	Nicht zutreffend.	0, 1, 2 <ul style="list-style-type: none"> • 0: Der Crawler ist entsprechend der Crawlerkonfigurationsdateien nicht für die Terminierung konfiguriert. • 1: Der Crawler ist für die Terminierung konfiguriert, aber die Terminierung wurde für die Sitzung inaktiviert. • 2: Der Crawler ist für die Terminierung konfiguriert, und die Terminierung ist für die Sitzung aktiviert. 	0, 1, 2 <ul style="list-style-type: none"> • 0: Der Crawler ist entsprechend der Crawlerkonfigurationsdateien nicht für die Terminierung konfiguriert. • 1: Der Crawler ist für die Terminierung konfiguriert, aber die Terminierung wurde für die Sitzung inaktiviert. • 2: Der Crawler ist für die Terminierung konfiguriert, und die Terminierung ist für die Sitzung aktiviert.
Server/ScheduleTime	Nicht zutreffend.	Planzeit, falls zutreffend.	Planzeit, falls zutreffend.
Server/TotalTime	Die Gesamtzeit, falls zutreffend.	Die Gesamtzeit, falls zutreffend.	Die Gesamtzeit, falls zutreffend.
Server/AggregationLevel	Nicht zutreffend.	0: Der Notes-Crawler durchsucht Dokumente im normalen Modus. (Die übrigen Crawler außer dem Notes-Crawler geben immer 0 zurück.)	0, 1: <ul style="list-style-type: none"> • 0: Der Notes-Crawler durchsucht Dokumente im normalen Modus. (Die übrigen Crawler außer dem Notes-Crawler geben immer 0 zurück.) • Der Notes-Crawler durchsucht Dokumente im Verzeichnismodus.

Table 13. Elements and Attributes for the Status Command of the Crawler Area for the Exchange Server-Crawler, the DB2 Content Manager-Crawler and the 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.

Tabelle 13. Elemente und Attribute für den Statusbefehl des Crawlerbereichs für den Exchange Server-Crawler, den DB2 Content Manager-Crawler und den Content Edition-Crawler (Forts.)

Element- und Attributname	Exchange Server-Crawler	DB2 Content Manager-Crawler	Content Edition-Crawler
Server/Status	Status (0, 1, 2, 3, 4, -1) <ul style="list-style-type: none"> 0: Keine Crawlersuche ausgeführt 1: Crawlersuche wird ausgeführt 2: Abgeschlossen (nicht terminiert) 3: Wartestatus 4: Angehalten -1: Fehler 	Status (0, 1, 2, 3, 4, 5, -1) <ul style="list-style-type: none"> 0: Keine Crawlersuche ausgeführt 1: Crawlersuche wird ausgeführt 2: Abgeschlossen (nicht terminiert) 3: Wartestatus 4: Angehalten 5: Terminiert -1: Fehler 	Status (0, 1, 2, 3, 4, 5, -1) <ul style="list-style-type: none"> 0: Keine Crawlersuche ausgeführt 1: Crawlersuche wird ausgeführt 2: Abgeschlossen (nicht terminiert) 3: Wartestatus 4: Angehalten 5: Terminiert -1: Fehler
Server/StatusMessage	<ul style="list-style-type: none"> 0: Keine Crawlersuche ausgeführt 1: Crawlersuche wird ausgeführt 2: Abgeschlossen (nicht terminiert) 3: Wartestatus 4: Angehalten 5: Terminiert -1: Fehler 	<ul style="list-style-type: none"> 0: Keine Crawlersuche ausgeführt 1: Crawlersuche wird ausgeführt 2: Abgeschlossen (nicht terminiert) 3: Wartestatus 4: Angehalten 5: Terminiert -1: Fehler 	<ul style="list-style-type: none"> 0: Keine Crawlersuche ausgeführt 1: Crawlersuche wird ausgeführt 2: Abgeschlossen (nicht terminiert) 3: Wartestatus 4: Angehalten 5: Terminiert -1: Fehler
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"> 0: Der Crawler ist entsprechend der Crawlerkonfigurationsdateien nicht für die Terminierung konfiguriert. 1: Der Crawler ist für die Terminierung konfiguriert, aber die Terminierung wurde für die Sitzung inaktiviert. 2: Der Crawler ist für die Terminierung konfiguriert, und die Terminierung ist für die Sitzung aktiviert. 	0, 1, 2 <ul style="list-style-type: none"> 0: Der Crawler ist entsprechend der Crawlerkonfigurationsdateien nicht für die Terminierung konfiguriert. 1: Der Crawler ist für die Terminierung konfiguriert, aber die Terminierung wurde für die Sitzung inaktiviert. 2: Der Crawler ist für die Terminierung konfiguriert, und die Terminierung ist für die Sitzung aktiviert. 	0, 1, 2 <ul style="list-style-type: none"> 0: Der Crawler ist entsprechend der Crawlerkonfigurationsdateien nicht für die Terminierung konfiguriert. 1: Der Crawler ist für die Terminierung konfiguriert, aber die Terminierung wurde für die Sitzung inaktiviert. 2: Der Crawler ist für die Terminierung konfiguriert, und die Terminierung ist für die Sitzung aktiviert.
Server/ScheduleTime	Planzeit, falls zutreffend.	Planzeit, falls zutreffend.	Planzeit, falls zutreffend.

Tabelle 13. Elemente und Attribute für den Statusbefehl des Crawlerbereichs für den Exchange Server-Crawler, den DB2 Content Manager-Crawler und den Content Edition-Crawler (Forts.)

Element- und Attributname	Exchange Server-Crawler	DB2 Content Manager-Crawler	Content Edition-Crawler
Server/TotalTime	Die Gesamtzeit, falls zutreffend.	Die Gesamtzeit, falls zutreffend.	Die Gesamtzeit, falls zutreffend.
Server/AggregationLevel	0: Der Notes-Crawler durchsucht Dokumente im normalen Modus. (Die übrigen Crawler außer dem Notes-Crawler geben immer 0 zurück.)	0: Der Notes-Crawler durchsucht Dokumente im normalen Modus. (Die übrigen Crawler außer dem Notes-Crawler geben immer 0 zurück.)	0: Der Notes-Crawler durchsucht Dokumente im normalen Modus. (Die übrigen Crawler außer dem Notes-Crawler geben immer 0 zurück.)

Tabelle 14. Elemente und Attribute für den Statusbefehl des Crawlerbereichs für den QuickPlace-Crawler, den Domino Document Manager-Crawler sowie für den Crawler für das UNIX- und Windows-Dateisystem.

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, -1) <ul style="list-style-type: none"> 0: Keine Crawlersuche ausgeführt 1: Crawlersuche wird ausgeführt 2: Abgeschlossen (nicht terminiert) 3: Wartestatus 4: Angehalten -1: Fehler 	Status (0, 1, 2, 3, 4, 5, -1) <ul style="list-style-type: none"> 0: Keine Crawlersuche ausgeführt 1: Crawlersuche wird ausgeführt 2: Abgeschlossen (nicht terminiert) 3: Wartestatus 4: Angehalten 5: Terminiert -1: Fehler 	Status (0, 1, 2, 3, 4, 5, -1) <ul style="list-style-type: none"> 0: Keine Crawlersuche ausgeführt 1: Crawlersuche wird ausgeführt 2: Abgeschlossen (nicht terminiert) 3: Wartestatus 4: Angehalten 5: Terminiert -1: Fehler
Server/StatusMessage	<ul style="list-style-type: none"> 0: Keine Crawlersuche ausgeführt 1: Crawlersuche wird ausgeführt 2: Abgeschlossen (nicht terminiert) 3: Wartestatus 4: Angehalten 5: Terminiert -1: Fehler 	<ul style="list-style-type: none"> 0: Keine Crawlersuche ausgeführt 1: Crawlersuche wird ausgeführt 2: Abgeschlossen (nicht terminiert) 3: Wartestatus 4: Angehalten 5: Terminiert -1: Fehler 	<ul style="list-style-type: none"> 0: Keine Crawlersuche ausgeführt 1: Crawlersuche wird ausgeführt 2: Abgeschlossen (nicht terminiert) 3: Wartestatus 4: Angehalten 5: Terminiert -1: Fehler
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.

Tabelle 14. Elemente und Attribute für den Statusbefehl des Crawlerbereichs für den QuickPlace-Crawler, den Domino Document Manager-Crawler sowie für den Crawler für das UNIX- und Windows-Dateisystem. (Forts.)

Element- und Attributname	QuickPlace-Crawler	Domino Document Manager-Crawler	Crawler für das UNIX- und Windows-Dateisystem
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"> 0: Der Crawler ist entsprechend der Crawlerkonfigurationsdateien nicht für die Terminierung konfiguriert. 1: Der Crawler ist für die Terminierung konfiguriert, aber die Terminierung wurde für die Sitzung inaktiviert. 2: Der Crawler ist für die Terminierung konfiguriert, und die Terminierung ist für die Sitzung aktiviert. 	0, 1, 2 <ul style="list-style-type: none"> 0: Der Crawler ist entsprechend der Crawlerkonfigurationsdateien nicht für die Terminierung konfiguriert. 1: Der Crawler ist für die Terminierung konfiguriert, aber die Terminierung wurde für die Sitzung inaktiviert. 2: Der Crawler ist für die Terminierung konfiguriert, und die Terminierung ist für die Sitzung aktiviert. 	0, 1, 2 <ul style="list-style-type: none"> 0: Der Crawler ist entsprechend der Crawlerkonfigurationsdateien nicht für die Terminierung konfiguriert. 1: Der Crawler ist für die Terminierung konfiguriert, aber die Terminierung wurde für die Sitzung inaktiviert. 2: Der Crawler ist für die Terminierung konfiguriert, und die Terminierung ist für die Sitzung aktiviert.
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, 1 <ul style="list-style-type: none"> 0: Der Notes-Crawler durchsucht Dokumente im normalen Modus. (Die übrigen Crawler außer dem Notes-Crawler geben immer 0 zurück.) Der Notes-Crawler durchsucht Dokumente im Verzeichnismodus. 	0, 1 <ul style="list-style-type: none"> 0: Der Notes-Crawler durchsucht Dokumente im normalen Modus. (Die übrigen Crawler außer dem Notes-Crawler geben immer 0 zurück.) Der Notes-Crawler durchsucht Dokumente im Verzeichnismodus. 	0, 1 <ul style="list-style-type: none"> 0: Der Notes-Crawler durchsucht Dokumente im normalen Modus. (Die übrigen Crawler außer dem Notes-Crawler geben immer 0 zurück.) Der Notes-Crawler durchsucht Dokumente im Verzeichnismodus.

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 15. 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

Tabelle 15. Auswahlmaskenwerte für den detaillierten Statusbefehl des Crawlerbereichs für den Web-Crawler (Forts.)

Maskenbit	Auswahl
64	Anzahl durchsuchter URL-Adressen
128	Anzahl nicht durchsuchter URL-Adressen
256	Anzahl überfälliger URL-Adressen
512	Verteilung des HTTP-Rückkehrcodes

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 16. Informationen, die durch den Befehl für den detaillierten Status für den Crawlerbereich des Web-Crawlers zurückgegeben wurden

Element	Attribute	Beschreibung
CrawlDetailsPerSite	<ul style="list-style-type: none"> LastActivationTime: LastActivationDuration: IPAddressList: RobotsContent: HTTPCodeDist: 	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 16. Informationen, die durch den Befehl für den detaillierten Status für den Crawlerbereich des Web-Crawlers zurückgegeben wurden (Forts.)

Element	Attribute	Beschreibung
NumURLsOverdueBy	Threshold, Value: Ganze Zahl (positiv oder negativ) Der Schwellenwert für die Suche nach erneuten Crawlersuchen, die ein angegebenes Zeitlimit (negativ = Sekunden vor dem jetzigen Zeitpunkt) überschritten haben oder die während eines angegebenen Zeitintervalls (positiv = Sekunden nach dem jetzigen Zeitpunkt) bevorstehen.	Die Anzahl URL-Adressen, die zumindest vor einigen Sekunden (Schwellenwert) erneut durchsucht werden konnten oder die in den nächsten Sekunden (Schwellenwert) ausgewählt werden können.
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 Serverhost dieser Site.
IPAddress	Value	IPv4-Adresse des Serverhosts 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-Rückkehrcode oder ein anderer interner Code.	Häufigkeit, in der ein bestimmter HTTP-Rückkehrcode 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 für den detaillierten Status für den Crawlerbereich 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 17. Elemente und Attribute für den Befehl für den detaillierten Status für den Crawlerbereich für den NNTP-Crawler, den DB2-Crawler und den Notes-Crawler

Element- und Attributname	NNTP-Crawler	DB2-Crawler	Notes-Crawler
Target@Name	Name der Newsgroup	Tabellenname	Sicht- oder Ordnername
Target@CrawlType	Nicht zutreffend.	0,1: <ul style="list-style-type: none"> 0: Aktive Crawlersuche (normal) 1: Passive Crawlersuche (DB2-Ereignisveröffentlichung) 	0
Target/Status	Status: (0, 1, 2, 3, 4, -1) <ul style="list-style-type: none"> 0: Keine Crawlersuche ausgeführt 1: Crawlersuche wird ausgeführt 2: Abgeschlossen (nicht terminiert) 3: Wartestatus 4: Angehalten -1: Fehler 	Status: (0, 1, 2, 3, 4, -1) <ul style="list-style-type: none"> 0: Keine Crawlersuche ausgeführt 1: Crawlersuche wird ausgeführt 2: Abgeschlossen (nicht terminiert) 3: Wartestatus 4: Angehalten -1: Fehler 	Status: (0, 1, 2, 3, 4, -1) <ul style="list-style-type: none"> 0: Keine Crawlersuche ausgeführt 1: Crawlersuche wird ausgeführt 2: Abgeschlossen (nicht terminiert) 3: Wartestatus 4: Angehalten -1: Fehler
Target/StatusMessage	<ul style="list-style-type: none"> 0: Keine Crawlersuche ausgeführt 1: Crawlersuche wird ausgeführt 2: Abgeschlossen (nicht terminiert) 3: Wartestatus 4: Angehalten -1: Fehler 	<ul style="list-style-type: none"> 0: Keine Crawlersuche ausgeführt 1: Crawlersuche wird ausgeführt 2: Abgeschlossen (nicht terminiert) 3: Wartestatus 4: Angehalten -1: Fehler 	<ul style="list-style-type: none"> 0: Keine Crawlersuche ausgeführt 1: Crawlersuche wird ausgeführt 2: Abgeschlossen (nicht terminiert) 3: Wartestatus 4: Angehalten -1: Fehler
Target/NumberOfRecords	Die letzte Artikelnummer auf dem Server.	Die Anzahl durchsuchter Datensätze.	Die Anzahl durchsuchter Dokumente.
Target/NumberOfCompletedRecords	Die Anzahl durchsuchter Artikel.	Die Anzahl durchsuchter Datensätze.	Die Anzahl durchsuchter Dokumente.
Target/NumberOfInsertedRecords	Die Anzahl neu veröffentlichter Artikel.	Die Anzahl eingefügter Datensätze.	Die Anzahl eingefügter Datensätze.
Target/NumberOfUpdatedRecords	Nicht zutreffend.	Die Anzahl aktualisierter Datensätze.	Die Anzahl aktualisierter Datensätze.
Target/NumberOfDeletedRecords	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.

Tabelle 17. Elemente und Attribute für den Befehl für den detaillierten Status für den Crawlerbereich für den NNTP-Crawler, den DB2-Crawler und den Notes-Crawler (Forts.)

Element- und Attributname	NNTP-Crawler	DB2-Crawler	Notes-Crawler
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 Notes-Crawler durchsucht Dokumente im normalen Modus. (Die übrigen Crawler außer dem Notes-Crawler geben immer 0 zurück.)	0: Der Notes-Crawler durchsucht Dokumente im normalen Modus. (Die übrigen Crawler außer dem Notes-Crawler geben immer 0 zurück.)	0, 1: <ul style="list-style-type: none"> 0: Der Notes-Crawler durchsucht Dokumente im normalen Modus. (Die übrigen Crawler außer dem Notes-Crawler geben immer 0 zurück.) Der Notes-Crawler durchsucht Dokumente im Verzeichnismodus.
Target/LastUpdatedTime	Nicht zutreffend.	Die Zeit der letzten Aktualisierung: <ul style="list-style-type: none"> 0: Aktive Crawlersuche (normal) 1: Passive Crawlersuche (DB2-Ereignisveröffentlichung) 	Nicht zutreffend.
Target/LastResetTime	Nicht zutreffend.	Statistikdaten zur letzten Zurücksetzung: <ul style="list-style-type: none"> 0: Aktive Crawlersuche (normal) 1: Passive Crawlersuche (DB2-Ereignisveröffentlichung) 	Nicht zutreffend.

Tabelle 18. Elemente und Attribute für den Befehl für den detaillierten Status für den Crawlerbereich für den Exchange Server-Crawler, den DB2 Content Manager-Crawler und den 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"> 0: Keine Crawlersuche ausgeführt 1: Crawlersuche wird ausgeführt 2: Abgeschlossen (nicht terminiert) 3: Wartestatus 4: Angehalten -1: Fehler 	Status (0, 1, 2, 3, 4, -1) <ul style="list-style-type: none"> 0: Keine Crawlersuche ausgeführt 1: Crawlersuche wird ausgeführt 2: Abgeschlossen (nicht terminiert) 3: Wartestatus 4: Angehalten -1: Fehler 	Status (0, 1, 2, 3, 4, -1) <ul style="list-style-type: none"> 0: Keine Crawlersuche ausgeführt 1: Crawlersuche wird ausgeführt 2: Abgeschlossen (nicht terminiert) 3: Wartestatus 4: Angehalten -1: Fehler

Tabelle 18. Elemente und Attribute für den Befehl für den detaillierten Status für den Crawlerbereich für den Exchange Server-Crawler, den DB2 Content Manager-Crawler und den 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"> • 0: Keine Crawlersuche ausgeführt • 1: Crawlersuche wird ausgeführt • 2: Abgeschlossen (nicht terminiert) • 3: Wartestatus • 4: Angehalten • -1: Fehler 	<ul style="list-style-type: none"> • 0: Keine Crawlersuche ausgeführt • 1: Crawlersuche wird ausgeführt • 2: Abgeschlossen (nicht terminiert) • 3: Wartestatus • 4: Angehalten • -1: Fehler 	<ul style="list-style-type: none"> • 0: Keine Crawlersuche ausgeführt • 1: Crawlersuche wird ausgeführt • 2: Abgeschlossen (nicht terminiert) • 3: Wartestatus • 4: Angehalten • -1: Fehler
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 Notes-Crawler durchsucht Dokumente im normalen Modus. (Die übrigen Crawler außer dem Notes-Crawler geben immer 0 zurück.)	0: Der Notes-Crawler durchsucht Dokumente im normalen Modus. (Die übrigen Crawler außer dem Notes-Crawler geben immer 0 zurück.)	0: Der Notes-Crawler durchsucht Dokumente im normalen Modus. (Die übrigen Crawler außer dem Notes-Crawler geben immer 0 zurück.)
Target/LastUpdatedTime	Nicht zutreffend.	Nicht zutreffend.	Nicht zutreffend.
Target/LastResetTime	Nicht zutreffend.	Nicht zutreffend.	Nicht zutreffend.

Tabelle 19. Elemente und Attribute für den Befehl für den detaillierten Status für den Crawlerbereich für den QuickPlace-Crawler, den Domino Document Manager-Crawler sowie für den Crawler für das UNIX- und Windows-Dateisystem.

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 19. Elemente und Attribute für den Befehl für den detaillierten Status für den Crawlerbereich für den QuickPlace-Crawler, den Domino Document Manager-Crawler sowie für den Crawler für das UNIX- und Windows-Dateisystem. (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"> • 0: Keine Crawlersuche ausgeführt • 1: Crawlersuche wird ausgeführt • 2: Abgeschlossen (nicht terminiert) • 3: Wartestatus • 4: Angehalten • -1: Fehler 	Status (0, 1, 2, 3, 4, -1) <ul style="list-style-type: none"> • 0: Keine Crawlersuche ausgeführt • 1: Crawlersuche wird ausgeführt • 2: Abgeschlossen (nicht terminiert) • 3: Wartestatus • 4: Angehalten • -1: Fehler 	Status (0, 1, 2, 3, 4, -1) <ul style="list-style-type: none"> • 0: Keine Crawlersuche ausgeführt • 1: Crawlersuche wird ausgeführt • 2: Abgeschlossen (nicht terminiert) • 3: Wartestatus • 4: Angehalten • -1: Fehler
Target/StatusMessage	<ul style="list-style-type: none"> • 0: Keine Crawlersuche ausgeführt • 1: Crawlersuche wird ausgeführt • 2: Abgeschlossen (nicht terminiert) • 3: Wartestatus • 4: Angehalten • -1: Fehler 	<ul style="list-style-type: none"> • 0: Keine Crawlersuche ausgeführt • 1: Crawlersuche wird ausgeführt • 2: Abgeschlossen (nicht terminiert) • 3: Wartestatus • 4: Angehalten • -1: Fehler 	<ul style="list-style-type: none"> • 0: Keine Crawlersuche ausgeführt • 1: Crawlersuche wird ausgeführt • 2: Abgeschlossen (nicht terminiert) • 3: Wartestatus • 4: Angehalten • -1: Fehler
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 durchsuchter Datenbanken für Bereiche und Räume.	Die Anzahl durchsuchter Aktenschrankdateien.	Die Anzahl Unterverzeichnisse im Crawlerbereich.
Target/NumberOf DeletedRecords	Die Anzahl aktualisierter Datensätze.	Die Anzahl aktualisierter Datensätze.	Die Anzahl aktualisierter 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 Notes-Crawler durchsucht Dokumente im normalen Modus. (Die übrigen Crawler außer dem Notes-Crawler geben immer 0 zurück.)	0: Der Notes-Crawler durchsucht Dokumente im normalen Modus. (Die übrigen Crawler außer dem Notes-Crawler geben immer 0 zurück.)	0: Der Notes-Crawler durchsucht Dokumente im normalen Modus. (Die übrigen Crawler außer dem Notes-Crawler geben immer 0 zurück.)

Tabelle 19. Elemente und Attribute für den Befehl für den detaillierten Status für den Crawlerbereich für den QuickPlace-Crawler, den Domino Document Manager-Crawler sowie für den Crawler für das UNIX- und Windows-Dateisystem. (Forts.)

Element- und Attributname	QuickPlace-Crawler	Domino Document Manager-Crawler	Crawler für das UNIX- und Windows-Dateisystem
Target/LastUpdatedTime	Nicht zutreffend.	Nicht zutreffend.	Nicht zutreffend.
Target/LastResetTime	Nicht zutreffend.	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>
  <SnapshotTimeStamp>1124318637564</SnapshotTimeStamp>
  <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>
  <NumberOfDocsToBeParsed>15881</NumberOfDocsToBeParsed>
  <NumberOfDocsAlreadyParsed>29</NumberOfDocsAlreadyParsed>
</CrawlerStatus>
<CrawlerStatus>
  <Name>Data Listener (server1)</Name>
  <Crawlerid>datalistener</Crawlerid>
  <Type>datalistener</Type>
  <ParserStatus>0</ParserStatus>
  <NumberOfDocsToBeParsed>0</NumberOfDocsToBeParsed>
  <NumberOfDocsAlreadyParsed>0</NumberOfDocsAlreadyParsed>
</CrawlerStatus>
</Monitor>
```

In der folgenden Tabelle werden die XML-Elemente für Informationen beschrieben, die durch den Befehl für den Parserstatus zurückgegeben werden:

Tabelle 20. Elemente für den Befehl für den Parserstatus

Element	Beschreibung
Status	1, falls die Parsersitzung für diese Objektgruppe aktiv ist, und 0, falls die Parsersitzung gestoppt wurde.

Tabelle 20. Elemente für den Befehl für den Parserstatus (Forts.)

Element	Beschreibung
State	<p>Möglicher Status: Idle, Restart, Parsing, Resuming, Stopped, Initializing, Pause, NoParserServiceIsAvailable.</p> <p>Der Status Idle gibt an, dass der Parser für N 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 Restart 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 Parsing gibt an, dass der Parser Dokumente verarbeitet.</p> <p>Der Status Pause gibt an, dass der Parser durch die Indexerstellungssitzung für diese Objektgruppe angehalten wurde. Der Status Initializing gibt an, dass der Parser gerade gestartet wird und seinen Status initialisiert.</p> <p>Der Status Resuming gibt an, dass der Parser durch die Indexerstellungssitzung für diese Objektgruppe aus dem Status Pause in den Status Parsing fortgesetzt wurde.</p> <p>Der Status NoParserServiceIsAvailable gibt an, dass keine JVMs für die Syntexanalyse und das Aufbereiten verfügbar sind, um die Dokumente für diese Objektgruppe zu verarbeiten. Dies bedeutet, dass alle JVMs für die Syntexanalyse und das Aufbereiten von anderen Objektgruppen verwendet werden.</p>
SnapshotTimeStamp	Die nächste Zeit in Sekunden seit 1970, wann der Parser gelesene Dokumente aus den Crawlertabellen abrufen.
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.
ParserServiceSession	Der Name der JVM für die Syntexanalyse und das Aufbereiten, die die Dokumente für diese Objektgruppe verarbeitet. Dieses Feld ist nur verfügbar, wenn sich der Parser im Status Parsing befindet.
Name	Der Name des Crawler.
Type	Der Crawlertyp (Web-, NNTP-, DB2-Crawler usw.)
ParserStatus	<ul style="list-style-type: none"> • 0: Die Dokumente aus diesem Crawler wurden noch nicht syntaktisch analysiert. • 1: Die Dokumente aus diesem Crawler werden zurzeit syntaktisch analysiert. • 2: Die Dokumente aus diesem Crawler wurden syntaktisch analysiert.
NumberOfDocsToBeParsed	Die Anzahl Dokumente aus diesem Crawler, die darauf warten, syntaktisch analysiert zu werden.
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:

```
FFQC5314I Ergebnis: <Monitor Type="Parser">
<Monitor Type="MainIndexHistory" Count="1">
  <IndexStatus Id="1"/>
  <StartTime>1123101789411</StartTime>
  <Progress>0</Progress>
  <CurrentPhase>0</CurrentPhase>
  <TotalPhase>0</TotalPhase>
  <IndexCopyProgress>0</IndexCopyProgress>
  <CurrentServer>0</CurrentServer>
  <TotalServer>0</TotalServer>
  <IndexCopyTime>0</IndexCopyTime>
  <IndexBuildTime>0</IndexBuildTime>
  <Status>0</Status>
  <StopTime>1123101789618</StopTime>
  <NumberOfDocuments>0</NumberOfDocuments>
</IndexStatus>
  <CurrentIndexWildcardSupport/>
  <NextIndexWildcardSupport/>
  <ScheduleStatus>
    <Status>1</Status>
</Monitor>
```

In der folgenden Tabelle wird jedes XML-Element für Informationen beschrieben, die durch den Statusbefehl für die Indexerstellung zurückgegeben werden:

Tabelle 21. Elemente für den Statusbefehl für die Indexerstellung

Element	Beschreibung
IndexStatusId	Die Indexstatus-ID.
StartTime	Die Startzeit in Sekunden seit 1970, als dieser Index erstellt wurde.
Progress	Der Prozentsatz der Fertigstellung dieser Indexerstellung.
CurrentPhase	<ul style="list-style-type: none"> • 1: Phase des erneuten Schreibens im Speicher • 2: Globale Analysephase • 3: Indexstellungsphase
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 alle Phasen der Indexerstellung.
Progress	0 für erfolgreiche Indexerstellung und erfolgreiches Kopieren und eine Zahl ungleich Null für den Code der Fehlermeldung.
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 None , QueryExpansion oder IndexExpansion .

Tabelle 21. Elemente für den Statusbefehl für die Indexerstellung (Forts.)

Element	Beschreibung
ScheduleStatus	<ul style="list-style-type: none"> 0, falls kein Zeitplan für diese Objektgruppe und diesen Indextyp aktiviert ist. 1, falls ein Zeitplan für diese Objektgruppe und diesen Indextyp aktiviert ist.
ScheduledTimeEnabled	Die nächste Zeit in Sekunden seit 1970, wann die Indexerstellung für diese Objektgruppe und diesen Indextyp ausgeführt wird.

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="Suchmanager (knoten1)" SearchID=
"searchmanager.node1" HostName="Arbeitsplatz.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 22. 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"> 0, falls der Suchindex für diese Objektgruppe nicht aktiv ist. 1, falls der Suchindex für diese Objektgruppe aktiv ist.

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 23. Elemente für den Befehl für den detaillierten Status für den Suchserver

Element	Beschreibung
CacheHits	Die Anzahl Ergebnisse, die aus dem Suchcache abgerufen wurden.

Tabelle 23. Elemente für den Befehl für den detaillierten Status für den Suchserver (Forts.)

Element	Beschreibung
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"> • 0, falls der Suchindex für diese Objektgruppe nicht aktiv ist. • 1, falls der Suchindex für diese Objektgruppe aktiv ist.
SessionName	Der Sitzungsname für den Suchindex dieser Objektgruppe.

Rückkehrcodes für die esadmin-Befehle

Die folgenden Codes können für die esadmin-Befehle zurückgegeben werden:

Tabelle 24. 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 auf Grund 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 25. 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

Tabelle 25. Beispiele für Sitzungsnamen, Ursprungsserver, Sitzungs-IDs und Sitzungsstatus (Forts.)

Sitzung	Server, auf dem die Sitzung aktiv ist	Sitzungs-ID	Sitzungsstatus
datalistener	Indexserver	10582	Gestartet
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

„Durchführen von Backup und Restore für ein System für die Unternehmenssuche“ auf Seite 283

Mit Backup- und Restore-Skripts können Sie das System für die Unternehmenssuche sichern und wiederherstellen.

„Überwachen der Aktivitäten der Unternehmenssuche“ auf Seite 249

Wenn Sie System- und Objektgruppenaktivitäten überwachen, können Sie den Status verschiedener Prozesse anzeigen, potenzielle Probleme aufspüren oder Konfigurationseinstellungen anpassen, um die Leistung zu verbessern.

Zugehörige Tasks

„Überwachen von Crawlern“ auf Seite 254

Sie können allgemeine Informationen zum Status der einzelnen Crawler in der Objektgruppe anzeigen oder Optionen auswählen, um detaillierte Informationen zur Crawleraktivität anzuzeigen.

„Starten der Server für die Unternehmenssuche“ auf Seite 243

Damit Benutzer eine Objektgruppe durchsuchen können, müssen Sie die Systemprozesse und anschließend die Server für die Crawleruche, die Syntaxanalyse, die Indexierung und die Suche für die Objektgruppe starten.

„Stoppen der Server für die Unternehmenssuche“ auf Seite 245

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.

Dokumentation zur Unternehmenssuche

Die Dokumentation zu WebSphere Information Integrator OmniFind Edition steht im PDF- oder HTML-Format zur Verfügung.

Das WebSphere Information Integrator OmniFind Edition-Installationsprogramm kann die Installation der Informationszentrale automatisch ausführen. Das Installationsprogramm installiert die Informationszentrale auf dem Suchserver. 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. Starten Sie die Informationszentrale, um HTML-Themen für die Unternehmenssuche anzuzeigen.

Wechseln Sie in das Verzeichnis `docs/ländereinstellung/pdf`, um die PDF-Dokumente anzuzeigen. Wenn Sie beispielsweise Dokumente in englischer Sprache suchen, wechseln Sie in das Verzeichnis `docs/en_US/pdf`. Sie haben auch die Möglichkeit, die PDF-Dokumentation sowie Downloads, Korrekturen, technische Hinweise und die Informationszentrale von der WebSphere Information Integrator OmniFind Edition-Unterstützungssite aus anzuzeigen.

In der folgenden Tabelle ist die verfügbare Dokumentation mit Dateinamen und Speicherposition aufgeführt.

Tabelle 26. PDF-Dokumentation zur Unternehmenssuche

Dokument	Dateiname	Speicherposition
<i>Installationshandbuch für die Unternehmenssuche</i> (Themen zu diesem Dokument stehen auch in der Informationszentrale zur Verfügung)	iiysi.pdf	docs/ländereinstellung/pdf/
<i>Verwaltung der Unternehmenssuche</i> (Themen zu diesem Dokument stehen auch in der Informationszentrale zur Verfügung.)	iiysa.pdf	docs/ländereinstellung/pdf/
<i>Programming Guide and API Reference for Enterprise Search</i> (Themen zu diesem Dokument stehen auch in der Informationszentrale zur Verfügung.)	iiysp.pdf	docs/ländereinstellung/pdf/
<i>Fehlernachrichten</i> (Themen zu diesem Dokument stehen auch in der Informationszentrale zur Verfügung.)	iiysm.pdf	docs/ländereinstellung/pdf/
<i>Installation Requirements for Enterprise Search</i> (Themen zu diesem Dokument stehen auch in der Informationszentrale zur Verfügung.)	iiysr.txt oder iiysr.htm	docs/ländereinstellung/ (Diese Datei kann auch über das Programm First Steps aufgerufen werden.)

|
|
|
|

Tabelle 26. PDF-Dokumentation zur Unternehmenssuche (Forts.)

Dokument	Dateiname	Speicherposition
<i>Release-Informationen</i>	iiysn.pdf	Ist nur auf der IBM WebSphere Information Integrator OmniFind Edition Documentation-Website verfügbar.
<i>Integration der Textanalyse</i>	iiyst.pdf	docs/ländereinstellung/pdf/

WebSphere II OmniFind Edition - Behindertengerechte Bedienung

Die IBM WebSphere Information Integrator OmniFind Edition-Benutzerschnittstellen und die zugehörige Dokumentation sind für behindertengerechte Bedienung geeignet.

Installationsprogramm

Sie können mit Hilfe von Direktaufrufen über die Tastatur im WebSphere II OmniFind Edition-Installationsprogramm navigieren. Die folgende Tabelle enthält eine Beschreibung einiger Direktaufrufe über die Tastatur.

Tabelle 27. 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

Administrationskonsole für die Unternehmenssuche und Informationszentrale

Die Administrationskonsole und die Informationszentrale sind browserbasierte Schnittstellen, die sich über Microsoft Internet Explorer oder Mozilla FireFox anzeigen lassen. 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.

PDF-Dokumentation

Die gesamte Dokumentation zur Unternehmenssuche liegt im PDF-Format vor. Sie können über Adobe Acrobat Version 6.0 auf die PDF-Dokumente zugreifen. Die PDF-Dokumente sind strukturiert und sollten von den meisten Sprachausgabeprogrammen gelesen werden können.

Glossar der Begriffe für die Unternehmenssuche

Dieses Glossar enthält Begriffe, die in den Schnittstellen für die Unternehmenssuche und in der zugehörigen Dokumentation verwendet werden.

Abfrage in natürlicher Sprache (Natural Language Query)

Eine Art der Suche, bei der geschriebene Ausdrücke (z. B. "Wer leitet die Finanzabteilung?") und nicht eine einfache Gruppe von Schlüsselwörtern analysiert 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 (Common Analysis Structure)

Eine Struktur, in der ein Dokument gespeichert wird, das von einer Textanalysesteuerkomponente analysiert wird. Die Informationen werden in einer allgemeinen Analysestruktur in Form von Annotationen und anderen Merkmalstrukturen gespeichert.

Allgemeine Übertragungsschicht (CCL - Common Communication Layer)

Die Kommunikationsinfrastruktur, die die verschiedenen Komponenten (Controller, Parser, Crawler, Indexierungskomponente) von WebSphere Information Integrator OmniFind Edition miteinander verbindet.

Analyseergebnisse (Analysis Results)

Die von Annotatoren erstellten Informationen. Analyseergebnisse, die den von Ihnen gesuchten Informationen entsprechen, werden in eine Datenstruktur geschrieben, die allgemeine Analysestruktur genannt wird.

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 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.

Archiv der Verarbeitungsteuerkomponente (Processing Engine Archive)

Eine komprimierte .pear-Archivdatei, die eine UIMA-Analysesteuerkomponente sowie sämtliche Ressourcen enthält, die zu ihrer Nutzung für eine benutzerdefinierte Analyse in der Unternehmenssuche erforderlich sind.

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.

Ausgangs-URL (Seed URL)

Der Ausgangspunkt einer Crawlersuche.

Bediener (Operator)

Ein Benutzer der Unternehmenssuche, der über die Berechtigung zum Beobachten, Starten und Stoppen von Prozessen auf Objektgruppenebene verfügt.

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.

Berechtigungsnaehweis (Credential)

Während der Authentifizierung zugewiesene detaillierte Informationen, die den Benutzer, gegebenenfalls vorhandene Gruppenzuordnungen und andere sicherheitsrelevante Identitätsattribute beschreiben. Mit Berechtigungsnaehweisen lassen sich eine Vielzahl von Services ausführen, wie z. B. Berechtigung, Prüfung und Delegation.

Bereich (Place)

Ein virtueller Ort, der im Portal angezeigt wird und in dem sich Einzelpersonen sowie Gruppen zur Zusammenarbeit online treffen. In einem Portal 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.

Bereich (Scope)

Eine Gruppe zusammengehöriger URIs, mit denen der Bereich einer Suchanforderung definiert wird.

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 mit Hilfe von Operatoren wie AND, NOT und OR kombiniert werden.

Boostklasse (Boost Class)

Eine Angabe, mit der der relative Rang eines Dokuments in den Suchergebnissen beeinflusst werden kann.

Boostwort (Boost Word)

Ein Wort, mit dem der relevante 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 Datenbanknamen, Dateisystempfaden,

Domänennamen, IP-Adressen und URLs) 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.

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.

Diakritische Zeichen (Diacritics)

Ein Zeichen (z. B. ein Akzent oder die Umlautzeichen in der deutschen Sprache), das einem Buchstaben angefügt wird, mit dem Ziel, die Aussprache eines Wortes zu ändern oder ähnliche Wörter voneinander abzugrenzen.

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 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 zu ermöglichen, die mit Notes-Java-Klassen und mit dem Domino-Server erstellt werden. Browserbenutzer und Domino-Server führen die Kommunikation und den Austausch von Objektdaten über DIIOP aus.

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 WebSphere Information Integrator OmniFind Edition durchsucht, syntaktisch analysiert oder indiziert wird. Das Durchsuchen von externen Datenquellen wird an die Anwendungsprogrammierschnittstelle für die Abfrage dieser Datenquellen delegiert.

Feld (Field)

Der kleinste identifizierbare Teil eines Datensatzes.

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.

Freiformatsuche (Free Text Search)

Eine Suche, in der der Suchbegriff als unformatierter Text dargestellt wird.

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.

Hybridsuche (Hybrid Search)

Eine Kombination aus Boolescher Suche und Freiformatsuche.

Identitätsmanagement (Identity Management)

Eine Funktionalität zur Verschlüsselung von Benutzerberechtigungen in einem sicheren Speicher.

In Warteschlange stellen (Enqueue)

Einträge in eine Warteschlange einfügen.

Index Siehe Volltextindex.**Indexierungswarteschlange (Index Queue)**

Eine Liste von zu verarbeitenden Anforderungen zur Indexreorganisation oder zur Indexaktualisierung.

Informationsextraktion (Information Extraction)

Eine Art der Begriffsextraktion, bei der signifikante Vokabularelemente (z. B. Namen, Begriffe und Ausdrücke) in Textdokumenten automatisch erkannt werden.

IP-Adresse (IP Address)

Eine eindeutige 32-Bit-Adresse, die einen Host im Netz identifiziert.

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.

Java Virtual Machine (JVM)

Softwareimplementierung eines Prozessors, die kompilierten Java-Code (Applets und Anwendungen) ausführt.

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 diese auszuführen, wenn die Seite bereitgestellt wird, um einem Client dynamischen Inhalt zurückzugeben.

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.

Kategorie (Category)

Eine Gruppe von Dokumenten mit ähnlichen Merkmalen.

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 zugegriffen wird.

Lemma

Die kanonische Form eines Worts. Lemmas spielen vor allem in stark flektierten Sprachen wie dem Tschechischen eine wichtige Rolle.

Lexikalische Affinität (Lexical Affinity)

Die Beziehung von Suchbegriffen, die eine ähnliche Bedeutung im Dokument haben. 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 (z. B. das Verbinden von f und i zur Ligatur fi).

Lightweight Directory Access Protocol (LDAP)

Ein offenes Protokoll, das mit Hilfe 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 aufweist.

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 Förderator (Local Federator)

Ein Clientförderator, der eine Gruppe durchsuchbarer Objekte föderiert.

Merkmalstruktur (Feature Structure)

Die zu Grunde 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.

Modellbasierte Kategorie (Model-Based Category)

Eine Taxonomie mit vordefinierten Begriffen, mit der das Thema eines Dokuments bestimmt wird, damit das Dokument zusammen mit Dokumenten ähnlichen Inhalts indexiert und durchsucht werden kann.

N-Gram-Segmentierung (N-Gram Segmentation)

Eine Analysemethode, bei der nicht wie bei der Unicode-basierten Leer-

zeichensegmentierung 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 in einer Webseite, die Roboter (z. B. den Web-Crawler) anweisen, den Links auf diesen Seiten nicht zu folgen.

No-Index-Anweisung (No-Index Directive)

Eine Anweisung in einer Webseite, die Roboter (z. B. den Web-Crawler) anweisen, den Inhalten auf diesen Seiten nicht in den Index einzuschließen.

Notes Remote Procedure Call (NRPC)

Die für die gesamte Notes-zu-Notes-Kommunikation verwendete architekturelle Schicht 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.

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.

Popularitätsrangfolge (Popular Ranking)

Ein Rangfolgetyp, der die vorhandene Rangfolge eines Dokuments gemäß der Popularität des Dokuments ergänzt.

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 einer URI und Schlüsselwörtern bzw. 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)

Der Prozess, mit dem das Lemma eines Worts in einem Wörterverzeichnis lokalisiert wird. Die Lemmatisierung unterscheidet sich von der Stammbildung darin, dass die Stammbildung algorithmisch ist und in der Regel nicht auf ein Verzeichnis zurückgreift, in dem die Wörter einer Sprache aufgelistet sind.

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 entsprechen, bestimmten Kategorien zugeordnet werden.

Schlüsselspeicherdatei (Keystore File)

Eine Schlüsseldatei mit öffentlichen Schlüsseln, die als Unterzeichnerzertifikat gespeichert werden, und mit privaten Schlüsseln, die als persönliches Zertifikat gespeichert werden.

Secure Sockets Layer (SSL)

Ein Sicherheitsprotokoll zur Gewährleistung von Datenschutz bei der Kommunikation.

Segmentierung (Segmentation)

Ein Prozess, bei dem die Pfadsteuerung Nachrichtenelemente (Basic Information Units) in kleinere Einheiten, so genannte BIU-Segmente aufteilt, damit diese auch von kleineren Puffergrößen in benachbarten Servern verarbeitet werden können.

Seite mit detaillierten Fehlerhinweisen (Soft Error Page)

Eine spezielle Seite, die eine detaillierte Erläuterung zu einem Fehler liefert, wenn ein HTTP-Server die von einem Client angeforderte Seite nicht zurückgeben kann, und den HTTP-Server so konfiguriert, dass dieser diese Seiten an Stelle einer Antwort zurückgibt, die lediglich aus einem Header mit einem Rückkehrcode besteht, der den Fehler näher angibt.

Servlet

Ein Java-Programm, das auf einem Web-Server ausgeführt wird und die Funktionalität des Servers erweitert, indem es auf Grund von Web-Client-Anforderungen dynamischen Inhalt generiert. Servlets werden gewöhnlich verwendet, für Datenbanken eine Verbindung zum Web herzustellen.

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.

Sprache XPath (XML Path) (XML Path Language (XPath))

Eine Sprache, die Teile eines XML-Quelldokuments eindeutig angibt oder adressiert. XPath bietet außerdem Basisfunktionen zur Bearbeitung von Zeichenfolgen, Zahlen und Booleschen Operatoren.

Spracherkennung (Language Identification)

Eine Funktion der Unternehmenssuche, die die Sprache eines Dokuments bestimmt.

Stammbildung (Stemming)

Siehe Wortstammbildung.

Stoppwort (Stop Word)

Ein häufig verwendetes Wort (z. B. *der, ein, und*), das von einer Suchanwendung ignoriert wird.

Stoppwortentfernung (Stop Word Removal)

Das Entfernen von Stoppwörtern aus der Abfrage, damit allgemeine Wörter ignoriert und auf diese Weise gezieltere Ergebnisse zurückgegeben werden.

Suchanwendung (Search Application)

Ein Programm, das Abfragen verarbeitet, den Index durchsucht, die Suchergebnisse zurückgibt und die Quelldokumente für Objektgruppen in einem System für die Unternehmenssuche 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 beigegeben 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.

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.

Token Die Basistexteinheiten, die von der Unternehmenssuche indexiert werden.

Tokens 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.

Überwachungsbeauftragter (Monitor)

Ein Benutzer der Unternehmenssuche, der über die Berechtigung zum Beobachten von Prozessen auf Objektgruppenebene verfügt.

Unicode-basierte Leerzeichensegmentierung (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)

Eine Zeichenfolge, die Informationsquellen auf einem Computer oder in einem Netz wie dem Internet darstellt. Diese Zeichenfolge enthält den abgekürzten Namen des Protokolls, mit dem auf die Informationsquelle zugegriffen wird, sowie die Informationen, mit denen das Protokoll die Informationsquelle lokalisiert.

Universal Resource Name (URN)

Ein aus einer kurzen Zeichenfolge bestehendes Element des Internetprotokolls, das einer bestimmten Syntax folgt. Die Zeichenfolge umfasst einen Namen oder eine Adresse, mit der auf eine Ressource verwiesen werden kann.

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 Art der Suche, bei der nach bestimmten Wörtern im selben Satz, Absatz oder Dokument gesucht wird.

Verwaltungsrolle (Administrative Role)

Die Klassifizierung eines Benutzers, die die Funktionen festlegt, die dieser Benutzer über die Administrationskonsole für die Unternehmenssuche ausführen kann. Die Rolle legt außerdem fest, welche Objektgruppen der Benutzer verwalten kann.

Web-Crawler (Web Crawler)

Eine Robotsoftwareklasse, die das Web durchsucht, indem sie ein Webdokument abrufen und den Links in diesem Dokument folgt.

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. Bei manchen Systemen sind mehrere Zeichen erforderlich.

Zertifikat (Certificate)

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.

Zertifizierungsstelle (Certificate Authority)

Eine Organisation, die Zertifikate ausstellt und die an elektronischen Transaktionen beteiligten Entitäten (Einzelpersonen bzw. Unternehmen) authentifiziert. Zertifizierungsstellen gewährleisten, dass die beiden Informationen austauschenden Parteien, auch tatsächlich sind, wer sie vorgeben zu sein.

Zugriffssteuerungsliste (Access Control List)

Eine Liste, in der die Benutzer, die auf das zugeordnete Objekt Zugriff haben, und deren Zugriffsberechtigungen für dieses Objekt angegeben sind.

Zusammenfassung (Summarization)

Das Einfügen von Sätzen in Suchergebnisse, die den Inhalt eines Dokuments kurz beschreiben. Siehe auch dynamische Zusammenfassung und statische Zusammenfassung.

Zugreifen auf Informationen zu WebSphere Information Integration

Informationen zu WebSphere Information Integration-Produkten sind telefonisch oder über das Web verfügbar.

Die hier angegebenen Telefonnummern gelten für Deutschland:

- Unter 0180 3 313233 erreichen Sie Hallo IBM, wo Sie Antworten zu allgemeinen Fragen erhalten.
- Unter 0180 5 5090 können Sie Handbücher telefonisch bestellen.

Informationen zu WebSphere Information Integration finden Sie auch im Web unter www.ibm.com/software/data/integration/db2ii/. Diese Site umfasst die folgenden aktuellen Informationen:

- Produktdokumentation
- Produktdownloads
- Fixpacks
- Release-Informationen und weitere Unterstützungsdokumentation
- Neuerungen zu WebSphere Information Integration
- Links zu Webressourcen wie White Papers und IBM Redbooks
- Links zu Newsgroups und Benutzergruppen
- Links zu Onlineinformationszentralen für WebSphere Information Integration-Produkte
- Bestellen von Handbüchern

Gehen Sie für den Zugriff auf Produktdokumentation wie folgt vor:

1. Rufen Sie die Website unter www.ibm.com/software/data/integration/db2ii/ auf.
2. Wählen Sie ein Produkt aus der Dropdown-Liste aus, zum Beispiel WebSphere Information Integrator OmniFind Edition.
3. Klicken Sie den Link **Support** links auf der Seite an.
4. Wählen Sie im Abschnitt **Learn** den gewünschten Link aus. Steht für das ausgewählte Produkt eine Informationszentrale zur Verfügung, können Sie den Link für die Informationszentrale auswählen. (Beispiel siehe Abb. 2 auf Seite 332)

Learn

- **Product documentation and manuals** (2 items)
- **Redbooks** (1 item)
- **V8.2 Documentation and release notes**

Information Center

Provides fast, online centralized access to product information.

- 1.0

Abbildung 2. Beispiel für Links zur Produktdokumentation auf einer WebSphere Information Integration-Unterstützungswebsite

Kommentare zur Dokumentation

Bitte senden Sie uns Ihre Kommentare zu diesen Informationen oder zu anderer Dokumentation von IBM WebSphere Information Integration.

Ihre Rückmeldung unterstützt IBM, hochwertige Informationen anzubieten. Bitte senden Sie uns Ihre Kommentare zu diesen Informationen oder zu anderer Dokumentation von WebSphere Information Integration. Zum Senden von Kommentaren stehen Ihnen die folgenden Möglichkeiten zur Verfügung:

1. Senden Sie Ihre Kommentare mit Hilfe des Kommentarformulars für Onlinedokumentation unter www.ibm.com/software/awdtools/rcf/.
2. Senden Sie Ihre Kommentare als E-Mail an comments@us.ibm.com. Geben Sie den Namen des Produkts, die Versionsnummer des Produkts sowie den Namen und die Teilenummer der Informationen (falls vorhanden) an. Wenn Sie Kommentare zu bestimmtem Text haben, geben Sie die Position des Texts (z. B. einen Titel, eine Tabellenummer oder eine Seitenzahl) an.

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 0190 7 72243 erreichen Sie die DB2 Helpline, wo Sie Antworten zu DB2-spezifischen Problemen erhalten.

Informationen zur nächsten IBM Niederlassung in Ihrem Land oder Ihrer Region finden Sie im IBM Verzeichnis für weltweite Kontakte, das Sie im Web unter www.ibm.com/planetwide abrufen können.

Marken

In diesem Abschnitt werden IBM Marken und bestimmte Marken anderer Hersteller aufgelistet.

Informationen zu IBM Marken finden Sie in <http://www.ibm.com/legal/copytrade.shtml>.

Die folgenden Begriffe sind Marken oder eingetragene Marken anderer Unternehmen:

Java und alle Java-basierten Marken und Logos sind in gewissen Ländern Marken oder eingetragene Marken von Sun Microsystems, Inc.

Microsoft, Windows, Windows NT und das Windows-Logo sind in gewissen Ländern Marken der Microsoft Corporation.

Intel, Intel Inside (Logos), MMX und Pentium sind in gewissen Ländern Marken der Intel Corporation.

UNIX ist in gewissen Ländern eine eingetragene Marke von The Open Group.

Linux ist in gewissen Ländern eine Marke von Linus Torvalds.

Andere Namen von Unternehmen, Produkten oder Services können Marken oder Servicemarken anderer Unternehmen sein.

Bemerkungen

Diese Informationen wurden für Produkte und Services entwickelt, die in Deutschland angeboten werden. Möglicherweise bietet IBM die in dieser Dokumentation beschriebenen Produkte, Services oder Funktionen nicht in allen Ländern 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 Europe, Director of Licensing, 92066 Paris La Defense Cedex, 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 unabhängigen, 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
U.S.A.

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 Handbuch aufgeführten Lizenzprogramms sowie des zugehörigen Lizenzmaterials erfolgt im Rahmen der Allgemeinen Geschäftsbedingungen der IBM, der Internationalen Nutzungsbedingungen der IBM für Programmpakete oder einer äquivalenten Vereinbarung.

Alle in diesem Dokument enthaltenen Leistungsdaten stammen aus einer gesteuerten 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.

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, Anwendungsprogramme zu entwickeln, verwenden, 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:

Outside In (®) Viewer Technology, © 1992-2005 Stellent, Chicago, IL., Inc. Alle Rechte vorbehalten.

IBM XSLT-Prozessor Lizenziertes Material - Eigentum der IBM © Copyright IBM Corporation, 1999-2005. Alle Rechte vorbehalten.

Index

A

- Abfragen mit vielen Ergebnissen
 - Beschreibung 177
 - Standardboostfaktoren 181
- Abfragen mit wenigen Ergebnissen
 - Beschreibung 177
 - Standardboostfaktoren 181
- Abfrageprüfung 216
- Adapter für die Unternehmenssuche 231
- Administrationskonsole
 - anmelden 19
 - Beschreibung 9
 - Schnittstelle 15
 - Taskübersicht 15
- Administrator für die Unternehmenssuche
 - Beschreibung 206
 - Kennwort auf einem Einzelserver ändern 20
 - Kennwort auf mehreren Servern ändern 22
 - Rollenkonfiguration 208
- Administratorkennwort
 - auf einem Einzelserver ändern 20
 - auf mehreren Servern ändern 22
- Adressenregeln für Web-Crawler 80
- AIX, Betriebssystem
 - Content Edition-Crawler, Konfiguration 45
 - DB2 Content Manager-Crawler, Konfiguration 57
 - Domino Document Manager-Crawler, Konfiguration 67
 - Ereignisveröffentlichung, Konfiguration 54
 - Notes-Crawler, Konfiguration 67
 - QuickPlace-Crawler, Konfiguration 67
- Aktive Websites überwachen 256, 257
- Aktivieren der Indexzeitpläne 145
- Alerts
 - Beschreibung 273
 - durchsuchte Dokumente 274
 - E-Mail-Empfang 279
 - E-Mail-Optionen 274, 276
 - freier Speicherbereich auf Servern 276
 - Indexbegrenzungen 250
 - indexierte Dokumente 274
 - Objektgruppenebene 250, 274
 - SMTP-Server, Konfiguration 278
 - Suchantwortzeit 274
 - Systemebene 276
- Alerts zu freiem Speicherbereich 276
- Allgemeine Analysestrukturen
 - Beschreibung 124
 - dem Index zuordnen 129
 - JDBC-Tabellen zuordnen 130
 - XML-Elemente zuordnen 128
- Ankertextanalyse
 - Beschreibung 205

- Ankertextanalyse (*Forts.*)
 - Dokumente indexieren 213
 - globale Analyse 212
 - Objektgruppensicherheit 212
- Anmelden an der Administrationskonsole 19
- Annotatoren 124
- Antwortzeitprotokoll überwachen 267
- Anzeigen
 - Protokolldateien 281
 - URI-Details 252
- APIs
 - Beschreibung 11
 - Datenempfangsfunktion 11
 - Search and Index 11, 183
- Ausgeblendete Suchergebnisse
 - Beschreibung 155
 - konfigurieren 156
- Ausgeblendete URIs
 - Beschreibung 155
 - konfigurieren 156
- Authentifizierung
 - Beschreibung 208
 - für Unternehmensanwendungen inaktivieren 210

B

- Backup-Scripts
 - ausführen 284, 285
 - Beschreibung 283
- Bearbeiten
 - Anwendungen für Datenempfangsfunktion 100
 - Crawlerbereiche 41
 - Crawlermerkmale 41
 - Objektgruppen 33
 - Suchanwendungsmerkmale 186, 195
- Bediener
 - Beschreibung 206
 - Rollenkonfiguration 208
- Befehle, Unternehmenssuche 287
- Behindertengerechte Bedienung 319
- Benutzeragenten 78
- Benutzerdefinierte Textanalyse
 - allgemeine Analysestrukturen zuordnen 129, 130
 - Analyseergebnisse dem Index zuordnen 129
 - Analyseergebnisse JDBC-Tabellen zuordnen 130
 - Beschreibung 124
 - Textanalysesteuerkomponenten 126, 127
 - XML-Elemente zuordnen 128
- Benutzerprofile
 - Beschreibung 217
 - Konfiguration 218
- Bereiche
 - Beschreibung 153
 - durchsuchen 153

- Bereiche (*Forts.*)
 - erstellen 154
 - URI-Formate 103
 - URI-Muster 153, 154

- Boostfaktoren
 - Boostklassenkonfiguration 177, 180
 - für Boostklassen 179, 181
 - für URI-Muster 175, 176
 - für Verzeichnisse von Boostwörtern 172
- Boostklassen
 - Abfragen mit vielen Ergebnissen 177, 181
 - Abfragen mit wenigen Ergebnissen 177, 181
 - Beschreibung 177
 - Erkennung von doppelten Dokumenten 177
 - Felder zuordnen 179
 - Konfiguration 179, 180
 - Standardwerte 181
- bos.iocp.rte, Modul 71

C

- CCLServer_date.log, Datei 26
- Chinesisch
 - N-Gram-Segmentierung 134
 - Syntaxanalyse von Zeilenvorschubzeichen 134
- config.properties, Datei
 - bearbeiten 195
 - Merkmalbeschreibungen 186
- Content Edition-Crawler
 - Direktmodus 44
 - in der Solaris-Betriebsumgebung konfigurieren 45
 - Konfiguration 43
 - Servermodus 44
 - unter dem Betriebssystem AIX konfigurieren 45
 - unter dem Betriebssystem Linux konfigurieren 45
 - unter Windows konfigurieren 46
 - URI-Formate 103
- Cookies für Web-Crawler-Suche
 - Beschreibung 91
 - Format 91
 - konfigurieren 92
- cookies.ini, Datei
 - Beschreibung 91
 - Format 91
 - konfigurieren 92
- crawl.rules, Datei 92
- Crawler
 - Anfangswerte 40
 - Anwendungen für Datenempfangsfunktion 100
 - Basiswerte 37
 - Beschreibung 4
 - Content Edition 43, 44

- Crawler (*Forts.*)
 - Crawlerbereiche bearbeiten 41
 - Crawlermerkmale bearbeiten 41
 - Crawlertypen kombinieren 37
 - DB2 47
 - DB2 Content Manager 56
 - Domino Document Manager 60
 - erstellen 40
 - Exchange Server 61, 62
 - Konfigurationsüberblick 37
 - löschen 42
 - NNTP 63
 - Notes 64, 66
 - Plug-ins 101
 - QuickPlace 73
 - Sicherheit auf Dokumentebene 214
 - Sicherheit auf Dokumentebene aktivieren 37
 - Standardmigrationseinstellungen 239
 - Systemstatus 254
 - terminieren 37
 - überwachen 254
 - UNIX-Dateisystem 77
 - Unterstützung für externe 11, 100
 - URI-Formate 103
 - Web 77
 - WebSphere Portal 96, 97, 98
 - Windows-Dateisystem 99
- Crawler, Beschreibung der Seite 15
- Crawler für das UNIX-Dateisystem
 - Konfiguration 77
 - URI-Formate 103
- Crawler für das Windows-Dateisystem
 - Konfiguration 99
 - Konfiguration der Sicherheit auf Dokumentebene 219
 - URI-Formate 103
- Crawler-Plug-ins 101
- Crawler-Server
 - starten 243, 254
 - stoppen 245, 254
- Crawlerbereich
 - Alerts 274
 - bearbeiten 41
 - Beschreibung 4
 - Web-Crawler-Konfiguration 80
- Crawlermerkmale
 - bearbeiten 41
 - Beschreibung 4
- Crawlertypen
 - Basiswerte 37
 - in Objektgruppe kombinieren 37
- Crawlerverlaufsbericht
 - Beschreibung 256
 - erstellen 259
 - HTTP-Rückkehrcodebericht 259
 - Sitebericht 259

D

- Dateierweiterungen
 - aus Web-Crawler-Bereichen ausschließen 80
 - von Objektgruppenparsern unterstützt 136, 137
 - von Stellant-Sitzungen unterstützt 138, 140

- Datenempfangsfunktion
 - erneut starten 100, 268
 - konfigurieren 100
 - überwachen 268
- Datenempfangsfunktions-API 11
- Datenfluss, System für die Unternehmenssuche 13
- Datenquellentypen
 - Content Edition-Repositorys 43, 44
 - DB2 Content Manager-Elementtypen 56
 - DB2-Datenbanken 47
 - Domino Document Manager-Datenbanken 60
 - Exchange Server, allgemein zugängliche Ordner 61
 - NNTP-Newsgruppen 63
 - Notes-Datenbanken 64, 66
 - QuickPlace-Datenbanken 73
 - relationale Datenbanken 47
 - UNIX-Dateisysteme 77
 - Unterstützung für externe 2, 11
 - von der Unternehmenssuche unterstützt 2
 - Websites 77
 - WebSphere-Portal-Sites 96
 - Windows-Dateisysteme 99
- DB2 Content Manager-Crawler
 - in der Solaris-Betriebsumgebung konfigurieren 57
 - Konfiguration 56
 - unter dem Betriebssystem AIX konfigurieren 57
 - unter dem Betriebssystem Linux konfigurieren 57
 - unter Windows konfigurieren 58
 - URI-Formate 103
- DB2-Crawler
 - Ereignisveröffentlichung 47
 - Ereignisveröffentlichung, Konfiguration 54, 55
 - Konfiguration 47
 - URI-Formate 103
 - WebSphere II Event Publisher Edition-Konfiguration 49
 - WebSphere MQ-Installation 54
 - WebSphere MQ-Installation unter Windows 55
 - WebSphere MQ-Konfiguration 52
- DF-Dokumentation 317
- Directory Assistance, Konfiguration 76
- Direktmodus, Content Edition-Repositorys 44
- Discovery 4
- Dokumentation 317
- Dokumentinhaltoptionen 148
- Dokumentrang
 - Boostklassen 177, 181
 - für Objektgruppe aktivieren 31
 - in migrierten Objektgruppen 237
 - statisch 172
 - URI-Muster 175, 176
 - Verzeichnisse von Boostwörtern 175
- Dokumenttypen
 - für Parser-Services 136, 137
 - für Stellant-Sitzungen 138, 140
 - Syntaxanalyse 135

- Dokumentüberwachung
 - aktivieren 270
 - Berichte 270
 - Beschreibung 269
 - inaktivieren 270
 - Protokolldateikonfiguration 270
- Dokumentzusammenfassungen
 - anpassen 166
 - Merkmale bearbeiten 167
- Domänenregeln für Web-Crawler 80
- Domino-Benutzerkonfiguration, QuickPlace-Crawler 75
- Domino Document Manager-Crawler
 - in der Solaris-Betriebsumgebung konfigurieren 67
 - IOCP-Konfiguration 71
 - Konfiguration 60
 - Konfiguration für Protokoll DIIOP 70
 - Protokoll NRPC 67, 69
 - unter dem Betriebssystem AIX konfigurieren 67
 - unter dem Betriebssystem Linux konfigurieren 67
 - unter Windows konfigurieren 69
 - URI-Formate 103
- Durchsuchen
 - externe Quellen 148
 - Objektgruppen 148
 - Quick Links 168
- Dynamische Rangfolge 171
- Dynamische Zusammenfassung 166

E

- E-Mail-Benachrichtigungen
 - Alerts 279
 - Nachrichten 279
 - SMTP-Server, Konfiguration 278
- EAR-Dateien
 - ESAdmin, Anwendung 210
 - ESPACServer.ear 97
 - ESSearchAdapter.ear 231
 - ESSearchApplication, Anwendung 210
 - ESSearchServer, Anwendung 210
- Ereignisveröffentlichung
 - Beschreibung 47
 - DB2-Crawler, Konfiguration 49, 52
 - 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 205
 - Boostklassenkonfiguration 177
 - globale Analyse 211
 - Sicherheit aktivieren 211
- Erstellen
 - Bereiche 154
 - Crawler 40
 - HTML-Suchfelder 123
 - Objektgruppen 30, 31
 - Quick Links 168

- Erstellen (*Forts.*)
 - regelbasierte Kategorien 118
 - Web-Crawler-Berichte 259
 - XML-Suchfelder 121
- es.cfg, Datei 20, 22
- ES_INSTALL_ROOT, Beschreibung 20, 22
- ES_NODE_ROOT, Beschreibung 20, 22
- es_special_field.default_field 181
- es_special_field.regular_text, Feld 181
- ESAdmin, Anwendung
 - anmelden 19
 - Sicherheit inaktivieren 210
- esadmin, Befehl 287
- esbackup.bat, Script 284, 285
- esbackup.sh, Script 284, 285
- eschangepw, Script 20, 22
- escrcm.sh, Script 57
- escrcm.vbs, Script 58
- escrdb2.sh, Script 54
- escrdb2.vbs, Script 55
- escrnote.sh, Script 67
- escrnote.vbs, Script 69
- escrvbr.sh, Script 45
- escrvbr.vbs, Script 46
- ESPACServer.ear, Datei 97
- esrestore.bat, Script 284, 285
- esrestore.sh, Script 284, 285
- ESSearchAdapter.ear, Datei 231
- ESSearchAdapterPortlet.war, Datei 232
- ESSearchApplication, Anwendung
 - config.properties, Datei 186, 195
 - erneut starten 195, 197
 - Sicherheit aktivieren 197
 - Sicherheit inaktivieren 210
- ESSearchPortlet.war, Datei 228
- ESSearchServer, Anwendung
 - Sicherheit inaktivieren 210
- Exchange Server-Crawler
 - Konfiguration 61
 - sichere Dokumente 62
 - URI-Formate 103
- Externe Crawler
 - Anwendungen für Datenempfangsfunktion 100
 - Datenempfangsfunktions-API 11
 - konfigurieren 100
- Externe Quellen
 - Beschreibung 201
 - durchsuchen 148
 - konfigurieren 201
 - Suchanwendungen zuordnen 202
 - Suchanwendungs-ID, Sicherheit 214

F

- Fehlernachrichten
 - E-Mail-Empfang 277, 279
 - Protokolldateien anzeigen 281
 - SMTP-Server, Konfiguration 278
- Felder zu Boostklassen zuordnen 179
- Feldspezifische Suche 148
- Firewalls, Exchange Server-Dokumente
 - durchsuchen 62
- Föderierte Objektgruppen 29
- followindex.rules, Datei
 - Beschreibung 94

- followindex.rules, Datei (*Forts.*)
 - konfigurieren 95
- Formularbasierte Authentifizierung 87, 89
- Freiformatsuche 148

G

- Gelöschte Dokumente
 - Berichte 270
 - Beschreibung 269
 - Protokolldateikonfiguration 270
- Gleichzeitig ablaufende Indexerstellung 146
- global.rules, Datei 92
- Globale Analyse
 - Ankertextanalyse 205, 212
 - Beschreibung 7
 - Erkennung von doppelten Dokumenten 205, 211
- Globale WebSphere-Sicherheit
 - inaktivieren 210
 - Suchanwendungsmerkmale 195
- Globaler Web-Crawler-Bereich 92

H

- Häufige Abfragen überwachen 267
- HTML-Dokumente suchen 123
- HTML-Suchfelder
 - Beschreibung 123
 - Elemente zuordnen 123
 - erstellen 123
- HTTP-Basisauthentifizierung 87, 88
- HTTP-Proxy-Server 90
- HTTP-Rückkehrcodes
 - von Web-Crawlern empfangen 259
 - Web-Crawler-Bericht 259

I

- I/O Completion Port-Modul, Crawlerkonfiguration 71
- Identitätsmanagement
 - Benutzerprofile 217
 - Konfiguration 218
- Implementierung
 - Adapter für die Unternehmenssuche 231
 - ESPACServer.ear, Datei 97
 - ESSearchAdapter.ear, Datei 231
 - ESSearchAdapterPortlet.war, Datei 232
 - ESSearchPortlet.war, Datei 228
 - Registrierungsportlet 232
 - Suchportlet 228, 229
- Inaktivieren der Indexzeitpläne 145
- Index, Beschreibung der Seite 15
- Indexaktualisierung
 - Beschreibung 7, 143
 - terminieren 144, 145
- Indexerstellungen
 - gleichzeitig ablaufend 146
 - parallel 146
 - starten 265
 - stoppen 265, 266

- Indexerstellungen (*Forts.*)
 - Systemstatus 266
- Indexierungswarteschlange 266
- Indexreorganisation
 - Beschreibung 7, 143
 - terminieren 144, 145
- Indexserver
 - starten 243
 - stoppen 245
- Indizes
 - Alerts 274
 - Ankertext 213
 - aus der Warteschlange löschen 266
 - ausgeblendete URIs 147, 155, 156
 - Bereiche 147, 153
 - Beschreibung 7, 143
 - gleichzeitig ablaufende Erstellungen 146
 - parallele Erstellungen 146
 - Platzhalterzeichen 147, 150, 152
 - terminieren 144
 - überwachen 265, 266
 - URI-Formate 103
 - URIs entfernen 147, 157
 - Zeitplan aktivieren 145, 265
 - Zeitplan inaktivieren 145, 265
 - Zeitplanänderung 145
- Indizes aktualisieren 7, 143
- Indizes reorganisieren 7, 143
- Intervalle für Web-Crawler für erneute Suche 85
- IOCP, Crawlerkonfiguration 71
- IP-Adressenregeln für Web-Crawler 80

J

- Japanisch
 - N-Gram-Segmentierung 134
 - Syntaxanalyse von Zeilenvorschubzeichen 134
- Java-Connector für DB2 Content Manager 57, 58
- JDBC, externe Quellen
 - bearbeiten 201
 - JDBC-Treiber 201
 - konfigurieren 201
 - löschen 201

K

- Kategoriebäume
 - aus WebSphere Portal migrieren 235, 237
 - Beschreibung 116
- Kategorien
 - aus WebSphere Portal migrieren 235, 237
 - Beschreibung 112
 - erstellen 118
 - Kategoriebäume 116
 - Kategorisierungstyp 117
 - modellbasiert 116, 117
 - regelbasiert 113, 117
 - suchen 112
 - Unterkategorien verschachteln 116
 - URI-Formate 103

- Kategorisierungsregeln
 - Dokumentinhalt 113, 118
 - konfigurieren 118
 - URI-Muster 113, 118
- Kategorisierungstyp
 - auswählen 31, 117
 - modellbasiert 116
 - regelbasiert 113
- Kennwort, Administrator für die Unternehmenssuche 20, 22
- Kennwortgeschützte Websites
 - formularbasierte Authentifizierung 89
 - HTTP-Basisauthentifizierung 88
- Koreanisch
 - N-Gram-Segmentierung 134
 - zusammengesetzte Begriffe, Analyse 132

L

- LDAP, externe Quellen
 - bearbeiten 201
 - konfigurieren 201
 - löschen 201
- Letzte Abfragen überwachen 267
- Linguistische Unterstützung
 - benutzerdefinierte Textanalyse 124
 - native XML-Suche 133
 - semantische Suche 124, 133
 - Synonymverzeichnisse 161
 - Verzeichnisse von Boostwörtern 172
 - Verzeichnisse von Stoppwörtern 163
- Linux, Betriebssystem
 - Content Edition-Crawler, Konfiguration 45
 - DB2 Content Manager-Crawler, Konfiguration 57
 - Domino Document Manager-Crawler, Konfiguration 67
 - Ereignisveröffentlichung, Konfiguration 54
 - Notes-Crawler, Konfiguration 67
 - QuickPlace-Crawler, Konfiguration 67
 - Solaris-Betriebsumgebung
 - Ereignisveröffentlichung, Konfiguration 54
- Löschen
 - Crawler 42
 - Indizes aus der Warteschlange 266
 - Objektgruppen 34
- Lotus Domino-Domänen 221, 222
- Lotus Domino-Trusted Server 222

M

- Maximales Intervall für erneute Suche 85
- Migration
 - modellbasierte Taxonomie 235
 - Objektgruppen 237
 - regelbasierte Taxonomie 237
- Migrationsassistent
 - Beschreibung 235
 - modellbasierte Taxonomien 235

- Migrationsassistent (*Forts.*)
 - Objektgruppen 237
 - Protokolldatei 241
 - regelbasierte Taxonomien 237
 - Standardcrawlereinstellungen 239
 - Standardobjektgruppeneinstellungen 239
 - starten 235, 237
- MIME-Typen in Web-Crawler-Bereiche einschließen 80
- Minimales Intervall für erneute Suche 85
- Modellbasierte Kategorien
 - Beschreibung 116
 - Kategorisierungstyp auswählen 117
- Modellbasierte Taxonomie aus WebSphere Portal migrieren 235
- Mustersuchanwendung
 - Beschreibung 12, 184
 - config.properties, Datei 186, 195
 - globale WebSphere-Sicherheit 197
 - Sicherheit aktivieren 197
 - Sicherheit inaktivieren 210
 - Standardimplementierung 195
 - starten 196
 - Suchfunktionen 183, 184
 - zugreifen 196

N

- N-Gram-Segmentierung 134
- Native XML-Suche 133
- NNTP-Crawler konfigurieren 63
- no-follow, Anweisung
 - Beschreibung 94
 - konfigurieren 95
- no-index, Anweisung
 - Beschreibung 94
 - konfigurieren 95
- Notes-Crawler
 - Feldzuordnungsregeln 66
 - in der Solaris-Betriebsumgebung konfigurieren 67
 - IOCP-Konfiguration 71
 - Konfiguration 64
 - Konfiguration der Sicherheit auf Dokumentenebene 221
 - Konfiguration für Protokoll DIIOP 70
 - Lotus Domino-Trusted Server 222
 - Protokoll NRPC 67, 69
 - Prüfung aktueller Berechtigungs-nachweise 222
 - Tipps zur Nutzung 66
 - unter dem Betriebssystem AIX konfigurieren 67
 - unter dem Betriebssystem Linux konfigurieren 67
 - unter Windows konfigurieren 69
 - URI-Formate 103

O

- Objektgruppen
 - Ankertextsicherheit 212

- Objektgruppen (*Forts.*)
 - aus WebSphere Portal migrieren 235, 237
 - bearbeiten 33
 - Beschreibung 3
 - durchsuchen 148
 - Entwurf 30
 - Erstellungsmöglichkeiten 29
 - Föderation von Datenquellen 29
 - Größe schätzen 31
 - löschen 34
 - MigrationWizard.log, Datei 237
 - mit dem Objektgruppenassistenten erstellen 30
 - mit der Sicht 'Objektgruppen' erstellen 31
 - Ressourcen schätzen 250
 - Sicherheit 211
 - Sicherheit für doppelte Dokumente 211
 - Standardmigrationseinstellungen 239
 - Suchanwendungen zuordnen 184
 - Suchanwendungs-ID, Sicherheit 214
 - Suchserver 159
 - Syntaxanalyse 111
 - Systemstatus 251
 - überwachen 251
 - Zugriffssteuerung auf Dokumentenebene umgehen 223
- Objektgruppen, Sicht
 - Beschreibung 15
 - Objektgruppen erstellen 31
- Objektgruppen-ID, Syntaxregeln 31
- Objektgruppenadministrator
 - Beschreibung 206
 - Rollenkonfiguration 208
- Objektgruppenassistent 30

P

- Parallele Indexerstellungen 146
- Parametrische Suche 148
- Parser
 - Beschreibung 5, 111
 - Datenanalysetasks 5
 - Dokumenttypen für Parser-Services 136, 137
 - Dokumenttypen für Stellent-Sitzungen 138, 140
 - Dokumenttypen syntaktisch analysieren 135
 - Leerzeichen 134
 - N-Gram-Segmentierung 134
 - native XML-Suche 133
 - starten 264
 - stoppen 264
 - Systemstatus 264
 - Threads 131
 - überwachen 264
 - Zeilenvorschubzeichen 134
 - zusammengesetzte Begriffe, Analyse 132
- Parser-Server
 - starten 243
 - stoppen 245
 - Threadkonfiguration 131
- parserTypes.cfg, Datei 136

- Platzhalterzeichen
 - Abfrageerweiterung 150, 152
 - Indexerweiterung 150, 152
- Plug-ins, für Crawler 101
- Portal Search Engine
 - Beschreibung 225
 - Sicherheit auf Dokumentenebene 227
 - Übergabe von Sicherheitstoken 227
- Portlet zur Taxonomieverwaltung 235
- Portlets
 - Beschreibung 225
 - ESearchAdapterPortlet.war, Datei 232
 - Portlet zur Taxonomieverwaltung 235
 - Registrierung 232
 - Suche 228, 229
 - Unternehmenssuche 225
 - unterstützte Versionen von WebSphere Portal 225
- Portnummer, Unternehmenssuche 26
- Präfixregeln für Web-Crawler 80
- Protokoll, Beschreibung der Seite 15
- Protokoll DIIOP, Crawlerkonfiguration 70
- Protokoll NRPC, Crawlerkonfiguration 67, 69
- Protokoll zum Ausschluss von Robots 78
- Protokolldateien
 - anzeigen 281
 - Beschreibung 273
 - Bewertungsstufen 277
 - E-Mail-Optionen 279
 - filtern 281
 - für Dokumentüberwachung 270
 - maximale Größe 277
 - Migrationsassistent 241
 - rollieren 277
 - SMTP-Server, Konfiguration 278
 - Standardposition 273
 - überwachen 281
- Proxy-Server 90
- Prüfung aktueller Berechtigungsnachweise 216, 219, 221, 222

Q

- Quick Links
 - Beschreibung 168
 - durchsuchen 168
 - erstellen 168
 - URI-Formate 103
- QuickPlace-Crawler
 - Directory Assistance, Konfiguration 76
 - Domino-Benutzerkonfiguration 75
 - in der Solaris-Betriebsumgebung konfigurieren 67
 - IOCP-Konfiguration 71
 - Konfiguration 73
 - Konfiguration für Protokoll DIIOP 70
 - Local User (Lokaler Benutzer), Sicherheit 75
 - Protokoll NRPC 67, 69

- QuickPlace-Crawler (*Forts.*)
 - unter dem Betriebssystem AIX konfigurieren 67
 - unter dem Betriebssystem Linux konfigurieren 67
 - unter Windows konfigurieren 69
 - URI-Formate 103

R

- Rangfolge von Suchergebnissen
 - Beschreibung 171
 - Boostklassen 177, 179, 180, 181
 - dynamisch 171
 - statisch 172
 - textbasierte Bewertung 171
 - URI-Muster 175, 176
 - Verzeichnisse von Boostwörtern 175
- Regelbasierte Kategorien
 - Beschreibung 113
 - erstellen 118
 - Kategorisierungstyp auswählen 117
- Regelbasierte Taxonomie aus WebSphere Portal migrieren 235, 237
- Restore-Scripts
 - ausführen 284, 285
 - Beschreibung 283
- robots.txt, Datei 78
- Rückkehrcodes, Unternehmenssuche 287
- runtime-generic.properties, Datei 167

S

- Schlüsselwörter in Quick Links 168
- Scripts
 - esbackup.bat 284, 285
 - esbackup.sh 284, 285
 - escrm.sh 57
 - escrm.vbs 58
 - esrdb2.sh 54
 - esrdb2.vbs 55
 - escrnote.sh 67
 - escrnote.vbs 69
 - escrvbr.sh 45
 - escrvbr.vbs 46
 - esrestore.bat 284, 285
 - esrestore.sh 284, 285
 - startcl 284
- Search and Index API 11, 183
- Seiten mit detaillierten Fehlerhinweisen, Web-Crawler 86
- Semantische Suche 124, 128, 133
- Servermodus, Content Edition-Repositories 44
- SI-API (Search and Index API) 11, 183
- Sicherheit
 - Ankertextanalyse 212
 - Authentifizierung 208, 210
 - Benutzerprofile 217
 - Beschreibung 205
 - Crawler-Plug-ins 101
 - Dokumentenebene 214, 215, 216, 218, 223
 - Erkennung von doppelten Dokumenten 211
 - für eine Objektgruppe aktivieren 205

- Sicherheit (*Forts.*)
 - für Objektgruppe aktivieren 31
 - für Unternehmensanwendung inaktivieren 210
 - globale WebSphere-Sicherheit 210
 - Identitätsmanagement 218
 - Lotus Domino-Dokumente 221, 222
 - Mustersuchanwendung 197
 - Objektgruppenebene 211, 223
 - Suchanwendungs-IDs 214
 - Verwaltungsrollen 208
 - Windows-Domänen 219
 - Zugriffssteuerung 208
 - Zugriffssteuerung auf Dokumentenebene umgehen 223
- Sicherheit, Beschreibung der Sicht 15
- Sicherheit "Local User", QuickPlace-Crawler 75
- Sicherheit auf Dokumentenebene
 - aktuellen Berechtigungsnachweis prüfen 216
 - Benutzerprofile 217
 - Beschreibung 205, 214
 - Crawler-Plug-ins 101
 - Crawlerkonfiguration 37
 - Echtzeitprüfung 216
 - für Lotus Domino-Dokumente 221
 - für Windows-Dateisysteme 219
 - Identitätsmanagement 218
 - indexierte Zugriffssteuerung 215
 - Lotus Domino-Dokumente 222
 - Portal Search Engine-Unterstützung 227
 - Sicherheitstoken 215
- Sicherheit auf Objektgruppenebene
 - aktivieren 31
 - Ankertextanalyse 212
 - Beschreibung 205, 211
 - Erkennung von doppelten Dokumenten 211
 - Suchanwendungs-IDs 214
- Sicherheitstoken
 - Crawlerkonfiguration 215
 - für eine Objektgruppe inaktivieren 223
 - Sicherheit auf Dokumentenebene 215
- Sicherheitstokens
 - Portal Search Engine-Verarbeitung 227
 - Sicherheit auf Dokumentenebene 227
- Siteverlaufsberichte
 - Beschreibung 256
 - erstellen 259
- Sitzungs-IDs, Unternehmenssuche 287
- SMTP-Server, Konfiguration 278
- Solaris-Betriebsumgebung
 - Content Edition-Crawler, Konfiguration 45
 - DB2 Content Manager-Crawler, Konfiguration 57
 - Domino Document Manager-Crawler, Konfiguration 67
 - Notes-Crawler, Konfiguration 67
 - QuickPlace-Crawler, Konfiguration 67
- Standardsuchanwendung 195

- Start-URL-Adressen für Web-Crawler 80, 85
 - startcl, Script 284
 - Starten
 - Crawler-Server 254
 - Datenempfangsfunktion 268
 - Indexerstellungen 265
 - Migrationsassistent 235, 237
 - Mustersuchanwendung 196
 - Parser 264
 - Server für die Unternehmenssuche 243
 - Suchserver 267
 - Statische Rangfolge
 - Beschreibung 172
 - für Objektgruppe aktivieren 31
 - in migrierten Objektgruppen 237
 - stellent.properties, Datei 138
 - Stellent-Sitzungen
 - Dokumenttypen syntaktisch analysieren 135
 - Dokumenttypen zuordnen 138
 - Standarddokumenttypen 140
 - stellentypes.cfg, Datei 138
 - Stoppen
 - Crawler-Server 254
 - Indexerstellungen 265, 266
 - Parser 264
 - Server für die Unternehmenssuche 243, 245
 - Suchserver 267
 - Suchantwortzeit
 - Alerts 274
 - überwachen 267
 - Suchanwendungen
 - benutzerdefiniert 183
 - Beschreibung 12
 - externen Quellen zuordnen 202
 - Muster 183, 184
 - Objektgruppen zuordnen 184
 - Sicherheit auf Objektgruppenebene 214
 - Suchanwendungs-IDs 214
 - Suchanwendungs-IDs 214
 - Suchcache
 - Beschreibung 160
 - konfigurieren 160
 - Suche, Beschreibung der Seite 15
 - Suchen
 - HTML-Dokumente 123
 - Kategorien 112
 - XML-Dokumente 120, 121, 128
 - Suchen nach Dokumentation zur Unternehmenssuche 317
 - Suchergebnisse
 - ausblenden 155, 156
 - Beschreibung 171
 - Boostklassenkonfiguration 177, 179, 180, 181
 - dynamische Rangfolge 171
 - dynamische Zusammenfassung gruppieren 155, 156
 - Konfiguration von URI-Mustern 175
 - Platzhalterzeichen 150
 - Platzhalterzeichenerweiterung 152
 - Rangfolge 176
 - statische Rangfolge 172
 - Suchergebnisse (Forts.)
 - textbasierte Bewertung 171
 - Zusammenfassungen 166, 167
 - Zusammenfassungen anpassen 166, 167
 - Suchoptionen
 - Dokumentinhalt 148
 - feldspezifische Suche 148
 - Freiformatsuche 148
 - für Suchergebnisse 148
 - parametrische Suche 148
 - Suchportlet, Implementierung 228
 - Suchportlet, Konfiguration 229
 - Suchrate überwachen 258
 - Suchserver
 - Antwortzeitprotokoll 267
 - Beschreibung 8, 159
 - häufige Abfragen 267
 - letzte Abfragen 267
 - starten 243, 267
 - stoppen 245, 267
 - Suchcache 160
 - Synonymverzeichnisse 161, 162
 - Synonymverzeichnisse zuordnen 163
 - Systemstatus 267
 - überwachen 267
 - Verzeichnisse von Boostwörtern 172
 - Verzeichnisse von Boostwörtern zuordnen 174, 175
 - Verzeichnisse von Stoppwörtern 163
 - Verzeichnisse von Stoppwörtern zuordnen 165
 - Suchzentrale für WebSphere Portal
 - Adapter für die Unternehmenssuche 225, 231
 - Beschreibung 225
 - Registrierungsportlet 225, 232
 - synonyms.xml, Datei 235
 - Synonymverzeichnisse
 - Beschreibung 161
 - dem System hinzufügen 162
 - einer Objektgruppe zuordnen 163
 - Syntaxanalyse, Beschreibung der Seite 15
 - System, Beschreibung der Sicht 15
 - System-Backup 283, 284
 - Systemressourcen
 - schätzen 250
 - überprüfen 250
 - Systemressourcen schätzen 250
 - Systemrestore 283, 284, 285
 - Systemstatus
 - Crawler 254
 - Indexerstellungen 266
 - Objektgruppen 251
 - Parser 264
 - Suchserver 267
 - Web-Crawler 256
 - Terminieren (Forts.)
 - Indexerstellungen 144, 145
 - Textanalyse
 - allgemeine Analysestrukturen 129, 130
 - Textanalysesteuerkomponenten 126, 127
 - XML-Elemente zuordnen 128
 - Textanalysesteuerkomponenten
 - Analyseergebnisse zuordnen 129, 130
 - Beschreibung 124
 - dem System hinzufügen 126
 - Objektgruppen zuordnen 127
 - XML-Elemente zuordnen 128
 - Textbasierte Bewertung 171
 - Textverarbeitung
 - allgemeine Analysestrukturen 124
 - Annotatoren 124
 - Textanalysesteuerkomponenten 124
 - Threaddetails überwachen 256
 - Threads
 - Parser 131
 - Web-Crawler 257
 - titles.xml, Datei 235
 - treenodes.xml, Datei 235
 - Trusted Server, Konfiguration 222
- ## U
- Überprüfen von Systemressourcen 250
 - Überwachen
 - aktive Web-Crawler-Sites 257
 - Antwortzeitprotokoll 267
 - Beschreibung 206
 - Crawler 254
 - Datenempfangsfunktion 268
 - Details von Web-Crawler-Threads 257
 - gelöschte Dokumente 270
 - häufige Abfragen 267
 - letzte Abfragen 267
 - Objektgruppen 251
 - Parser 264
 - Protokolldateien 281
 - Rollenkonfiguration 208
 - Suchrate des Web-Crawlers 258
 - Suchserver 267
 - Unternehmenssuche 249
 - URI-Details 252
 - Web-Crawler 256
 - Überwachung, Beschreibung der Sicht 15
 - UIMA
 - allgemeine Analysestrukturen 129, 130
 - Analyseergebnisse dem Index zuordnen 129
 - Analyseergebnisse JDBC-Tabellen zuordnen 130
 - Beschreibung 124
 - hinzufügen von Textanalysesteuerkomponenten zum System 126
 - Objektgruppen zuordnen 127
 - XML-Elemente zuordnen 128
 - Unternehmensanwendungen
 - ESAdmin, Anwendung 210
 - ESPACServer.ear, Datei 97

- Unternehmensanwendungen (*Forts.*)
 - ESSearchAdapter.ear, Datei 231
 - ESSearchApplication, Anwendung 210
 - ESSearchServer, Anwendung 210
- Unternehmenssuche
 - Administrationskonsole 9
 - APIs 11
 - aus Backup wiederherstellen 284
 - Backup-Scripts 283
 - Befehle 287
 - Crawler-Server 4, 37
 - Datenflussdiagramm 13
 - Indexserver 7, 143
 - Integration in WebSphere Portal 225
 - Komponenten 3
 - Parser 5, 111
 - Portnummerkonfiguration 26
 - Protokolldateien 273
 - Restore-Scripts 283
 - Rückkehrcodes 287
 - Server starten 243
 - Server stoppen 243, 245
 - Sicherheit 205
 - Sicherheit auf Dokumentenebene 214
 - Sicherheit auf Objektgruppenebene 211
 - sichern 284
 - Sitzungs-IDs 287
 - Suchanwendungen 12
 - Suchserver 8, 159
 - Systemdateien wiederherstellen 285
 - Überblick 1
 - überwachen 249
 - URI-Formate 103
 - Verwaltungsrollen 206
- Unternehmenssuche sichern 283, 284
- Unternehmenssuche wiederherstellen 283
- Unterstützung für JavaScript in Web-Crawlern 79
- URI-Details
 - gelöschte Dokumente 270
 - überwachen 252
- URIs
 - aus dem Index entfernen 157
 - Beeinflussung von statischen Bewertungen 175, 176
 - Bereiche 153, 154
 - Details anzeigen 252
 - Formate in der Unternehmenssuche 103
 - in Suchergebnissen ausgeblendet 155, 156
 - Kategorisierungsregeln 113, 118
 - Quick Links 168
- URIs aus dem Index entfernen 157
- URL-Adressen so schnell wie möglich aufrufen 85
- URL-Adressen so schnell wie möglich erneut aufrufen 85
- URL-Pfadtiefe 80

V

- vbr_access_services.jar, Datei 45, 46

- Verwaltungsrollen
 - Administrator für die Unternehmenssuche 206, 208
 - Bediener 206, 208
 - Beschreibung 206
 - konfigurieren 208
 - Objektgruppenadministrator 206, 208
 - überwachen 206, 208
- Verzeichnisse von Boostwörtern
 - Beschreibung 172
 - dem System hinzufügen 174
 - einer Objektgruppe zuordnen 175
- Verzeichnisse von Stoppwörtern
 - Beschreibung 163
 - dem System hinzufügen 165
 - einer Objektgruppe zuordnen 165

W

- Web-Crawler
 - aktive Sites 256, 257
 - Benutzeragenten 78
 - Berichte erstellen 259
 - Cookieformat 91
 - Cookiekonfiguration 92
 - Cookies 91
 - Crawlerbereich begrenzen 80
 - Crawlersuchregeln 80
 - Crawlerverlauf 256
 - followindex.rules, Datei 94, 95
 - globaler Crawlerbereich 92
 - Intervalle für erneute Suche 85
 - kennwortgeschützte Websites 87, 88, 89
 - Konfiguration 77
 - no-follow, Anweisung 94, 95
 - no-index, Anweisung 94, 95
 - Proxy-Server 90
 - Seiten mit detaillierten Fehlerhinweisen 86
 - Siteverlauf 256
 - Start-URL-Adressen 80, 85
 - Suchrate 258
 - Systemstatus 256
 - Threaddetails 256, 257
 - überwachen 256
 - Unterstützung für JavaScript 79
 - URL-Adressen so schnell wie möglich aufrufen 85
 - URL-Status 256
 - zuletzt durchsuchte URLs 256
- Web-Crawler-Bereich begrenzen 80
- WebSphere II Event Publisher Edition, Konfiguration des DB2-Crawlers 49
- WebSphere II OmniFind Edition 319
 - Administrationskonsole 9
 - APIs 11
 - Befehle 287
 - behindertengerechte Bedienung 319
 - Crawler-Server 4
 - Datenflussdiagramm 13
 - Identitätsmanagement 218
 - Indexserver 7
 - Integration in WebSphere Portal 225
 - Kennwort auf einem Einzelservers ändern 20

- WebSphere II OmniFind Edition (*Forts.*)
 - Kennwort auf mehreren Servern ändern 22
 - Komponenten 3
 - Parser 5
 - Portnummerkonfiguration 26
 - Rückkehrcodes 287
 - Sitzungs-IDs 287
 - Suchanwendungen 12
 - Suchserver 8
 - Überblick 1
- WebSphere MQ, Konfiguration des Crawler-Servers 54, 55
- WebSphere MQ, Konfiguration des DB2-Crawlers 52
- WebSphere Portal
 - Kategoriebaummigration 235
 - modellbasierte Taxonomien 235
 - Objektgruppenmigration 235
 - Portlet zur Taxonomieverwaltung 235
 - Standardmigrationseinstellungen 239
 - Suchportlet 229
 - Suchportlet, Implementierung 228
 - Taxonomiemigration 235
 - Unternehmenssuche, Integration 225
- WebSphere Portal-Crawler
 - Implementierung der Datei "ESPAC-Server.ear" 97
 - Konfiguration 96
 - kopieren, Site-URL-Adressen 98
 - Unternehmensanwendung, Implementierung 97
 - URI-Formate 103
- WebSphere Portal-Suchzentrale
 - Adapter für die Unternehmenssuche 225, 231
 - Beschreibung 225
 - Registrierungsportlet 225, 232
- Wiederherstellen
 - aus Backup 284
 - Systemdateien auf neuen Servern 285
- Windows-Betriebssystem
 - Content Edition-Crawler, Konfiguration 46
 - Crawlerkonfiguration 69
 - DB2 Content Manager-Crawler, Konfiguration 58
 - Ereignisveröffentlichung, Konfiguration 55
- Windows-Domänen 219
- WpsMigratorLog.log, Protokolldatei 241

X

- XML-Abfragesyntax, nativ 133
- XML-Dokumente
 - native XML-Suche 133
 - suchen 121
- XML-Elemente
 - allgemeinen Analysestrukturen zuordnen 128
 - suchen 120, 128
 - Suchfeldern zuordnen 121
- XML-Fragmente, native XML-Suche 133

XML-Suchfelder
 Beschreibung 120, 128
 Elemente zuordnen 120, 121, 128
 erstellen 121
XPath, native XML-Suche 133

Z

Zugriffssteuerung
 aktuelle Prüfung 216
 Anforderungen für Lotus Domino 221, 222
 Anforderungen für Windows-Dateisysteme 219
 Benutzerprofile 217
 Beschreibung 208
 für eine Objektgruppe inaktivieren 223
 Portal Search Engine-Unterstützung 227
 Sicherheit auf Dokumentebene 215, 227
Zuletzt durchsuchte URLs überwachen 256
Zuordnen
 allgemeine Analysestrukturen zu JDBC-Tabellen 130
 allgemeine Analysestrukturen zum Index 129
 Felder zu Boostklassen 179
 HTML-Suchfelder 123
 XML-Elemente zu allgemeinen Analysestrukturen 128
 XML-Suchfelder 121
Zusammenfassungen
 anpassen 166, 167
 dynamische 166
Zusammengesetzte Begriffe, Syntaxanalyse 132

IBM



Java[™]
COMPATIBLE

SC12-3432-02

