

IBM DB2 Information Integrator  
OmniFind Edition



# Administering Enterprise Search

*Version 8.2*



IBM DB2 Information Integrator  
OmniFind Edition



# Administering Enterprise Search

*Version 8.2*

Before using this information and the product it supports, be sure to read the general information under "Notices."

This document contains proprietary information of IBM. It is provided under a license agreement and Copyright law protects it. The information contained in this publication does not include any product warranties, and any statements provided in this manual should not be interpreted as such.

You can order IBM publications online or through your local IBM representative:

- To order publications online, go to the IBM Publications Center at [www.ibm.com/shop/publications/order](http://www.ibm.com/shop/publications/order)
- To find your local IBM representative, go to the IBM Directory of Worldwide Contacts at [www.ibm.com/planetwide](http://www.ibm.com/planetwide)

When you send information to IBM, you grant IBM a nonexclusive right to use or distribute the information in any way it believes appropriate without incurring any obligation to you.

© Copyright International Business Machines Corporation 2004. All rights reserved.

US Government Users Restricted Rights – Use, duplication or disclosure restricted by GSA ADP Schedule Contract with IBM Corp.

---

# Contents

## About this book . . . . . xvii

Who should read this book . . . . . xvii

## Chapter 1. What is enterprise search?. . . 1

Data source types supported by enterprise search . . . 1

Enterprise search component overview . . . . . 2

Enterprise search crawlers . . . . . 3

Enterprise search parsers . . . . . 4

Enterprise search indexes . . . . . 5

Search servers for enterprise search. . . . . 6

Enterprise search administration console . . . . . 7

Monitoring an enterprise search system . . . . . 8

Enterprise search log files . . . . . 8

Enterprise search APIs . . . . . 9

Sample search application for enterprise search. . . 9

The enterprise search data flow. . . . . 10

## Chapter 2. Enterprise search administration overview . . . . . 13

Logging in to the administration console . . . . . 15

Changing the enterprise search administrator password . . . . . 16

## Chapter 3. Enterprise search collections . . . . . 19

Creating a collection by using the Collection wizard 19

Creating a collection by using the Collections view 20

Editing a collection. . . . . 21

Deleting a collection . . . . . 22

## Chapter 4. Enterprise search crawler administration . . . . . 23

Creating a crawler . . . . . 25

Editing crawler properties . . . . . 26

Editing a crawl space . . . . . 26

Deleting a crawler . . . . . 27

Content Manager crawlers . . . . . 27

Configuring the crawler server on AIX for DB2

Content Manager . . . . . 28

Configuring the crawler server on Linux for DB2

Content Manager . . . . . 30

Configuring the crawler server on Windows for

DB2 Content Manager. . . . . 31

DB2 crawlers . . . . . 32

Exchange Server crawlers. . . . . 33

NNTP crawlers . . . . . 34

Notes crawlers . . . . . 35

Configuring the crawler server on UNIX for

Lotus Notes . . . . . 36

Configuring the crawler server on Windows for

Lotus Notes . . . . . 38

Configuring the DIOP protocol for Notes

crawlers . . . . . 39

UNIX file system crawlers . . . . . 40

VeniceBridge crawlers . . . . . 41

Configuring the crawler server on UNIX for

VeniceBridge . . . . . 42

Configuring the crawler server on Windows for

VeniceBridge . . . . . 43

Web crawlers . . . . . 44

Rules to limit the Web crawl space . . . . . 45

Recrawl interval settings in the Web crawler . . . 49

Options for visiting URLs with the Web crawler 49

Windows file system crawlers . . . . . 49

Configuring support for external crawlers . . . . . 50

## Chapter 5. Enterprise search index administration . . . . . 53

Scheduling index builds . . . . . 54

Changing the index schedule . . . . . 55

Enabling and disabling the index schedules . . . 55

Configuring concurrent index builds . . . . . 56

Removing URIs from the index. . . . . 57

## Chapter 6. Search applications for enterprise search . . . . . 59

Configuring default search application properties. . 61

Accessing the sample search application. . . . . 62

Associating search applications with collections . . 63

## Chapter 7. Linguistic capabilities in enterprise search . . . . . 65

Language identification . . . . . 66

Stop word removal. . . . . 66

Character normalization . . . . . 66

Linguistic support for dictionary-based

segmentation . . . . . 67

Linguistic support for nondictionary-based

segmentation . . . . . 68

Word segmentation in Japanese. . . . . 69

Orthographic variants in Japanese. . . . . 69

## Chapter 8. How enterprise search searches collections. . . . . 71

Query techniques . . . . . 71

Working with quick links. . . . . 73

Quick links . . . . . 73

Configuring quick links . . . . . 74

Working with categories . . . . . 75

Rule-based categories . . . . . 75

Model-based categories . . . . . 76

Category trees . . . . . 77

Selecting the categorization type . . . . . 78

Configuring categories. . . . . 79

Working with scopes . . . . . 80

Scopes . . . . . 80

Configuring scopes. . . . . 81

Working with XML search fields . . . . . 82

XML search fields . . . . .	82
Mapping XML elements to search fields . . . . .	82

**Chapter 9. How enterprise search returns search results . . . . . 85**

Text-based scoring . . . . .	85
Static ranking. . . . .	85
Dynamic summarization . . . . .	86
Collapsing results from the same Web site . . . . .	86

**Chapter 10. Starting and stopping enterprise search . . . . . 89**

Starting the enterprise search components . . . . .	89
Stopping the enterprise search components. . . . .	90

**Chapter 11. Monitoring enterprise search activity . . . . . 93**

Alerts . . . . .	93
Configuring collection-level alerts . . . . .	94
Configuring system-level alerts. . . . .	95
Monitoring a collection . . . . .	96
Viewing details about a URI. . . . .	96
Monitoring crawlers . . . . .	97
Viewing details about Web crawler activity. . . . .	98
Web crawler thread details . . . . .	99
Web crawler active sites . . . . .	100
Web crawler crawl rate . . . . .	100
Creating Web crawler reports . . . . .	101
Web crawler HTTP return codes . . . . .	102
Monitoring the parser . . . . .	105
Monitoring index activity for a collection . . . . .	106
Monitoring the enterprise search index queue . . . . .	106
Monitoring the search servers . . . . .	107
Search cache. . . . .	109
Configuring a search cache. . . . .	109

**Chapter 12. Enterprise search security 111**

Administrative roles . . . . .	112
Configuring administrative users . . . . .	113
Authentication versus access control. . . . .	113
Document-level security . . . . .	114
Collection-level security . . . . .	115
Anchor text analysis . . . . .	115
Enabling anchor text-only document processing	116
Duplicate document analysis . . . . .	117
Security with search application IDs. . . . .	117
Allowing users to access all documents. . . . .	118

**Chapter 13. Enterprise search integration with WebSphere Portal . . . 119**

Deploying the enterprise search portlets in WebSphere Portal . . . . .	121
Configuring the Portal Search Engine to search enterprise search collections . . . . .	121
Document-level security with the Portal Search Engine . . . . .	121

**Chapter 14. Migration from WebSphere Portal . . . . . 123**

Migrating a model-based taxonomy from WebSphere Portal . . . . .	123
Migrating a collection from WebSphere Portal . . . . .	124
Migrated collection settings. . . . .	126
Migration wizard log file . . . . .	127

**Chapter 15. Backup and restore for enterprise search. . . . . 129**

Backing up the enterprise search system . . . . .	129
Restoring the enterprise search system . . . . .	130
Restoring enterprise search system files to new servers . . . . .	130

**Chapter 16. Enterprise search log files 133**

Configuring log files . . . . .	133
Receiving e-mail about logged messages . . . . .	134
Viewing log files . . . . .	136

**Chapter 17. URI formats in enterprise search . . . . . 139**

**Chapter 18. Enterprise Search messages . . . . . 143**

General system messages . . . . .	143
FFQO0164E: Unable to load properties file <i>file_name</i> . . . . .	143
FFQO0165E: Unable to load included properties file <i>file_name</i> . . . . .	143
FFQO0168E: Unable to save properties to file <i>file_name</i> . . . . .	143
Migration wizard messages. . . . .	143
FFQW0001E: No message was found for the <i>message_key</i> key. Contact the enterprise search administrator. . . . .	143
FFQW0002E: File <i>file_name</i> cannot be read. Change your selection. . . . .	143
FFQW0003E: No collection name is set for file <i>file_name</i> . . . . .	144
FFQW0008E: The user ID or password is invalid. . . . .	144
FFQW0011E: An error occurred while importing the WebSpherePortal Search Engine collection in file <i>file_name</i> . At line <i>line_number</i> and column <i>column_number</i> , the following error occurred: <i>error_message</i> . The file could not be imported. . . . .	144
FFQW0012E: An error occurred while creating collection <i>collection_name</i> . The following error was returned: <i>error_code</i> . . . . .	144
FFQW0013E: An error occurred while creating a Web crawler for collection <i>collection_name</i> . The following error occurred: <i>error_code</i> . . . . .	144
FFQW0014E: An error occurred while creating collection <i>collection_name</i> . No session could be found for the ID <i>user_name</i> . . . . .	145
FFQW0016E: An error occurred while creating collection <i>collection_name</i> : <i>error_message</i> . . . . .	145

FFQW0019E: An error occurred while creating a crawler configuration file - Exception Message: <i>error_message</i> . . . . .	145	FFQD1036E: ERROR: The environment variable ES_NODE_ROOT is not defined. \nVerify that DB2 Information Integrator OmniFind Edition was installed correctly and that this script was run by the enterprise search administrator user. . . . .	148
FFQW0120I: The migration completed successfully. See the log file for more information. . . . .	145	FFQD1037E: ERROR: The environment variable ES_INSTALL_ROOT is not defined. \nVerify that DB2 Information Integrator OmniFind Edition was installed correctly and this script was run by the enterprise search administrator user. . . . .	148
FFQW0140I: The collection was imported from the WebSphere Portal configuration file <i>file</i> . . . . .	145	FFQD1101E: ERROR: Undefined platform. . . . .	148
FFQW0141I: Collection <i>collection_name</i> was created successfully. . . . .	145	FFQD1106E: ERROR: Run this setup shell as enterprise search administrator user, not as the root user. . . . .	149
FFQW0142I: The Web crawler for collection <i>collection_name</i> was created successfully. . . . .	145	FFQD1107E: ERROR: ES_CFG environment variable is not defined. Check that DB2 Information Integrator OmniFind Edition is correctly installed on this computer. . . . .	149
FFQW0154I: Creating the crawler data for collection <i>collection_name</i> . . . . .	145	FFQD1108E: ERROR: The enterprise search configuration file does not exist. . . . .	149
FFQW0155I: The following settings for the maximum number of crawl threads were found: <i>list_thread_numbers</i> . Imported value is <i>maximum_threads</i> . . . . .	146	FFQD1123E: ERROR: Lotus Notes directory was not found. . . . .	149
FFQW0156I: The following settings for the maximum crawl path depths were found: <i>list_path_depths</i> . Imported value is <i>maximum_path_depth</i> . . . . .	146	FFQD1128E: ERROR: Lotus Notes data directory was not found. . . . .	149
FFQW0157I: The following settings for the crawl document timeout were found: <i>list_timeout_limits</i> . Imported value is <i>maximum_timeout_limit</i> . . . . .	146	FFQD1138E: ERROR: Failed to remove <i>filename</i> . Remove it and run this script again. . . . .	149
FFQW0158I: The following settings for default character sets were found: <i>list_charsets</i> . Imported value is <i>charset</i> . . . . .	146	FFQD1139E: ERROR: Failed to create <i>filename</i> . . . . .	150
FFQW0159I: The configuration file <i>file</i> was created. Return code: <i>return_code</i> . . . . .	146	FFQD1142E: ERROR: <i>filename</i> was not found. . . . .	150
Configuration manager messages. . . . .	146	FFQD1407E: ERROR: Venetica VeniceBridge directory was not found. . . . .	150
FFQC1000E: Collection name <i>collection_name</i> already exists. . . . .	146	FFQD1420E: ERROR: WebSphere Application Server directory was not found. . . . .	150
FFQC1003E: Session name <i>session_name</i> already exists. . . . .	146	FFQD2005E: The crawler server is not configured for the specified crawler type. Confirm that the crawler server is configured by the appropriate setup script. . . . .	150
FFQC1008E: Session <i>session_id</i> is already configured. Unable to add session. . . . .	146	FFQD2007E: An error occurred while registering the <i>api_name</i> API for operation <i>operation_class</i> . . . . .	150
FFQC2034E: No index available to do content refresh. . . . .	147	FFQD2100E: The SQL exception occurred while accessing the database. Error code: <i>error_message</i> . Error message: <i>error_code</i> . . . . .	151
FFQC5217E: An error occurred while loading property file <i>file_name</i> . . . . .	147	FFQD2102E: User name or password is incorrect. . . . .	151
FFQC5228W: File name <i>file_name</i> has file size of <i>file_size</i> . This exceeds file replacement size limit of <i>file_limit</i> . . . . .	147	FFQD2103E: The following error occurred while executing the DB2 discovery API. Return code: <i>return_code</i> . Error message: <i>error_message</i> . . . . .	151
Crawler messages. . . . .	147	FFQD2104E: A connection cannot be established. . . . .	151
FFQD1003E: ERROR: User <i>user</i> does not exist. Check <i>file</i> . . . . .	147	FFQD2105E: The table or view <i>table_or_view_name</i> was not found for schema name <i>schema_name</i> . . . . .	151
FFQD1011E: ERROR: <i>file</i> was not found. . . . .	147	FFQD2106E: The DB2 instance <i>server_name:port_number</i> was not found. . . . .	151
FFQD1031E: ERROR: This script is not running on the crawler server. \nRun this script on a crawler server. . . . .	147		
FFQD1034E: ERROR: The environment variable CMBR00T is not defined. \nVerify that DB2 Information Integrator for Content was installed correctly. . . . .	148		
FFQD1035E: ERROR: The environment variable CMCOMMON is not defined. \nVerify that DB2 Information Integrator for Content was installed correctly. . . . .	148		

FFQD2130E: A Notes exception occurred while accessing the server. ID: <i>error_id</i> . Error message: <i>error_message</i> . . . . .	151	FFQD2243E: An error occurred while validating the truststore file <i>truststore_file</i> . . . . .	155
FFQD2131E: The specified ID file does not exist or cannot be read. . . . .	152	FFQD2244E: The truststore file format or password is incorrect. The specified truststore is <i>truststore_file</i> . . . . .	155
FFQD2132E: A Notes session cannot be created. . . . .	152	FFQD2245E: An unknown host name <i>host_name</i> was detected. . . . .	155
FFQD2133E: Other Notes crawlers are running. They must be stopped because they are holding the lock <i>lock_name</i> . Path: <i>lock_filepath</i> . Name: <i>lock_filename</i> . . . . .	152	FFQD2246E: The URL <i>public_folder_URL</i> is incorrect. . . . .	156
FFQD2135E: The crawler server is not configured for NRPC. Check that the crawler server is configured by the appropriate setup script. . . . .	152	FFQD2247E: An I/O error occurred while connecting to <i>public_folder_URL</i> . . . . .	156
FFQD2160E: The directory <i>directory name</i> is not valid. . . . .	152	FFQD2248E: A protocol exception occurred while sending a request to <i>public_folder_URL</i> . . . . .	156
FFQD2161E: There are more than <i>levels of subdirectories to discover</i> available subdirectories. To reduce this number, specify a Main directory name or pattern or specify fewer levels of subdirectories to discover. . . . .	152	FFQD2249E: An I/O error occurred while sending a request to <i>public_folder_URL</i> . . . . .	156
FFQD2162E: The specified name <i>directory name</i> is not a directory name but a file name. . . . .	153	FFQD2250E: The number of folders ( <i>actual_folder_count</i> ) exceeds the limit <i>folder_count_limit</i> . . . . .	156
FFQD2163E: The specified pattern for directory <i>directory name or pattern</i> is not valid. . . . .	153	FFQD2251E: The user ID, password, or both are incorrect. The specified user name is <i>user_name</i> . . . . .	156
FFQD2164E: An internal error occurred while discovering file directories. . . . .	153	FFQD2252E: The specified folder <i>public_folder_URL</i> does not exist. . . . .	156
FFQD2190E: The network path for directory <i>network path name</i> was not found ( <i>directory=network path name</i> ). Error message: <i>detail error message</i> . . . . .	153	FFQD2253E: An unexpected HTTP response was received. HTTP State Code: <i>HTTP_state_code</i> . HTTP Response Message: <i>HTTP_response_message</i> . . . . .	157
FFQD2191E: The specified drive letter is not valid. . . . .	153	FFQD2254E: An I/O error occurred while receiving a response from <i>public_folder_URL</i> . . . . .	157
FFQD2210E: The user ID, password, or both are incorrect. The specified user ID is <i>user_id</i> . . . . .	153	FFQD2255E: The server response is incorrect XML. . . . .	157
FFQD2211E: The connection to the Content Manager server <i>server</i> failed. . . . .	153	FFQD2256E: An I/O error occurred while parsing the received XML. . . . .	157
FFQD2212I: Connected to Content Manager server <i>server</i> . . . . .	154	FFQD2257E: An XML parser error occurred. . . . .	157
FFQD2213E: CMCOMMON keyword is not found in <i>discovery.properties</i> file. . . . .	154	FFQD2258E: The specified URL <i>public_folder_URL</i> is valid, but it is not a public folder. . . . .	157
FFQD2214E: CM INI file <i>file</i> was not found. . . . .	154	FFQD2270E: The specified JNDI property is not valid. . . . .	157
FFQD2215E: Cannot read the CM INI file, or the CM INI file <i>file</i> was not found. . . . .	154	FFQD2271E: The specified repository name is not valid. . . . .	158
FFQD2216E: The item type <i>item_type</i> does not exist. . . . .	154	FFQD2272E: The specified user name or password is not valid. . . . .	158
FFQD2217E: CM DKException occurred. . . . .	154	FFQD2273E: The specified item class name is not valid. . . . .	158
FFQD2218E: CM Exception occurred. . . . .	155	FFQD2274E: A VeniceBridge API error occurred. Error message: <i>error_message</i> . . . . .	158
FFQD2240E: A truststore file must be specified to use SSL. . . . .	155	FFQD2275E: The VeniceBridge server instance cannot be created. Error message: <i>error_message</i> . . . . .	158
FFQD2241E: A truststore password must be specified to use SSL. . . . .	155	FFQD2276E: The VeniceBridge server cannot be located. Error message: <i>error_message</i> . . . . .	158
FFQD2242E: The truststore file <i>truststore_file</i> does not exist. . . . .	155	FFQD2277E: A communication error occurred. Error message: <i>error_message</i> . . . . .	158
		FFQD2278E: The specified repository <i>repository_name</i> was not found. Error message: <i>error_message</i> . . . . .	159



FFQD2279E: The VeniceBridge user instance was not initialized. Error message: <i>error_message</i> .. . . . .	159	FFQD3023W: The server <i>server_name</i> is not scheduled. . . . .	162
FFQD2280E: Logon to <i>repository_name</i> repository failed. Error message: <i>error_message</i> .. . . . .	159	FFQD3024E: The return code of fetch is incorrect. Return code: <i>return_code</i> . State: <i>state</i> . . . . .	162
FFQD2282E: The specified item class <i>item_class_name</i> was not found. Repository: <i>repository_name</i> . . . . .	159	FFQD3025E: An exception occurred while fetching data.. . . . .	162
FFQD3000I: <i>crawler_name</i> crawler started successfully.. . . . .	159	FFQD3026E: An exception occurred while processing thread <i>thread_id</i> .. . . . .	162
FFQD3001E: <i>crawler_name</i> crawler failed to start.. . . . .	159	FFQD3027E: An exception occurred for the crawler.. . . . .	162
FFQD3002I: <i>crawler_name</i> crawler stopped successfully.. . . . .	159	FFQD3028E: The specified server <i>server_name</i> is not found. . . . .	162
FFQD3003E: <i>crawler_name</i> crawler did not stop successfully. . . . .	159	FFQD3029E: an exception occurred while initializing the connection manager.. . . . .	163
FFQD3004I: <i>crawler_name</i> crawler was paused successfully... . . . .	159	FFQD3030E: An exception occurred while initializing the state manager.. . . . .	163
FFQD3005E: <i>crawler_name</i> crawler failed to pause.. . . . .	160	FFQD3031E: An exception occurred while crawler status was being requested. . . . .	163
FFQD3006I: <i>crawler_name</i> crawler resumed successfully.. . . . .	160	FFQD3032E: An exception occurred while the crawler space status was being requested. . . . .	163
FFQD3007E: <i>crawler_name</i> crawler failed to resume. . . . .	160	FFQD3033E: An exception occurred while the crawl space status was being requested. . . . .	163
FFQD3008E: Factory exception occurred. Verify that the modules are installed properly. . . . .	160	FFQD3034I: Deleting obsolete documents. . . . .	163
FFQD3009I: Start the crawl for <i>crawler_name</i> on <i>server_name</i> . . . . .	160	FFQD3035W: Interrupted to delete obsolete documents. Will try again next time. . . . .	163
FFQD3010I: Finished the crawl for <i>crawler_name</i> on <i>server_name</i> .. . . . .	160	FFQD3036I: Finished deleting obsolete documents for <i>crawler_name</i> on <i>server_name</i> .. . . . .	163
FFQD3011E: Failed to stop for <i>target_host_name</i> on <i>target_name</i> . . . . .	160	FFQD3037E: An exception occurred while updating internal status.. . . . .	164
FFQD3012I: The crawler is already crawling on server <i>server_name</i> .. . . . .	160	FFQD3038E: An error occurred because of inconsistency of status. Information: <i>additional_data</i> . . . . .	164
FFQD3013W: The crawled data was truncated. URI: <i>URI</i> . Server: <i>target_host_name</i> . Target: <i>target_name</i> . . . . .	160	FFQD3039E: An error occurred while accessing internal status. Information: <i>additional_data</i> . . . . .	164
FFQD3014E: An exception occurred while handling crawled history.. . . . .	161	FFQD3040E: An exception occurred while deleting obsolete records. . . . .	164
FFQD3015E: An exception occurred while handling crawled history. URI: <i>uniform_resource_identifier</i> . Server: <i>server_name</i> . Target: <i>crawler_name</i> . . . . .	161	FFQD3041I: The schedule component started successfully. . . . .	164
FFQD3016E: An exception occurred while accessing the database. . . . .	161	FFQD3042E: The schedule component failed to start.. . . . .	164
FFQD3017E: An exception occurred while accessing the database. URI: <i>uniform_resource_identifier</i> . Server: <i>server_name</i> . Target: <i>crawler_name</i> . . . . .	161	FFQD3043I: The schedule component stopped successfully. . . . .	164
FFQD3018E: an exception occurred while handling the crawler configuration. . . . .	161	FFQD3044E: The schedule component failed to stop. . . . .	164
FFQD3019I: The server schedule is set for <i>server_name</i> server. . . . .	161	FFQD3045I: Starting the scheduled crawl for <i>crawler_name</i> .. . . . .	165
FFQD3020I: The schedule is disabled for <i>server_name</i> server. . . . .	161	FFQD3046E: An exception occurred in the schedule component. . . . .	165
FFQD3021W: The schedule for specified server <i>server_name</i> is already set.. . . . .	162	FFQD3047I: The internal repository was initialized successfully.. . . . .	165
FFQD3022W: The schedule for specified server <i>server_name</i> is already disabled.. . . . .	162	FFQD3048E: An unknown internal repository type <i>internal_repository_type</i> was detected. . . . .	165
		FFQD3049E: An internal repository initialization error. . . . .	165
		FFQD3050E: Access violation. . . . .	165
		FFQD3051E: Another repository is open.. . . . .	165
		FFQD3052E: There is no database connection. . . . .	165
		FFQD3053E: A database error occurred while opening the internal repository. . . . .	166

FFQD3054E: A database error occurred while closing the internal repository. . . . .	166
FFQD3055E: A database error occurred while writing a record to the internal repository. The record hash is <i>record_hash</i> . . . . .	166
FFQD3056E: A database error occurred while getting the last modified date. The record hash is <i>record_hash</i> . . . . .	166
FFQD3057E: A database error occurred while updating the last modified date. The record hash is <i>record_hash</i> . . . . .	166
FFQD3058E: A database error occurred while detecting a record change. The record hash is <i>record_hash</i> . . . . .	166
FFQD3059E: A database error occurred while getting checksum. The record hash is <i>record_hash</i> . . . . .	167
FFQD3060E: An I/O error occurred while getting checksum. The record hash is <i>record_hash</i> . . . . .	167
FFQD3061E: A database error occurred while checking a record. The record hash is <i>record_hash</i> . . . . .	167
FFQD3062E: A database error occurred while writing an excluded record. The record hash is <i>record_hash</i> . . . . .	167
FFQD3063E: A database error occurred while deleting a record from the repository. The record hash is <i>record_hash</i> . . . . .	167
FFQD3064E: A database error occurred while getting obsolete records. The target ID is <i>record_hash</i> , and the threshold time is <i>threshold_time</i> . . . . .	167
FFQD3065I: The internal repository was opened successfully. . . . .	167
FFQD3066I: The internal repository was closed successfully. . . . .	168
FFQD3067I: Loading crawler configuration files <i>general_config_file</i> and <i>crawler_config_file</i> . . . . .	168
FFQD3068I: Loading of the configuration files completed successfully. . . . .	168
FFQD3069E: An error occurred while creating the configuration factory. . . . .	168
FFQD3070E: Loading of configuration file <i>general_config_file</i> failed. . . . .	168
FFQD3071E: No configuration file was loaded. . . . .	168
FFQD3072E: An unknown configuration object was detected. . . . .	168
FFQD3073E: Getting document builder failed. . . . .	169
FFQD3074E: A configuration file format error was detected. . . . .	169
FFQD3075E: An I/O error occurred while parsing the configuration. . . . .	169
FFQD3076E: The crawler configuration file <i>crawler_config_file</i> was not found. . . . .	169
FFQD3077E: The expected the root element name in the crawler configuration file is <i>expected_root_element</i> , but <i>detected_root_element</i> was detected. . . . .	169

FFQD3078E: Attribute <i>attribute_name</i> is required in <i>element_name</i> element. . . . .	169
FFQD3079E: Duplicate name <i>value_of_name_attribute</i> was detected in element <i>element_name</i> . . . . .	169
FFQD3080E: Incorrect date string format was detected in <i>element_name</i> element. . . . .	170
FFQD3081E: Incorrect interval string format was detected in <i>element_name</i> element. . . . .	170
FFQD3082E: <i>element_name</i> element's value <i>actual_value</i> is out of bounds. . . . .	170
FFQD3083W: The XML parser reported a warning. . . . .	170
FFQD3084E: The XML parser reported an error. . . . .	170
FFQD3085E: The XML parser reported an unrecoverable error. . . . .	170
FFQD3086E: The formatter is not initialized. . . . .	170
FFQD3087E: Cannot convert to XML. . . . .	171
FFQD3088W: Parametric search is set to be enabled for a nonparametric field. The parametric search flag was disabled. . . . .	171
FFQD3089W: Metadata was truncated. . . . .	171
FFQD3090W: Temporary file <i>filename</i> is not deleted. . . . .	171
FFQD3091W: The formatted data record is null. No action is required. . . . .	171
FFQD3102E: Cannot open internal data store. DBName: <i>database_name</i> . DBSchema: <i>schema_name</i> . DBUser: <i>database_user</i> . TableName: <i>table_name</i> . . . . .	171
FFQD3103E: Cannot close internal data store. DBName: <i>database_name</i> . DBSchema: <i>schema_name</i> . DBUser: <i>database_user</i> . TableName: <i>table_name</i> . . . . .	172
FFQD3104E: Cannot write metadata to file <i>file_name</i> because of FileNotFoundException. . . . .	172
FFQD3105E: Cannot write to internal data store. DBName: <i>database_name</i> . DBSchema: <i>schema_name</i> . DBUser: <i>database_user</i> . TableName: <i>table_name</i> . . . . .	172
FFQD3106E: Cannot write to internal data store. DBName: <i>database_name</i> . DBSchema: <i>schema_name</i> . DBUser: <i>database_user</i> . TableName: <i>table_name</i> . . . . .	172
FFQD3107E: Cannot update by deletion flag to internal data store. DBName: <i>database_name</i> . DBSchema: <i>schema_name</i> . DBUser: <i>database_user</i> . TableName: <i>table_name</i> . . . . .	172
FFQD3108E: Cannot update by deletion flag to internal data store. DBName: <i>database_name</i> . DBSchema: <i>schema_name</i> . DBUser: <i>database_user</i> . TableName: <i>table_name</i> . . . . .	172
FFQD3109W: The crawled file <i>file_name</i> is updated by other programs during updating internal data store. . . . .	173
FFQD3110E: Cannot close file stream. . . . .	173
FFQD3111E: Password decryption failed. . . . .	173

FFQD3500E: JDBC class <i>jdbc_driver_class</i> is not found. . . . .	173	FFQD3523W: The connection manager was already initialized. . . . .	177
FFQD3501W: Exceptions occurred while altering <i>table_name</i> table. . . . .	173	FFQD3524E: JNDI look-up failed for <i>jndi_name</i> . . . . .	177
FFQD3502W: Exceptions occurred while creating <i>table_name</i> table. . . . .	173	FFQD3534E: Exceptions occurred while <i>component_name</i> was changing state from <i>from_state</i> to <i>to_state</i> . . . . .	177
FFQD3503W: Transaction rollback failed. . . . .	173	FFQD3540E: Failed to get file lock <i>file_name</i> because of IOException. . . . .	177
FFQD3504W: Transaction commit failed. . . . .	174	FFQD3541E: Failed to get file lock <i>file_name</i> because of FileNotFoundException. . . . .	178
FFQD3505W: Renaming table <i>orig_table_name</i> to table <i>dest_table_name</i> failed. Crawler will try to copy and drop it instead. . . . .	174	FFQD3543E: Failed to release file lock <i>file_name</i> . . . . .	178
FFQD3506W: Exceptions occurred while deleting all data from <i>table_name</i> table. . . . .	174	FFQD3544E: Failed to release file lock <i>file_name</i> because file lock is null. . . . .	178
FFQD3507W: Exceptions occurred while inserting data into <i>table_name</i> table. . . . .	174	FFQD3545E: Algorithm <i>algorithm_name</i> for loading truststore file was not found. . . . .	178
FFQD3508W: Exceptions occurred while dropping <i>table_name</i> table. . . . .	174	FFQD3546E: No certificate was loaded. . . . .	178
FFQD3509W: This document is rejected because the URI <i>URI</i> for <i>table_name</i> table ( <i>uri_length</i> byte) is too long. Maximum length is <i>maximum_length</i> byte. . . . .	174	FFQD3547E: Unknown keystore type <i>keystore_type</i> is detected. . . . .	178
FFQD3510W: This document is rejected because it is too large for <i>table_name</i> table (URI : <i>URI</i> , <i>document_size</i> byte). Maximum length is <i>maximum_size</i> byte. . . . .	175	FFQD3548E: Truststore file format or password is not valid. The specified truststore file is <i>truststore_file</i> . . . . .	178
FFQD3511W: This document is rejected because it has too much metadata for <i>table_name</i> table (URI : <i>URI</i> , <i>metadata_size</i> byte). Maximum length is <i>maximum_size</i> byte. . . . .	175	FFQD3549E: An I/O error occurred while closing the truststore file <i>truststore_file</i> . . . . .	179
FFQD3512W: An unexpected row count was returned while inserting into <i>table_name</i> table. Expected count was <i>expected_count</i> , but it was <i>actual_count</i> . . . . .	175	FFQD3550E: Failed to initialize javax.net.ssl.TrustManager. . . . .	179
FFQD3513W: An unexpected row count was returned while updating <i>table_name</i> table. Expected count was <i>expected_count</i> , but it was <i>actual_count</i> . . . . .	175	FFQD3551E: Failed to initialize javax.net.ssl.SSLContext. . . . .	179
FFQD3514W: An unexpected row count was returned while deleting from <i>table_name</i> table. Expected count was <i>expected_count</i> , but it was <i>actual_count</i> . . . . .	175	FFQD3553E: Exceptions occurred while creating connection to the database <i>database_name</i> . . . . .	179
FFQD3515W: Exceptions occurred while querying <i>table_name</i> table. . . . .	176	FFQD4000E: Notes class is not found. Setup for Notes Crawler is required. . . . .	179
FFQD3516W: Exceptions occurred while inserting into <i>table_name</i> table. . . . .	176	FFQD4001E: Notes NRPC class (Notes.jar) is not found. Setup for Notes NRPC crawler is required. . . . .	179
FFQD3517W: Exceptions occurred while updating <i>table_name</i> table. . . . .	176	FFQD4002E: Notes native library <i>library_name</i> for NRPC access was not found. Setup for Notes NRPC crawler is required. . . . .	180
FFQD3518W: Exceptions occurred while deleting from <i>table_name</i> table. . . . .	176	FFQD4003E: Notes DIIOP class (NCS0.jar) is not found. Setup for Notes DIIOP crawler is required. . . . .	180
FFQD3519W: Exceptions occurred while preparing query <i>table_name</i> table. SQL is <i>sql_statement</i> . . . . .	176	FFQD4004E: Cannot determine Notes library name from detected platform <i>platform</i> . . . . .	180
FFQD3520W: Exceptions occurred while preparing insert into <i>table_name</i> table. SQL is <i>sql_statement</i> . . . . .	176	FFQD4005E: Information for server, target, or both were not specified. . . . .	180
FFQD3521W: Exceptions occurred while preparing update <i>table_name</i> table. SQL is <i>sql_statement</i> . . . . .	177	FFQD4006E: Server name <i>server_string</i> in the configuration file does not represent the Lotus Notes server and database. . . . .	180
FFQD3522W: Exceptions occurred while preparing delete from <i>table_name</i> table. SQL is <i>sql_statement</i> . . . . .	177	FFQD4007E: Server name <i>server_string</i> in the configuration file does not represent Lotus Notes database <i>database_name</i> . Database name must be specified. . . . .	180
		FFQD4008E: Connection to the database <i>database_name</i> on the server <i>server_name</i> is already active. . . . .	181
		FFQD4009E: Cannot connect to the Lotus Notes database <i>database_name</i> on server <i>server_name</i> . . . . .	181

FFQD4010E: Cannot change the ID to <i>id_file_name</i> . . . . .	181	FFQD4035E: An error occurred while connecting to database <i>database_name</i> on <i>server_name</i> . Check that you can access the database and that the crawl target is available to user <i>user_name</i> . . . . .	185
FFQD4011E: Cannot disconnect from Lotus Notes database <i>database_name</i> on the server <i>server_name</i> . . . . .	181	FFQD4036E: Cannot connect to database <i>database_name</i> on <i>server_name</i> . Check that you can access the database and that the crawl target is available to user <i>user_name</i> . . . . .	185
FFQD4012E: Cannot change the ID to <i>id_file_name</i> . . . . .	181	FFQD4037E: Cannot select formula-selected documents. Check that you can access the documents and that the formula is based on a Notes formula. Server: <i>server_name</i> . User: <i>user_name</i> . Database: <i>database_name</i> . Formula: <i>formula</i> . . . . .	185
FFQD4013E: Cannot verify next record existence. Server: <i>server_name</i> . Database: <i>database_name</i> . View/Folder: <i>view_folder_name</i> . . . . .	181	FFQD4038E: An error occurred while getting View or DocumentCollection <i>view_or_folder_name</i> from database <i>database_name</i> on server <i>server_name</i> . Check that you can access the database and that the crawl target is available to specified user <i>user_name</i> . . . . .	185
FFQD4014E: Cannot change the ID to <i>id_file_name</i> . . . . .	181	FFQD4039E: Cannot get View or DocumentCollection from database <i>database_name</i> on server <i>server_name</i> . Check that you can access the database and that the crawl target is available to user <i>user_name</i> . View/Folder: <i>view_or_folder_name</i> .	186
FFQD4015E: Cannot create the URI. . . . .	182	FFQD4040E: Cannot get View or DocumentCollection from database <i>database_name</i> on server <i>server_name</i> . Check that you can access the database and that the crawl target is available to user <i>user_name</i> . View/Folder: <i>view_or_folder_name</i> .	186
FFQD4016E: Cannot gather the unchanged attachment names since the last crawl. . . . .	182	FFQD4041E: Cannot initialize Notes thread. 186	
FFQD4017E: Cannot fetch the target content. Server: <i>server_name</i> . Database: <i>database_name</i> . View/Folder: <i>view_folder_name</i> . . . . .	182	FFQD4042E: Cannot obtain current Notes user ID <i>id_filename</i> for server <i>server_name</i> . . . . .	186
FFQD4018E: Undefined record status <i>status_code</i> was found in the internal repository. . . . .	182	FFQD4043E: Connection for server <i>server_name</i> is already active. ID file: <i>id_filename</i> . . . . .	186
FFQD4019E: Cannot obtain the attachment file name. . . . .	182	FFQD4044E: An error occurred while connecting to database <i>database_name</i> on server <i>server_name</i> . Check that you can access the database and that the crawl target is available to user <i>id_filename</i> . . . . .	187
FFQD4020E: Cannot write to the temporary file <i>filename</i> . . . . .	182	FFQD4045E: Cannot connect to database <i>database_name</i> on server <i>server_name</i> . Check that you can access the database and that the crawl target is available to user <i>id_filename</i> . . . . .	187
FFQD4021E: Cannot fetch the attachment file. . . . .	182	FFQD4046E: Cannot select formula-selected documents. Check that you can access the documents and that the formula is based on a Notes formula. Server: <i>server_name</i> . ID file: <i>id_filename</i> . Database: <i>database_name</i> . Formula: <i>formula</i> . . . . .	187
FFQD4022E: An I/O error occurred while reading or writing binary content. . . . .	183		
FFQD4023E: Undefined record status <i>status_code</i> . . . . .	183		
FFQD4024E: Cannot evaluate the string length because of an I/O error. . . . .	183		
FFQD4025W: Cannot obtain last modified date from the Lotus Notes document. 0 will be used instead. . . . .	183		
FFQD4026W: Cannot obtain the creation date from the Lotus Notes document. 0 will be used instead. . . . .	183		
FFQD4027W: Cannot obtain the security field value. . . . .	183		
FFQD4028W: Cannot obtain file name for the data source metadata. . . . .	183		
FFQD4029W: Cannot obtain last modified date from the Lotus Notes document. 0 will be used instead. . . . .	184		
FFQD4030W: Cannot obtain creation date from the Lotus Notes document. 0 will be used instead. . . . .	184		
FFQD4031E: The connection is already active for server <i>server_name</i> using user <i>user_name</i> .	184		
FFQD4032E: Access to the server <i>server_name</i> is denied. Check the server setting that allows Java/CORBA access for user <i>user_name</i> .	184		
FFQD4033E: User ID <i>user_name</i> , password, or both are not valid for server <i>server_name</i> .	184		
FFQD4034E: Cannot create a Notes session for server <i>server_name</i> by user <i>user_name</i> .	184		

FFQD4047E: An error occurred while getting View or DocumentCollection from database <i>database_name</i> on server <i>server_name</i> . Check that you can access the database and that the crawl target is available to user <i>id_filename</i> . View/Folder: <i>view_or_folder_name</i> . . . . .	187	FFQD4074E: Current view or DocumentCollection is null. . . . .	191
FFQD4048E: Cannot get View or DocumentCollection from database <i>database_name</i> on server <i>server_name</i> . Check that you can access the database and that the crawl target is available to user <i>id_filename</i> . View/Folder: <i>view_or_folder_name</i> . . . . .	188	FFQD4075E: Cannot get the universal ID for the View. . . . .	191
FFQD4049E: Cannot get View or DocumentCollection from database <i>database_name</i> on server <i>server_name</i> . Check that you can access the database and that the crawl target is available to user <i>id_filename</i> . View/Folder: <i>view_or_folder_name</i> . . . . .	188	FFQD4076E: Current database is null. . . . .	191
FFQD4050E: View or DocumentCollection is null. . . . .	188	FFQD4077E: Cannot get database replica ID. . . . .	192
FFQD4051E: Cannot fetch Document from View. . . . .	188	FFQD4078E: Cannot gather items to be crawled. . . . .	192
FFQD4052E: Cannot fetch the Document from the View or Folder while comparing with formula-selected Document universal ID. . . . .	188	FFQD4079E: ID file <i>id_filename</i> does not exist. . . . .	192
FFQD4053E: Cannot fetch Document from DocumentCollection. . . . .	189	FFQD4080E: ID authorization failed for the ID file <i>id_filename</i> . . . . .	192
FFQD4054E: Cannot fetch the Document from DocumentCollection while comparing with formula-selected Document universal ID. . . . .	189	FFQD4081E: Cannot create Notes session. . . . .	192
FFQD4055E: Current Document is null. . . . .	189	FFQD4082E: Cannot recycle registration. . . . .	192
FFQD4056E: Cannot obtain created date from current Document. . . . .	189	FFQD4083E: Cannot change ID to <i>id_filename</i> . . . . .	192
FFQD4057E: Current document is null. . . . .	189	FFQD4084E: This session is on DIIOp. . . . .	192
FFQD4058E: Cannot obtain last modified date from current document. . . . .	189	FFQD4085E: Cannot recycle session. . . . .	193
FFQD4059E: Cannot obtain current document size. . . . .	189	FFQD4086E: Cannot recycle current document. . . . .	193
FFQD4060E: Cannot fetch field from current document. . . . .	189	FFQD4087E: Cannot recycle View/DocumentCollection. . . . .	193
FFQD4061E: Cannot recycle Item. . . . .	190	FFQD4088E: Cannot recycle current document. . . . .	193
FFQD4062E: Cannot fetch field from current document. . . . .	190	FFQD4089E: Cannot recycle current EmbeddedObject. . . . .	193
FFQD4063E: Cannot recycle item. . . . .	190	FFQD4090E: Cannot recycle current EmbeddedObject. . . . .	193
FFQD4064E: Current EmbeddedObject is null. . . . .	190	FFQD4091W: A URI that was too long was generated. This row will be ignored. Generated URI: <i>uri</i> . . . . .	193
FFQD4065E: Cannot get InputStream from current EmbeddedObject. . . . .	190	FFQD4200I: Connected to <i>database_URL</i> using user <i>user_name</i> . . . . .	193
FFQD4066E: Current EmbeddedObject is null. . . . .	190	FFQD4201I: Disconnected from <i>database_URL</i> . . . . .	194
FFQD4067E: Cannot get file name from current EmbeddedObject. . . . .	190	FFQD4202W: The metadata is truncated because it exceeds the size limit <i>data_size_limit</i> . . . . .	194
FFQD4068E: Current EmbeddedObject is null. . . . .	190	FFQD4203W: The content is ignored because it exceeds the size limit <i>data_size_limit</i> . . . . .	194
FFQD4069E: Cannot get the file size from current EmbeddedObject. . . . .	191	FFQD4204W: A row in table <i>table_name</i> on database <i>database_URL</i> was ignored because null key is detected in the column <i>column_name</i> . . . . .	194
FFQD4070E: Current document is null. . . . .	191	FFQD4205W: A row was ignored because the URI is too long. URI: <i>URI</i> . . . . .	194
FFQD4071E: Cannot get the form name from current document. . . . .	191	FFQD4207E: The fields specification is not valid. . . . .	194
FFQD4072E: Current document is null. . . . .	191	FFQD4208E: User ID, password, or both are incorrect. The connection to <i>database_URL</i> cannot be established. . . . .	194
FFQD4073E: Cannot get document universal Notes ID. . . . .	191	FFQD4209E: Database <i>database_URL</i> was not found. . . . .	195
		FFQD4210E: Table or view <i>database_URL</i> was not found. . . . .	195
		FFQD4211E: Column was not found. . . . .	195
		FFQD4212E: SQL syntax error was detected. Database: <i>database_URL</i> . SQL: <i>SQL_statement</i> . . . . .	195
		FFQD4213E: User <i>user_name</i> does not have CONNECT or SELECT privilege. Database: <i>database_URL</i> . Table: <i>table_name</i> . . . . .	195
		FFQD4214E: Database connection to <i>database_URL</i> is full. . . . .	195
		FFQD4215E: Connection was terminated from the database <i>database_URL</i> . . . . .	195

FFQD4216E: Connection to <i>database_URL</i> failed. . . . .	196	FFQD4311E: The attribute <i>attribute</i> does not exist or the attribute is BLOB type. . . . .	200
FFQD4217E: Disconnection from <i>database_URL</i> failed. . . . .	196	FFQD4312E: The component <i>component</i> does not exist.. . . . .	200
FFQD4218E: An I/O error occurred. Database: <i>database_URL</i> . Table: <i>table_name</i> . . . . .	196	FFQD4313E: An error occurred while fetching attributes.. . . . .	200
FFQD4219E: Database error occurred. Database: <i>database_URL</i> . SQLCODE: <i>SQLCODE</i> . SQLSTATE: <i>SQLSTATE</i> . . . . .	196	FFQD4314E: An error occurred while retrieving the last modified date.. . . . .	200
FFQD4220E: There is no database connection.	196	FFQD4315E: An error occurred while retrieving a creation date. . . . .	201
FFQD4221E: Columns to be crawled are not specified. Database: <i>database_URL</i> . Table: <i>table_name</i> . . . . .	196	FFQD4316E: An error occurred while parsing a date <i>date</i> .. . . . .	201
FFQD4222E: Security field must be a string column. Database: <i>database_URL</i> . Table: <i>table_name</i> . Column: <i>column_name</i> . Type: <i>security_type</i> .. . . . .	196	FFQD4317E: An error occurred while retrieving an item type <i>item_type</i> .. . . . .	201
FFQD4223E: Date field must be a DATE or TIMESTAMP column. Database: <i>database_URL</i> . Table: <i>table_name</i> . Column: <i>column_name</i> . Type: <i>date_type</i> . . . . .	197	FFQD4318E: An error occurred while retrieving an original file name. . . . .	201
FFQD4224E: Unsupported data type was found. Database: <i>database_URL</i> . Table: <i>table_name</i> . Column: <i>column_name</i> . Type: <i>data_type</i> . . . . .	197	FFQD4319E: The attribute <i>attribute</i> specified as the field to control access does not exist. . . . .	201
FFQD4225E: DATE or TIMESTAMP column is specified as content. Database: <i>database_URL</i> . Table: <i>table_name</i> . Column: <i>column_name</i> . . . . .	197	FFQD4320E: The attribute <i>attribute</i> specified as the document date is incorrect. . . . .	201
FFQD4226E: A nonnumeric column (including BIGINT) is specified as a parametric searchable field. Database: <i>database_URL</i> . Table: <i>table_name</i> . Column: <i>column_name</i> . Type: <i>data_type</i> . . . . .	197	FFQD4321E: The attribute <i>attribute</i> specified as the document date does not exist.. . . . .	202
FFQD4227E: A non-binary column is specified as content. Database: <i>database_URL</i> . Table: <i>table_name</i> . Column: <i>column_name</i> . Type: <i>data_type</i> . . . . .	198	FFQD4322E: An error occurred while retrieving a document date. Attribute: <i>attribute</i> . . . . .	202
FFQD4228E: Multiple content is not supported. Database: <i>database_URL</i> . Table: <i>table_name</i> . . . . .	198	FFQD4323W: A URI that was too long was generated. This row will be ignored. Generated URI: <i>uri</i> . . . . .	202
FFQD4300E: An internal error occurred (another connection exists).. . . . .	198	FFQD4324I: Connected to server. Database: <i>database</i> . User: <i>user</i> . . . . .	202
FFQD4301E: The item type <i>item_type</i> does not exist or the item type is a part item type. . . . .	198	FFQD4325I: Disconnected from server. . . . .	202
FFQD4302E: Connection failed. Database: <i>database</i> . User: <i>user</i> . . . . .	198	FFQD4326E: Resource Manager is not online or available. . . . .	202
FFQD4303E: Disconnection failed.. . . . .	199	FFQD4327W: An error occurred while fetching content. Skip this document. PID: <i>pid</i> . . . . .	202
FFQD4304E: An error occurred while fetching content. PID: <i>pid</i> .. . . . .	199	FFQD4402W: Property <i>property_name</i> cannot be changed now. Restart the crawler. . . . .	203
FFQD4305E: An error occurred while creating a temporary file.. . . . .	199	FFQD4409W: NNTP crawler will be paused because the number of articles exceeds the limit <i>max_num_of_articles</i> . . . . .	203
FFQD4306E: Database connection does not exist.. . . . .	199	FFQD4413W: Crawler will sleep <i>sleep_interval</i> seconds because of an error. . . . .	203
FFQD4307E: An error occurred while executing a query command <i>cmd</i> . . . . .	199	FFQD4414E: Updating the server status failed. Crawler will try again later.. . . . .	203
FFQD4308E: An error occurred while fetching a record. PID: <i>pid</i> . . . . .	199	FFQD4415E: Updating the newsgroup status failed. Crawler will try again later.. . . . .	203
FFQD4309E: An error occurred while fetching a part. PID: <i>pid</i> . . . . .	200	FFQD4416E: Inserting the message ID failed. Crawler will try again later. . . . .	203
FFQD4310E: The attribute <i>attribute</i> does not exist.. . . . .	200	FFQD4417E: Inserting the news article failed. The crawler will try again later. . . . .	204
		FFQD4418E: Inserting deleted news article failed. Crawler will try again later.. . . . .	204
		FFQD4419W: Crawler will be stopped despite failure to update the server status. . . . .	204
		FFQD4420W: Crawler will be stopped despite failure to update newsgroup status. This might cause to recrawl some articles.. . . . .	204

FFQD4421W: Crawler will be stopped despite failure to update message id. This might cause to recrawl some articles. . . . .	204
FFQD4422W: Crawler will be stopped despite failure to insert article. . . . .	204
FFQD4423W: Crawler will be stopped despite failure to insert deleted article. . . . .	205
FFQD4427E: The newsgroup list could not be retrieved because of an unknown host <i>host_name</i> . . . . .	205
FFQD4428E: The newsgroup list could not be retrieved because of a network error or server error on <i>host_name</i> . . . . .	205
FFQD4433W: Cleaning was interrupted. This process will be restarted next time. . . . .	205
FFQD4434W: Message ID was not found in this article <i>message_id</i> . This article will be ignored. . . . .	205
FFQD4435W: Message ID is inconsistent ( <i>message_id1</i> and <i>message_id2</i> must be identical). This article will be ignored. . . . .	205
FFQD4436W: Unexpected date format <i>date</i> . . . . .	206
FFQD4437E: I/O Error occurred while parsing article. . . . .	206
FFQD4449E: Errors occurred in <i>api_name</i> . . . . .	206
FFQD4450E: File <i>file_name</i> was not found. . . . .	206
FFQD4451E: Network I/O error occurred. . . . .	206
FFQD4452E: Unexpected response <i>server_response</i> was returned for command <i>nntp_command</i> . . . . .	206
FFQD4453W: The status of article pointer could not be retrieved. Return code was <i>return_code</i> . . . . .	206
FFQD4454W: The article could not be retrieved. Return code was <i>return_code</i> . . . . .	207
FFQD4455W: The crawler could not move to the next article. Return code was <i>return_code</i> . . . . .	207
FFQD4456W: The crawler could not switch to newsgroup <i>newsgroup_name</i> on <i>server_name</i> . Return code was <i>return_code</i> . . . . .	207
FFQD4457W: The newsgroup list could not be retrieved. Return code was <i>return_code</i> . . . . .	207
FFQD4458E: Server name <i>server_name</i> could not be resolved. . . . .	207
FFQD4461E: The crawler received service disconnected message from server. . . . .	207
FFQD4500E: The <i>directory_name</i> directory was not found. . . . .	208
FFQD4501E: The <i>directory_name</i> directory was not found. Error message : <i>detail_message</i> . . . . .	208
FFQD4502E: Failed to open <i>filename</i> file. The file might be used by another program. . . . .	208
FFQD4503E: An exception occurred while reading <i>filename</i> file. . . . .	208
FFQD4504E: The subdirectory level <i>levels of subdirectories</i> is incorrect. . . . .	208
FFQD4505W: The length of directory name is too long to be stored. Directory: <i>directory_name</i> . . . . .	208

FFQD4506W: The length of the URI name is too long to be stored. URI: <i>directory_name</i> . . . . .	208
FFQD4507E: The network drive cannot be used. Error message : <i>detail_message</i> . . . . .	209
FFQD4508E: The user ID <i>user_name</i> or the password is incorrect. Network folder: <i>directory_name</i> . . . . .	209
FFQD4509E: The temporary file <i>temporary_file</i> was not created. . . . .	209
FFQD4510E: The temporary file <i>temporary_file</i> failed to be loaded. . . . .	209
FFQD4511E: The temporary file <i>temporary_file</i> failed to be stored. . . . .	209
FFQD4600E: An I/O error occurred while receiving a response from the URL <i>public_folder_URL</i> . . . . .	209
FFQD4601E: An XML parser configuration error occurred. . . . .	209
FFQD4602E: WebDAV response from URL <i>public_folder_URL</i> is an incorrect XML document. . . . .	210
FFQD4603I: Connected to <i>public_folder_URL</i> using user <i>user_name</i> . . . . .	210
FFQD4604I: Connected to <i>public_folder_URL</i> using user <i>user_name</i> and truststore file <i>truststore_file</i> . . . . .	210
FFQD4605I: Disconnected from <i>public_folder_URL</i> . . . . .	210
FFQD4607E: Connecting to <i>public_folder_URL</i> failed. . . . .	210
FFQD4608E: Checking next record failed. . . . .	210
FFQD4609E: Error occurred while getting content from URL <i>uniform_resource_locator</i> . . . . .	210
FFQD4610E: An I/O error occurred while getting content from the URL <i>public_folder_URL</i> . . . . .	211
FFQD4611E: Could not crawl the public folder <i>public_folder_URL</i> . The HTTP Status is <i>HTTP_state_code</i> . . . . .	211
FFQD4612E: WebDAV request to <i>public_folder_URL</i> failed. The returned status code and message is <i>HTTP_state_code</i> , <i>HTTP_response_message</i> . . . . .	211
FFQD4613W: Time format error was detected while parsing the value <i>string</i> . . . . .	211
FFQD4614W: The specified URL <i>document_URL</i> already exists. . . . .	211
FFQD4616E: Temporary file <i>temporary_file</i> does not exist. . . . .	211
FFQD4617E: An I/O error occurred while writing objects into the temporary file <i>temporary_file</i> . . . . .	211
FFQD4618E: An I/O error occurred while closing object output stream to the file <i>temporary_file</i> . . . . .	212
FFQD4619E: An I/O error occurred while closing the file <i>temporary_file</i> . . . . .	212
FFQD4620E: An I/O error occurred while reading objects from the temporary file <i>temporary_file</i> . . . . .	212

FFQD4621E: An I/O error occurred while closing object input stream from the file <i>temporary_file</i> .	212
FFQD4622E: An unknown host name <i>host_name</i> was detected.	212
FFQD4623E: The URL <i>public_folder_URL</i> is incorrect.	212
FFQD4624E: Could not refresh connection to <i>public_folder_URL</i> .	212
FFQD4625E: An I/O error occurred while sending a request to <i>public_folder_URL</i> .	213
FFQD4626E: Connecting to <i>public_folder_URL</i> failed.	213
FFQD4627E: A truststore file must be specified to use SSL.	213
FFQD4628E: A truststore password must be specified to use SSL.	213
FFQD4629E: The truststore file <i>truststore_file</i> does not exist.	213
FFQD4631E: Truststore file format or password is incorrect. The specified truststore is <i>truststore_file</i> .	213
FFQD4632W: A document was ignored because the URI is too long. URI: <i>URI</i> .	213
FFQD4633W: A folder was ignored because the URI is too long. URI: <i>URI</i> .	214
FFQD4707E: Server name <i>server_string</i> in configuration file does not represent VeniceBridge server and repository name.	214
FFQD4708E: VeniceBridge API error occurred.	214
FFQD4709E: VeniceBridge API is not licensed.	214
FFQD4710E: VeniceBridge server instance cannot be created.	214
FFQD4711E: VeniceBridge API configuration error occurred.	214
FFQD4712E: VeniceBridge server cannot be located.	214
FFQD4713E: Communication related error occurred.	214
FFQD4714E: The repository <i>repository_name</i> is not found.	215
FFQD4715E: The VeniceBridge user instance was not initialized.	215
FFQD4716E: Logon to <i>repository_name</i> repository failed.	215
FFQD4718E: The item class <i>item_class</i> was not found for repository <i>repository_name</i> .	215
FFQD4719E: The property <i>property</i> was not found. Repository: <i>repository_name</i> . Item class: <i>item_class</i> .	215
FFQD4720E: The property <i>property</i> for ranking date was not found. Repository: <i>repository_name</i> . Item class: <i>item_class</i> .	215
FFQD4721E: The property <i>property</i> for security field is not found. Repository: <i>repository_name</i> . Item class: <i>item_class</i> .	215
FFQD4722E: The specified selection criteria <i>condition_statement</i> has a problem.	216

FFQD4723E: Ranking date field must be DATE or DATETIME data type. Repository: <i>repository_name</i> . Item class: <i>item_class</i> . Property: <i>property</i> . Type: <i>data_type</i> .	216
FFQD4724E: An unsupported data type <i>data_type</i> was found. Repository: <i>repository_name</i> . Item class: <i>item_class</i> . Property: <i>property</i> .	216
FFQD4725E: The executed query has a problem.	216
FFQD4726E: No order key is found. Key: <i>order_key_property</i> .	216
FFQD4728W: A URI that was too long was generated. This record will be ignored. URI: <i>uri</i> .	216
FFQD4729W: The property <i>property</i> was not found.	217
FFQD4730E: An I/O error occurred while writing content.	217

## Chapter 19. Glossary . . . . . 219

### DB2 Information Integrator documentation . . . . . 225

Documentation about event publishing function for DB2 Universal Database on z/OS	225
Documentation about event publishing function for IMS and VSAM on z/OS	226
Documentation about event publishing and replication function on Linux, UNIX, and Windows	226
Documentation about federated function on Linux, UNIX, and Windows	227
Documentation about federated function on z/OS	228
Documentation about replication function on z/OS	229
Documentation about enterprise search function on Linux, UNIX, and Windows	230
Release notes and installation requirements	230
Viewing release notes and installation requirements	231
Viewing and printing PDF documentation.	232
Accessing DB2 Information Integrator documentation	232

### Accessibility . . . . . 235

Keyboard input and navigation	235
Keyboard focus.	235
Keyboard input.	235
Keyboard navigation.	235
Accessible display.	235
Font settings.	236
Non-dependence on color	236
Compatibility with assistive technologies	236
Accessible documentation	236

### Notices . . . . . 237

Trademarks	239
------------	-----

### Contacting IBM . . . . . 241

### Obtaining product information . . . . . 243



**Providing comments on the  
documentation. . . . . 245**

**Index . . . . . 247**



---

## About this book

Use this book for help with administering DB2 Information Integrator OmniFind Edition, Version 8.2 (enterprise search) collections.

DB2 Information Integrator OmniFind Edition provides a technology called enterprise search. The enterprise search components are installed when you install IBM DB2 Information Integrator OmniFind Edition DB2 II OmniFind). The term *enterprise search* is used in the documentation for DB2 II OmniFind unless references are made to CD labels or specific product components.

This book covers the following topics:

- An introduction to enterprise search concepts, components, and capabilities
- Instructions on how to create, monitor, and administer collections
- Information about enterprise search linguistic capabilities
- Information about how enterprise search searches for data and returns search results
- Information about enabling custom search applications to search enterprise search collections
- An overview of the different levels of security that are available in enterprise search
- Information about how enterprise search integrates with IBM WebSphere Portal
- Instructions on how to migrate WebSphere Portal collections to enterprise search
- Instructions for backing up and restoring the system
- Instructions for creating and viewing log files
- Explanations of enterprise search messages

---

## Who should read this book

This book is intended for system administrators and system operators who are responsible for creating, monitoring, and administering enterprise search collections.

Use this book to create collections, select content for the collection, and configure options for making the content searchable. Also use this book to monitor collection and system activity, enroll users as enterprise search administrators, and associate search applications with collections.

This book assumes that you are familiar with Web applications, and that you have experience with the data sources that you want to make searchable with enterprise search.



---

## Chapter 1. What is enterprise search?

An enterprise search system provides extensive capabilities for searching any number of structured and unstructured data sources with a single query. The enterprise search system provides fast query response times and a consolidated, ranked result set that enables you to quickly and easily locate the information that you need.

The enterprise search components, which are installed with IBM® DB2® Information Integrator OmniFind™ Edition (DB2 II OmniFind), collect information from throughout your enterprise and make that information available for searching. By entering a query in a Web browser, you can search local and remote databases, collaboration systems, content management systems, file systems, and internal and external Web sites at the same time.

Designed to integrate seamlessly with your existing systems, the enterprise search components handle the logistics that are required to collect data from diverse sources and index the data for fast retrieval. By applying linguistic analysis and other types of analysis to the data, enterprise search can deliver highly relevant search results. You do not need to learn different interfaces to search various repository types.

Security is an integral element for enterprise search. Only users who are authorized to administer the system can do so. With the security mechanisms available in IBM WebSphere® Application Server, you can configure administrative roles and authenticate all administrative users. By configuring administrative roles, you control which users have access to various administrative functions.

You can also specify options to associate security tokens with data when the data is being collected. If your search applications enable security, you can use these tokens to enforce access controls and ensure that only users with the proper credentials are able to query the data and view search results.

### **Related concepts**

Chapter 12, “Enterprise search security,” on page 111

Security mechanisms in enterprise search enable you to protect sources from unauthorized searching and restrict administrative functions to specific users.

---

## Data source types supported by enterprise search

Predefined support is available for searching a variety of data source types.

After installing IBM DB2 Information Integrator OmniFind Edition, you can begin collecting data from the following types of data sources:

- IBM DB2 Content Manager item types (documents, resources, and items)
- IBM DB2 Information Integrator nickname tables for IBM DB2 Universal Database™ for z/OS®, IBM Informix®, Oracle, and federated databases
- IBM DB2 Universal Database databases
- IBM Lotus Notes® databases
- Microsoft® Exchange Server public folders
- Microsoft Windows® file systems

- Network news transfer protocol (NNTP) news groups
- UNIX<sup>®</sup> file systems
- Venetica VeniceBridge repositories (item classes for Documentum and FileNet Panagon Content Services)
- Web sites (on the Internet or within your intranet)

With the application programming interface (API) for enterprise search, you can extend support to other types of sources, such as proprietary and legacy databases that are not listed here.

#### **Related concepts**

Data listener API

---

## **Enterprise search component overview**

The enterprise search components collect data from throughout your enterprise; analyze, parse, and categorize the information; and create an index that users can search.

An enterprise search *collection* represents the set of sources that users can search with a single query. When you create a collection, you specify which sources you want to include and configure options for how users can search the indexed data.

You can create multiple collections, and each collection can contain data from a variety of data sources. For example, you might create a collection that includes documents from IBM DB2 Universal Database, IBM Lotus Notes, and IBM DB2 Content Manager databases. When users search this collection, the search results potentially include documents from each of the data sources.

Creating and administering a collection involves the following activities:

#### **Collecting data**

The *crawler* components collect documents from data sources, either on a continual basis or according to a schedule that you specify. Frequent crawling ensures that users always have access to the latest information.

#### **Analyzing data**

The *parser* components extract text from documents, and do linguistic analysis and other types of analysis on each document that a crawler crawls. The detailed content analysis improves the quality of search results.

#### **Indexing data**

The *index* components run on a regularly scheduled basis to add information about new and changed documents to the index. The index components also do global analysis of the documents in a collection to enhance the quality of the search results.

#### **Searching data**

The *search* components search the index and work with your search applications to process search requests and return search results.

An IBM DB2 Information Integrator OmniFind Edition installation includes several other components that enable you to easily configure the sources that you want to search, monitor system activity, and troubleshoot any problems that occur. The product also provides a working sample search application that you can use as a template for creating your own search applications.

#### **Related concepts**

Chapter 4, “Enterprise search crawler administration,” on page 23  
You configure crawlers for the different types of data that you want to include in a collection. A single collection can contain any number of crawlers.

Chapter 7, “Linguistic capabilities in enterprise search,” on page 65  
Enterprise search offers linguistic search support for text documents in most Indo-European languages and Asian languages such as Japanese.

Chapter 5, “Enterprise search index administration,” on page 53  
To ensure that users always have access to the latest information, enterprise search creates an index for each collection and maintains that index by periodically refreshing and reorganizing the content.

Chapter 8, “How enterprise search searches collections,” on page 71  
Search applications for enterprise search enable users to specify simple, free-text queries and more complex queries that provide much finer document matching criteria.

## Enterprise search crawlers

Enterprise search crawlers collect documents from data sources so that the documents can be analyzed, indexed, and searched.

The crawler component that is provided with IBM DB2 Information Integrator OmniFind Edition has the following functions:

- When you configure a crawler, the *discovery* processes find information about the sources that are available to be crawled, such as the names of all the views and folders in a Lotus Notes database or the names of all file systems on a UNIX server.
- After you select the sources that you want to crawl and start the crawler, the crawler components collect data from the sources so that the data can be analyzed and indexed.

A single collection can have multiple crawlers, and each crawler is designed to gather data from a particular type of data source. For example, you might create three crawlers to combine data from file systems, Notes® databases, and relational databases in the same collection. Or, you might create several crawlers of the same type, and set up different crawling schedules for them according to how frequently the data that is being crawled by each crawler changes.

Some crawlers, such as those for Web and NNTP sources, run continuously. After you specify which uniform resource locators (URLs) or NNTP news groups you want to crawl, the crawler returns periodically to check for data that is new and changed. You can start and stop other types of crawlers manually, or you can set up crawling schedules. If you schedule a crawler, you specify when it is to run initially and how often it needs to visit the data sources to crawl new and changed documents.

*Crawler properties* are a set of rules that govern the behavior of a particular crawler when it crawls. For example, you specify rules to control how the crawler uses system resources. The set of sources that is eligible to be crawled constitutes the *crawl space* of a crawler. After you create a crawler, you can edit the crawler properties at any time to alter how the crawler collects data. You can also edit the crawl space to change the crawler schedule, add new sources, or remove sources that you no longer want to search.

### Related concepts

Chapter 4, “Enterprise search crawler administration,” on page 23  
You configure crawlers for the different types of data that you want to include in a collection. A single collection can contain any number of crawlers.

## Enterprise search parsers

An enterprise search parser analyzes documents that were collected by a crawler and prepares them for indexing.

The parser component that is provided with IBM DB2 Information Integrator OmniFind Edition analyzes document content and document metadata. It stores the results of the analysis in a data store for access by the indexing component. The parser does the following tasks:

- Extracts text from whatever the format a document is in. For example, the parser extracts text from the tags in XML and HTML documents. By using Stellent for IBM DB2 Information Integrator OmniFind Edition Outside In Viewer Technology, the parser also extracts text from binary formats such as Microsoft Word and Adobe Acrobat portable document format (PDF) documents.
- Detects the character set encoding of each document. Before doing any linguistic analysis, the parser uses this information to convert all text to Unicode.
- Detects the source language of each document.
- Applies parsing rules that you specify for the collection. When you configure the parser, you can configure:

### Field mapping rules for XML documents

This feature enables users to search structured and unstructured content in XML documents. If you map XML elements to search fields, users can specify the field names in queries and search specific parts of XML documents. (Queries that search specific fields can provide more precise search results than free text queries that search all document content.)

### Categories

This feature enables users to search documents by the categories that the documents belong to. Users can also select categories in the search results and browse only documents that belong to that same category.

When you create a collection, you choose the type of categories that you want to use, if any. If you use *rule-based* categories, documents are associated with categories according to rules that you define. You can configure rule-based categories with enterprise search collections that you create and with collections that you migrate from IBM WebSphere Portal.

If you use *model-based* categories, documents are associated with model-based categories that exist in your WebSphere Portal system. To use this option, WebSphere Portal must be installed on the enterprise search index server. You must also use the categorization tools in WebSphere Portal to administer the categories.

- Extracts text and adds tokens to enhance the retrievability of data. During this phase, the parser does the following tasks:
  - Character normalization, such as normalizing capitalization and diacritical marks such as the German umlaut.
  - Analyzing the structure of paragraphs, sentences, words, and white space. Through linguistic analysis, the parser decomposes compound words and assigns tokens that enable dictionary and synonym lookup.

### Related concepts



Chapter 7, “Linguistic capabilities in enterprise search,” on page 65  
Enterprise search offers linguistic search support for text documents in most Indo-European languages and Asian languages such as Japanese.

“Working with categories” on page 75

Categories enable you to group documents that share common characteristics, and search and retrieve only documents that meet the criteria for being members of that group.

“XML search fields” on page 82

XML search fields enable users to query specific parts of XML documents.

#### **Related tasks**

“Monitoring the parser” on page 105

Monitor the parser for a collection when you need to start the parser or stop it.

## **Enterprise search indexes**

The enterprise search indexing components run on regular schedules to add information about new and changed documents to the index.

The indexing components for enterprise search are installed with IBM DB2 Information Integrator OmniFind Edition.

To ensure that users always have access to the latest information in the sources that they search, building an index involves two stages:

### **Reorganizing the index**

When an index is reorganized, the entire index is rebuilt so that the structure has an optimal organization. The indexing processes read all of the data that was collected by crawlers and analyzed by the parser.

### **Refreshing the index**

When an index is refreshed, information that was crawled since the last time the index was reorganized is added to the index.

When you configure index options for a collection, you can specify schedules for reorganizing and refreshing the index. The frequency with which you reorganize and refresh the index depends on your system resources and whether the sources being indexed contain static or dynamic content.

To ensure the availability of new information, schedule the index to be refreshed frequently. Periodically schedule a reorganization of the index to consolidate all of the new information, analyze new content, and optimize the performance of the index.

Only one index per collection can be in the index queue at a time, but you can queue multiple indexes from multiple collections and allow the indexes to be built concurrently. Building indexes concurrently helps ensure that the reorganization of a very large index does not block the refreshing of other indexes. Index building can be a resource-intensive process, so for large systems, you must monitor system loads to adjust reorganization and refresh rates.

You can also start the indexing processes without scheduling them. For example, if you change certain parsing rules and want those changes to become available to your search applications, you can start an index reorganization after the data is recrawled and parsed instead of waiting for the time that the index reorganization is scheduled to start.

When building an index, the indexing processes do global document analysis. During this phase, algorithms are applied to identify duplicate documents, to analyze the link structure of documents, and to do special processing on anchor text (the text that describes the target page in a hypertext link) in Web documents.

After an index is built, you can configure scopes. A *scope* enables you to limit what users can see in the collection. For example, you might create one scope that includes the URIs for documents in your Technical Support department and another scope for the URIs of documents in your Human Resources department. If the search application supports this feature, users can search and retrieve documents from only those subsets of the collection.

#### **Related concepts**

Chapter 5, “Enterprise search index administration,” on page 53

To ensure that users always have access to the latest information, enterprise search creates an index for each collection and maintains that index by periodically refreshing and reorganizing the content.

“Scopes” on page 80

Scopes enable you to restrict the documents that users can search when they search a collection.

#### **Related tasks**

“Scheduling index builds” on page 54

You can specify schedules for reorganizing an index and refreshing an index with new content.

“Configuring concurrent index builds” on page 56

Multiple indexes can share the indexing processes. By sharing the processes, you can ensure that the reorganization of a very large index does not block the availability of other indexes that are waiting in the queue to be built.

“Monitoring index activity for a collection” on page 106

Monitor the index for a collection when you need to see the progress of an index that is being built, enable or disable the index schedule, or start and stop indexing activity.

“Monitoring the enterprise search index queue” on page 106

You can view the status of all index builds in the index queue, stop an index that is being built, or delete an index from the queue.

## **Search servers for enterprise search**

The search servers for enterprise search work with your search applications to process queries, search the index, and return search results.

The search servers for enterprise search are installed when you install IBM DB2 Information Integrator OmniFind Edition. When you configure the search servers for a collection, you can specify options for how the collection is to be searched:

- You can configure a search cache to hold frequently requested search results. A search cache can improve search and retrieval performance.
- If you predetermine that certain documents are relevant to certain queries, you can configure quick links. A *quick link* associates a specific URI with specific keywords and phrases. If a query contains any of the keywords or phrases that specify in a quick link definition, the associated URI is returned automatically in the search results.
- If your crawlers add security tokens to documents (security tokens enable search applications to enforce access controls), you can disable security if you decide that you want to allow all users to search the collection.

- You can specify a default language for searching documents in the collection.

#### **Related concepts**

Chapter 6, “Search applications for enterprise search,” on page 59

A search application enables you to search collections in your enterprise search system. You can create any number of search applications, and a single search application can search any number of collections.

“Search cache” on page 109

When the load on the search servers is relatively high, you can enhance performance by caching search results.

“Document-level security” on page 114

Document-level security ensures that users who search collections are able to access only the documents that they are allowed to see.

“Quick links” on page 73

Quick links enable you to provide users with links to specific documents.

#### **Related tasks**

“Monitoring the search servers” on page 107

You can view detailed status information about search server activity for a specific collection, or view detailed status information for the search servers throughout your enterprise search system.

“Allowing users to access all documents” on page 118

You can specify an option that enables all users to search all documents in the collection, regardless of whether security tokens are associated with the documents.

## **Enterprise search administration console**

The enterprise search administration console runs in a browser, which means administrative users can access it from any location at any time. Security mechanisms ensure that only those users who are authorized to access administrative functions do so.

The administration console for enterprise search is installed on the search servers when you install IBM DB2 Information Integrator OmniFind Edition.

The administration console includes wizards that can help you do several of the primary administrative tasks. For example, the Collection wizard helps you create a collection and allows you to save your work in draft mode. Crawler wizards are specific to a data source type and help you select the sources that you want to include and make searchable in your collections.

For other administrative tasks, you can select individual items that you want to administer. For example, when you edit a collection, you can go directly to the Index page to change the index schedule or go directly to the Parse page to modify a rule for searching XML documents.

#### **Related concepts**

Chapter 2, “Enterprise search administration overview,” on page 13

You use the enterprise search administration console to create and administer collections, start and stop components, monitor system activity and log files, configure administrative users, and associate search applications with collections.

“Administrative roles” on page 112

Enterprise search uses the concept of roles to control access to various functions in the administration console.

### **Related tasks**

“Logging in to the administration console” on page 15

To administer an enterprise search system, you specify a URL in a Web browser and then log in to the administration console.

## **Monitoring an enterprise search system**

You can use the enterprise search administration console to monitor system activities and adjust operations as needed.

After you install IBM DB2 Information Integrator OmniFind Edition and create at least one collection, you can view detailed statistics for each major activity (crawling, parsing, indexing, and searching). The information includes average response times and progress information, such as how many documents were crawled or indexed during a specific crawl or index building session.

You can stop and start most activities. For example, you can pause an activity, change its configuration or troubleshoot a problem, and restart processing when you are ready to allow the activity to proceed.

### **Related concepts**

Chapter 11, “Monitoring enterprise search activity,” on page 93

When you monitor system and collection activities, you can view the status of various processes, watch for potential problems, or adjust configuration settings to enhance performance.

Chapter 10, “Starting and stopping enterprise search,” on page 89

After you create a collection, you must start the components for crawling, parsing, indexing, and searching data. Stop and restart the components when you need to notify the system about changes in the collection.

## **Enterprise search log files**

Log files are created for individual collections and for the system-level components.

When you configure logging options for an enterprise search collection or for the system, you specify the types of messages that you want to log (such as error messages and warning messages). You also specify how often you want the system to rotate older log files to make room for recent messages. You can choose options to receive e-mail about specific messages, or all error messages, whenever they occur.

When you view log files, you select the log file that you want to view (the file name includes information about when the file was created and which component issued the messages). You can also specify viewing filters. For example, you can choose to see only error messages or only messages from a particular component.

### **Related concepts**

Chapter 16, “Enterprise search log files,” on page 133

You can choose the types of messages that you want to log for a collection and for the system, specify options for creating and viewing log files, and specify options for receiving e-mail about messages.

### **Related tasks**

“Configuring log files” on page 133

You can specify the types of messages that you want to log and specify options for creating log files.

“Receiving e-mail about logged messages” on page 134

You can specify options to receive e-mail whenever certain messages, or certain types of messages, are logged.

“Viewing log files” on page 136

You can view log messages that the system and collection components write to a common log file. You can also specify filters to view messages of a specific severity level and messages from specific components.

## Enterprise search APIs

The application programming interfaces for enterprise search enable you to create custom search applications and use external crawlers to update the content of collections.

After installing IBM DB2 Information Integrator OmniFind Edition, the following families of APIs are available for extending enterprise search collections:

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

Use this API to build custom search applications.

### Data listener API

Use this API to receive data from external crawlers. The external crawlers can connect to the enterprise search data listener component, then add data to a collection or remove data from a collection.

### Related concepts

Search and Index API (SI-API)

Data listener API

Chapter 6, “Search applications for enterprise search,” on page 59

A search application enables you to search collections in your enterprise search system. You can create any number of search applications, and a single search application can search any number of collections.

### Related tasks

“Configuring support for external crawlers” on page 50

You can extend enterprise search by using the data listener API to create an external crawler. To enable the crawler to access enterprise search collections, you configure the data listener.

## Sample search application for enterprise search

You can use the sample search application for enterprise search as a template for developing custom search applications.

A sample search application is installed when you install IBM DB2 Information Integrator OmniFind Edition. This sample search application demonstrates most of the search and retrieval functions that are available for enterprise search. The application is a working example that enables you to search all active collections in your enterprise search system. With this feature, you can easily test new collections before you make them available to users.

The sample search application includes options for specifying simple queries or queries with advanced options, such as options for searching categories or specifying how many documents can appear on a result page. If a collection includes documents in multiple languages, you can restrict the result set by specifying which languages you want to search. You can also choose to see a summary of the results or details about each result document.

For information about using the sample search application, click **Help** while you are using the application. To create your own search applications, you use the search and index API for enterprise search.

#### **Related concepts**

Chapter 6, “Search applications for enterprise search,” on page 59

A search application enables you to search collections in your enterprise search system. You can create any number of search applications, and a single search application can search any number of collections.

Search and Index API (SI-API)

#### **Related tasks**

“Configuring default search application properties” on page 61

The sample search application for enterprise search can search all active collections in your system. You can edit a properties file to specify options for your Web server environment or to specify a different search application as the default application.

“Accessing the sample search application” on page 62

You access the sample search application for enterprise search by specifying a URL in a Web browser.

---

## **The enterprise search data flow**

A collection consists of data that is collected by crawlers, analyzed by a parser, and built into a searchable index.

The enterprise search components that you install with IBM DB2 Information Integrator OmniFind Edition closely interact to ensure the flow of data through the system. Crawlers gather documents from data sources throughout your enterprise. The parser extracts useful information from the crawled documents and generates tokens that can, for example, associate documents with categories and help determine the relevance of documents to the terms in a search request. The index stores the data for efficient retrieval.

By using a Web browser and a search application, users search indexed collections. The search application can display a list of results that users can click in a browser, or the application can be more sophisticated and return dynamically generated content that is based on information in different sources.

For example, a catalog search application can customize the display of products that satisfies a search request. A single query can search through documents from different types of data sources, such as a combination of documents from IBM DB2 Content Manager and Lotus Notes repositories.

Administrators determine what data will be collected and how it will be crawled, parsed, indexed, and searched. By monitoring system activity, administrators also make adjustments to optimize data throughput.

The following diagram shows the flow of information through an enterprise search system.

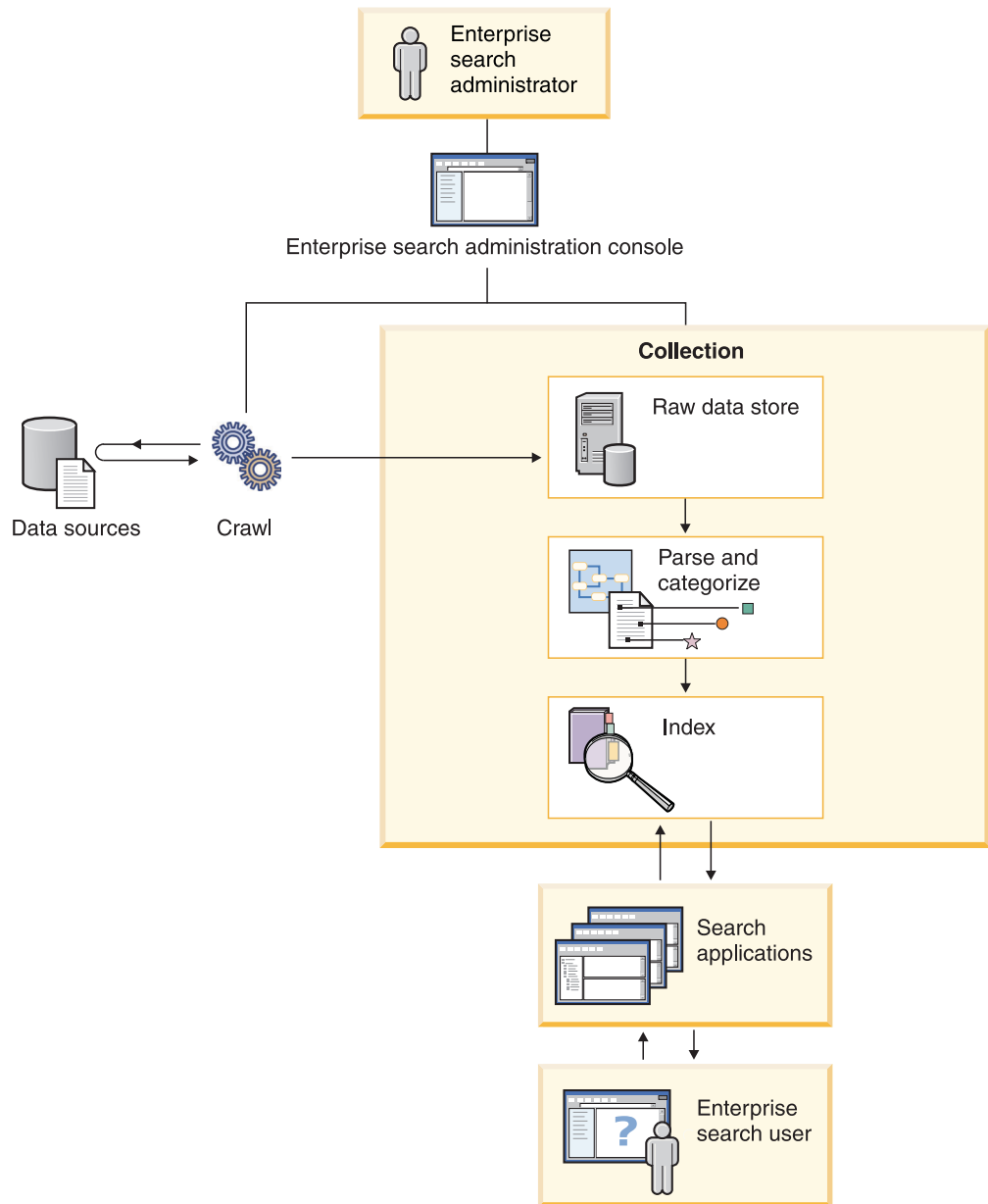


Figure 1. How data flows through an enterprise search system





---

## Chapter 2. Enterprise search administration overview

You use the enterprise search administration console to create and administer collections, start and stop components, monitor system activity and log files, configure administrative users, and associate search applications with collections.

### Collections view

Use the following steps as a guideline to creating your first collection and administering the system.

1. Log in to the enterprise search administration console.

The Collections view is the entry point for creating and administering collections.

2. To create a collection by using the collection wizard, click **Collection Wizard** and follow the wizard prompts.
3. To create a collection by using the Collections view, click **Create Collection**, fill in the fields on the Create a collection page, then click **OK**.

You must edit a new, empty collection to add content to it. You must also specify options for how you want to crawl data and make the data searchable.

4. Click  **Edit** for your new collection, then select pages to configure the collection.

- Optional: On the General page, you can specify additional information for this collection. For example, if your collection contains documents in Japanese, Chinese, or Korean, you can select a different type of lexical analysis. You can select other tasks on the General page:

- If you need to prevent users from seeing certain documents after the collection is created, click **Remove URIs from the index**.
- You can click **Configure alerts** and specify options for receiving alerts about collection activity. For example, an alert can inform you when the average search response time is exceeding a specified limit.

- Required: On the Crawl page, configure at least one crawler.

A single collection can contain data from a variety of data sources. You must configure at least one crawler for each type of data source that you want to include. When you create a crawler, a wizard that is specific to the type of data being crawled helps you configure the crawler.

- Optional: On the Parse page, configure options for how the crawled data is to be parsed so that it can be effectively searched.

To better search the collection, you can configure rules to automatically associate certain documents with categories, create logical views of the collections (scopes), and map elements in XML documents to search fields.

- On the Index page, specify schedules for reorganizing and refreshing the index. Schedule the index builds to occur frequently so that your users always have access to the latest information.

- Optional: On the Search page, specify options for searching the index, such as options to set aside cache space for search results.

If you want specific URIs to appear automatically in the search results whenever a query includes particular keywords or phrases, click **Configure quick links** and specify the URIs and keywords.

- Optional: On the Log page, specify options for the types of messages that you want to log and how often you want the log files to be recycled.  
If you want to receive e-mail whenever certain messages or certain types of messages are logged, click **Configure e-mail options for messages**.

### Start the components

After you specify which data sources you want to crawl and options for collecting and searching the data, you can start the processes for building the collection.

### System view

If you are a member of the enterprise search administrator role, you can click **System** on the toolbar to configure alerts for system-level events, specify how many indexes can be built concurrently, and specify options for logging messages that are produced by system-level processes. Collection administrators, operators, and monitors cannot access this view.


### Security view

If you are a member of the enterprise search administrator role, you can click **Security** on the toolbar to specify access controls for collections and the administration console. Collection administrators, operators, and monitors cannot access this view.

Until you create your own search applications, you can use the sample search application to search all collections. After you create a search application, use the Security view to associate your application with the collections that it can search.

If you enable security in IBM WebSphere Application Server, you can also use the Security view to configure administrative roles. By configuring administrative roles, you can allow more users to administer the system, yet restrict each user's access to specific functions and collections.

### Monitor views

You can click  **Monitor** to monitor the system or collection components at any time. If your administrative role permits, you can also start and stop component processes while you monitor them.

#### Related tasks

“Starting the enterprise search components” on page 89

To enable users to search a collection, you must start the components that crawl, parse, index, and search the collection.

“Stopping the enterprise search components” on page 90

You might need to stop and restart an enterprise search component if you make changes to its configuration or if you need to troubleshoot problems.

“Creating a collection by using the Collections view” on page 20

Use the Collections view to create an empty collection. You can then select specific pages to specify how you want to add content to the collection and make the collection searchable.

“Creating a collection by using the Collection wizard” on page 19

If you are new to enterprise search, a wizard can help you with creating a collection. The wizard provides details about each step in the process and enables you to save your settings as you progress.

“Creating a crawler” on page 25

When you create a crawler, you specify the type of crawler that you want to create. A wizard helps you specify information about the data that you want to include in the collection.

“Monitoring a collection” on page 96

You can view general information about the status of each component in a collection or select options to view detailed information about individual components and URIs.

---

## Logging in to the administration console

To administer an enterprise search system, you specify a URL in a Web browser and then log in to the administration console.

### Prerequisites

You must log in with a user ID that is authorized to access the enterprise search administration console:

- If you do not enable global security in WebSphere Application Server, only the enterprise search administrator that was specified during the installation of DB2 II OmniFind can access the administration console.
- If you enable global security in WebSphere Application Server, you can use the enterprise search administration console to configure administrative roles for users. The user IDs that you configure must exist in a WebSphere Application Server user registry. When you configure administrative roles, you allow more users to log in to the administration console, but you can control the functions and collections that each administrative user can access.

### Procedure

To log in to the enterprise search administration console:

1. Type the URL for the administration console in your Web browser. For example:

`http://SearchServer.com/ESAdmin/`

*SearchServer.com* is the host name of the search server for enterprise search.

Depending on your Web server configuration, you might also need to specify the port number. For example:

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

2. On the Welcome page, type your user ID and password and click **Log in**.

The Collections view, which is your entry point for administering the system and collections, is displayed. If you use administrative roles, the actions that you can take and the collections that you see depend on your administrative role.

If your session is inactive for a period of time, the system logs you out automatically. To continue administering the system, log in again.

### Related tasks

“Starting the enterprise search components” on page 89

To enable users to search a collection, you must start the components that crawl, parse, index, and search the collection.

---

## Changing the enterprise search administrator password

The password for the initial enterprise search administrator is stored in an encrypted form. To change the password, use the `EncryptProperty` command.

### Prerequisites

The enterprise search administrator ID and password must be valid on your operating system and must have authority to access and configure DB2 Universal Database.

### About this task

The user ID for the initial enterprise search administrator is specified during the installation of DB2 II OmniFind. The password for this user ID is stored in the `es.cfg` file in an encrypted format. This file is installed in the enterprise search data directory on all enterprise search servers.

If you want to change the password, you must run the `EncryptProperty` command to change it on each server (you cannot use a text editor to edit the `es.cfg` file). The enterprise search administrator ID and password must be the same on all enterprise search servers.

The installation program creates two environment variables that you can use with the `EncryptProperty` command:

- `ES_INSTALL_ROOT` is the enterprise search installation directory.
- `ES_NODE_ROOT` is the enterprise search data directory.

### Procedure

To change the enterprise search administrator password, complete the following steps on each enterprise search server:

1. Log in as the enterprise search administrator.
2. Enter the following command:

- On a UNIX computer:

```
java -cp es.oss.jarPath:. com.ibm.es.util.EncryptProperty -f filename
-p property newValue
```

- On a Windows computer:

```
java -cp es.oss.jarPath;. com.ibm.es.util.EncryptProperty -f filename
-p property newValue
```

*es.oss.jarPath*

The classpath to the enterprise search `es.oss.jar` file. This file is installed in the `lib` subdirectory. For example:

```
ES_INSTALL_ROOT/lib/es.oss.jar
```

*filename*

Optional: The fully qualified name of the property file. If you omit this option, the default value is `es.cfg`.

*property*

Optional: The name of a property that you want to change. If you omit this option, the default value is `Password`.

*newValue*

The new value for the property to be encrypted (the new password). This value cannot include spaces.

**Examples:**

To run the EncryptProperty command from the same directory as the es.cfg file on a UNIX computer, enter the following commands:

```
cd $ES_NODE_ROOT/nodeinfo
java -cp
$ES_INSTALL_ROOT/lib/es.oss.jar:. com.ibm.es.util.EncryptProperty newValue
```

To run the EncryptProperty command from anywhere other than the default directory on a Windows computer, enter the following commands (enter the EncryptProperty command on one line):

```
java -cp
%ES_INSTALL_ROOT%\lib\es.oss.jar;. com.ibm.es.util.EncryptProperty
-f %ES_NODE_ROOT%\nodeinfo\es.cfg -p Password newValue
```



---

## Chapter 3. Enterprise search collections

An enterprise search collection contains the entire set of sources that users can search with a single query.

When you create a collection, you specify options that apply to the entire collection. The collection is empty until you add content to it.

You can add collections to an enterprise search system in two ways:

- If you are not familiar with the enterprise search administration console, or if you are still learning how the collection components work together, you might want to use the Collection wizard to create a collection. The Collection wizard helps you progress through the tasks and allows you to save your work as a draft collection while it is being created.
- When you are more familiar with the administration console, you might prefer to create collections by selecting the specific pages that you want to administer in the Collections view.

After you create a collection, you use controls in the Collections view to edit, monitor, and delete the collection.

### Related tasks

“Monitoring a collection” on page 96

You can view general information about the status of each component in a collection or select options to view detailed information about individual components and URIs.

“Migrating a collection from WebSphere Portal” on page 124

To migrate collections from WebSphere Portal to enterprise search, prepare the collections in WebSphere Portal, then use the migration wizard to migrate them.

---

## Creating a collection by using the Collection wizard

If you are new to enterprise search, a wizard can help you with creating a collection. The wizard provides details about each step in the process and enables you to save your settings as you progress.

### Prerequisites

To create a collection, you must be a member of the enterprise search administrator role.

To add content to a collection or to specify options for how content in the collection can be parsed, indexed, or searched, you must be an enterprise search administrator or a collection administrator for the collection.

### About this task

While you create a collection, you can save it in a draft state. While it is in a draft state, any administrator who has authority to administer the collection can make changes to it. For example, you might want a collection administrator who has experience with Lotus Notes sources to configure a Notes crawler. Later, a

collection administrator who has experience with UNIX systems might edit the draft collection to configure a UNIX file system crawler.

### Procedure

To use the Collection wizard to create a collection:

1. Click **Collections** to open the Collections view.
2. Click **Collection Wizard**.
3. Follow the instructions in the wizard to create an empty collection and add content to it.

You must configure general information about the collection and create at least one crawler. You can accept the default values for the remaining configuration options, or specify options for your new collection.

4. To save a collection before you finish creating it, click **Save as Draft**.  
Your collection is listed with other draft collections on the Collections view.
5. To return to a collection that you are still creating, click **Return to wizard** on the Collections view.
6. Click **Finish** to create the collection.

Your new collection is listed with other collections on the Collections view.

After you create a collection, you must start the processes for crawling, parsing, indexing, and searching the collection. Until you are ready to associate the collection with the search applications that can search it, you can use the sample search application to search your new collection.

---

## Creating a collection by using the Collections view

Use the Collections view to create an empty collection. You can then select specific pages to specify how you want to add content to the collection and make the collection searchable.

### Prerequisites

To create a collection, you must be a member of the enterprise search administrator role.

To add content to a collection or to specify options for how content in the collection can be parsed, indexed, or searched, you must be an enterprise search administrator or a collection administrator for the collection.

### About this task

For information about the values that you can specify for a new collection, click **Help** while you are creating the collection.

### Procedure

To create a collection from the Collections view:

1. On the Collections view, click **Create Collection**.
2. On the Create a collection page, provide information or make selections in the following fields:
  - **Collection name**



- **Document importance** (the strategy for assigning a static ranking factor to rank documents in some order)
  - **Categorization type**
  - **Default language**
3. Accept the default values for the following fields, or specify options that you want to use with this collection:
    - **Description**
    - **Index location** (by default, the collection is created on the index server)
    - **Enable collection security** (by default, collection security is enabled)
  4. Click **OK**.

The Collections view lists your new collection with other collections in your enterprise search system.

The collection is empty until you add content to it. To add content to a new collection, select the collection in the Collections view, edit it, create at least one crawler, and specify options for how you want data to be parsed, indexed, and searched.

You must then start the processes for crawling, parsing, indexing, and searching the collection. You can use the sample search application to search your new collection until you are ready to make it available to users with your own search applications.

---

## Editing a collection

You edit collections to specify information about the documents that you want to include in a collection.

### Prerequisites

To edit a collection, be a member of the enterprise search administrator role or be a collection administrator for the collection.


### About this task

When you edit a collection, you specify options for crawling data sources, parsing documents, reorganizing and refreshing the index, searching the indexed content, and logging error messages. When you create a collection, you must edit it to add content to it. Later, you can edit the collection to update the content or to change the way information is crawled, parsed, indexed, searched, or logged.

If you make changes to fields on the General, Index, Search, and Log pages, the changes are not saved until you click **Apply**. If you move to another page and make changes, the changes that you make on all pages are saved when you click **Apply**. If you click **Reset**, the changes that you make on all pages are returned to their previous settings.

### Procedure

To edit a collection:

1. Click **Collections** in the toolbar to open the Collections view.
2. Locate the collection that you want to edit in the list of collections, and click  **Edit**.

3. Make changes to any of the following pages:

#### **General**

Specify general information about the collection, remove URIs from the index, or configure alerts so that you can receive e-mail about certain states or events.

**Crawl** Specify the data sources that you want to crawl and specify options for how the content is to be crawled. Every collection must contain at least one crawler, and a single collection can contain data from multiple data sources. You must configure at least one crawler for each type of data source that you want to include in the collection.

**Parse** Specify options for how you want documents that were crawled to be parsed and analyzed. You can configure categories and scopes, which enable users to search subsets of the collection, and you can configure rules that enable users to search specific elements in XML documents.

**Index** Specify schedules for reorganizing the entire index and refreshing the index with new and changed content.

#### **Search**

Specify options for searching the collection, such as configuring a search cache and selecting a search language. You can also configure quick links, which is a feature that ensures the return of predetermined URIs whenever a user includes specific words or phrases in a query.

**Log** Specify the types of messages that you want to log and options for creating and rotating log files. You can also specify options for receiving e-mail whenever certain messages, or certain types of messages, are logged.

4. Click **Apply**.

---

## **Deleting a collection**

Deleting a collection completely removes all information about the collection from your enterprise search system.

### **Prerequisites**

To delete a collection, you must be a member of the enterprise search administrator role.

You must stop all processes associated with the collection before you can delete it.

### **About this task**

Deleting a collection can be a time-consuming process. After you confirm that you want to delete the collection, the system deletes all data in the system that relates to the collection.

### **Procedure**

To delete a collection:

1. Click **Collections** to open the Collections view.
2. In the list of collections, locate the collection that you want to delete and click  **Delete** .

---

## Chapter 4. Enterprise search crawler administration

You configure crawlers for the different types of data that you want to include in a collection. A single collection can contain any number of crawlers.

### Configuring crawlers

You use the enterprise search administration console to create, edit, and delete crawlers. Typically, an expert in the type of data being crawled configures the crawler. For example, to set up a crawler to crawl Lotus Notes data sources, the collection administrator should either be a Notes administrator or work closely with someone who is knowledgeable about the databases that are being crawled.

You can create a crawler based on the system default values or based on values that are specified for an existing crawler. If you use an existing crawler as the base for a new crawler, you can quickly create multiple crawlers that have similar properties, and then configure them to crawl different sources, for example, or operate on different crawling schedules.

When you create a crawler, a wizard for the type of data that is being crawled helps you specify properties that control how the crawler uses system resources. The wizard also helps you select the sources that you want to search.

You can make changes to existing crawlers at any time. You can edit crawler properties or parts of the crawl space as needed. Crawler wizards also help you to make these changes.

### Populating a new crawler with base values

You can create a crawler based on the system default values or based on values that are specified for an existing crawler. If you use an existing crawler as the base for a new crawler, you can quickly create multiple crawlers that have similar properties, and then configure them to crawl different sources, for example, or operate on different crawling schedules.

By copying a crawler, you can divide the crawling workload among multiple crawlers that use the same crawling rules. For example, you might copy a Notes crawler because you want to use the same properties and field crawling rules with a different Notes server. The only differences might be the databases that each crawler crawls and the crawling schedules.

### Combining crawler types in a collection

Enterprise search crawlers are designed to gather information from specific types of data sources. When you configure crawlers for a collection, you must decide how to combine these different data source types so that users can easily search your enterprise data. For example, if you want users to be able to search Microsoft Windows file systems and Microsoft Exchange Server public folders with a single query, create a collection that includes Windows file system crawlers and Exchange Server crawlers.

When you combine multiple types of crawlers in a single collection, ensure that all of the crawlers can use the same static ranking method. (You specify the static

ranking method when you create the collection.) For example, if you combine Web sources (which use document links as a ranking factor) and NNTP sources (which typically use the document date as a ranking factor), the quality of the search results might be degraded.

### **Document-level security**

Each crawler can associate security tokens with the documents that it crawls. If you specify that you want to use security tokens when you configure the crawler, the crawler associates the tokens that you specify with each document.

If you enable security in your custom search applications, your applications can use the security tokens that the crawlers associated with documents to authenticate users. This feature enables you to restrict access to some documents in a collection and to allow other documents to be searched by all users. For example, in one collection you might allow all users to access all documents in your Microsoft Exchange Server public folders, but allow only users with specific user IDs to access documents in your Lotus Notes databases.

### **Scheduling crawlers**

Crawlers that you create for Web sources and NNTP news groups run continuously. After you start such crawlers, you typically do not need to stop them unless you change the crawler's configuration.

For all other crawler types, you specify a crawling schedule when you configure the crawler. For some data source types, a single schedule controls when the crawler visits all data sources in the crawl space. For other data source types, you can specify different schedules for specific data sources. For example, you can specify different schedules for crawling each Notes database that the crawler crawls.

By creating multiple crawler schedules, you can better control when the crawler visits the target sources. For example, to crawl databases in different time zones, you can schedule the crawler for times when users are most likely to be finished with their work for the day.

#### **Related concepts**

"Enterprise search crawlers" on page 3

Enterprise search crawlers collect documents from data sources so that the documents can be analyzed, indexed, and searched.

#### **Related tasks**

"Monitoring crawlers" on page 97

You can view general information about the status of each crawler in a collection or select options to view detailed information about a crawler activity.

#### **Related reference**

Chapter 17, "URI formats in enterprise search," on page 139

The uniform resource identifier (URI) of each document in an enterprise search collection indicates the type of crawler that added the document to the collection.

---

## Creating a crawler

When you create a crawler, you specify the type of crawler that you want to create. A wizard helps you specify information about the data that you want to include in the collection.

### Prerequisites


To create a crawler, you must be a member of the enterprise search administrator role or be a collection administrator for the collection that owns the crawler.

### About this task

You must create at least one crawler for a collection. The type of crawler that you create depends on the type of data that you want to include in the collection. A wizard for the type of crawler that you create helps you specify options for the crawler. For example, the wizard helps you specify options for how the crawler is to use system resources. The wizard also helps you select the data sources that you want to include in the collection.

### Procedure

To create a crawler:

1. Click **Collections** to open the Collections view.
2. In the list of collections, locate the collection that you want to change and click  **Edit**.
3. On the Crawl page, click **Create Crawler**.
4. Select the crawler type and base values for the crawler:
  - a. Select the type of crawler that supports the type of data that you want to crawl, such as Web sites, Lotus Notes databases, or UNIX file systems.  
After you select a crawler type, options for how you want to create it are displayed.
  - b. Select the base values for the crawler:

#### **Create a crawler based on the system default values**

Populates the default crawler settings with the installation default values.

If you select this option, click **Next** to begin configuring your new crawler.

#### **Create a crawler based on an existing crawler**

Populates the default crawler settings with values that are configured for another crawler of this type.

If you select this option, a list of crawlers that match this crawler type is displayed. Select the crawler that you want to use for the new crawler, then click **Next** to begin configuring your new crawler.

A wizard for the type of crawler that you are creating opens. Follow the wizard prompts to create the crawler. Click **Help** on any page in the wizard to learn more about the options that you can specify for that type of crawler.

Your new crawler is listed on the Crawl page with other crawlers that belong to the collection. You can click options to edit the crawler properties and the crawl space any time that you need to make changes to the crawler.

---

## Editing crawler properties

You can change information about the crawler and how it crawls data. For example, you can change how the crawler uses system resources.

### Prerequisites



To edit crawler properties, you must be a member of the enterprise search administrator role or be a collection administrator for the collection that owns the crawler.

### About this task

When you edit crawler properties, click **Help** to learn about the types of changes that you can make. The properties that you can edit depend on the crawler type.

### Procedure

To edit the properties for a crawler:

1. Click **Collections** to open the Collections view.
2. In the list of collections, locate the collection that you want to add a crawler to and click  **Edit**.
3. On the Crawl page, locate the crawler that you want to edit and click  **Crawler properties**.
4. Change the crawler properties, then click **OK**.
5. For the changes to become effective, stop and restart the crawler.

---

## Editing a crawl space

You can change information about the data sources that a crawler crawls. For example, you can add data sources, remove data sources, change the crawling schedule, and change the rules for crawling documents in a specific data source.

### Prerequisites


To edit a crawl space, you must be a member of the enterprise search administrator role or be a collection administrator for the collection that owns the crawler.


### About this task

To learn about the changes that you can make for the type of crawler that you are administering, click **Help** while you edit the crawl space.

### Procedure

To edit a crawl space:

1. Click **Collections** to open the Collections view.
2. In the list of collections, locate the collection that you want to change and click  **Edit**.

3. On the Crawl page, locate the crawler that you want to edit and click  **Crawl space**.
4. Change the crawl space by selecting the options that you want to change.  
The options that you can choose depend on the crawler type. For some options, such as adding data sources to the collection, a wizard for the crawler type opens to help you change the crawl space.
5. For the changes to become effective, stop and restart the crawler.

---

## Deleting a crawler



Deleting a crawler removes all information about the crawler from your enterprise search system. Information that was previously crawled by the crawler remains in the index until you reorganize the index.

### Prerequisites

To delete a crawler, you must be a member of the enterprise search administrator role or be a collection administrator for the collection that owns the crawler.

### Procedure

To delete a crawler:

1. Click **Collections** to open the Collections view.
2. In the list of collections, locate the collection that you want to change and click  **Edit**.
3. On the Crawl page, locate the crawler that you want to delete and click  **Delete**.

---

## Content Manager crawlers

To include IBM DB2 Content Manager item types in an enterprise search collection, you must configure a Content Manager crawler.

### Setting up the crawler server

Before you can crawl a DB2 Content Manager server, you must run a script on the crawler server. This script, which is provided with IBM DB2 Information Integrator OmniFind Edition, enables the Content Manager crawler to communicate with DB2 Content Manager servers.

Complete the following tasks, as appropriate for your environment, before you use the enterprise search administration console to configure a Content Manager crawler:

- “Configuring the crawler server on AIX for DB2 Content Manager” on page 28.
- “Configuring the crawler server on Linux for DB2 Content Manager” on page 30.
- “Configuring the crawler server on Windows for DB2 Content Manager” on page 31.

### Configuring a Content Manager crawler

You can use the Content Manager crawler to crawl any number of DB2 Content Manager servers. When you configure the crawler, you specify options for how the

crawler is to crawl all DB2 Content Manager servers in the crawl space. You also select the specific item types that you want to crawl on each server.

To create or change a Content Manager crawler, you must log in to the enterprise search administration console. You must be a member of the enterprise search administrator role or be a collection administrator for the collection that owns the crawler.

When you create the crawler, a wizard helps you do these tasks:

- Specify properties that control how the crawler operates and uses system resources. The crawler properties control how the crawler crawls all the item types on all DB2 Content Manager servers in the crawl space.
- Select the DB2 Content Manager servers that you want to crawl.
- Specify user IDs and passwords that enable the crawler to access content on the DB2 Content Manager servers.
- Set up a schedule for crawling the servers.
- Select the item types that you want to crawl on each DB2 Content Manager server.
- Specify options for making the attributes in some item types searchable. For example, you can exclude certain types of documents from the crawl space or specify security tokens to control access to content.

Click **Help** while you are creating the crawler to learn about the fields in the wizard and how to provide the crawler with the information that it needs to crawl data.

After you create a crawler, you can edit the properties or crawl space as needed. Wizards help you make these changes:

- Change the crawler properties. For example, you might want to change the number of threads that the crawler can activate at a time.
- Add DB2 Content Manager servers and item types to the crawl space.
- Change the password that the crawler uses to access a specific server.
- Change the schedule for crawling specific servers.
- Change the rules for crawling documents in an item type.
- Remove DB2 Content Manager servers and item types from the crawl space.

## Configuring the crawler server on AIX for DB2 Content Manager

If you install DB2 II OmniFind on an IBM AIX® computer, you must run a script to configure the crawler server. The script enables the Content Manager crawler to communicate with IBM DB2 Content Manager servers.

### About this task

The Content Manager crawler uses the Java™ connector for Content Manager Version 8 to access DB2 Content Manager servers. You install this connector by installing IBM DB2 Information Integrator for Content Version 8.2 on the crawler server. To ensure that the Content Manager crawler can work with DB2 Content Manager, you run a setup script that DB2 II OmniFind provides on the crawler server after you install the connector.

### Procedure



To configure the crawler server so that it can crawl DB2 Content Manager servers:

1. Install the Java connector for Content Manager Version 8 on the crawler server:
  - a. On the crawler server, log in as the root user:

```
su - root
```
  - b. Run the db2profile file:

```
./home/db2inst/sqllib/db2profile
```
  - c. Export the JAVAHOME environment variable. For example:

```
export JAVAHOME=/usr/IBMJava2-141
```
  - d. Add the Java directory to the PATH environment variable:

```
export PATH=$PATH:$JAVAHOME/bin
```
  - e. Insert the DB2 Information Integrator for Content installation CD and run the installation script:

```
./frnxsetup.sh
```

The DB2 Content Manager Enterprise Information Portal installation wizard opens.
  - f. In the Component Selection window, take the following actions:
    - 1) Select **Local connectors** from the **Components** list, then select **Content Manager V8 connector** from the **Subcomponents** list.
    - 2) Select **Connector toolkits and samples** from the **Components** list, then select **Content Manager V8 connector** from the **Subcomponents** list.
  - g. Specify a password for the ICMNLSDB database when you are prompted, and accept the default settings for the remaining windows.
2. On the crawler server, log in with a user ID that is in the DB2 administration group.
3. Catalog the remote DB2 Content Manager library server database, and verify that the crawler server can connect to the DB2 Content Manager server:

```
db2 catalog tcpip node node_name remote hostname server port  
db2 catalog database database_name as alias at node node_name
```
4. Optional: Log in as the root user and test the database connection:

```
./usr/lpp/cmb/bin/cmbenv81.sh  
cd /usr/lpp/cmb/samples/java/icm  
javac *.java  
java SConnectDisconnect ICMdatabase_name CMadmin_ID CMadmin_password
```
5. Run the setup script for the Content Manager crawler on the crawler server:
  - a. Change to the DB2 II OmniFind bin directory:

```
cd $ES_INSTALL_ROOT/bin
```
  - b. Start the script and answer the prompts:

```
escrcm.sh
```
6. Stop and restart the enterprise search administration console and all sessions on the enterprise search common communications layer (CCL):

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

## Configuring the crawler server on Linux for DB2 Content Manager

If you install DB2 II OmniFind on a Linux™ computer, you must run a script to configure the crawler server. The script enables the Content Manager crawler to communicate with IBM DB2 Content Manager servers.

### About this task

The Content Manager crawler uses the Java connector for Content Manager Version 8 to access DB2 Content Manager servers. You install this connector by installing the IBM DB2 Content Manager Linux Toolkit 8.2 on the crawler server. To ensure that the Content Manager crawler can work with DB2 Content Manager, you run a setup script that DB2 II OmniFind provides on the crawler server after you install the connector.

### Procedure

To configure the crawler server so that it can crawl DB2 Content Manager servers:

1. Install the Java connector for Content Manager Version 8 on the crawler server:
  - a. Log in as the root user:

```
su - root
```
  - b. Run the db2profile file:

```
./home/db2inst/sqllib/db2profile
```
  - c. Export the JAVAHOME environment variable. For example:

```
export JAVAHOME=/opt/IBMJava2-141
```
  - d. Add the Java directory to the PATH environment variable:

```
export PATH=$PATH:$JAVAHOME/bin
```
  - e. Insert the DB2 Content Manager Linux Toolkit installation CD and change to the appropriate language directory. For example, to change to the English directory, enter:

```
cd English
```
  - f. Run the installation script:

```
./setuplinux.bin
```

The DB2 Content Manager Enterprise Information Portal installation wizard opens.
  - g. In the Setup Type window, select **custom**.
  - h. In the Features List window, select **Content Manager V8 connector**.
  - i. Specify a password for the ICMNLSDB database when you are prompted, and accept the default settings for the remaining windows.
2. On the crawler server, log in with a user ID that is in the DB2 administration group.
3. Catalog the remote DB2 Content Manager library server database, and verify that the crawler server can connect to the DB2 Content Manager server:

```
db2 catalog tcpip node node_name remote hostname server port  
db2 catalog database database_name as alias at node node_name
```

4. Optional: Log in as the root user and test the database connection:

```
./opt/IBMcmb/bin/cmbenv81.sh  
cd /opt/IBMcmb/samples/java/icm
```

- ```
javac *.java
java SConnectDisconnect ICMdatabase_name CAdmin_ID CAdmin_password
```
5. Run the setup script for the Content Manager crawler on the crawler server:
    - a. Change to the DB2 II OmniFind bin directory:
 

```
cd $ES_INSTALL_ROOT/bin
```
    - b. Start the script and answer the prompts:
 

```
./escrcm.sh
```
  6. Stop and restart the enterprise search administration console and all sessions on the enterprise search common communications layer (CCL):
 

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

## Configuring the crawler server on Windows for DB2 Content Manager

If you install DB2 II OmniFind on a Microsoft Windows computer, you must run a script to configure the crawler server. The script enables the Content Manager crawler to communicate with IBM DB2 Content Manager servers.

### About this task

The Content Manager crawler uses the Java connector for Content Manager Version 8 to access DB2 Content Manager servers. You install this connector by installing IBM DB2 Information Integrator for Content Version 8.2 on the crawler server. To ensure that the Content Manager crawler can work with DB2 Content Manager, you run a setup script that DB2 II OmniFind provides on the crawler server after you install the connector.

### Procedure

To configure the crawler server so that it can crawl DB2 Content Manager servers:

1. Install the Java connector for Content Manager Version 8 on the crawler server:
  - a. Insert the DB2 Information Integrator for Content installation CD. The installation program begins automatically.
 

The DB2 Content Manager Enterprise Information Portal installation wizard opens.
  - b. In the Select Machine Type window, select **EIP development workstation**.
  - c. In the Component Selection window, take the following actions:
    - 1) Select **Local connectors** from the **Components** list, then select **Content Manager V8 connector** from the **Subcomponents** list.
    - 2) Select **Connector toolkits and samples** from the **Components** list, then select **Content Manager V8 connector** from the **Subcomponents** list.
  - d. Specify a password for the ICMNLSDB database when you are prompted, and accept the default settings for the remaining windows.
2. Catalog the remote DB2 Content Manager library server database and verify that the crawler server can connect to the DB2 Content Manager server. Enter the following commands at a command prompt on the crawler server:

```
db2 catalog tcpip node node_name remote hostname server port
db2 catalog database database_name as alias at node node_name
```

3. Optional: Test the database connection:
  - a. Select **Start** → **Programs** → **Enterprise Information Portal for Multiplatforms and Development**.
  - b. Enter the following commands:

```
cd \cmbroot\samples\java\icm
javac *.java
java SConnectDisconnect ICMdatabase_name CAdmin_ID CAdmin_password
```

4. Run the setup script for the Content Manager crawler on the crawler server:
  - a. Change to the DB2 II OmniFind bin directory:

```
cd %ES_INSTALL_ROOT%\bin
```
  - b. Start the script and answer the prompts:

```
escrcm.vbs
```
5. Stop and restart the enterprise search administration console and all sessions on the enterprise search common communications layer (CCL):
  - a. At a command prompt, stop the administration console:

```
esadmin stop
```
  - b. Select **Start** → **Programs** → **Administrative Tools** → **Services**, then restart the IBM DB2 Information Integrator OmniFind Edition service.
  - c. At a command prompt, start the administration console:

```
esadmin start
```

---

## DB2 crawlers

You use the DB2 crawler to include IBM DB2 Universal Database databases in a collection. You can also use the DB2 crawler to include nickname tables that you create for IBM DB2 Universal Database for z/OS, IBM Informix, Oracle, and federated databases.

You must configure a separate crawler for each database server that you want to crawl. When you configure the crawler, you specify options for how the crawler is to crawl all databases on the same server. You also select the specific tables that you want to crawl in each database.

To create or change a DB2 crawler, you must log in to the enterprise search administration console. You must be a member of the enterprise search administrator role or be a collection administrator for the collection that owns the crawler.

When you create the crawler, a wizard helps you do these tasks:

- Specify properties that control how the crawler operates and uses system resources. The crawler properties control how the crawler crawls all the databases on a particular database server.
- Specify information about the types of databases that you want to crawl.

If you plan to crawl remote databases that are not cataloged on the local database server, you must start the DB2 Administration Server on the remote server before you can use the DB2 crawler to crawl those databases. You must also specify the host name and port of the remote database server when you configure the crawler.

- Select the database that you want to crawl.
- Specify user IDs and passwords that enable the crawler to access databases that use access controls.
- Set up a schedule for crawling the databases.
- Select the tables that you want to crawl in each database.  
To optimize the performance of the discovery processes, choose to crawl all tables only if the database does not contain many tables or if the tables do not contain many columns. If you select some tables to crawl now, you can edit the crawl space later and add more tables to the collection.
- Specify options for making the columns in specific tables searchable. For example, you can enable certain columns to be used in parametric queries or specify security tokens to control access to content.

Click **Help** while you are creating the crawler to learn about the fields in the wizard and how to provide the crawler with the information that it needs to crawl data.

After you create a crawler, you can edit the properties or crawl space as needed. Wizards help you make these changes:

- Change the crawler properties. For example, you might want to change the number of threads that the crawler can activate at a time.
- Add databases and tables to the crawl space.
- Change the password that the crawler uses to access specific databases.
- Change the schedule for crawling specific databases.
- Change the rules for crawling the columns in a table.
- Remove databases and tables from the crawl space.

---

## Exchange Server crawlers

To include Microsoft Exchange Server public folders in an enterprise search collection, you must configure an Exchange Server crawler.

You can use the Exchange Server crawler to crawl any number of folders and subfolders on Exchange Server public folder servers. When you create a crawler, you select the content that you want to crawl on a public folder server. Later you can edit the crawl space to add content from another public folder server.

To create or change an Exchange Server crawler, you must log in to the enterprise search administration console. You must be a member of the enterprise search administrator role or be a collection administrator for the collection that owns the crawler.

When you create the crawler, a wizard helps you do these tasks:

- Specify properties that control how the crawler operates and uses system resources. The crawler properties control how the crawler crawls all subfolders on all servers in the crawl space.
- Specify information about the Exchange Server public folder server that you want to crawl.

You must specify a user ID and password so that the crawler can access content on the server. If the server uses the Secure Sockets Layer (SSL) protocol, you can specify options that enable the crawler to access the keystore file on the crawler server.

- Set up a schedule for crawling the public folder server.
- Select folders and subfolders to crawl.
- Specify options for making documents in subfolders searchable. For example, you can exclude certain types of documents from the crawl space or specify security tokens to control access to content.

Click **Help** while you are creating the crawler to learn about the fields in the wizard and how to provide the crawler with the information that it needs to crawl data.

After you create a crawler, you can edit the properties or crawl space as needed. Wizards help you make these changes:

- Change the crawler properties. For example, you might want to change the number of threads that the crawler can activate at a time.
- Add another public folder server to the crawl space.
- Add subfolders to the crawl space.
- Change the passwords (and keystore file information) that the crawler uses to access a server.
- Change the schedule for crawling servers.
- Change the rules for crawling documents in a subfolder.
- Remove servers and subfolders from the crawl space.

---

## NNTP crawlers

To include articles from NNTP news groups in an enterprise search collection, you must configure an NNTP crawler.

You can use the NNTP crawler to crawl any number of NNTP servers. When you configure the crawler, you select the news groups on each server that you want to crawl. You can also specify patterns for the news groups that you want to exclude. With this feature, you can easily allow the crawler to crawl the majority of news groups on a server, and forbid the crawler from crawling a few news groups that you do not want users to search.

For example, you can specify rules to include all the news groups on a specific NNTP server, then specify that you want to exclude news groups on that server if their names include the string `private`.

To create or change an NNTP crawler, you must log in to the enterprise search administration console. You must be a member of the enterprise search administrator role or be a collection administrator for the collection that owns the crawler.

When you create the crawler, a wizard helps you do these tasks:

- Specify properties that control how the crawler operates and uses system resources. The crawler properties control how the crawler crawls all news groups in the crawl space. You can also specify security tokens to control access to content.
- Specify patterns to include news groups, and specify patterns to exclude certain news groups from the crawl space.

Click **Help** while you are creating the crawler to learn about the fields in the wizard and how to provide the crawler with the information that it needs to crawl data.

After you create a crawler, you can edit the properties or crawl space as needed. Wizards help you make these changes:

- Change the crawler properties. For example, you might want to change the number of threads that the crawler can activate at a time.
- Add news groups to the crawl space.
- Remove news groups from the crawl space.

---

## Notes crawlers

To include IBM Lotus Notes databases in an enterprise search collection, you must configure a Notes crawler.

### Setting up the crawler server

If a Lotus Notes server that you plan to crawl uses the Notes Remote Procedure Call (NRPC) protocol, you must run a script on the crawler server. This script, which is provided with IBM DB2 Information Integrator OmniFind Edition, enables the Notes crawler to communicate with the servers that use NRPC.

If a Lotus Notes server that you plan to crawl uses the Domino Internet Inter-ORB Protocol (DIIOP), you do not need to run a setup script on the crawler server. However, you must configure the Lotus Notes server so that the Notes crawler can access the server.

Complete the following tasks, as appropriate for your environment, before you use the enterprise search administration console to configure a Notes crawler:

- “Configuring the crawler server on UNIX for Lotus Notes” on page 36.
- “Configuring the crawler server on Windows for Lotus Notes” on page 38.
- “Configuring the DIIOP protocol for Notes crawlers” on page 39.

**Important:** After you configure a Notes crawler that uses NRPC for a collection, you must stop it before you configure any additional Notes crawlers that use NRPC. The discovery processes that use NRPC (which provide information about the sources that are available for crawling) and the crawler processes that use NRPC cannot run concurrently.

### Configuring a Notes crawler

You can use the Notes crawler to crawl any number of standard Lotus Notes databases (.nsf files). You must configure a separate crawler for each Lotus Notes server that you want to crawl. When you configure the crawler, you can specify whether you want to crawl all databases on the server or specific views and folders.

To create or change a Notes crawler, you must log in to the enterprise search administration console. You must be a member of the enterprise search administrator role or be a collection administrator for the collection that owns the crawler.

When you create the crawler, a wizard helps you do these tasks:

- Specify properties that control how the crawler operates and uses system resources. The crawler properties control how the crawler crawls all the databases on a particular server.
- Identify the Lotus Notes server host name, port, and communications protocol.
- Select the databases on the Lotus Notes server that you want to crawl.
- Set up a schedule for crawling the databases.
- Select the documents that you want to crawl. You can crawl all documents in a database or documents in selected views and folders.
- Specify options for making the fields in various databases, views, and folders searchable. For example, you can exclude certain types of documents from the crawl space or specify security tokens to control access to content.

Click **Help** while you are creating the crawler to learn about the fields in the wizard and how to provide the crawler with the information that it needs to crawl data.

After you create a crawler, you can edit the properties or crawl space as needed. Wizards help you make these changes:

- Change the crawler properties. For example, you might want to change the number of threads that the crawler can activate at a time.
- Add Lotus Notes databases, views, and folders to the crawl space.
- Change the password that the crawler uses to access the Lotus Notes server.
- Change the schedule for crawling databases.
- Change the rules for crawling documents in a databases, views, and folders.
- Edit a set of default rules for how fields are to be crawled and searched. The Notes crawler uses these rules with all databases, views, and folders that you do not create rules for.
- If a database, view, or folder uses the default rules for the crawl space, specify that you want to configure rules for the documents in that data source. If you no longer want to use the rules that you configured for a database, view, or folder, start using the default rules for the crawl space.
- Remove Lotus Notes databases, views, and folders from the crawl space.

## Configuring the crawler server on UNIX for Lotus Notes

If you install DB2 II OmniFind on an AIX computer or on a Linux computer, and you plan to crawl Lotus Notes servers that use the Notes Remote Procedure Call (NRPC) protocol, you must run a script to configure the crawler server. The script enables the Notes crawler to communicate with Lotus Notes servers.

### Restrictions

Lotus Domino<sup>®</sup> Server cannot run at the same time, on the same computer, with a Notes crawler that is configured to use NRPC. If you attempt to start the Notes crawler while Domino is running, an error occurs and the crawler stops.

### About this task

The Notes crawler for NRPC uses Domino libraries as a Lotus Notes client. You install these libraries by installing Lotus Domino Server Version 6.0.2 or later on the crawler server. To ensure that the Notes crawler can work with the Domino libraries, you run a setup script that DB2 II OmniFind provides on the crawler server after you install the Domino libraries.



## Procedure

To configure the crawler server so that it can crawl Lotus Notes servers:

1. Create the user notes and the group notes on the crawler server:
  - a. Log in as the root user:  
`su - root`
  - b. Add a user:  
`useradd notes`
  - c. Add a password for this user:  
`passwd notes`  
You will be prompted to change the password.
2. Install the Domino server on the crawler server:
  - a. Insert the Domino Server CD, version 6.0.2 or later, and mount it. (If you do not have a CD, you can download the image.)
  - b. Change to the folder for your operating system.  
  
On an AIX computer, enter: `cd /mnt/cdrom/aix`  
On a Linux computer, enter: `cd /mnt/cdrom/linux`
  - c. Start the installation program:  
`./install`
  - d. Answer the prompts and accept the default values or specify your preferred installation settings (such as paths for the installation directory and data directory).  
  
Consult the Domino documentation if you need assistance with installing Domino Server.
3. Run the setup script for the Notes crawler on the crawler server:
  - a. Log in as the enterprise search administrator (this user ID was specified during the installation of DB2 II OmniFind):
  - b. Start the setup script this script is installed in the ES\_INSTALL\_ROOT/bin directory):  
`escrnote.sh`
  - c. Answer the prompts:
    - For the following prompt, answer Y if Domino Server is installed in the default directory, and answer N if it is not:  
The Lotus Notes directory path /opt/lotus/notes/latest/linux was found.  
Is this the correct Lotus Notes directory path?  
  
The default path on an AIX computer is /opt/lotus/notes/latest/ibmpow. The default path on a Linux computer is /opt/lotus/notes/latest/linux.
    - If Domino Server is not installed into the default directory on the crawler server, specify where Domino is installed in response to the following prompt:  
Enter the path for the Lotus Notes directory  
  
For example, on a Linux computer you might specify /opt/lotus/notes/latest/linux
    - For the following prompt, answer Y if the Domino Server data directory is installed in the default directory, and answer N if it is not:

The Lotus Notes data directory path /local/notesdata was found.  
Is this the correct Lotus Notes data directory path?.

The default path on Linux and AIX computers is /local/notesdata.

- If the Domino Server data directory is not deployed in the default location on the crawler server, specify the Domino data path in response to the following prompt:

Enter the path for Lotus Notes data directory.

4. Stop and restart the enterprise search administration console and all sessions on the enterprise search common communications layer (CCL):

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

## Configuring the crawler server on Windows for Lotus Notes

If you install DB2 II OmniFind on a Microsoft Windows computer, and you plan to crawl Lotus Notes servers that use the Notes Remote Procedure Call (NRPC) protocol, you must run a script to configure the crawler server. The script enables the Notes crawler to communicate with Lotus Notes servers.

### Restrictions

Lotus Domino Server and the Lotus Notes client cannot run at the same time, on the same computer, with a Notes crawler that is configured to use NRPC. If you attempt to start a Notes NRPC crawler while Domino is running, an error occurs and the crawler stops.

### About this task

The Notes crawler for NRPC uses the Lotus Notes client libraries. You install these libraries by installing Lotus Notes Release 6.0.2 or later on the crawler server. To ensure that the Notes crawler can work with the Lotus Notes client libraries, you run a setup script that DB2 II OmniFind provides on the crawler server after you install the Lotus Notes client libraries.

### Procedure

To configure the crawler server so that it can crawl Lotus Notes servers:

1. On the crawler server, log in with a user ID that is a member of the Administrators group. Ensure that the user ID has authority to install Lotus Notes.
2. Install Lotus Notes:
  - a. Insert the Lotus Notes CD, Release 6.0.2 or later. (If you do not have a CD, you can download the image.)
  - b. Start the installation program: setup.exe
  - c. Answer the prompts and accept the default values or specify your preferred installation settings (such as paths for the installation directory and data directory).

Consult the Lotus Notes documentation if you need assistance with installing Lotus Notes.

3. Run the setup script for the Notes crawler on the crawler server:

- a. Log in as the enterprise search administrator (this user ID was specified during the installation of DB2 II OmniFind):
- b. Start the setup script (this script is installed in the ES\_INSTALL\_ROOT\bin directory):  
escrnte.vbs
- c. Answer the prompts:
  - For the following prompt, answer Y if Lotus Notes is installed in the default directory, and answer N if it is not:  
The Lotus Notes directory path c:\lotus\notes was found.  
Is this the correct Lotus Notes directory path?  
  
The typical installation path on a Windows computer is c:\lotus\notes or c:\lotus\domino.
  - If Lotus Notes is not installed in the default directory on the crawler server, specify where Lotus Notes is installed in response to the following prompt:  
Enter the path for the Lotus Notes directory
  - For the following prompt, answer Y if the Lotus Notes data directory is deployed in the default location, and answer N if it is not:  
The Lotus Notes data directory path c:\lotus\notes\data was found.  
Is this the correct Lotus Notes data directory path?  
  
The typical path on a Windows computer is c:\lotus\notes\data or c:\lotus\domino\data.
  - If the Lotus Notes data directory is not deployed in the default location on the crawler server, specify the data directory path in response to the following prompt:  
Enter the path for Lotus Notes data directory.
4. Stop and restart the enterprise search administration console and all sessions on the enterprise search common communications layer (CCL):
  - a. At a command prompt, stop the administration console:  
esadmin stop
  - b. Select **Start** → **Programs** → **Administrative Tools** → **Services**, then restart the IBM DB2 Information Integrator OmniFind Edition service.
  - c. At a command prompt, start the administration console:  
esadmin start

## Configuring the DIIOP protocol for Notes crawlers

To crawl a Lotus Notes server that uses the Domino Internet Inter-ORB Protocol (DIIOP), you must configure the Lotus Notes server so that the Notes crawler can use the protocol.

### Prerequisites

The Lotus Notes server that you want to crawl must be running the DIIOP and HTTP tasks.

### Procedure

To configure server settings for a Lotus Notes server that uses the DIIOP protocol:

1. Configure the server document:

- a. Open the server document on the Lotus Notes server that you want to crawl. This document is stored in the Domino directory.
- b. On the Configuration page, expand the **server** section.
- c. On the Security page, in the **Programmability Restrictions** area, specify the appropriate security restrictions for your environment in the following fields:

- **Run restricted Lotus Script/Java agents**
- **Run restricted Java/Javascript/COM**
- **Run unrestricted Java/Javascript/COM**

For example, you might specify an asterisk (\*) to allow unrestricted access by Lotus<sup>®</sup> Script/Java agents, and specify user names that are registered in the Domino Directory for the Java/Javascript/COM restrictions.

**Important:** The Notes crawler that you configure to crawl Lotus Notes servers with the DIIOP protocol must be able to use the user names that you specify in these fields.

- d. Open the Internet Protocol page, then open the HTTP page, and set the **Allow HTTP clients to browse database** option to **Yes**.
2. Configure the user document:
    - a. Open the user document on the Lotus Notes server that you want to crawl. This document is stored in the Domino directory.
    - b. On the Basics page, in the **Internet password** field, specify a password.  
When you use the enterprise search administration console to set up this Lotus Notes server for crawling, specify this user ID and password on the Specify the Lotus Notes server to crawl page. The crawler uses this information to access the Lotus Notes server.
  3. Restart the DIIOP task on the Lotus Notes server.

---

## UNIX file system crawlers

To include documents that are stored in UNIX file systems in an enterprise search collection, you must configure a UNIX file system crawler.

You can use the UNIX file system crawler to crawl any number of UNIX file systems. When you configure the crawler, you select the local and remote directories and subdirectories that you want to crawl.

If you install the crawler server on a UNIX computer, you cannot use that server to crawl Microsoft Windows file system sources (the Windows file system crawler does not appear in the list of available crawler types).

To create or change a UNIX file system crawler, you must log in to the enterprise search administration console. You must be a member of the enterprise search administrator role or be a collection administrator for the collection that owns the crawler.

When you create the crawler, a wizard helps you do these tasks:

- Specify properties that control how the crawler operates and uses system resources. The crawler properties control how the crawler crawls all subdirectories in the crawl space.
- Set up a schedule for crawling the file systems.

- Select the subdirectories, and the levels of subdirectories, that you want the crawler to crawl.
- Specify options for making documents in subdirectories searchable. For example, you can exclude certain types of documents from the crawl space or specify security tokens to control access to content.

Click **Help** while you are creating the crawler to learn about the fields in the wizard and how to provide the crawler with the information that it needs to crawl data.

After you create a crawler, you can edit the properties or crawl space as needed. Wizards help you make these changes:

- Change the crawler properties. For example, you might want to change the number of threads that the crawler can activate at a time.
- Add subdirectories to the crawl space.
- Change the schedule for crawling file systems.
- Change the rules for crawling documents in a subdirectory.
- Remove subdirectories from the crawl space.

---

## VeniceBridge crawlers

To include Venetica VeniceBridge repositories in an enterprise search collection, you must configure a VeniceBridge crawler.

### Setting up the crawler server

Before you can crawl a VeniceBridge server, you must run a script on the crawler server. This script, which is provided with IBM DB2 Information Integrator OmniFind Edition, enables the VeniceBridge crawler to communicate with VeniceBridge servers.

Complete the following task, as appropriate for your environment, before you use the enterprise search administration console to configure a VeniceBridge crawler:

- “Configuring the crawler server on UNIX for VeniceBridge” on page 42.
- “Configuring the crawler server on Windows for VeniceBridge” on page 43.

### Configuring a VeniceBridge crawler

You can use the VeniceBridge crawler to crawl Documentum and FileNet Panagon Content Services repositories that are managed by VeniceBridge. You must configure a separate crawler for each VeniceBridge server that you want to crawl.

When you configure the crawler, you specify options for how the crawler is to crawl all repositories in the crawl space. You also select the specific item classes that you want to crawl in each repository.

To create or change a VeniceBridge crawler, you must log in to the enterprise search administration console. You must be a member of the enterprise search administrator role or be a collection administrator for the collection that owns the crawler.

When you create the crawler, a wizard helps you do these tasks:

- Specify properties that control how the crawler operates and uses system resources. The crawler properties control how the crawler crawls all the item classes on the VeniceBridge server.
- Select the VeniceBridge server that you want to crawl.
- Select the FileNet or Documentum repositories that you want to crawl on that server.
- Specify user IDs and passwords that enable the crawler to access content in the selected repositories.
- Set up a schedule for crawling the repositories.
- Select the item classes that you want to crawl in each repository.
- Specify options for making the properties of item classes searchable. For example, you can exclude certain types of documents from the crawl space or specify security tokens to control access to content.

Click **Help** while you are creating the crawler to learn about the fields in the wizard and how to provide the crawler with the information that it needs to crawl data.

After you create a crawler, you can edit the properties or crawl space as needed. Wizards help you make these changes:

- Change the crawler properties. For example, you might want to change the number of threads that the crawler can activate at a time.
- Add item classes to the crawl space.
- Change the password that the crawler uses to access a specific repository.
- Change the schedule for crawling repositories.
- Change the rules for crawling documents in an item class.
- Remove item classes from the crawl space.

## Configuring the crawler server on UNIX for VeniceBridge

If you install DB2 II OmniFind on an AIX computer or on a Linux computer, you must run a script to configure the crawler server. The script enables the VeniceBridge crawler to communicate with VeniceBridge servers.

### About this task

The VeniceBridge crawler uses Java libraries of VeniceBridge as a VeniceBridge Java client. You install these libraries by installing Venetica VeniceBridge 5.0.5 or later on the crawler server. To ensure that the VeniceBridge crawler can work with the Java libraries, you run a setup script that DB2 II OmniFind provides on the crawler server after you install the VeniceBridge libraries.

### Procedure

To configure the crawler server so that it can crawl VeniceBridge servers:

1. Install Venetica VeniceBridge 5.0.5 on the crawler server:
  - a. Log in as the enterprise search administrator (this user ID was specified when DB2 II OmniFind was installed).
  - b. Insert the VeniceBridge CD and untar the vbr\_unix.tar file to a temporary installation location.
  - c. Change to the GenericUnix directory.
  - d. Start the VeniceBridge installer:

- ```
./install.bin
```
- e. In the Choose Product Features window, select **Bridges Only** for the installation option.
  - f. In the Choose Bridges To Install window, select **Sample File System Bridge**.
2. Run the setup script for the VeniceBridge crawler on the crawler server:
    - a. Log in as the enterprise search administrator (this user ID was specified during the installation of DB2 II OmniFind):
    - b. Start the script and answer the prompts:

```
escrvbr.sh
```
  3. Stop and restart the enterprise search administration console and all sessions on the enterprise search common communications layer (CCL):

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

4. If the VeniceBridge server to be crawled uses WebSphere Application Server, copy the `vbr_access_services.jar` file from the VeniceBridge server to the crawler server.

**Copy from:**

The `vbr_access_services.jar` file is in the following default location:

*was\_install\_root*/installedApps/*server\_name*/*application\_name*

*was\_install\_root* is the WebSphere Application Server installation directory, *server\_name* is the name that you specified for the server, and *application\_name* is the name that you specified for the VeniceBridge application in WebSphere Application Server.

**Copy to:**

The target directory on the crawler server is *vbr\_install\_root*/lib, where *vbr\_install\_root* is the VeniceBridge installation directory on the crawler server.

## Configuring the crawler server on Windows for VeniceBridge

If you install DB2 II OmniFind on a Microsoft Windows computer, you must run a script to configure the crawler server. The script enables the VeniceBridge crawler to communicate with VeniceBridge servers.

### About this task

The VeniceBridge crawler uses Java libraries of VeniceBridge as a VeniceBridge Java client. You install these libraries by installing Venetica VeniceBridge 5.0.5 or later on the crawler server. To ensure that the VeniceBridge crawler can work with the Java libraries, you run a setup script that DB2 II OmniFind provides on the crawler server after you install the VeniceBridge libraries.

### Procedure

To configure the crawler server so that it can crawl VeniceBridge servers:

1. Install Venetica VeniceBridge 5.0.5 on the crawler server:
  - a. Insert the VeniceBridge CD.
  - b. Change to the Windows directory.

- c. Start the VeniceBridge installer:  
install.exe
  - d. In the Choose Product Features window, select **Bridges Only** for the installation option.
  - e. In the Choose Bridges To Install window, select **Sample File System Bridge**.
2. Run the setup script for the VeniceBridge crawler on the crawler server:
    - a. Log in with the enterprise search administrator user ID (this user ID was specified during the installation of DB2 II OmniFind):
    - b. Start the script and answer the prompts:  
escrvbr.vbs
  3. Stop and restart the enterprise search administration console and all sessions on the enterprise search common communications layer (CCL):
    - a. At a command prompt, stop the administration console:  
esadmin stop
    - b. Select **Start** → **Programs** → **Administrative Tools** → **Services**, then restart the IBM DB2 Information Integrator OmniFind Edition service.
    - c. At a command prompt, start the administration console:  
esadmin start
  4. If the VeniceBridge server to be crawled uses WebSphere Application Server, copy the vbr\_access\_services.jar file from the VeniceBridge server to the crawler server.

**Copy from:**

The vbr\_access\_services.jar file is in the following default location:

*was\_install\_root*\installedApps\*server\_name*\*application\_name*

*was\_install\_root* is the WebSphere Application Server installation directory, *server\_name* is the name that you specified for the server, and *application\_name* is the name that you specified for the VeniceBridge application in WebSphere Application Server.

**Copy to:**

The target directory on the crawler server is *vbr\_install\_root*\lib, where *vbr\_install\_root* is the VeniceBridge installation directory on the crawler server.

---

## Web crawlers

To include data from Web sites in an enterprise search collection, you must configure a Web crawler.

You can use the Web crawler to crawl any number of Hypertext Transfer Protocol (HTTP) servers and secure HTTP (HTTPS) servers. The crawler visits a Web site and reads the data and information on the site. It can also follow links in documents to crawl additional documents.

The crawled data can be in one of many common formats, and comes from various sources within your intranet or the Internet. Common formats include HTML, PDF, Microsoft Word, Lotus WordPro, Extensible Markup Language (XML), and so on.



To create or change a Web crawler, you must log in to the enterprise search administration console. You must also be a member of the enterprise search administrator role or be a collection administrator for the collection that owns the crawler.

When you create the crawler, a wizard helps you do these tasks:

- Specify properties that control how the crawler operates and uses system resources. The crawler properties control how the crawler crawls all Web pages in the crawl space. You can also specify security tokens to control access to content.
- Specify patterns and rules to allow and forbid visits to certain Web sites or certain types of documents.

Click **Help** while you are creating the crawler to learn about the fields in the wizard and how to provide the crawler with the information that it needs to crawl data.

After you create a crawler, you can edit the properties or crawl space as needed. Wizards help you make these changes:

- Change the crawler properties. For example, you might want to change the number of threads that the crawler can activate at a time.
- Add rules for visiting new or different Web sites.
- Change the rules for crawling certain types of documents, or change the depth of the crawl.
- Remove Web sites or certain types of documents from the crawl space.

## Rules to limit the Web crawl space

To ensure that users access only the Web sites that you want them to search, you specify rules to limit what the Web crawler can crawl.

When a Web crawler crawls a Web page, it discovers links to other pages and puts those links in a queue to be crawled next. Crawling and discovery can be repeated as long as time and memory resources permit. When you configure a Web crawler, you specify where the crawler is to begin crawling. From these initial URLs (which are called *seed URLs*) the Web crawler can reach any document on the Web that is connected by direct or indirect links from the seed URLs.

To limit the crawl space, configure the Web crawler to crawl certain URLs thoroughly and ignore links that point outside the area of interest. Because the crawler, by default, accepts any URL that it discovers, you must specify rules that identify which URLs you want to include in the collection, and eliminate the rest of the pages.

You can specify in several ways what you want the Web crawler to crawl and not crawl. You can specify:

- Three types of crawling rules: domain, Internet Protocol (IP) address, and URL prefix
- A list of file extensions for documents that you want to exclude
- A list of MIME types for documents that you want to include
- The maximum number of directories in a URL path
- A list of seed URLs where the crawler is to begin crawling

Crawling rules have the form:

action type target

action is forbid or allow; type is domain, IP address, or URL prefix (HTTP or HTTPS); and target depends on the value of type. You can specify an asterisk (\*) as a wildcard character, in limited ways, to specify targets that match a pattern.

### Domain rules

The target of a domain rule is a DNS domain name. For example, you can specify that the entire `www.ibm.com` domain is to be crawled:

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

The first character in the target can be an asterisk, which causes the rule to apply to any host name that ends with the rest of the pattern. For example, you can specify that no domains that match the following pattern are to be crawled:

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

Host name matching is case-sensitive, whether you specify an explicit domain name or a domain name pattern. For example, `*.user.ibm.com` matches `joe.user.ibm.com` and `mary.smith.user.ibm.com`, but not `joe.user.IBM.com`.

If a Web site's port is not 80 (the standard for HTTP), you must specify the port number in the domain name. For example, to enable crawling of secure pages, specify the port number for HTTPS requests (the standard is 443):

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

### Address rules

An address rule enables you to control the crawling of entire hosts or networks by specifying an IP address (IPv4 only) and netmask as the target. For example:

```
allow address 9.0.0.0 255.0.0.0
```

The netmask enables you to specify pattern matching. For an address rule to apply to a candidate IP address, the IP address in the rule and the candidate IP address must be identical, except where masked off by zeros in the netmask. The address rule defines a pattern, and the netmask defines the significant bits in the address pattern. A zero in the netmask acts as a wildcard and signifies that any value that is specified in that same bit position in the address matches.

In the preceding example, the allow rule applies to any IP address with 9 in the first octet, and any value at all in the last three octets.

A useful rule to include as the final address in your list of rules is `forbid address 0.0.0.0 0.0.0.0`. This rule matches any IP address because the netmask makes all bits insignificant (the rule forbids all addresses that are not allowed by a preceding rule in your list of rules).

### Prefix rules

A prefix rule controls the crawling of URLs that begin with a specified string. The target is a single URL, which typically contains one or more asterisks to signify a pattern. For example, an asterisk is often specified as the final character in the prefix string.

A prefix rule enables you to crawl all or part of a Web site. You can specify a directory path or pattern, and then allow or forbid everything from that point on in the directory tree. For example, the following rules work together to allow the crawler to crawl everything in the public directory at foo.ibm.com, but forbid the crawler from accessing any other pages on the site:

```
forbid prefix http://foo.ibm.com/public/*
forbid prefix http://foo.ibm.com/*
```

When you specify prefix rules, you can specify more than one asterisk and you can specify them anywhere in the prefix string, not just in the last position. For example, the following rule forbids the crawler from crawling any documents in a top-level directory of the foo.ibm.com site if the directory name ends in fs. (For example, you might have file system mounts that do not contain information that would be useful in the search index.)

```
http://foo.ibm.com/*fs/*
```

### Crawling rule order

The crawler applies the crawling rules at various times during the process of discovering and crawling URLs. The order of the rules is important, but only within the rules of a each type. It makes a difference whether an address rule comes before or after another address rule, but it makes no difference whether an address rule comes before or after a prefix rule, because the crawler does not apply the rules at the same time.

Within the set of rules for a single type, the crawler tests a candidate domain, address, or URL against each rule, from the first specified rule to the last, until it finds a rule that applies. The action specified for the first rule that applies is used.

**Important:** If no rule applies, then the default action is that the candidate domain, address, or URL is allowed to be crawled.

The dependency on order leads to a typical structure for most crawling rules:

- The set of domain rules typically begins with forbid rules that eliminate single domains from the crawl space. For example, the collection administrator might determine that certain domains do not contain useful information.
- The list of forbid rules is typically followed by a series of allow rules (with wildcard characters) that enable the crawler to visit any domain that ends in one of the high-level domain names that define an enterprise intranet (such as \*.ibm.com and \*.lotus.com).

The set of domain rules typically ends with the following rule, which eliminates domains that were not allowed by a preceding rule:

```
forbid domain *
```

This rule also prevents the crawl space from including the entire Internet, for example.

- The set of address rules typically begins with a small number of allow rules that enable the crawler to crawl the high-level (class-A, class-B, or class-C) networks that span an enterprise intranet.

The address rules typically end with the following rule, which prevents the crawler from crawling Web sites that are outside the corporate network.

```
0.0.0.0 0.0.0.0
```

- The set of prefix rules is usually the largest, because it contains arbitrarily detailed specifications of allowed and forbidden regions that are specified as

trees and subtrees. A good approach is to allow or forbid more tightly localized regions first, and then specify the opposite rule, in a more general pattern, to allow or forbid everything else.

The prefix section does not typically end with a typical rule. The suggested final domain and address rules can ensure that the crawler does not crawl beyond the enterprise network more efficiently than by testing URLs.

The crawler can apply prefix rules more efficiently if you group the rules by action (forbid or allow). For example, instead of specifying short sequences of allow and forbid rules that alternate with each other, specify a long sequence of rules that stipulate one action and then specify a long sequence of rules that stipulate the other action. You can interweave allow and forbid rules to achieve the goals of your crawl space. But grouping the allow rules together and the forbid rules together can improve crawler performance.

### **File extensions, MIME types, and maximum crawl depth**

These options provide additional ways for you to specify content for the crawl space. You can exclude certain types of documents based on document's file extension, and you can include certain types of documents based on the document's MIME type. When you specify which MIME types you want the crawler to crawl, consider that the MIME type is often set incorrectly in Web documents.

The maximum crawl depth is the number of slashes in a URL from its site root. This option enables you to prevent the crawler from being drawn into recursive file system structures of infinite depth. The crawl depth does not correspond to the levels that the crawler traverses when it follows links from one document to another.

### **Seed URLs**

Seed URLs are the URLs that the crawler begins crawling with, and these URLs are inserted into the crawl every time the crawler is started. If the seed URLs were already discovered, they will not be crawled or recrawled sooner than other Web sites that you allow in the crawling rules.

A seed URL is important the first time that a Web crawler is started and the crawl space is empty. A seed URL is also important when you add a seed URL that was not previously discovered to an existing crawl space.

Seed URLs must be fully qualified URLs, not just domain names. You must specify the protocol and, if the port is not 80, the port number.

The following URLs are valid seed URLs:

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

The following URL is not a valid seed URL:

```
www.ibm.com
```

You must also include the seed URLs in your crawling rules. For example, the crawler cannot begin crawling with a specified seed URL if the crawling rules do not allow that URL to be crawled.

## Recrawl interval settings in the Web crawler

To influence how frequently the Web crawler revisits URLs, you specify options in the Web crawler properties.

Most of the other crawler types in an enterprise search system run according to schedules that an administrator specifies. In contrast, after you start a Web crawler, it typically runs continuously. To control how often it revisits URLs that it previously crawled, you specify minimum and maximum recrawl intervals.

When you use the enterprise search administration console to create a Web crawler or to edit Web crawler properties, you can select an option to configure advanced properties. On the Advanced Web crawler properties page, you specify minimum recrawl interval and maximum recrawl interval options. The Web crawler uses the values that you specify to calculate an interval for recrawling data.

The first time that a page is crawled, the crawler uses the date and time that the page is crawled and an average of the specified minimum and maximum recrawl intervals to set a recrawl date. The page will not be recrawled before that date. The time that the page will be recrawled after that date depends on the crawler load and the balance of new and old URLs in the crawl space.

Each time that the page is recrawled, the crawler checks to see if the content has changed. If the content has changed, the next recrawl interval will be shorter than the previous one, but never shorter than the specified minimum recrawl interval. If the content has not changed, the next recrawl interval will be longer than the previous one, but never longer than the specified maximum recrawl interval.

## Options for visiting URLs with the Web crawler

You can force the Web crawler to visit specific URLs as soon as possible.

If you need to refresh the crawl space with information from certain Web sites, you can monitor the crawler, select the **URLs to visit or revisit** option, then specify the URLs or URL patterns of the pages that need to be crawled or recrawled.

For example, if your Communications department adds a Web page to your intranet or revises a page to reflect an important policy change, you can specify the URL of the new or changed page. If the crawler is running, the crawler queues the specified URL for crawling the next time that it checks for pages that are waiting to be visited (typically every ten minutes). If the crawler is not running, it queues the specified URL so that it can be crawled the next time that the crawler is started.

Ensure that the crawling rules include a rule that allows the crawler to visit the URLs that you specify. The crawler can visit the URLs that you specify sooner than it normally would. However, for a URL to be crawled at all, a crawling rule must exist that allows the URL to be crawled.

The newly crawled data becomes available for searching the next time that the index is reorganized or refreshed.

---

## Windows file system crawlers

To include documents that are stored in Microsoft Windows file systems in an enterprise search collection, you must configure a Windows file system crawler.

You can use the Windows file system crawler to crawl any number of Windows file systems. When you configure the crawler, you select the local and remote directories and subdirectories that you want to crawl.

If you install the crawler server on a Windows computer, you cannot use that server to crawl UNIX file system sources (the UNIX file system crawler does not appear in the list of available crawler types).

To create or change a Windows file system crawler, you must log in to the enterprise search administration console. You must be a member of the enterprise search administrator role or be a collection administrator for the collection that owns the crawler.

When you create the crawler, a wizard helps you do these tasks:

- Specify properties that control how the crawler operates and uses system resources. The crawler properties control how the crawler crawls all subdirectories in the crawl space.
- Set up a schedule for crawling the file systems.
- Select subdirectories to crawl.

You can specify how many levels of subdirectories that you want the crawler to crawl. To crawl remote file systems, you also specify a user ID and password that enables the crawler to access data.

- Specify options for making documents in subdirectories searchable. For example, you can exclude certain types of documents from the crawl space or specify security tokens to control access to content.

Click **Help** while you are creating the crawler to learn about the fields in the wizard and how to provide the crawler with the information that it needs to crawl data.

After you create a crawler, you can edit the properties or crawl space as needed. Wizards help you make these changes:

- Change the crawler properties. For example, you might want to change the number of threads that the crawler can activate at a time.
- Add subdirectories to the crawl space.
- Change the schedule for crawling file systems.
- Change the rules for crawling documents in a subdirectory.
- Remove subdirectories from the crawl space.

---

## Configuring support for external crawlers

You can extend enterprise search by using the data listener API to create an external crawler. To enable the crawler to access enterprise search collections, you configure the data listener.

### Procedure

To configure the data listener:

1. On the index server, log in as the enterprise search administrator. This user ID was specified during the installation of DB2 II OmniFind.
2. Change to the ES\_NODE\_ROOT/master\_config/datalistener directory.
3. Use a text editor to edit the dlConfig.prp file.

- In `DL_Port`, specify the port number that the data listener listens to. The default value is 6668.
- In `DL_Num_Of_Threads`, specify the number of working threads that the data listener can create. The default value is 6.

4. Use a text editor to edit the `dlAuthentication.prp` file.

This file specifies the authentication information for the remote client. Each line in the file specifies an ID and password that enables a data listener application to access enterprise search. The client IDs must be unique within the enterprise search system. For example:

```
clientID_1 = password_1
clientID_2 = password_2
```

Specify an ID and password for each data listener application that you want to authorize to access the enterprise search system. The data listener application owner must use the credentials that you specify.

5. Use a text editor to edit the `dlAuthorization.prp` file.

This file specifies authorization data. Each line in the file specifies the ID of a data listener application that can add data to or remove data from a collection. The client IDs must match the IDs that you specified in the `dlAuthentication.prp` file. That means each client ID is allowed to update content in one collection only. For example:

```
clientID_1 = collectionID_1
clientID_2 = collectionID_2
```

6. The data listener is started automatically when the enterprise search system is started. For the changes in the property files to become effective sooner, enter the following commands to stop and restart the data listener:

```
esadmin datalistener stop
esadmin datalistener start
```

### **Related concepts**

Data listener API





---

## Chapter 5. Enterprise search index administration

To ensure that users always have access to the latest information, enterprise search creates an index for each collection and maintains that index by periodically refreshing and reorganizing the content.

To make the data that is collected by crawlers searchable, you must create indexes. When you first create a collection, enterprise search creates an index for all of the data that was initially crawled. When the crawlers crawl new and changed data sources, enterprise search refreshes the index with new content. Eventually, the refreshed content needs to be merged into the base index. This merging process is called reorganization. Whenever the index is refreshed or reorganized, the new content is copied to the search servers and made available for searching.

Crawlers collect data continuously or on a regularly scheduled basis. If you refresh the index frequently, you enable users to search the most current data. Eventually, an index that is continuously refreshed must be reorganized. As a refreshed index grows larger, it consumes more system resources. Therefore, to maintain optimal performance, you should reorganize indexes regularly.

How often you reorganize an index depends on:

- System resources (disk space, processor speed, and memory)
- How many documents are being crawled and recrawled
- The type of data that is being crawled
- How often you change category rules (the rules do not become effective until the index is reorganized)
- How often you force a crawler to start instead of running at a scheduled time
- How often external crawlers remove or add URIs (these types of crawlers interact with enterprise search through the data listener API)

For collections with approximately 4 million documents that are built with mostly Web documents, you should reorganize the index approximately once a day, and refresh the index every one or two hours.

To maintain a current, searchable index, you do the following tasks:

- Specify schedules for refreshing and reorganizing the index
- Change the index schedule
- Enable and disable the index schedule
- Configure concurrent index builds
- Remove URIs from the index

### **Related concepts**

“Scopes” on page 80

Scopes enable you to restrict the documents that users can search when they search a collection.

### **Related tasks**

“Monitoring index activity for a collection” on page 106

Monitor the index for a collection when you need to see the progress of an index that is being built, enable or disable the index schedule, or start and stop indexing activity.

“Monitoring the enterprise search index queue” on page 106

You can view the status of all index builds in the index queue, stop an index that is being built, or delete an index from the queue.

---

## Scheduling index builds

You can specify schedules for reorganizing an index and refreshing an index with new content.

### Prerequisites

To schedule an index build, you must be a member of the enterprise search administrator role or a collection administrator for that collection.


### About this task

To ensure that users always have access to the latest information in the sources that they search, schedule the index to be reorganized and refreshed on a regular basis. When an index is reorganized, the entire index is rebuilt. The indexing processes read all of the data that was collected by crawlers and analyzed by the parser. When an index is refreshed, information that was crawled since the last time the index was reorganized is made searchable.

By default, the option to schedule index builds is selected. This option tells the scheduler process to schedule tasks to refresh and reorganize the index whenever the enterprise search system is started. You can clear the **Enable when system starts** check box at any time if you need to prevent a scheduled index build from running. For example, you might need to disable the schedule to troubleshoot problems.

### Procedure

To schedule index builds:

1. Click **Collections** to open the Collections view.
2. In the list of collections, locate the collection that you want to change and click  **Edit**.
3. To specify how often the index is to be refreshed with new content, specify the following options on the Index page in the **Specify a schedule to refresh content** area:
  - a. In the **Start on** area, in the **Month**, **Day**, **Year**, and **Hour** fields, specify when you want the index to be refreshed the first time.
  - b. In the **Update every** area, in the **days** and **hours** fields, specify how often you want the index to be refreshed.

Typically, you should refresh the index frequently, such as every hour or two. Depending on how often the source content changes, specify a lower or higher interval. For example, you might specify every hour (0 days and 1 hour) or every 12 hours (0 days and 12 hours).
4. To specify how often the index is to be completely rebuilt, specify the following options in the **Specify a schedule to reorganize the index** area:
  - a. In the **Start on** area, in the **Month**, **Day**, **Year**, and **Hour** fields, specify when you want the index to be reorganized the first time.
  - b. In the **Update every** area, in the **days** and **hours** fields, specify how often you want the index to be reorganized.

Typically, you should reorganize the index regularly, such as every 24 hours. Depending on how often the source content changes, specify a lower or higher interval. For example, you might specify every 12 hours (0 days and 12 hours) or every two and a half days (2 days and 12 hours).

5. Click **Apply**.

## Changing the index schedule


You can change the schedule for reorganizing or refreshing an index.

### Prerequisites

To change an index schedule, you must be a member of the enterprise search administrator role or be a collection administrator for that collection.

### Procedure

To change the index schedule:

1. Click **Collections** to open the Collections view.
2. In the list of collections, locate the collection that you want to change and click  **Edit**.
3. On the Index page, change the appropriate values in the **Month**, **Day**, **Year**, and **Hour** fields to specify how often the index is to be refreshed with new content or how often the index is to be reorganized.
4. Click **Apply**.

## Enabling and disabling the index schedules

You can enable and disable the schedules for refreshing and reorganizing the index.

### Prerequisites


To enable or disable an index schedule, you must be a member of the enterprise search administrator role or be a collection administrator for that collection.





### About this task

You can disable a schedule for an index if you need to prevent a scheduled index build from running. For example, you might want to disable the schedule to prevent an index from being built at the scheduled date and time so that you can troubleshoot problems.

You can enable and disable the schedule while you are editing a collection, and you can enable or disable the schedule while you are monitoring a collection.

### Procedure

1. To enable or disable the schedule for an index by editing a collection, take the following steps:
  - a. Click **Collections** to open the Collections view.
  - b. In the list of collections, locate the collection that you want to change and click  **Edit**.
  - c. On the Index page, select or clear the **Enable when system starts** check box to enable or disable the schedule for refreshing the index.

- d. To enable or disable the schedule for reorganizing the index, select or clear the **Enable when system starts** check box.
      - e. Click **Apply**.
    2. To enable or disable the schedule for an index by monitoring a collection, take the following steps:
      - a. Click **Collections** to open the Collections view.
      - b. In the list of collections, locate the collection that you want to change and click  **Monitor**.
      - c. On the Index page, if an index is scheduled, and you do not want it to be built at the scheduled date and time, click  **Disable schedule**. The index is not built until you enable the schedule or click  **Start** to start the index building process.
      - d. If an index is scheduled, but the schedule for building it is disabled, click  **Enable schedule**.

The index is queued for building at the date and time that you specified in the index schedule.

---

## Configuring concurrent index builds

Multiple indexes can share the indexing processes. By sharing the processes, you can ensure that the reorganization of a very large index does not block the availability of other indexes that are waiting in the queue to be built.

### Prerequisites

To specify how many indexes can be in the index queue concurrently, you must be a member of the enterprise search administrator role.

### Restrictions

The number of concurrent index builds cannot exceed the number of collections in your enterprise search system. For example, if you have five collections, you must enter a number that is less than or equal to five.

### About this task

Index building is a resource-intensive process. A large amount of system memory and disk space is consumed while an index is being built.

Enterprise search can build multiple indexes at a time, but only one index per collection can be in the queue at a time. When an index build is requested or scheduled, it enters the index queue and waits for its turn to be processed. Because each collection has its own index, there might be several indexes in the index queue at the same time. When you configure the number of concurrent index builds, you specify how many indexes can share the queue and the index building resources at the same time.

For example, with concurrent builds, the indexing processes can work on the index for one collection, then move on to work on other indexes (up to the maximum number of concurrent builds that you allow), before returning to work again on the first index. An index build can progress without waiting for the completion of other indexes that precede that index in the queue.

### Procedure

To specify how many indexes can share the index building queue concurrently:

1. Click **System** to open the System view.
2. On the Index page, type the number of indexes that can share resources and be built at the same.
3. Click **Apply**.

---

## Removing URIs from the index

To prevent users from searching documents in a collection, you can remove the URIs for those documents from the index.

### Prerequisites

To remove URIs from the index, you must be a member of the enterprise search administrator role or a collection administrator for that collection.

### About this task


If you specify a fully qualified URI, users stop seeing the URI in the search results. However, if a user submits the same query, and result documents for that query are in the search cache, then the cached result page for the URI that you removed continues to be returned in the search results. The search cache is not refreshed, and the URI is not removed from the index, until the next time the index is refreshed or reorganized.

If you specify a pattern to remove multiple URIs, users continue to see the URIs that match that pattern in the search results until the next time the index is refreshed or reorganized.

When you remove a URI from the index, you do not remove it from the crawl space. The next time that the crawler crawls the document, the URI is built into the index and becomes available for searching again. To remove a URI from the crawl space, you must update the crawling rules to exclude the document, and then stop and restart the crawler.

### Procedure

To remove URIs for specific documents from the index:

1. Click **Collections** to open the Collections view.
2. In the list of collections, locate the collection that you want to change and click  **Edit**.
3. On the General page, click **Remove URIs from the index**.
4. On the Remove URIs from the index page, type the URIs (or the URI patterns) that you want to remove from the index.

For example:

```
http://domain.org/hr/*  
db2://knowledgeManagement/ROI*  
cm://enterprise/finance*
```

5. Click **OK**.

### Related reference

Chapter 17, "URI formats in enterprise search," on page 139  
The uniform resource identifier (URI) of each document in an enterprise search collection indicates the type of crawler that added the document to the collection.

---

## Chapter 6. Search applications for enterprise search

A search application enables you to search collections in your enterprise search system. You can create any number of search applications, and a single search application can search any number of collections.

### Sample search application

The sample search application for enterprise search demonstrates most of the available search and retrieval functions. This sample application is also a working example that demonstrates how you can use the IBM search and index API (SI-API) to build interactive, Web-based search applications that reflect the goals of your enterprise.

Unless you change its configurations settings, the sample search application enables you to search all active collections in your enterprise search system. You can use the sample search application to test new collections before you make them available to users. In a production environment, enterprise search administrators control which search applications are allowed to search specific collections.

### Search and retrieval functions

The sample search application demonstrates the following search and retrieval functions.

#### Collection-level functions

With these functions you can:

- Select the collection that you want to search (the list of collections includes only the collections that an enterprise search administrator associates with your search application).
- Refresh the list of collections to access new collections and changes to existing collections.

#### Query functions

With these functions you can:

- Specify simple, free-text queries.
- Specify more complex queries to improve the precision of search results. For example, you can search specific fields in a document or use syntax notations to search for documents that include specific phrases or exclude specific words.
- Search specific types of documents. If the collection includes multiple data source types, you can also search specific types of data, such as only Microsoft Word documents or only portable document format (PDF) documents.
- Specify which language your query terms are in.
- Be prompted with suggestions for spelling corrections if possibly misspelled words are detected in the query string.
- Search specific subsets of a collection. For example, a search application can limit your view to a predefined range of documents (a scope), or you can submit a query that searches only the documents that belong to a specific category.

## Search result functions

With these functions you can:

- See your search results on the query page. You can clear the search results from the display at any time.
- Control how many result documents appear on each search results page, and browse forward and backward through the result set.
- Hide and display details about the result documents. For example, you can browse the document titles or view summaries and additional fields for each document.
- Collapse documents from the same Web site. For example, if one site returns 100 documents, the top two documents are grouped together in the result set. You can see the remaining 98 documents by selecting an option to view them.
- Restrict documents in the search results to a specific language.
- Sort documents by relevance or by document date, and specify whether the sort order is ascending or descending.
- View information about the categories that a result document belongs to (if the collection uses categories), and browse only the documents that belong to a specific category.
- Retrieve a document by clicking its URI.
- Retrieve documents by clicking quick links. A quick link associates keywords with URIs. If a query includes the specified keywords, the associated URIs (which were predetermined to be highly relevant for those keywords) appear in the search results.

## Custom search applications

You can run the search applications that you create as stand-alone Web applications in an IBM WebSphere Application Server environment, or you can launch them as portlets in an IBM WebSphere Portal environment. By using the search and index API, you can design search applications that, like the sample search application, work seamlessly in both environments.

### Related concepts

Chapter 8, “How enterprise search searches collections,” on page 71

Search applications for enterprise search enable users to specify simple, free-text queries and more complex queries that provide much finer document matching criteria.

“Query techniques” on page 71

Enterprise search collections support a range of query techniques, many of which are familiar to users who are accustomed to searching the Internet.

Chapter 9, “How enterprise search returns search results,” on page 85

When a user enters a query in a search application, the search processes return the most relevant results for the terms and conditions of the query.

“Security with search application IDs” on page 117

To provide collection-level security, you specify which search applications can search each collection.

“Document-level security” on page 114

Document-level security ensures that users who search collections are able to access only the documents that they are allowed to see.

“Document-level security with the Portal Search Engine” on page 121

You can use the IBM WebSphere Portal Search Engine to enforce document-level security when users search enterprise search collections.



---

## Configuring default search application properties

The sample search application for enterprise search can search all active collections in your system. You can edit a properties file to specify options for your Web server environment or to specify a different search application as the default application.

### About this task

The DB2 II OmniFind installation program deploys a sample search application for enterprise search into IBM WebSphere Application Server on the search servers for enterprise search. To configure this search application, you edit a properties file, `config.properties`, that is deployed with it.

When you edit the properties file, you specify information about using the sample search application with your Web server.

### Procedure

To configure default search application properties:

1. Edit the `config.properties` file with a standard text editor.

The `config.properties` file is installed in the following location, where `ES_INSTALL_ROOT` is the DB2 II OmniFind installation directory on the search server:

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

2. Edit the following properties for your environment:

#### **applicationName**

Specify the name of a valid search application. The default value is `Default`.

Change the default value if you do not want to use the sample search application for enterprise search as the default application and want to use a different application instead. For example, if you enabled global security in WebSphere Application Server and want to use a secure application as your default search application, specify the name of that application.

#### **applicationPassword**

Specify a password for the search application. The default value is `password`.

Change this value only if the `applicationName` property specifies an application that requires a password. This value is ignored by the Default sample search application.

#### **hostName**

Specify the fully qualified host name of the Web server that is configured to support your WebSphere Application Server instance. The default value is `localhost`.

To ensure that the search application functions correctly, change the default value to the fully qualified host name that your Web server and WebSphere Application Server are configured to use. For example, if

the computer host name is MyMachine and the Web server is configured to use www.ibm.com as its host name, specify www.ibm.com.

**portNumber**

Specify the port number of the Web server that is configured to support your WebSphere Application Server instance. The default value is 80.

Change the default value only if you change the port for your Web server instance. If you change this value, also update the portNumber value in the following file:

*ES\_INSTALL\_ROOT/master\_config/searchmanager/setup.ini*

**categorizationType**

Specify the type of categories that the search application supports:

**rulebased**

Specify this value if the search application uses categories that you configure by specifying rules in the enterprise search administration console. This is the default value.

**modelbased**

Specify this value if the search application uses model-based categories that exist in IBM WebSphere Portal. If you select this option, ensure that WebSphere Portal is installed on the enterprise search index server. You must also use WebSphere Portal to administer the categories.

3. Using the WebSphere administrative console, stop and restart the ESSearchApplication application.

---

## Accessing the sample search application

You access the sample search application for enterprise search by specifying a URL in a Web browser.

**Prerequisites**

The sample search application must be configured for your Web server environment.

**About this task**

The sample search application is installed on the search servers for enterprise search. You can use this application as provided to test collections before you make them available to users. You can also use the application as a model for creating your own search applications.

**Procedure**

To start the sample search application:

1. Type the URL for the search application in a Web browser. For example:

`http://SearchServer.com/ESSearchApplication/`

*SearchServer.com* is the host name of the search server.

Depending on your Web server configuration, you might also need to specify the port number. For example:

`http://SearchServer.com:9080/ESSearchApplication/`

2. On the **Search** page, select the collection that you want to search and submit a query.

---

## Associating search applications with collections

Before you can use a new search application, it must be associated with the collections that it can search.

### Prerequisites

To associate search applications with the collections that they can search, you must be a member of the enterprise search administrator role.

### Procedure

To associate a search application with one or more collections:

1. Click **Security** in the toolbar of the administration console.
2. On the Search Applications page, click **Add Search Application**.
3. Type the name of the search application.
4. Select the collections that the application can search:
  - Click **All collections** if you want the search application to access all collections.
  - Click **Specific collections** if you want the search application to access only the collections that you specify.

When you select this option, a list of collection names is displayed. Select the **Select** check box for each collection that the application can search.



---

## Chapter 7. Linguistic capabilities in enterprise search

Enterprise search offers linguistic search support for text documents in most Indo-European languages and Asian languages such as Japanese.

The purpose of linguistic support is to improve the quality of search results by finding the best possible collection of documents that match the query.

Linguistic processing is performed during two stages of processing: when a document is processed to be added to the index and when a user issues a query.

When a document is processed, enterprise search determines the language of a document, segments the stream of input text into distinct units, also called tokens, and then creates an index that contains these tokens that forms the basis of the search.

When a query is submitted, the user or the application must manually specify the query language. The query string is segmented and analyzed, and then searched for in the index.

Documents and query strings can be analyzed by two types of analysis, depending on the language:

### **Dictionary-based linguistic analysis**

This type of analysis processes common languages. The analysis includes word and sentence segmentation and lemmatization. *Lemmatization* is a form of linguistic processing that analyzes the base form of each word that occurs in text, and determines inflected forms that share the same part of speech (for example, lemmatization of the word goose finds the word geese).

Linguistic processing also involves lexical analysis, which is the process of creating alternative representations of the input text that associates all available dictionary data to the tokens that are recognized in the input text. Search quality is greatly enhanced by using advanced language processing.

### **Nondictionary-based analysis**

This type of analysis processes less common languages and is required less often. The analysis includes Unicode-based white space segmentation and n-gram segmentation.

#### **Related concepts**

“Enterprise search parsers” on page 4

An enterprise search parser analyzes documents that were collected by a crawler and prepares them for indexing.

#### **Related tasks**

“Monitoring the parser” on page 105

Monitor the parser for a collection when you need to start the parser or stop it.

---

## Language identification

Before word and sentence segmentation, character normalization, or lemmatization can occur, enterprise search must determine the language of the source document.

Enterprise search can automatically detect the following languages:

Arabic	French	Korean
Chinese (Traditional and Simplified)	German	Polish
Czech	Greek	Portuguese
Danish	Hebrew	Russian
Dutch	Hungarian	Spanish
English	Italian	Swedish
Finnish	Japanese	Turkish

The linguistic processes for enterprise search detect the language of a source document during indexing, not during query processing.

Documents for which the language cannot be detected automatically are processed with basic language-independent technology.

The enterprise search language detection technology is best suited for monolingual documents. If a document is multilingual, an attempt is made to determine the most dominant language used in the document. However, the analysis results are not always satisfactory.

The language detection technology in enterprise search can be used to restrict your search results to only documents that are in a particular language. For example, if you search for documents about Jacques Chirac, you can specify that only documents that are written in French are to be included in the search results.

---

## Stop word removal

In enterprise search, all stop words, for example, common words like *a* and *the*, are removed from multiple word queries to increase search performance.

Stop word recognition in Japanese is based on grammatical information. For example, enterprise search recognizes whether the word is a noun or a verb, whereas the other languages work with special lists.

---

## Character normalization

Character normalization is a process that can improve recall. Improving recall by character normalization means that more documents are retrieved even if the documents do not exactly match the query.

Enterprise search uses Unicode compatibility normalization that includes the normalization of Asian full-width and half-width characters.

For example, in Japanese, a full-width alphanumeric character is normalized to the half-width character, a half-width Katakana character to the full-width character, and so on. Enterprise search also removes Katakana middle dots, which are used as compound word delimiters in Japanese.

Other forms of character normalization include:

### Case normalization

For example, finding documents with *USA* when searching for *usa*.

**Umlaut expansion**

For example, finding documents containing *schoen* when searching for *schön*.

**Accent removal**

For example, finding documents containing *é* when searching for *e*.

**Other diacritics removal**

For example, finding documents containing *ç* when searching for *c*.

**Ligature expansion**

For example, finding documents containing *Æ* when searching for *ae*.

All normalizations work both ways. You can find documents that contain *usa* when you search for *USA*, documents that contain words with *e* when you search for *é*, and so on. These normalizations can also be combined. For example, you can find documents that contain *météo* when you search for *METEO*.

The normalizations are based on Unicode character properties and are not language-dependent. For example, enterprise search supports diacritic removal for Hebrew and ligature expansion for Arabic.

---

## Linguistic support for dictionary-based segmentation

If the language of a document is correctly detected, and language-specific dictionaries are available, then appropriate linguistic processing is applied.

Segmentation is the process by which input text is broken down into distinct lexical units. This process comprises some of the following linguistic processing activities:

**Word segmentation**

Word segmentation is used for languages that do not use white spaces (or delimiters) between words, such as Japanese and Chinese.

**Lemmatization**

Lemmatization is a form of linguistic processing that determines the lemma for each word form that occurs in text. The *lemma* of a word encompasses its base form plus inflected forms that share the same part of speech. For example, the lemma for *go* encompasses *go*, *goes*, *went*, *gone*, and *going*. Lemmas for nouns group singular and plural forms (such as *calf* and *calves*). Lemmas for adjectives group comparative and superlative forms (such as *good*, *better*, and *best*). Lemmas for pronouns group different cases of the same pronoun (such as *I*, *me*, *my*, and *mine*).

Lemmatization requires a dictionary for both indexing and searching.

Enterprise search indexes the lemmas and the inflected words, and lemmatizes all inflected words in a query. Lemmatization enhances search quality by finding documents that contain variants of an inflected word in the query. For example, documents that contain the word *mice* are found when a query includes the word *mouse*.

**Contraction splitting**

Search quality is improved by identifying contractions and splitting them up into their component parts. For example:

*wouldn't* is split into *would* + *not*

*Horse's* is split into *Horse* + *is* or *'s* (to account for query ambiguity)

### Clitic identification

Clitics are a special form of contractions, and search quality is improved by determining their component parts. A *clitic* is an element that behaves like an affix and a word. However, clitics are difficult to identify because they are also part of word formation. Unlike other morphological (word structure) phenomena, clitics occur in a syntactic structure and their attachment to words is not part of the word formation rules. For example:

*reparti-lo-emos* has the components *repartir* + *lo* + *emos*  
*l'avenue* has the components *le* + *avenue*  
*dell'arte* has the components *dello* + *arte*.

### Nonalphabetic character recognition

The linguistic processes recognize nonalphabetic characters. Depending on the internal language-dependent logic, some nonalphabetic characters are returned as separate lexical units of different types, and some are grouped.

For example, apostrophes or hyphens in the case of clitics are considered word parts, and they are considered full stops (or periods) in the case of unknown abbreviations. The linguistic processing can also recognize some special sequences of characters as tokens, for example, URLs, e-mail addresses, and dates.

### Abbreviation recognition

The linguistic processes recognize abbreviations that are in the dictionary as one lexical unit. If the abbreviation is not in the dictionary, then the abbreviation is recognized as a lexical item, but the abbreviation will not have any associated dictionary information.

Recognizing abbreviations correctly is vital for sentence recognition. For example, the period at the end of an abbreviation is not necessarily the end of a sentence.

### End-of-sentence marker recognition

The linguistic processes correctly identify end-of-sentence markers for sentence segmentation.

Dictionary-based linguistic support is available for the following languages:

Chinese (Simplified and Traditional)	Italian
Czech	Japanese
Danish	Korean
Dutch	Norwegian (Bokmal and Nynorsk)
English	Polish
Finnish	Portuguese (National and Brazilian)
French (National and Canadian)	Russian
German (National and Swiss)	Spanish
Greek	Swedish

---

## Linguistic support for nondictionary-based segmentation

For documents in languages that are not supported by language detection and lexical analysis technology, enterprise search provides basic support in the form of Unicode-based white space and n-gram segmentation.

### Unicode-based white space segmentation

This method of linguistic processing uses the white space (or blank space) between words as a word delimiter.



### **N-gram segmentation**

This method of linguistic processing treats overlapping sequences of  $n$  characters as a single word. This simple method of segmentation is sufficient for many retrieval tasks.

These methods are independent of any language dictionary and do not include sophisticated linguistic processing technology, such as base-form reduction.

N-gram segmentation is used for languages such as Thai that have no blank spaces to use as delimiters. The same method applies to Hebrew and Arabic. Although these two languages use white space delimiters, n-gram segmentation returns better results than the basic form of Unicode-based white space segmentation does.

## **Word segmentation in Japanese**

If the text document or the query string is recognized as being in Japanese, enterprise search performs relevant word segmentation by using morphological analysis technology that is optimized for the Japanese language.

An example of this optimization is word decomposition. Japanese uses a large number of compound words. These words are decomposed into tokens of optimal size to achieve better search results. Inflected words and prepositions are also decomposed to improve search performance.

## **Orthographic variants in Japanese**

Japanese uses many orthographic variants. Katakana variants are the most important because Katakana is often used to spell and pronounce foreign words. Many Katakana variants are commonly used in Japanese.

Enterprise search uses a variant dictionary to map typical Katakana variants to their base forms (similar to a lemma) so that all documents, including those with orthographic variants of the Katakana word in the query string, are found.

Enterprise search also supports typical Okurigana variants, which are Kanji word endings that are written in Hiragana.



---

## Chapter 8. How enterprise search searches collections

Search applications for enterprise search enable users to specify simple, free-text queries and more complex queries that provide much finer document matching criteria.

When a user submits a query, the search servers use the index to quickly locate relevant documents. The search servers use the enterprise search data store, which contains the parsed and tokenized data, to retrieve metadata for the relevant documents. Metadata can include but is not limited to the document's URI, title, description, date, data type, and so on.

When you create a collection, you specify options for how you want data and metadata to be parsed, stored in the index, and searched by search applications. The options that you can specify for making data searchable include:

### **Configuring quick links**

You can predetermine URIs to be returned for certain keywords and phrases. When users specify the keywords or phrases, the predefined URI is returned with the search results. The quick link URIs are returned in addition to URIs that the search servers return by searching the index.

### **Configuring categories**

You can group documents that share a similar URI pattern or that contain specific words into categories. When users search the collection, they can limit the search results to only the documents that belong to specific categories.

### **Configuring scopes**

You can specify a range of URIs that are in the index and limit the user's view of the collection to only those pages.

### **Configuring searchable fields**

For data sources that include fields, you can specify whether you want users to be able to search the data by specifying field names in queries. You can also map elements in XML documents to field names, and enable users to query specific parts of XML documents.

### **Related concepts**

Chapter 9, "How enterprise search returns search results," on page 85

When a user enters a query in a search application, the search processes return the most relevant results for the terms and conditions of the query.

---

## Query techniques

Enterprise search collections support a range of query techniques, many of which are familiar to users who are accustomed to searching the Internet.

### **Free text search**

The enterprise search index is a full text index with content from various data sources. You can search all of the content by specifying a simple query in natural language. The search processes search all fields and all content to find documents that are relevant to the query.

**Example:**

A free text search can be as simple as the following query:

```
bicycle chain
```

To indicate which words must or must not appear in a document, you can include special notations. For example, you can precede a word by a plus sign (+) to specify that a document must contain that word for a match to occur. Precede a word by a minus sign (-) to exclude documents that contain that word from the search results. Enclose two or more words in quotation marks (") to search for an exact phrase.

**Example:**

In the following free text query, a match occurs only if a document contains the exact phrase science fiction and does not contain the word robot:

```
+"science fiction" -robot
```

**Fielded search**

A fielded search enables you to constrain the object of the query to specific fields and metadata of a document. For example, you can specify that certain words must exist in the title of a document.

To specify a fielded search in enterprise search, include the field name and the word or phrase that must exist in that field in your query.

**Example:**

The following query searches for documents that must contain the word `ibm` and the phrase `enterprise search` in the title field:

```
title:ibm title:"enterprise search"
```

To search a field by field name, you must enable the field for fielded searching when you configure the crawler.

**Parametric search**

A parametric search is a type of fielded search that enables you to do comparative or evaluative queries on numeric and date fields and metadata. For example, you can search for documents that are of a certain size or that were written after a certain date. You can also search for documents with attributes that are greater than, less than, or equal to a specified value.

**Example 1:**

The following query searches for items that cost exactly 50 dollars (or whatever currency unit is indexed for the price field):

```
#price::=50
```

**Example 2:**

The following query searches for documents that have a file size greater than 1024 but less than or equal to 2048:

```
#filesize::>1024<=2048
```

To search a field with a parametric query, you must enable the field for parametric searching when you configure the crawler.

### Advanced query syntax

The search techniques supported in enterprise search further provide a level of granularity that enables you to query specific documents attributes. For example, you can search for documents that were written in a specific language, documents of a particular document type, and documents that come from a specific type of data source.

#### Example:

The following example searches for text documents that were written in English and added to the collection by the Web crawler:

```
$doctype::text/html $language::en $source::web
```

### Spell checking

When you specify a query, enterprise search provides an option for checking the spelling of query terms. If a user misspells a term in the query, the search server can provide suggestions for how to spell the term correctly.

For example, if you specify saerch as a query term, you would see an option to specify search as a possible correction to the spelling of your original term.

The spelling detection support varies by query language and by the documents that are in the collection that is being searched.

#### Related concepts

Query syntax

Enterprise search uses the following syntax to process queries.

---

## Working with quick links

Quick links are documents that are returned in the search results whenever a user submits a query that includes specific words and phrases.

You use the enterprise search administration console to configure quick links for a collection.

### Quick links

Quick links enable you to provide users with links to specific documents.

A quick link is a URI that enterprise search automatically includes in the search results when a query includes certain words or phrases. Typically, the quick link URIs appear at the top of the result list, which helps ensure that users see the documents that you predetermined to be relevant to the query.

Quick links are returned in addition to other search results. The search processes search the index for documents that match the query terms, and return URIs for those documents in addition to the quick link URIs.

When you configure a quick link, you can specify a descriptive title and summary for the URI to help users recognize the document and quickly determine whether it is a document that they want to retrieve.

For example, for the URI <http://www.ibm.com/education/us/>, you might use a title such as IBM Education in the United States, and provide the summary Solutions, products, and resources for professionals, educators, and students in the United States.

To use quick links in enterprise search collections, the option for showing quick links must be available in the search application. In some search applications, users might have the ability to enable and disable the return of quick links when they search the collection.

## Configuring quick links

To create a quick link for an enterprise search collection, you associate the URI of a document with the keywords that trigger its inclusion in the search results.

### Prerequisites


To configure quick links, you must be a member of the enterprise search administrator role or be a collection administrator for the collection that the quick link belongs to.

### About this task

For examples of how to specify keywords and URIs for quick links, click **Help** while you are creating or editing a quick link.

### Procedure

To configure a quick link:

1. Click **Collections** to open the Collections view.
2. In the list of collections, locate the collection that you want to configure a quick link for and click  **Edit**.
3. On the Search page, click **Configure quick links**.
4. On the Quick links page, click **Create Quick Link**.
5. Specify the keywords and phrases that cause this quick link to be returned in the search results, the URI for the document that you predetermined is relevant to this query, and other options for this quick link.

You can specify one keyword, several keywords, or one phrase (two or more words enclosed in quotation marks) per line. Separate keywords with a space (you cannot use a comma to delimit keywords). Press the Enter key to start a new line.

6. Click **OK**.

Your new quick link is listed on the Quick links page with the other quick links that belong to this collection.

### Related reference

Chapter 17, “URI formats in enterprise search,” on page 139

The uniform resource identifier (URI) of each document in an enterprise search collection indicates the type of crawler that added the document to the collection.

---

## Working with categories

Categories enable you to group documents that share common characteristics, and search and retrieve only documents that meet the criteria for being members of that group.

If you associate documents with categories, and your search applications support this feature, users can search a subset of the collection by specifying the category name. If they search the entire collection, users can refine the search results and browse only the documents that are in the same category as one of the result documents.

Enterprise search supports two approaches for categorizing documents:

### Rule-based

Use this approach if you want to associate documents with categories according to rules that you specify. For example, you can group documents that share a URI pattern or group documents that contain or exclude specific words and phrases.

### Model-based

Use this approach if you use IBM WebSphere Portal and you want to use the predefined categories in WebSphere Portal to search enterprise search collections.

To create and administer categories, you use the enterprise search administration console:

- You select the categorization type when you create a collection. You can choose to use no categories, use rule-based categories, or use model-based categories.
- When you configure parsing rules for the collection, you can change the categorization type, if necessary. If you change the categorization type after documents are crawled and indexed, search quality will be degraded until you recrawl the documents and reorganize the index.
- If you choose to use rule-based categories, you use the administration console to administer the category tree, categories, and category rules.

### Related concepts

“Enterprise search parsers” on page 4

An enterprise search parser analyzes documents that were collected by a crawler and prepares them for indexing.

### Related tasks

“Monitoring the parser” on page 105

Monitor the parser for a collection when you need to start the parser or stop it.

## Rule-based categories

You can configure rules to control which documents are associated with categories in an enterprise search collection.

You can create categories and category rules for collections that you create in enterprise search and for rule-based categories that you migrate from IBM WebSphere Portal collections. To configure rules for categorizing documents, you

must specify that you want to use rule-based categories when you create the collection or when you specify parsing rules for the collection.

The parser uses the rules that you specify to associate documents with one or more categories:

- If a document passes at least one rule in a category, the parser associates the document with the category.
- If a document passes at least one rule in several categories, the parser associates the document with all of the categories.
- If a document does not pass any of the rules for a category, the parser does not associate the document with a category. Users can search for this document and retrieve it when they search the collection, but they cannot search a category and expect to retrieve the document.

When you administer the category tree (or taxonomy) for a collection, you decide where in the hierarchy of categories that you want to add a new category. You also use the category tree to select a category that you want to edit, and then add rules for categorizing documents, delete rules, or change the content of individual rules.

When you configure a rule for categorizing documents, you choose whether enterprise search is to use the URI of a document or content in the document to determine whether the document belongs to the category:

#### **URI pattern**

A URI rule applies to the document's URI. You specify a partial URI (a pattern), and documents that have the specified pattern in their URIs pass the rule.

For example, if you specify that the rule text is `/hr/`, then the first URI below passes the rule, and the second URI does not:

```
file:///corporate/hr/medicalform.doc
http://company.com/human_resources/medicalform.htm
```

#### **Document content**

A content rule applies to the text of the document. You express the rule in the same format as a query. If the document is valid for the query, it passes the rule. When you configure the rule, you specify the words and phrases that documents must contain or exclude, and you choose a language for applying word stemming rules.

For example, the following rule specifies that if a document contains either the word `hr` or the phrase `human resources`, the document passes the rule:

```
hr "human resources"
```

For another example, the following rule specifies that if a document contains the word `hr` but not the word `benefits`, the document passes the rule:

```
+hr -benefits
```

#### **Related tasks**

"Migrating a collection from WebSphere Portal" on page 124

To migrate collections from WebSphere Portal to enterprise search, prepare the collections in WebSphere Portal, then use the migration wizard to migrate them.

## **Model-based categories**

If you use model-based categories in your IBM WebSphere Portal system, you can continue to use those categories with enterprise search collections.



WebSphere Portal provides a predefined taxonomy that includes over 2,300 subjects. These subjects are grouped into main business category areas, such as Computers, Finance, and Transportation. Portal users can create applications that automatically determine which documents match these subject areas, and they can customize the categories for their own business needs.

If you want to use the WebSphere Portal categories with enterprise search, you must:

- Use the migration wizard to import model-based taxonomy files to enterprise search.
- Specify that you want to use model-based categories when you create a collection or when you configure parsing rules for a collection.
- Ensure that WebSphere Portal is installed on the enterprise search index server.
- Use the categorization tools in WebSphere Portal to administer the categories. You cannot administer model-based categories with the enterprise search administration console.

#### **Related tasks**

“Migrating a model-based taxonomy from WebSphere Portal” on page 123  
You can select which model-based taxonomy you want to use with an enterprise search collection by using the Taxonomy Management Portlet in your WebSphere Portal installation. Collections that you already migrated to enterprise search are not affected by a new taxonomy selection.

## **Category trees**

A category tree enables you to view all the rule-based categories in a collection. You use the category tree to create categories, delete categories, and edit the rules that associate documents with categories.

A category tree, which is also called a taxonomy, is arranged in a hierarchy. The tree starts with the root category, and all other categories stem from the root category. You can nest any number of categories and subcategories to provide users with different choices for browsing and retrieving documents.

For example, if a document passes the rules in several categories, it is associated with all of those categories. When users search a category, or browse documents that belong to a category when they browse search results, the fact that a document belongs to multiple categories enhances the likelihood that users will find it.

When you administer the category tree, you can control which documents belong to one or more categories by nesting new categories under existing categories. When you create a category, you specify whether it is to be created at the root level or as a subcategory of another category. You also use the category tree to delete categories from the collection and to change the rules for associating documents with categories. When you edit a category, you can rename the category, add or delete categorization rules, or modify the content of individual rules.

When you administer the category tree, use the following descriptions of search and browse behavior as a guideline:

- If a user searches a high-level category, that category and all of its subcategories are searched for documents that match the search criteria. If a user searches a category that has no additional subcategories, only that category is searched.

- If a user is browsing search results and selects an option to browse documents that belong to a specific category, only the documents in that category are displayed. The names of any subcategories are also displayed in the search results, so that the user can navigate between categories and view subsets of documents at a time.

#### **Related tasks**

“Migrating a collection from WebSphere Portal” on page 124

To migrate collections from WebSphere Portal to enterprise search, prepare the collections in WebSphere Portal, then use the migration wizard to migrate them.

## **Selecting the categorization type**

When you select a categorization type, you specify the approach that you want to use to associate documents with categories in the collection.

### **Prerequisites**

To change the categorization type, you must be a member of the enterprise search administrator role or be a collection administrator for the collection that you are changing.


### **About this task**

The categorization type is specified when the collection is created. If necessary, you can change how documents are categorized for a collection. You can use rule-based categories that you configure specifically for a collection, use model-based categories that exist in your IBM WebSphere Portal system, or use no categories.

**Important:** If you change the categorization type after you crawl data and create an index for a collection, the index becomes inconsistent. To ensure the accuracy of search results, recrawl all documents in the collection and reorganize the index.

### **Procedure**

To select the categorization type:

1. Click **Collections** to open the Collections view.
2. In the list of collections, locate the collection that you want to change and click  **Edit**.
3. On the Parse page, click **Select a categorization type**.
4. On the Select a categorization type page, select one of the following options:

**None** Select this option if you do not want to categorize documents in this collection.

#### **Rule-based (category rules that you configure for this collection)**

Select this option if you want to use a taxonomy that contains rules that you configure specifically for this collection.

- If you are configuring a collection that you created for enterprise search, select this option to specify category names and rules for categorizing documents.
- If you are configuring a collection that you migrated from WebSphere Portal, select this option to use or change the rule-based categories that you imported.

### **Model-based (model-based categories that exist in WebSphere Portal)**

Select this option if you want to associate documents with model-based categories that exist in a WebSphere Portal system. To use this option, WebSphere Portal must be installed on the enterprise search index server. You must also use the categorization tools in WebSphere Portal to administer the categories.

5. Click **OK**.

## **Configuring categories**

You can create any number of categories for a collection, and each category can contain any number of rules. The rules determine which documents are associated automatically with the category.

### **Prerequisites**

To configure categories, you must be a member of the enterprise search administrator role or be a collection administrator for the collection that the categories belong to.

The option to use rule-based categories must be selected as the categorization type.

For examples of how to specify rules for associating documents with categories, click **Help** while you are creating or editing a category.


### **About this task**

If your search applications enable support for categories, users can search a subset of the collection by specifying the category name. Users can also select a category in the search results and browse only the documents that belong to the selected category.

New categories, and changes to existing categories, do not become effective until the index is reorganized.

### **Procedure**

To configure a category:

1. Click **Collections** to open the Collections view.
2. In the list of collections, locate the collection that you want to configure a category for and click  **Edit**.
3. On the Parse page, click **Configure the category tree**.
4. On the Configure the category tree page, select the node in the tree where you want to add the category and click **Create a category**.

If you select the root node, the new category is created at the root level. If you select an existing category name, the new category is nested below the selected category in the category tree.

A wizard opens to help you specify rules for associating documents with the new category:

- a. On the Create a category page, type a descriptive name for the category, then click **Next**.
- b. On the Create category rules page, click **Add Rule**.

- c. On the Create a category rule page, type a unique name for the rule in the **Rule name** field. This name must be unique across all categories in the collection.
- d. Specify the rule that you want to use for associating documents with this category, then click **OK**.
  - If you want enterprise search to use the URI of a document when determining whether the document belongs to the category, click **URI pattern** and then specify the URI pattern.

If the text that you specify exists in the URI, the document is associated with the category.

For example: `cm://company.com/public`
  - If you want enterprise search to analyze words in a document when determining whether the document belongs to the category, click **Document content**, select the language of the documents, and then specify the words that must or must not appear in the document content. You express the rule in the same format as a query.

If a document includes or excludes the words that you specify, the document is associated with the category.

For example: `+finance -accounting`
- e. Click **Finish**.

Your new category is listed on the Configure the category tree page with the other categories that belong to this collection.

#### **Related reference**

Chapter 17, “URI formats in enterprise search,” on page 139

The uniform resource identifier (URI) of each document in an enterprise search collection indicates the type of crawler that added the document to the collection.

---

## **Working with scopes**

Configure a scope when you want to present users with a logical view of a collection.

A scope is a group of related URLs in an index. When you configure a scope, you limit the documents that users can see in the collection. When users search the collection, they search only the documents in the scope.

To use this feature, support for searching scopes must be included in your search applications.

You use the enterprise search administration console to configure scopes for a collection.

## **Scopes**

Scopes enable you to restrict the documents that users can search when they search a collection.

When you create scopes, you specify a set of URIs that are in the index for a collection. Limiting the range of documents that users can search helps ensure that documents in the search results are specific to the information users seek.

For example, you might create one scope that includes the URIs for your Technical Support department and another scope that includes the URIs for your Human

Resources department. If your search application supports scopes, users in the Technical Support department will retrieve documents from the Technical Support scope, and users in the Human Resources department will retrieve documents from the Human Resources scope.

You can create as many scopes as you want, although creating too many scopes can affect performance. Configure scopes so that most search requests need to filter only on one or two scopes. Because scopes can contain entire URIs or URI patterns, the same document can belong to more than one scope.

When you create, edit, or delete a scope, the change becomes effective the next time the index is reorganized.

## Configuring scopes

When you configure a scope for an enterprise search collection, you specify the URIs, or URI patterns, for a range of documents in the index that users are allowed to search.

### Prerequisites

To configure scopes, you must be a member of the enterprise search administrator role or be a collection administrator for the collection that the scopes belong to.


### About this task

If your search applications enable support for scopes, users can search only the documents that match the URIs that define the boundaries of the scope when they search the collection.

New scopes and changes to existing scopes do not become effective until the index is reorganized.

### Procedure

To configure a scope:

1. Click **Collections** to open the Collections view.
2. In the list of collections, locate the collection that you want to configure a scope for and click  **Edit**.
3. On the Parse page, click **Configure scopes**.
4. On the Scopes page, click **Create Scope**.
5. Specify a name for the scope and the URIs and URI patterns that define the boundaries of the scope. You can also specify URIs and URI patterns that you want to exclude from the scope.
6. Click **OK**.

Your new scope is listed on the Scopes page with the other scopes that belong to this collection.

### Related reference

Chapter 17, “URI formats in enterprise search,” on page 139

The uniform resource identifier (URI) of each document in an enterprise search collection indicates the type of crawler that added the document to the collection.

---

## Working with XML search fields

Map XML elements to search fields if you want to enable users to search specific parts of XML documents.

You use the enterprise search administration console to map XML elements to search fields.

### XML search fields

XML search fields enable users to query specific parts of XML documents.

XML documents are becoming more common because they contain both semi-structured and unstructured text. The structure of XML is encapsulated and uses a context that is explicitly defined by XML elements that surround the text. For example, an author's name might appear as follows:

```
<author>John Smith</author>
```

In this context, the text John Smith identifies the author of an XML document.

Enterprise search can associate, or map, the text inside XML elements with search field names. When you configure parsing options for a collection, you specify which XML elements are to be mapped to which search field names. By mapping XML elements to search fields, you enable users to search those elements by specifying the mapped field names in queries. Queries that search specific fields can provide more precise search results than free-text queries that search all document content.

For example, if your collection includes XML documents, and you specify that the title and author elements are to be marked as search fields in the index, users can query these specific elements. A search for `author:Smith` finds XML documents that have Smith in the author element.

#### Related concepts

“Enterprise search parsers” on page 4

An enterprise search parser analyzes documents that were collected by a crawler and prepares them for indexing.

#### Related tasks

“Monitoring the parser” on page 105

Monitor the parser for a collection when you need to start the parser or stop it.

### Mapping XML elements to search fields

When you map an XML element to a search field, you specify which XML elements users can search by specifying a field name in a query.

#### Prerequisites

To map XML elements to search fields, you must be a member of the enterprise search administrator role or a be collection administrator for the collection that the XML documents belong to.

#### About this task

When you create an XML field mapping, or add, change, or delete fields in an existing XML field mapping, the change becomes effective after you stop and


restart the parser. The new and changed mappings apply to data that is parsed after you restart the parser. The new and changed mappings have no effect on data that is already parsed and indexed.

This task uses the following sample XML document to show how you might map personnel records and enable users to directly query certain elements.

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<personnel>
  <personnelrecord>
    <phone>5555</phone>
    <email>joe@us.ibm.com</email>
    <jobroles>Manager, architect
      <jobrole>Managing Search Development Group</jobrole>
      <jobrole>Architecting Search Technology</jobrole>
    </jobroles>
    <location>New York</location>
    <section id="expertise">
      <text>Linguistics</text>
    </section>
  </personnelrecord>
</personnel>
```

### Procedure

To map XML elements in this example to search fields:

1. Click **Collections** to open the Collections view.
2. In the list of collections, locate the collection that you want to map XML elements for and click  **Edit**.
3. On the Parse page, click **Map XML elements to fields**.
4. On the XML field mappings page, click **Create XML mapping**. The Create an XML field mapping page opens.
5. In the **XML root element name** field, type the root element name: `personnel`. Ensure that the name that you specify here exactly matches the root element in the XML documents that you want to search. When parsing and indexing XML documents, enterprise search selects which mapping to use according to the root element name.
6. In the **XML mapping name** field, type a name for this set of XML field mapping rules.  
After you create a set of XML mapping rules, this name is displayed on the XML field mappings page, and you select this name to add, delete, or change the mapping rules.
7. Map the XML element `jobrole` to a search field named `jobrole`:
  - a. In the **Field name** field, type `jobrole`.
  - b. In the **XML element name** field, type `jobrole`.
  - c. To enable users to query the `jobrole` field and view the job roles in search results, select the **Search by field name** and **Show in search results** check boxes.
8. Map the XML element `jobroles` to the same search field:
  - a. Click **Add field** to add a blank line to the list of field mapping rules.
  - b. In the **Field name** field, type `jobroles`.
  - c. In the **XML element name** field, type `jobrole`.

**Tip:** The XML element names do not need to match the search field names, and multiple XML elements can map to the same search field.

- d. To enable users to query the `jobrole` field and view the job roles in search results, select the **Search by field name** and **Show in search results** check boxes.
9. Map the XML element section with the attribute `expertise` to a search field named `expertise`:
  - a. Click **Add field** to add a blank line to the list of field mapping rules.
  - b. In the **Field name** field, type `expertise`.
  - c. In the **Field name** field, type `section`.
  - d. In the **XML attribute name** field, type `id`.
  - e. In the **XML attribute value** field, type `expertise`.
  - f. To enable users to query the `expertise` field and view the expertise values in search results, select the **Search by field name** and **Show in search results** check boxes.
10. Click **OK**.

**Examples:**

To find everyone in an organization who work on search products, specify the following query:

```
jobrole:search
```

To find everyone in an organization who has expertise in linguistics, specify the following query:

```
expertise:linguistics
```



---

## Chapter 9. How enterprise search returns search results

When a user enters a query in a search application, the search processes return the most relevant results for the terms and conditions of the query.

The search servers for enterprise search use several techniques to produce the most relevant search results:

- Text-based scoring
- Static ranking results
- Dynamically summarizing document content
- Collapsing results from the same Web site

### **Related concepts**

Chapter 8, “How enterprise search searches collections,” on page 71

Search applications for enterprise search enable users to specify simple, free-text queries and more complex queries that provide much finer document matching criteria.

---

### **Text-based scoring**

Enterprise search dynamically calculates a score for each document that matches the terms in a query.

To calculate the text score of each document that matches a query, enterprise search considers many factors, such as:

- The frequency of each query term in the entire collection. In general, query terms that appear in most documents contribute less to a document’s score than query terms that appear in a more selective set of documents.
- The number of appearances of each query term in the matching document. In general, the more occurrences of query terms within a document, the higher its score is.
- The proximity with which query terms appear in each matching document. In general, query terms that appear in close proximity to each other in a document contribute more to that document’s score than the same terms with more distant occurrences.
- The context in which query terms appear in each matching document. For example, a query term that appears in the title of a document contributes more to the document’s score than the same term that appears in the plain text of the document.

The length of each document and the richness of its vocabulary are also factors in determining its score.

---

### **Static ranking**

For certain types of documents, you can associate a static ranking factor that increases the importance of those documents in the search results.

When you create a collection, you specify whether you want to associate a static ranking factor with the documents in the collection. For Web content, the number

of links to a document from other documents, and the origins of those links, can increase the relevance of that document in the search results.

For documents that include date fields or date metadata, you can use the date of the document to increase its relevance. For example, recent articles in NNTP news groups might be more relevant than older articles. If a data source includes multiple date values, you can choose which one is most important for determining the relevance of documents in the data source.

If you use static ranking with a collection, ensure that you do not mix data sources that use different ranking types in the same collection. For example, if you want to use the links to a document as the static ranking factor, ensure that the collection contains only Web documents. Search quality can be degraded when sources with different ranking models are combined in the same collection.

You must also ensure that the documents contain fields and values that enable static ranking to be applied. For example, if you specify that you want to use the document date as a factor for ranking documents in the collection, and the documents do not contain date fields or attributes, the search quality might be degraded.

---

## Dynamic summarization

Dynamic summarization is a technique that determines which phrases of a result document best represent the concepts that the user is searching for.

For enterprise search, dynamic summarization tries to capture sentences in documents that contain a large variety of the search terms. A few sentences, or parts of sentences, are selected and displayed in the search results. The search terms are highlighted through HTML rendering of the search results.

---

## Collapsing results from the same Web site

You can specify options for grouping result documents from the same Web site in the search results. To specify options for collapsing results from specific Web sites, you edit the SiteDefs.txt file.

### About this task

Enterprise search can organize the search results so that individual results from the same Web site are grouped together. When results are collapsed, the top result from the Web site typically appears flush left. One or more lower ranking results are grouped and indented below the top result.

In the sample search application for enterprise search, the top two search result documents from each Web site are displayed. If more than two result documents are returned from the same Web site, you can specify that you want to see the collapsed results.

You use the site definition file, SiteDefs.txt, to specify how you want the results from Web sites to be grouped and collapsed. Each collection has its own SiteDefs.txt file, which enables you to specify search result options for individual collections.

For most Web sites, the host name in a URL identifies the site. For example, the following URLs have the same main site, <http://www.mysite.com>:

http://www.mysite.com/hr/us  
http://www.mysite.com/news/index.html  
http://www.mysite.com/persona/users

You might want to treat all the URLs that start with `http://www.mysite.com/hr/` as belonging to the same site, group the results from those pages in the search results, and collapse the display of those results. To do this, you add the URL prefix, `http://www.mysite.com/hr/`, to the `SiteDefs.txt` file.

The changes that you make to the `SiteDefs.txt` do not take effect until the next time that the index is reorganized.

### Procedure

To specify options for collapsing and grouping search results from specific Web sites:

1. Log in as the enterprise search administrator (this user ID was specified during the installation of DB2 II OmniFind):
2. Change to the following directory on the index server, where *collectionID* is the ID that enterprise search assigned to your collection when you created it:

UNIX: `cd $ES_NODE_ROOT/master_config/collectionID.indexer`  
Windows: `cd %ES_NODE_ROOT%\master_config\collectionID.indexer`

3. Use a text editor to edit the `SiteDefs.txt` file.
4. Add the URL prefixes for the Web sites that you want to collapse. Specify one URL per line.

For example:

```
http://www.mysite.com/hr/  
http://www.mysite.com/news/  
http://www.mysite.com/persona/  
http://www.mysite.com/help/  
http://www.mysite.com/global/  
http://www.mysite.com/services/
```

5. Enter the following command to notify the system of the changes:

```
esadmin configmanager syncComponent -sid collectionID.indexer
```



---

## Chapter 10. Starting and stopping enterprise search

After you create a collection, you must start the components for crawling, parsing, indexing, and searching data. Stop and restart the components when you need to notify the system about changes in the collection.

Most enterprise search components can run continuously or in accordance with schedules that you specify. For example, you can specify schedules for reorganizing or refreshing the index. After you start the components for parsing data and searching the index, you typically do not need to stop them except for troubleshooting.

If you make changes to the content of a collection, or if you change the rules for how the crawlers are to collect data from the sources in your enterprise, you typically need to stop and restart the crawlers for the changes to become effective. If you do not change the crawling rules, the crawlers either run continuously (in the case of Web and NNTP crawlers) or according to schedules that you specify.

---

### Starting the enterprise search components

To enable users to search a collection, you must start the components that crawl, parse, index, and search the collection.

#### Prerequisites



Configure the data sources that you want to crawl and specify options for how you want that data to be parsed, indexed, and searched. For example, if you want users to be able to view category details in the search results, configure categories before you start the parser.

To start the enterprise search components, you must be a member of the enterprise search administrator role, a collection administrator for the collection, or an operator with authority to administer that collection.


You must start the enterprise search components for a collection in the correct order. For example, you must start a crawler and crawl data before you can parse and index documents.

#### Procedure



To start the enterprise search components:

1. Start the enterprise search administration console and log in as the enterprise search administrator. If you use administrative roles, you can log in as a collection administrator or operator who has authority for the collection that you want to start.
2. On the Collections view, locate the collection that you want to administer and click  **Monitor**.
3. On the Crawl page, for each crawler that you want to start, click  **Start**.
  - If you start a Web or NNTP crawler, the crawler begins crawling data immediately. These types of crawlers run continuously to crawl and recrawl documents on Web sites and NNTP news groups.


- If you start one of the other crawler types, the crawler begins crawling at its scheduled date and time. If you did not schedule the crawler, or if you want to start the crawler sooner:

- Click  **Details**.
- In the crawl space details area of the Details page for that crawler, click the **Start** icon (or the **Force start** icon) for each data source that you want to crawl (such as a server, database, or subfolder).

After the crawler starts, you can let it run continuously. If you scheduled the crawler, the crawler will run again at the scheduled dates and times.

4. After data is crawled, open the Parse page and click  **Start** to start the parser. You can let the parser run continuously. You typically do not need to stop the parser unless you make changes to categories or XML field mappings.
5. Optional: To force the indexing processes to start, instead of waiting for indexing to begin at the scheduled date and time, open the Index page and, in the **Reorganization** area, click  **Start**.

You can let the indexing processes run continuously. The index will be refreshed and reorganized at the scheduled dates and times.

6. On the Search page, click  **Start**.

You can let the search servers run continuously. You typically do not need to stop the search servers unless you make changes to their configuration or for troubleshooting purposes.

#### **Related concepts**

Chapter 2, “Enterprise search administration overview,” on page 13

You use the enterprise search administration console to create and administer collections, start and stop components, monitor system activity and log files, configure administrative users, and associate search applications with collections.

“Administrative roles” on page 112

Enterprise search uses the concept of roles to control access to various functions in the administration console.

#### **Related tasks**

“Logging in to the administration console” on page 15

To administer an enterprise search system, you specify a URL in a Web browser and then log in to the administration console.

---

## **Stopping the enterprise search components**

You might need to stop and restart an enterprise search component if you make changes to its configuration or if you need to troubleshoot problems.

### **Prerequisites**







To stop the enterprise search components, you must be a member of the enterprise search administrator role, a collection administrator for the collection, or an operator with authority to administer that collection.

### **About this task**

You can stop the enterprise search components independently of each other. For example, if you stop and restart a crawler to incorporate changes that you made to its configuration, you do not need to stop and restart the parser.

## Procedure

To stop the enterprise search components:

1. Click **Collections** to open the Collections view.
2. In the list of collections, locate the collection that you want to administer and click  **Monitor**.
3. On the Crawl page, locate the crawler that you want to administer, and stop or pause it.  
If you change the crawl space, or if you change the crawler properties, stop and restart the crawler to incorporate the changes.
4. On the Parse page, click  **Stop** to stop the parser.  
If you change categories or XML field mappings, stop and restart the parser to incorporate the changes.
5. On Index page, click  **Stop** to stop an index that is currently being built (refresh or reorganization).  
You can also stop an index build while you are monitoring the index queue. To do this, select **System** on the toolbar, click  **Monitor**, open the Index page, and then click  **Stop** for the index that you want to stop building.
6. On the Search page, click  **Stop** to stop the search processes. Typically, you do not need to stop the search processes except for troubleshooting.

### Related concepts

Chapter 2, “Enterprise search administration overview,” on page 13  
You use the enterprise search administration console to create and administer collections, start and stop components, monitor system activity and log files, configure administrative users, and associate search applications with collections.





---

## Chapter 11. Monitoring enterprise search activity

When you monitor system and collection activities, you can view the status of various processes, watch for potential problems, or adjust configuration settings to enhance performance.

With the enterprise search administration console, you can monitor the system and adjust operations as needed. You can view detailed statistics for each major activity (crawling, parsing, indexing, and searching). The statistics include average response times and progress information, such as how many documents were crawled or indexed during a crawl or index-building session.

By clicking icons, you can stop and start most activities. These operations enable you to pause an activity, make changes to its configuration or troubleshoot a problem, and restart processing when you are ready to allow the activity to proceed.

### Related tasks

“Starting the enterprise search components” on page 89

To enable users to search a collection, you must start the components that crawl, parse, index, and search the collection.

“Stopping the enterprise search components” on page 90

You might need to stop and restart an enterprise search component if you make changes to its configuration or if you need to troubleshoot problems.

---

## Alerts

You can configure enterprise search to write messages to the log file whenever it detects that certain events occurred.

Messages that are triggered by events, called alerts, inform you about conditions that you might want to address, such as a resource that is running out of free space. When you configure alerts for enterprise search, you specify the conditions that you want the system to monitor. Whenever the condition occurs, the system automatically writes a message to the log file.

If you want to be notified directly about a condition, you can specify options to receive e-mail whenever one of the monitored messages is logged.

You can configure alerts for collection-level events and for events that occur at the system level. At the collection level, the system can:

- Monitor the number of documents that each crawler crawls, and issue an alert message when the maximum number of documents allowed is about to be reached.
- Monitor the number of documents being added to the index for your collections, and issue an alert message when the maximum number of documents allowed is about to be reached.
- Inform you when the time that is required to respond to search requests is exceeding a limit that you specify.

At the system level, the system can monitor the disk space on each enterprise search server and issue an alert message when the amount of free space is low.

## Configuring collection-level alerts


By configuring alerts, you can ensure that messages are written to the log file whenever certain collection-level events occur. You can also receive e-mail whenever messages about these events are logged.

### Prerequisites

To configure alerts for a collection, you must be a member of the enterprise search administrator role or be a collection administrator for the collection.

### Procedure

To configure collection-level alerts:

1. Click **Collections** to open the Collections view.
2. In the list of collections, locate the collection that you want to configure alerts for and click  **Edit**.
3. On the General page, click **Configure alerts**.
4. If you want the system to monitor the number of documents that each crawler is crawling, take the following steps:
  - a. Select the **When the number of documents crawled by any crawler reaches a percentage of the maximum allowed** check box.
  - b. In the **Percentage** field, specify when you want a message to be logged. Specify this number as a percentage of the maximum number of documents that the crawlers can crawl (you specify the **Maximum number of documents** when you configure the crawler properties). The default value is 90 percent.
5. If you want the system to monitor the number of documents that are being added to the index, take the following steps:
  - a. Select the **When the number of documents in the index exceeds a limit** check box.
  - b. In the **Limit** field, type the number of documents that the index can contain. When this number is reached, the system writes a log message about the event.
6. If you want the system to inform you when the time required to respond to search requests is exceeding a limit, take the following steps:
  - a. Select the **When the search response time exceeds a limit** check box.
  - b. In the **Limit** field, type the number of seconds that you consider acceptable as a maximum search response time.

When this number is exceeded, the system writes a log message about the event. For example, if you keep the default value, then the system creates a log message whenever a search server averages five seconds or longer to respond to search requests.

Typical response times are less than a half a second. Averages greater than one second might indicate that your operating system needs tuning for better performance or that a problem exists in the search server

configuration settings. For example, you might want to increase the amount of space that you allocate for the search cache.

7. Click **OK**.

If you want to receive e-mail when the system logs messages about these events, open the Log page, then click **Configure e-mail options for messages** so that you can specify an e-mail address.

Before you can receive e-mail, you must also ensure that information about your mail server is configured. To do this, an enterprise search administrator must select **System** on the toolbar, open the Log page, then click **Configure e-mail options for messages**.

#### **Related tasks**

“Receiving e-mail about logged messages” on page 134

You can specify options to receive e-mail whenever certain messages, or certain types of messages, are logged.

## **Configuring system-level alerts**

By configuring alerts, you can ensure that messages are written to the log file whenever certain system-level events occur. You can also receive e-mail whenever messages about these events are logged.

#### **Prerequisites**

To configure system-level alerts, you must be an enterprise search administrator.

#### **Procedure**

To configure system-level alerts:

1. Click **System** to open the System view.
2. On the General page, click **Configure alerts**.
3. If you want the system to monitor the amount of free space that is available on the enterprise search servers, select the **When the amount of available file system space reaches a percentage of the total space** check box.
4. In the **Percentage** field, specify when you want the system to notify you that the amount of free space on a server is low. Specify this number as a percentage of total file system space. The default value is 80 percent.  
If your enterprise search system is set up on multiple servers, the system creates a separate log message for each server. For example, a message informs you when the space on the Crawler server is low; and separate messages inform you about space constraints on the Index and search servers.
5. Click **OK**.

If you want to receive e-mail whenever the system logs a message about this event, open the Log page, then click **Configure e-mail options for messages** so that you can specify an e-mail address and information about your mail server.

#### **Related tasks**

“Receiving e-mail about logged messages” on page 134

You can specify options to receive e-mail whenever certain messages, or certain types of messages, are logged.

---

## Monitoring a collection


You can view general information about the status of each component in a collection or select options to view detailed information about individual components and URIs.


### Prerequisites


All enterprise search administrative users can monitor collections. To start or stop components, or to enable or disable schedules, you must be a member of the enterprise search administrator role, a collection administrator for the collection, or an operator for the collection.

### Procedure

To monitor a collection:

1. Click **Collections** to open the Collections view.
2. In the list of collections, locate the collection that you want to monitor to and click  **Monitor**. Information about the current status of each collection component is displayed.

**Tip:** If you are editing a collection and are already on the General page, you can click  **Monitor** to change to the view for monitoring the collection.

3. To see detailed information about a specific URI, click  **URI details**.

For example, you might want to see whether a specific URI is in the index, or whether the index that the URI is in was copied to the search servers.

4. To monitor an individual component and see detailed statistics about that component's activity, click the **Status** icon.

### Related concepts

Chapter 3, "Enterprise search collections," on page 19

An enterprise search collection contains the entire set of sources that users can search with a single query.

---

## Viewing details about a URI

You can view detailed information about a URI. You can see current and historical information about how the document that is represented by this URI is crawled, indexed, and searched.

### Prerequisites

Before you submit a request to view a URI report or send a report to an e-mail address, ensure that the component that you want to receive information from is active. For example, to view details about how a document is crawled, indexed, and searched, ensure that the Web crawler, index server, and search servers are running.


### About this task


Collecting information about a URI is a time-consuming process. You can choose an option to view the information that you request, then wait for it to be displayed. A more efficient option is to send the report to an e-mail address that you specify.


The index server and search servers can provide information about all URIs (such as whether a URI is in the index and whether it has been copied to the search servers). To view information about how a document was crawled, you must specify the URI for a document that was crawled by a Web crawler.

### Procedure

To view details about a URI:

1. Click **Collections** to open the Collections view.
2. In the list of collections, locate the collection that you want to monitor to and click  **Monitor**.

**Tip:** If you are editing a collection and are already on the General page, you can click  **Monitor** to change to the view for monitoring the collection.

3. Click  **URI details**.
4. On the URI details page, type the URI that you want to view information for.
5. Select the check boxes for the type of information that you want to see:

#### Crawler details

Select this check box to see information about how a document was crawled by a Web crawler, and information about its current status in the crawl space.

#### Index details

Select this check box to see whether a document was indexed and copied to the search servers.

#### Search details

Select this check box to see information about how the document can be searched and whether the document is available for searching.

6. To wait for the report to be displayed, click **View report**.
7. To send the report to an e-mail address so that you can view at a later time, click **Send report**.
  - a. On the Send a detailed URI report page, type an e-mail address for receiving the report.
  - b. Click **Send Report**.

#### Related reference

Chapter 17, “URI formats in enterprise search,” on page 139

The uniform resource identifier (URI) of each document in an enterprise search collection indicates the type of crawler that added the document to the collection.

---

## Monitoring crawlers


You can view general information about the status of each crawler in a collection or select options to view detailed information about a crawler activity.


### Prerequisites


If your administrative role limits you to monitoring collections, you can view crawler statistics but you cannot change a crawler’s behavior (such as starting or stopping the crawler).

### Procedure

To monitor a crawler:

1. Click **Collections** to open the Collections view.
2. In the list of collections, locate the collection that you want to monitor and click  **Monitor**.
3. Open the Crawl page.

**Tip:** If you are editing a collection and are already on the Crawl page, you can click  **Monitor** to change to the view for monitoring crawlers.

4. If the crawler is running or paused and you want to see detailed status information about the crawler, click  **Details**. The types of statistics that you see vary with the crawler type.

If your administrative role allows you to administer processes for a collection, you can start, stop, and pause the crawler while you view details about crawler activity. If the crawler can be scheduled, you can also enable and disable the crawling schedule.

5. If the crawler is stopped or paused and you want to start a crawler session, click  **Start**.



**For Web and NNTP crawlers:**

If the crawler was stopped, the crawler begins crawling again and crawls the entire crawl space. If the crawler was paused, it resumes crawling at the beginning of the target where it was paused.

**For all other crawler types:**

If the crawler was stopped, the crawler begins crawling at its scheduled date and time. If you did not schedule the crawler, or if you want to start the crawler sooner, click the **Details** icon. Then, in the crawl space details area, click the **Start** icon (or the **Force start** icon) for each data source that you want to crawl (such as a server, database, or subfolder).

If the crawler was paused, the crawler resumes crawling at the beginning of the target where it was paused. For example, the DB2 crawler resumes crawling at the first row in the table that was being crawled when you paused the crawler.

6. If the crawler is running and you want to stop it, click  **Stop** or  **Pause**. The crawler stops crawling data until you restart it.

**Related concepts**

Chapter 4, “Enterprise search crawler administration,” on page 23

You configure crawlers for the different types of data that you want to include in a collection. A single collection can contain any number of crawlers.

**Related reference**

Chapter 17, “URI formats in enterprise search,” on page 139

The uniform resource identifier (URI) of each document in an enterprise search collection indicates the type of crawler that added the document to the collection.

---

## Viewing details about Web crawler activity


By viewing details about Web crawler activity, you can assess overall performance and adjust the Web crawler properties and crawl space definitions as necessary.


**Prerequisites**


All enterprise search administrative users can monitor crawler activities. To start or stop a crawler, you must be a member of the enterprise search administrator role, a collection administrator for the collection, or an operator for the collection.

## Procedure

To view details about a Web crawler's activity:

1. Click **Collections** to open the Collections view.
2. In the list of collections, locate the collection that owns the Web crawler that you want to monitor to and click  **Monitor**.
3. Open the Crawl page.

**Tip:** If you are editing a collection and are already on the Crawl page, you can click  **Monitor** to change to the view for monitoring crawlers.

4. If the Web crawler that you want to monitor is running or paused, click  **Details**.
5. On the details page for the Web crawler, view or select the following options to see detailed statistics about the crawler's current and past activity.
  - a. Click **Thread details** to see how many threads are actively crawling Web sites and how many are in an inactive state.
  - b. Click **Active sites** to see information about the Web sites that the crawler is currently crawling.
  - c. Click **Most recently crawled URLs**. This information shows what the crawler is currently crawling. If the items in the list do not change as you refresh the view, then no crawling is occurring.
  - d. Click **Crawler history** to view reports about past crawler activity.
  - e. In the **URL status** area, type the URL for a Web site that you want to see information about, then click **View**.

For example, use this option to see whether a URL is in the crawl space, whether it has been crawled or only discovered, when it should be crawled again, and information about the last attempt to crawl the Web site.

After details about the URL are displayed, you can click **Site history** to see additional statistical information about crawler activity at that URL.

## Web crawler thread details

You can monitor the Web crawler to see how many threads are actively crawling Web sites and how many are in an inactive state.

When you view details about a Web crawler while monitoring a collection, you can view the status of the crawler threads. The states that you are most likely to see include:

### Waiting

Indicates that the thread does not have a URL to crawl. This condition can occur when a thread finishes a crawl and the crawler cannot find more URLs to crawl fast enough. For example, if the crawler property that controls how long the crawler must wait before it can retrieve another page from same site is too high, it can prevent URLs from being supplied fast enough.

### Fetching

Indicates that the thread is downloading a page from a Web site.

### Completed

Indicates that the thread is sending the pages that it crawled to the rest of the crawler, but is not yet ready to crawl another URL.

### Suspended

Indicates that the crawler is paused

Ideally, all threads are fetching pages all the time. If threads are often in a completed state, then the database might be having throughput problems.

If threads are often in a waiting state, review the value specified for the **Maximum number of active hosts** field in the crawler properties. If the value is low, there might not be enough sites in the crawl space to keep the threads busy, or there might not be enough URLs eligible to be crawled. Conditions that can cause low activity include DNS lookup failures and robot lookup failures.

## Web crawler active sites

You can monitor the Web crawler to see information about the Web sites that the crawler is actively crawling.

When you view details about a Web crawler while monitoring a collection, you can view statistics about active sites. The statistics show:

- How many URLs the crawler brought from its internal database to memory for crawling at this time
- How many URLs the crawler has attempted to crawl so far
- How much time remains before a site is deactivated and removed from memory for this iteration of the crawler
- How much time a site has been in memory so far

This information changes from moment to moment as the crawler progresses through the crawling rules that are configured for it. Ideally, the number of activated URLs is close to the value that is configured for the **Maximum number of active hosts** field in the crawler properties.

If the number of activated URLs is near zero, then the crawler is not finding eligible URLs. Conditions that can cause such low activity include DNS lookup failures, network connectivity issues, database errors, and crawl space definition problems. For example:

- If many sites have been in memory for a long time, and few URLs have been crawled, look for network connectivity problems.
- If not enough sites are in the list, look for crawl space definition problems or DNS lookup problems.
- If sites are being crawled at a reasonable rate, but are leaving memory with many URLs not being crawled, edit the crawler properties and adjust the timeout value in the **Maximum time that URLs can remain in memory** field to keep the sites in memory longer.

## Web crawler crawl rate

You can monitor the Web crawler to see information about how fast the crawler is downloading pages from Web sites.



When you view details about a Web crawler while monitoring a collection, you can view statistics about how fast the crawler is crawling data (the crawl rate). You can also view statistics about how many URLs the crawler crawled since the current session began.

The crawl rate is the number of pages that are being crawled per second. This number correlates to several properties that you can configure for the Web crawler:

- The number of crawler threads
- The number of active sites
- The amount of time that the crawler must wait before it can retrieve another page from the same Web server

If the crawler has one active site per crawler thread, and the crawler must wait two seconds before it can retrieve another page from the same Web server, then the crawler cannot crawl faster than one page per thread per two seconds. For example, if the crawler uses the default number of threads (200), then crawler can crawl 100 pages per second for 200 threads.

If there are twice as many active sites as crawler threads, and the crawler must wait two seconds before it can retrieve another page from the same Web server, then the crawler could reach one page per thread per second. However, network download speeds and database throughput would then become limiting factors. An indication of strong crawler performance is when the crawl rate aligns with the number of crawler threads, active sites, and crawler wait time.

Another factor to review when you monitor Web crawler performance is the number of URLs that the crawler crawled since the start of the current crawler session. Divide that number by the total amount of the time that the crawler has been running to calculate an average of the long-term throughput. If this number is not increasing, the crawler is either finished, or it is unable to proceed. For example, network connectivity errors, database errors, and DNS lookup failures can block the progress of the crawler.

## Creating Web crawler reports

By viewing reports about past Web crawler activity, you can assess overall performance and adjust the Web crawler properties and crawl space definitions as necessary.

### Prerequisites

If your administrative role limits you to monitoring collections, you can view crawler statistics and create reports about crawler activity, but you cannot change the crawler's behavior (such as starting or stopping the crawler).


### About this task


Different types of reports can provide you with information about Web crawler activity. For certain types of reports, information is returned as fast as it can be collected from the crawler's internal database. The Site report and HTTP return code reports take time to create. If you create these types of reports, you can specify an e-mail address for receiving the report instead of waiting for results to be returned to the enterprise search administration console.


For information about how to interpret statistics in the reports, click **Help** while you are monitoring the Web crawler and creating the reports.

## Procedure

To create Web crawler reports:

1. Click **Collections** to open the Collections view.
2. In the list of collections, locate the collection that owns the Web crawler that you want to monitor to and click  **Monitor**.
3. Open the Crawl page.

**Tip:** If you are editing a collection and are already on the Crawl page, you can click  **Monitor** to change to the view for monitoring crawlers.

4. If the Web crawler that you want to create reports for is running or paused, click  **Details**.
5. On the details page for the Web crawler, select an option for the type of report that you want to create:

- In the **Crawler status summary** area, click **Crawler history** to create reports about the crawler and all the sites that it discovers or crawls.
- In the **URL status** area, specify the URL of specific site that you want to create a report for, click **View**, and then click **Site history**.

6. For both crawler history and site history reports, select the check box of each statistic that you want to see in a report, then click **View report**.

For these types of statistics, the crawler returns a report to the administration console as fast as it can retrieve information from its internal database.

7. If you are creating a crawler history report, specify options for creating a Site report, then click **Run report**.

This report is created with the statistics that you choose to include and saved in a file that you specify (the file name must be absolute). You can specify that you want to receive e-mail after the report is created.

8. If you are creating a crawler history report, specify options for creating an HTTP return code report, then click **Run report**.

This report provides information about the number of HTTP return codes distribution per site. The report is saved in a file that you specify (the file name must be absolute). You can specify that you want to receive e-mail after the report is created.

Use this report to see which sites return a large number of 4xx return codes (which indicate that pages were not found), 5xx return codes (which indicate a server problem), 6xx return codes (which indicate connectivity problems), and so on.

This report is most useful when the crawler has been active for some time (for example, a crawler that has been active for weeks). It can help you identify vanished sites, newly arrived sites, sites with huge numbers of URLs (which might indicate redundant crawling of a Lotus Notes database), and sites with a recursive file system served by the HTTP server. If the sites with large numbers of HTTP return codes are not contributing to the index, you can improve the performance of the crawler by removing the sites from the crawl space.

## Web crawler HTTP return codes

When you monitor a Web crawler, you can view information about the HTTP return codes that the crawler receives from the pages that it attempts to crawl.

## Purpose

When you monitor the Web crawler history, or monitor the status of a specific URL, you can see information about the HTTP return codes that were returned to the crawler. You can use this information to manage the crawl space and optimize crawler performance. For example, if the crawler receives a large number of HTTP return codes for a URL, and the return codes indicate that pages at that location cannot be crawled, you can improve performance by removing that URL from the crawl space.

The following table lists the HTTP return codes and how the Web crawler interprets them. Values from 100 to 505 are standard HTTP return codes (see <http://www.w3.org/Protocols/rfc2616/rfc2616.html> for more information). The remaining HTTP returns codes are proprietary to enterprise search and the Web crawler.

*Table 1. HTTP return code from the Web crawler*

Code	Description	Code	Description	Code	Description	Code	Description
NULL	Uncrawled	401	Unauthorized	500	Internal server error	700	Parse error (no header end)
100	Continue	402	Payment required	501	Not implemented	710	Parse error (header)
200	Successful	403	Forbidden	502	Bad gateway	720	Parse error (no HTTP code)
201	Created	404	Not found	503	Service unavailable	730	Parse error (body)
202	Accepted	405	Method not allowed	504	Gateway timeout	740	Excluded by robots.txt file
203	Non-authoritative information	406	Not acceptable	505	HTTP version not supported	741	Robots temporarily unavailable
204	No content	407	Proxy authentication required	611	Read error	760	Excluded by crawl space definition
205	Reset content	408	Request timeout	612	Connect error	770	Bad protocol or nonstandard system port
206	Partial content	409	Conflict	613	Read timeout	780	Excluded by file type exclusions
300	Multiple choices	410	Gone	614	SSL handshake failed	2004	No index META tag
301	Moved permanently	411	Length required	615	Other read error	4044	Excluded by robots.txt file
302	Found	412	Precondition failed	680	DNS lookup failure		
303	See other	413	Request entity too large				
304	Not modified	414	Request URI is too long				
305	Use proxy	415	Unsupported media type				

Table 1. HTTP return code from the Web crawler (continued)

Code	Description	Code	Description	Code	Description	Code	Description
306	(Unused)	417	Expectation failed				
307	Temporary redirect						

**Table notes:**

**4xx return codes**

You will rarely see a 400 (bad request) code. According the HTTP return code standard, 4xx codes are supposed to be indicate that the client (the crawler) failed. However, the problem is usually at the server or in the URL that the crawler received as a link. For example, some Web servers do not tolerate URLs that try to navigate up from the site root (such as `http://xyz.ibm.com/../../foo`). Others Web servers have no problem with this upward navigation and ignore the parent directory operator (`..`) when the crawler is already at the root.

Some servers treat a request for the site root as an error, and some obsolete links might request operations that are no longer recognized or implemented. When asked for a page that it no longer serves, the application server throws an exception, which causes the Web server to return the HTTP return code 400 because the request is no longer considered valid.

- 615** Indicates that the crawler component that downloads data from Web sites encountered an unexpected exception. A large number of this type of return code might indicate that there is a problem with the crawler.

**6xx return codes**

Except for 615, the 6xx return codes indicate problems that can be expected in crawling, such as timeouts. The following return codes might require corrective action:

**611, 612, and 613**

Indicate slow sites or poor network performance.

- 614** Indicates that the crawler is unable to crawl secure (HTTPS) sites. If you believe that these sites should be accessible, verify that the certificates are set up correctly on the crawler server and on the target Web server. For example, if a site is certified by a recognized certificate authorities (CAs), you can add new CAs to the trust store that is used by the crawler.

Also look at how self-signed certificates are configured on the sites that you are trying to crawl. The crawler is configured to accept self-signed certificates. Some sites create a self-signed certificate for a root URL (such as `http://foo.ibm.com/`), and then try to use that certificate on subdomains (such as `http://bar.foo.ibm.com/`). The crawler cannot accept certificates that are used in this manner. It accepts self-signed certificates only if the domain name of the subject (`foo.ibm.com`) and the signer of the certificate match the domain name of the page that is being requested.

- 680** Indicates that the crawler was not able to obtain IP addresses for hosts in the crawl space, perhaps because of network access problems. This type of error means that the crawler is not able to

crawl entire sites, not just that it was unable to crawl some URLs. A large number of this type of return code greatly reduces throughput.

### 7xx return codes

The 7xx codes are mostly due to rules in the crawl space:

#### 710 - 730

Indicate that problems prevented the crawler from doing a complete download, or that the crawler encountered invalid HTML data at a site. If you see a large number of these types of return codes, contact your enterprise search support representative for assistance.

#### 740 or 4044

Indicate that the content of a file cannot be indexed because the document is excluded by restrictions in the site's robots.txt file.

**740** Indicates that anchor links that point to the excluded document can be included in the index.

**4044** Indicates that the anchor links in documents that point to the excluded document are also excluded from the index.

**741** Indicates that a site has a robots.txt file that allows the crawl, but the download failed. If it is repeatedly unable to crawl the URL, the URL is removed from the crawl space. If you seen a large number of this type of return code, check to see whether the target site is temporarily or permanently unavailable. If the target site is no longer available, remove it from the crawl space.

The remaining 7xx return codes mostly occur when you make changes to the crawl space after the crawler has been running for awhile. These return codes typically do not indicate problems that you need to address.

---

## Monitoring the parser


Monitor the parser for a collection when you need to start the parser or stop it.


### Prerequisites


If your administrative role limits you to monitoring collections, you can view the status of the parser, but you cannot start the parser or stop it.

### Procedure


To monitor the parser:

1. Click **Collections** to open the Collections view.
2. In the list of collections, locate the collection that you want to monitor and click  **Monitor**.
3. Open the Parse page.

**Tip:** If you are editing a collection and are already on the Parse page, you can click  **Monitor** to change to the view for monitoring the collection.

4. If the parser is stopped and you want to start it, click  **Start**.

When you first create a collection, start the parser only after the crawler begins crawling data. This ensures that the parser has data to analyze and categorize. After you start the parser, you can let it run continuously.

5. If the parser is running and you want to stop it, click  **Stop**.  
For example, stop the parser to incorporate changes that you make to categories or XML field mappings.

---

## Monitoring index activity for a collection


Monitor the index for a collection when you need to see the progress of an index that is being built, enable or disable the index schedule, or start and stop indexing activity.


### Prerequisites





All enterprise search administrative users can monitor index activities. To start or stop an index build, or to enable or disable the index schedule, you must be a member of the enterprise search administrator role, a collection administrator for the collection, or an operator for the collection.

### Procedure

To monitor the index for a collection:

1. Click **Collections** to open the Collections view.
2. In the list of collections, locate the collection that you want to monitor and click  **Monitor**.
3. Open the Index page.

**Tip:** If you are editing a collection and are already on the Index page, you can click  **Monitor** to change to the view for monitoring the collection.

4. If an index is scheduled, and you do not want it to be built at the scheduled date and time, click  **Disable schedule**. The index will not be built until you enable the schedule or start the index building process.
5. If an index is scheduled, but the schedule for building it is currently disabled, click  **Enable schedule**. The index will be queued for building at the date and time that you specified in the index schedule.
6. If an index is stopped and you want to start it, click  **Start**.  
Typically, indexing occurs on a regularly scheduled basis. If you stop an index while it is being built, or if you disable the schedule for an index, you can click **Start** to force the index build to begin.
7. If an index build is active and you want to stop it, click  **Stop**.  
You might need to stop an index build, for example, to force an index reorganization after you change the change the type of categorization used in the collection.

---

## Monitoring the enterprise search index queue

You can view the status of all index builds in the index queue, stop an index that is being built, or delete an index from the queue.

### Prerequisites

To administer the index queue, you must be a member of the enterprise search administrator role.

### About this task

Enterprise search can build multiple indexes at a time, but only one index per collection can be in the queue at a time. When you configure index options for the system, you specify how many indexes can share the queue and the index building resources concurrently.

### Procedure

To monitor the index queue:

1. Click **System** to open the System view.
2. Click **Monitor**, then select the Index page.


A list of the collections that currently have indexes in the index queue is displayed. For each index, you can see the type of index that is being built (refresh or reorganization), the time that the index entered the index queue, and the time that a build of the index began (if a build is in progress).

3. To administer an individual index, click the **Status** icon.

For example, you might want to see how close an index is to being completed, see how many documents are in the index, or disable the index schedule.

4. To stop an index that is currently being built, click  **Stop**.

For example, if you changed category rules, you might want to stop an index refresh so that you can force an index reorganization to start instead.

To start an index build after you stop it, either wait for the index to enter the index queue at its next scheduled start time, or click the **Status** icon to monitor the index, then click  **Start** to refresh or reorganize the index.

5. To delete an index from the index queue, click  **Delete**.

---

## Monitoring the search servers


You can view detailed status information about search server activity for a specific collection, or view detailed status information for the search servers throughout your enterprise search system.


### Prerequisites

All enterprise search administrative users can monitor search servers for the collections that they are authorized to administer. To monitor all the search servers in your enterprise search system, you must be a member of the enterprise search administrator role.

To start or stop a search server, you must be a member of the enterprise search administrator role, a collection administrator for the collection, or an operator for the collection.

### Procedure

1. To monitor the search servers for a single collection:
  - a. Click **Collections** to open the Collections view.
  - b. In the list of collections, locate the collection that you want to monitor and click  **Monitor**.
  - c. Select the Search page.

**Tip:** If you are editing a collection and are already on the Search page, you can click  **Monitor** to change to the view for monitoring the collection.

2. To monitor all the search servers in your enterprise search system:
  - a. Click **System** to open the System view.
  - b. Click **Monitor**, then select the Search page.
3. If a search server is stopped and you want to start it, click  **Start**.
4. If a search server is running and you want to stop it, click  **Stop**.
5. To see a summary of how much time a search server spends processing search requests, click **Response time (history)**.

The report shows, in milliseconds, the average amount of time that the search server spent responding to search requests on a particular date.

The average response time is an indicator for how well the system is performing, and corresponds to quality of service. An increase in response time might indicate that the system is under heavy load. For example, the number of collections being searched and the collection size might be overwhelming the system.

6. To see a list of the most frequently submitted queries, click **Popular queries**.

The report shows you the keywords in the 50 most frequently submitted queries and how many times users submitted a particular query.

By reviewing the most frequent queries, you can identify candidates for quick links. By creating quick links, you can positively impact the search quality for many users. You can ensure that highly relevant documents are always returned in the search results.

You might also want to create links to the resources that answer those queries from the enterprise portal. For example, if users frequently search for information about expense accounts, include a link to the page that discusses expense account procedures on your intranet home page.

7. To see a list of the most recently submitted queries, click **Recent queries**.

The report shows you the keywords in the 50 most recently submitted queries.

By reviewing the most recent queries, you can identify current trends and urgent situations in the organization. For example, you might see a surge of interest being shown for some topic. That surge in interest might indicate that a quick link for that topic is needed or that you need to make that topic available to users in other ways (such as providing a link on the enterprise portal).

#### **Related concepts**

“Search servers for enterprise search” on page 6

The search servers for enterprise search work with your search applications to process queries, search the index, and return search results.

Chapter 6, “Search applications for enterprise search,” on page 59

A search application enables you to search collections in your enterprise search system. You can create any number of search applications, and a single search application can search any number of collections.

“Search cache” on page 109

When the load on the search servers is relatively high, you can enhance performance by caching search results.

“Document-level security” on page 114

Document-level security ensures that users who search collections are able to access only the documents that they are allowed to see.

“Quick links” on page 73

Quick links enable you to provide users with links to specific documents.



## Search cache

When the load on the search servers is relatively high, you can enhance performance by caching search results.

When the search servers process search requests, they first check if results for the same query already exist in the cache. If the search servers find the appropriate result documents, they can quickly return search results to the user. If the search servers do not find the appropriate result documents, they search the index.

When the search cache fills, the oldest result documents and result documents for infrequent queries are cycled out to make room for new search results.

From the enterprise search administration console, you can enable search caching and also specify the capacity of the cache (the number of queries whose results can be cached simultaneously).

## Configuring a search cache


You can enable or disable the search cache for a collection. You can also specify options to control the size of the search cache.

### Prerequisites

To configure a search cache for a collection, you must be a member of the enterprise search administrator role or be a collection administrator for the collection that you are administering.

### Procedure

To configure the search cache:

1. Click **Collections** to open the Collections view.
2. In the list of collections, locate the collection that you want to configure the search cache for and click  **Edit**.
3. On the Search page, select the **Use the search cache** check box.
4. Specify the maximum number of search result sets that the search cache can hold in the **Maximum number of entries in the search cache** field.
5. Click **Apply**.



---

## Chapter 12. Enterprise search security

Security mechanisms in enterprise search enable you to protect sources from unauthorized searching and restrict administrative functions to specific users.

With enterprise search, users can search a wide range of data sources. To ensure that only users who are authorized to access content do so, and to ensure that only authorized users are able to access the administration console, enterprise search coordinates and enforces security at several levels.

### Web server

The first level of security is the Web server. If you enable global security in IBM WebSphere Application Server, you can assign users to administrative roles and authenticate users who attempt to administer the system. When a user logs in to the administration console, only the functions and collections that the user is authorized to administer are available to that user.

Search applications can also use the authentication support in WebSphere Application Server to authenticate users who access the search application or search collections.

### Collection-level security

When you create a collection, you can enable security at the collection level. If you enable collection security, the global analysis processes apply special rules:

- To ensure that the security controls for each document are evaluated, documents with duplicate (or near duplicate) content are indexed independently instead of having their content indexed jointly in a canonical representation.
- The anchor text processing phase of global analysis normally associates text that appears in one document (the source document) with another document (the target document) in which that text does not necessarily appear. This enables the target document to be retrieved by queries that specify text that appears in the source document. This type of anchor text processing presents a security risk if users are allowed to view the target document but not the source document. When collection security is enabled, anchor text processing is disabled. This means that a document is returned in the search results only if its own content or metadata matches query.

There is a trade-off between enabling collection security and search quality. Enabling collection security reduces the information that is indexed for each document. A side effect is that fewer results will be found for some queries.

Collection-level security is also available to your search applications through an application ID. To search collections, an enterprise search administrator must associate your search application with the specific collections that it can search. You can then use standard access control mechanisms to permit or deny users access to specific search applications.

### Document-level security

When you configure crawlers for a collection, you can enable document-level security. If you choose this option, the crawler associates

security tokens with each document that it crawls. The security tokens are stored with the documents in the index.

If you enable global security in WebSphere Application Server, your search applications can use these security tokens to enforce access controls. To ensure that users search and retrieve only the documents that they have permission to access, a search application can include credentials from the logged in user on the queries that it passes to the search servers.

Security for your collections extends beyond the authentication and access control mechanisms that enterprise search can use to protect indexed content. Safeguards also exist to prevent a malicious and unauthorized user from gaining access to data while it is in transit. For example, the search servers use protocols such as the Secure Sockets Layer (SSL), the Secure Shell (SSH), and the Secure Hypertext Transfer Protocol (HTTPS) to communicate with the index server and the search application user interface.

Additional security is provided through encryption. For example, the password for the enterprise search administrator, which is specified during product installation, is stored in an encrypted format.

For increased security, you need to ensure that the server hardware is appropriately isolated and secure from unauthorized intrusion. By installing a firewall, you can protect the enterprise search servers from intrusion through another part of your network. Also ensure that there are no spare open ports on the enterprise search servers. Configure the system so that it listens for requests only on ports that are explicitly assigned to enterprise search activities and applications.

---

## Administrative roles

Enterprise search uses the concept of roles to control access to various functions in the administration console.

During the installation of IBM DB2 Information Integrator OmniFind Edition (DB2 II OmniFind), the installer configures a user ID and password for the enterprise search administrator. The first time that you access the administration console, you must log in as this user. If you do not enable global security in IBM WebSphere Application Server, this user ID is the only user ID that you can use to access the administration console.

If you enable global security in IBM WebSphere Application Server, you can enroll additional users as enterprise search administrative users. By assigning users to roles, you can restrict access to specific collections and control the functions that each administrative user can do. The user IDs that you assign to administrative roles in enterprise search must exist in a WebSphere Application Server user registry.

When an administrative user logs in, enterprise search authenticates the user ID. Only the collections and functions that the user is allowed to administer are available in the console.

You can enroll users in the following administrative roles:

### **Enterprise search administrator**

These users create collections and have the authority to administer all aspects of your enterprise search system. When DB2 II OmniFind is

installed, the installer specifies the user ID and password for the first enterprise search administrative user. This user can assign other users to the enterprise search administrator role.

#### **Collection administrator**

These users can edit, monitor, and control the operation of specific collections or all collections. These users cannot create collections or administer components that span collections.

#### **Operator**

These users can monitor system activity and control the operation of specific collections or all collections. These users can start and stop collection activity, for example, but they cannot create collections, edit collections, or administer components that span collections.

#### **Monitor**

These users can monitor system activity for specific collections or all collections. They cannot control operations (such as starting and stopping components), create collections, edit collections, or administer components that span collections.

#### **Related tasks**

“Logging in to the administration console” on page 15

To administer an enterprise search system, you specify a URL in a Web browser and then log in to the administration console.

---

## **Configuring administrative users**

By configuring administrative roles, you can restrict access to specific collections and control the functions that each administrative user can do.

#### **Prerequisites**

Before you assign a user to an administrative role, ensure that security is enabled in IBM WebSphere Application Server. Also ensure that the user ID exists in a WebSphere Application Server user registry.

To configure administrative users, you must be a member of the enterprise search administrator role.

#### **Procedure**

To assign users to administrative roles:

1. Click **Security** to open the Security view.
2. On the Administrative Roles page, click **Add User**.
3. Type the user ID of the user that you want to enroll and select an appropriate administrative role.
4. If you are not enrolling this user as an enterprise search administrator, select the collections that this user can administer.

You can select the check boxes for individual collections or enable the user to administer all collections.

---

## **Authentication versus access control**

To protect content from unauthorized users, and to control access to administrative functions, enterprise search supports user authentication and access control.

## Authentication

Authentication is the process by which a system verifies that users are who or what they declare themselves to be. Because access is typically based on the identity of the user who requests the resource, authentication is essential to effective security.

To authenticate users who attempt to access the administration console, enterprise search leverages the authentication support that is provided with IBM WebSphere Application Server.

To authenticate users who search enterprise search collections, your search applications can leverage security in WebSphere Application Server and implement your preferred methods for authenticating user credentials. Typically, user credentials consist of a user ID and a password that are passed to the search application when a user logs in or attempts to access the search application.

User authentication can be implemented in other ways, depending on the resources and protocols that are available in your enterprise. For example, you might force users to identify themselves by using smart cards, by managing digital certificates and a public key infrastructure, or by assigning tickets when users log in to track their authentication state.

## Access control

Access control refers to limiting what users can do after they identify themselves and are authenticated. An access control list (ACL) is the most common way in which access to resources is limited. An ACL is a list of user identifications (user names, group names, user roles, and so on). Each user identification is associated with a set of permissions that define the user's rights and privileges.

For example, access controls can allow or deny access to files on a file server and control whether a user who is allowed access can read, create, edit, or delete files on that server.

In enterprise search, all access control depends on whether a user has permission to read data in the index. Depending on how you enable access controls in your search applications and the rules that you specify for the collection and for crawlers when you administer enterprise search, you can:

- Allow all users to search all documents in a collection.
- Allow all users to search all documents that were crawled by certain crawlers, and restrict access to documents that were crawled by other crawlers.
- Allow specific users to search specific documents. For example, when you specify the databases that you want to crawl with a Notes crawler, you can specify options that enable certain users to access certain views and folders, and prevent other users from searching those documents.

---

## Document-level security

Document-level security ensures that users who search collections are able to access only the documents that they are allowed to see.

Document-level access control is achieved by associating one or more security tokens with the documents in the system. By default, each document is considered a public document, which means that it can be searched by all users. When you

configure a crawler, you can specify that you want to use security tokens to limit which users can access the documents that are crawled by that crawler.

If a data source type includes fields, you can specify that you want to use the security data in that field to enforce access controls. If the data source does not have fields, if you do not want to use a specific field as a security field, or if the field that you specify does not include security data that enables access controls to be enforced, you can define security tokens for the crawler to associate with documents.

The administrator for each collection decides the security tokens that the crawler is to associate with documents. For example, a security token might represent a user ID, a group ID, a user role, or any other value that you determine is valid for the data source.

Security tokens accompany documents as the documents pass through the stages of parsing, analysis, and indexing. If your search applications enable global security in IBM WebSphere Application Server, you can use the security tokens to control access to documents. Users who search the collection are able to search only the documents that their credentials permit them to see. If a user's credentials do not pass the rules of the security tokens, the documents that are protected by the tokens are not available for searching.

For more information about incorporating security controls into your custom search applications, see the *Programming Guide and API Reference for Enterprise Search*.

#### **Related concepts**

Chapter 6, "Search applications for enterprise search," on page 59

A search application enables you to search collections in your enterprise search system. You can create any number of search applications, and a single search application can search any number of collections.

"Document-level security with the Portal Search Engine" on page 121

You can use the IBM WebSphere Portal Search Engine to enforce document-level security when users search enterprise search collections.

---

## **Collection-level security**

To provide collection-level security, you configure options for indexing content and options for allowing search applications to search specific collections.

When you create a collection, you can choose an option to enable collection security. If you choose this option, the enterprise search global analysis processes apply different rules for indexing anchor text and duplicate documents.

After you create a search application, a search application ID enables you to specify which collections the search application can search, and which users can access the search application.

## **Anchor text analysis**

If you enable collection security, the global analysis processes apply special rules for indexing the anchor text in documents crawled by Web crawlers.

Anchor text is the information within a hypertext link that describes the page that the link connects to. For example, in the following link, the text Query Syntax is the anchor text in a link that connects to the syntax.htm page on a Web site:

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

Typically, the Web crawler follows links in documents to crawl additional documents and includes these linked pages in the index. During global analysis, the index processes the collection by associating the anchor text not only with the document in which it is embedded (the source document) but also with the target document. In the example above, the anchor text Query Syntax is associated with the target page syntax.htm and with the page that contains the anchor construct (<a href="...>).

If you enable collection security when you create a collection, the analysis and indexing processes do not do anchor text processing. By disabling anchor text processing, text is no longer associated with a document unless it actually appears in the document or the document's metadata. Users will not be exposed to information in documents that they are not allowed to access because the anchor text will never be associated with other documents that they might have access to.

Enabling collection security can enhance the security of Web documents by allowing users to search only the documents with security tokens that match their credentials. However, by not processing anchor text, the search results might not include all of the documents that are potentially relevant to a query.

If you do not enable collection security, the analysis and indexing processes index the anchor text for all pages that were retrieved by the Web crawler. This does not necessarily include all pages in the crawl space. Some pages might not have been retrieved because rules in a site's robots.txt file prohibited the Web crawler from visiting those pages.

By default, the analysis and indexing processes respect the robots.txt file by not performing anchor text processing on prohibited pages. If that leads to unsatisfactory search results, the behavior can be changed to ignore the robots.txt file by propagating and indexing anchor text for those pages, even though their content is not available. These pages are called *anchor text-only* documents.

You specify how you want to process anchor text-only documents by setting properties for the index server and Web crawler.

## Enabling anchor text-only document processing

You can enable anchor text-only document processing for Web documents if you need to remove documents from the index because a directive in a robots.txt file prohibits the crawler from visiting those documents.

### About this task

The Web crawler can store a metadata record that indicates that a URL should not be searchable because of a directive in a robots.txt file. If the URL was added to the index during a previous crawl that used different robots.txt rules, the content is removed from the index.

You must edit two configuration files: anchortext.ini and crawl.properties.

### Procedure

To enable the anchor text-only document processing:



1. On the index server, log in as the enterprise search administrator. This user ID was specified during the installation of DB2 II OmniFind.
2. Change to the `ES_INSTALL_ROOT/configurations/indexer` directory, open the `anchortext.ini` file with a text editor, and ensure that the `AnchorTextAnchorOnly` property is set to `true`.

When you change this value, the change becomes effective the next time that the index is reorganized.

3. On the crawler server, log in as the enterprise search administrator.
4. Change to the `ES_NODE_ROOT/master_config/crawler session ID` directory, open the `crawl.properties` file with a text editor, and specify one of the following values for the `robots_exclusion_return_code` property:

**740** Allows anchor text references to remain in the index even though the document content is removed. If this property is not in the file, the default value is 740.

**4044** Removes the anchor text references from the index along with the document content.

Changes that you make in the `crawl.properties` file take effect when you stop and restart the Web crawler, and then recrawl the documents.

## Duplicate document analysis

If you enable collection security, the global analysis processes do not identify duplicate documents in the collection.

During global analysis, the indexing processes identify documents that are duplicates, or near duplicates, of each other. They then associate all of these documents with one canonical representation of the content. By allowing duplicate documents to be identified, you can ensure that search results do not contain multiple documents with the same (or nearly the same) content.

If you enable collection security when you create a collection, duplicate documents are not identified, and so they are not associated with a common canonical representation. Instead, each document is indexed independently. This ensures that users search only the documents with security tokens that match their credentials. For example, two documents might be nearly identical in content, but use different access control lists to enforce security.

Disabling duplicate document analysis can enhance the security of documents in a collection, but search quality might be degraded if users receive multiple copies of the same document in the search results for one query.

## Security with search application IDs

To provide collection-level security, you specify which search applications can search each collection.

All search applications are required to pass an application identifier to the enterprise search APIs. An enterprise search administrator and your search applications can use this identifier to enforce collection-level security:

Before a search application can access a collection and search it, an enterprise search administrator must associate the search application with the specific collections that it can search. A search application can have access to all collections in an enterprise search system or be restricted to specific collections.

To enforce access controls, you can associate a set of user IDs or group IDs with your search application and allow only those users to access the application and search collections. For example, you can restrict access to the URL that launches your search application.

For more information about search application IDs, and how to incorporate security controls into your custom search applications, read about the search and index API available for enterprise search.

#### **Related concepts**

Chapter 6, “Search applications for enterprise search,” on page 59

A search application enables you to search collections in your enterprise search system. You can create any number of search applications, and a single search application can search any number of collections.

Search and Index API (SI-API)

---

## **Allowing users to access all documents**

You can specify an option that enables all users to search all documents in the collection, regardless of whether security tokens are associated with the documents.

### **Prerequisites**

To allow all users to access all documents in a collection, you must be a member of the enterprise search administrator role or be a collection administrator for the collection.


### **About this task**

When you configure a crawler, you can associate security tokens with the documents that are being crawled. Your search applications can use these tokens to enforce access controls when users search the collection. If you want to open up the collection for searching by all users, you can specify an option that forces the search servers to ignore any security tokens that are passed with a query, and allow all users to access all documents in the collection.

You might want to select this option if you are testing a new collection or if you need to troubleshoot a problem with the security processing in a search application.

### **Procedure**

To allow all users to search all documents in a collection:

1. Click **Collections** to open the Collections view.
2. In the list of collections, locate the collection that you want to administer and click  **Edit**.
3. On the Search page, select the **Enable all users to search all documents (do not use access controls)** check box.

Crawlers continue to add security tokens to documents, but the search servers ignore the tokens and allow all users to search all documents in the collection.

4. Click **Apply**.

---

## Chapter 13. Enterprise search integration with WebSphere Portal

You can expand the search capabilities of IBM WebSphere Portal by deploying enterprise search portlets in WebSphere Portal and the WebSphere Portal Search Center.

### Integration points

The enterprise search portlets integrate with WebSphere Portal in several ways:

#### WebSphere Portal

WebSphere Portal provides users with a single access point for interacting with applications, content, processes, and people. The WebSphere Portal framework enables new applications, called portlets, to be integrated and deployed without impacting other applications in the portal.

If you deploy the enterprise search portlet into WebSphere Portal, you can search enterprise search collections from the WebSphere Portal interface. Through WebSphere Portal configuration settings, you can ensure that the enterprise search portlet has the same look and feel as other portlets in your WebSphere Portal environment.

#### Portal Search Engine

The WebSphere Portal Search Engine crawls Web sites, Lotus Notes databases, and file systems. Administration portlets enable administrators to build indexed collections, and search portlets enable users to search those collections.

If you use WebSphere Portal Version 5.0.2 or later, you can migrate configuration information for Portal Search Engine collections and taxonomies to enterprise search.

If you use WebSphere Portal Version 5.1 or later, you can use the Portal Search Engine Document Search portlet to search enterprise search collections. Configuration properties enable administrators to easily switch between the two search capabilities as needed.

#### WebSphere Portal Search Center

The WebSphere Portal Search Center provides a central starting point for searching all sources that are made available for searching through WebSphere Portal. The Search Center and the Universal search portlet enable you to search WebSphere Portal content and any other collections that an administrator registers with the Search Center.

The Search Center has a paged interface. You can search all available collections through one common page, or you can select a page to search an individual collection. For example, there is a page for Portal Search Engine indexes and a page for Portal Document Management libraries.

To enable enterprise search collections to be searched from the Search Center, IBM DB2 Information Integrator OmniFind Edition (DB2 II OmniFind) provides an adapter and a registration portlet. The adapter adds a page for enterprise search to the Search Center interface, and the registration portlet registers the enterprise search portlet with the Search Center.

## Benefits of integrating

Enterprise search enhances the WebSphere Portal search environment by providing support for searching a wider range of data source types. The Portal Search Engine portlets search Web, Notes, and file system sources only. With the enterprise search portlet, you can search these types of sources and all the other data source types that enterprise search supports.

Enterprise search also offers benefits in scalability. The Portal Search Engine is useful for small-sized or medium-sized businesses where a single server is sufficient to support the search and retrieval workload. To support enterprise-level capacities, enterprise search distributes the workload over four servers (one for crawling data, one for parsing and indexing data, and two to support search and retrieval processes).

## Portlet deployment overview

The portlets that you can use for enterprise search depend on the version of WebSphere Portal that you use:

### WebSphere Portal Version 5.0.2

You can deploy the enterprise search portlet and use it to search enterprise search collections. This portlet can coexist with the Portal Search Engine portlets.

### WebSphere Portal Version 5.1

- You can deploy the enterprise search portlet and use it to search enterprise search collections. This portlet can coexist with the Portal Search Engine portlets.
- After you deploy the enterprise search portlet, you can configure the Document Search portlet in the Portal Search Engine to search enterprise search collections.
- After the enterprise search portlet is registered with the WebSphere Portal Search Center, you can use the Search Center to search enterprise search collections. You can select a page to search only enterprise search collections, or you can enter a query that searches enterprise search collections and all the other collections that are available in the Search Center.

To deploy the enterprise search portlets, you download them from the WebSphere Portal catalog Web site, then use standard WebSphere Portal procedures for portlet deployment. Readme files that are stored with the portlets in the catalog provide additional information about the use and implementation of the portlets.

For the latest information about the enterprise search portlets for WebSphere Portal, see the readme file for DB2 II OmniFind.

### Related concepts

Chapter 14, “Migration from WebSphere Portal,” on page 123

Enterprise search provides a migration wizard that you can use to migrate taxonomies and collections from IBM WebSphere Portal to enterprise search.

---

## Deploying the enterprise search portlets in WebSphere Portal

You can download enterprise search portlets from the IBM WebSphere Portal portlet catalog Web site.

### Procedure

To deploy the enterprise search portlets for the WebSphere Portal:

1. Read the DB2 II OmniFind readme file for the latest information about the enterprise search portlets. The readme file contains the latest information about the portlets and describes how they are listed in the WebSphere Portal portlet catalog.
2. Access the WebSphere Portal portlet catalog at:  
<http://catalog.lotus.com/wps/portal/portalworkplace>
3. Read the readme files that accompany the enterprise search portlets. The readme files contain the most current information about system requirements, configuration, and deployment.
4. Deploy the enterprise search portlets by using the standard WebSphere Portal procedures for portlet deployment.

---

## Configuring the Portal Search Engine to search enterprise search collections

You can configure the Document Search portlet in the WebSphere Portal Search Engine to search enterprise search collections.

### Prerequisites

The enterprise search portlet must be deployed in WebSphere Portal before you can configure the Portal Search Engine to use the portlet.

### About this task

After you deploy the enterprise search portlet, you can continue to use the Portal Search Engine to search indexed data in WebSphere Portal. To enable users to search enterprise search collections, a WebSphere Portal administrator edits properties in the Portal Search Engine configuration. The coexistence of these two search capabilities enables you to switch between them and use the most appropriate search solution for your needs.

### Procedure

See the DB2 II OmniFind readme file for specific instructions on how to configure the Portal Search Engine to use the enterprise search portlet.

---

## Document-level security with the Portal Search Engine

You can use the IBM WebSphere Portal Search Engine to enforce document-level security when users search enterprise search collections.

If an enterprise search crawler associates a group ID security token with the documents that it crawls, and if you configure the Document Search portlet for the Portal Search Engine to search enterprise search collections, then the Portal Search Engine can derive the group ID for the logged in user, and pass the security token

for that group ID with the query to enterprise search. The security token ensures that only the documents that a user is authorized to see are returned in the search results.

If the crawler associated another type of security token with documents, such as a user ID or a user role, and you want to enforce document-level security when you search enterprise search collections, then you must create a custom search portlet. The Portal Search Engine derives security tokens for group IDs only.

**Related concepts**

Chapter 6, “Search applications for enterprise search,” on page 59

A search application enables you to search collections in your enterprise search system. You can create any number of search applications, and a single search application can search any number of collections.

“Document-level security” on page 114

Document-level security ensures that users who search collections are able to access only the documents that they are allowed to see.

---

## Chapter 14. Migration from WebSphere Portal

Enterprise search provides a migration wizard that you can use to migrate taxonomies and collections from IBM WebSphere Portal to enterprise search.

To migrate taxonomies and collections, you run the migration wizard on the enterprise search index server. After you migrate a taxonomy, you can use it with enterprise search collections. You can also use enterprise search to administer and search collections that you migrate from WebSphere Portal.

In enterprise search, a taxonomy is called a category tree. After you migrate a rule-based taxonomy, you can use the enterprise search administration console to administer the category tree. To use model-based taxonomies with enterprise search, WebSphere Portal must be installed on the index server.

If you want to migrate taxonomies and collections, always migrate the model-based taxonomy files first, before you migrate collections. If you do not do this, model-based categorization will not work with the collections that you migrate from WebSphere Portal.

### Related concepts

Chapter 13, “Enterprise search integration with WebSphere Portal,” on page 119  
You can expand the search capabilities of IBM WebSphere Portal by deploying enterprise search portlets in WebSphere Portal and the WebSphere Portal Search Center.

---

## Migrating a model-based taxonomy from WebSphere Portal

You can select which model-based taxonomy you want to use with an enterprise search collection by using the Taxonomy Management Portlet in your WebSphere Portal installation. Collections that you already migrated to enterprise search are not affected by a new taxonomy selection.

### About this task

To migrate a model-based taxonomy, you must select and export the taxonomy in WebSphere Portal. Then, use the enterprise search migration wizard to migrate the taxonomy to enterprise search.

### Procedure

To migrate a model-based taxonomy from WebSphere Portal to enterprise search:

1. Export your current model-based taxonomy from the WebSphere Portal Taxonomy Management portlet. The taxonomy comprises the following XML files:

- synonyms.xml
- titles.xml
- treenodes.xml

2. Copy these files to the enterprise search index server.

3. On the enterprise search index server, log in as the enterprise search administrator. This user ID was specified during the installation of DB2 II OmniFind.

4. Change to the enterprise search installation directory:

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

5. Enter the following command to start the migration wizard, then click **Next**.

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

Ensure that you do not create an enterprise search collection while the migration wizard is running.

6. Select **Migrate model-based taxonomy files from a WebSphere Portal installation**, then click **Next**.
7. If you are starting the migration wizard for the first time, specify the paths to where WebSphere Application Server and WebSphere Portal are installed. Click **Next**.
8. Browse for the directory that contains the model-based taxonomy files, select the XML files that you must migrate to use the model-based taxonomy (synonyms.xml, titles.xml, and treenodes.xml), then click **Next**.

If errors occur, see the MigrationWizard.log file that is in the directory where the migration wizard is installed.

#### Related concepts

“Model-based categories” on page 76

If you use model-based categories in your IBM WebSphere Portal system, you can continue to use those categories with enterprise search collections.

---

## Migrating a collection from WebSphere Portal

To migrate collections from WebSphere Portal to enterprise search, prepare the collections in WebSphere Portal, then use the migration wizard to migrate them.

### Procedure

To migrate a collection from WebSphere Portal to enterprise search:

1. In WebSphere Portal Search Engine, stop all of the crawler processes in the collections that you want to migrate, and approve or reject all pending documents. (Enterprise search does not support the concept of pending documents.)
2. For each collection that you want to migrate, use the Portal Search Engine portlets to export the settings to XML files.
3. If the enterprise search index server is installed on a separate server, copy the exported XML files to the index server.
4. On the enterprise search index server, log in as the enterprise search administrator. This user ID was specified during the installation of DB2 II OmniFind.
5. Change to the enterprise search installation directory:

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



6. Enter the following command to start the migration wizard, then click **Next**.

UNIX: `./eswpsmigrate.sh`

Windows: `eswpsmigrate.bat`

7. Select **Migrate Portal Search Engine collections of a WebSphere Portal installation**, then click **Next**.
8. Browse to the directory that contains the exported Portal Search Engine configuration files, select the files that you want to migrate, then click **Next**. The selected configuration files are analyzed and validated.
9. Enter the following information for each collection, then click **Next** to start migrating the collections to enterprise search:
  - The name of the collection as you want to use it in enterprise search.
  - The criterion by which the document importance is determined for the collection. The static ranking factor can be none, based on document dates, or based on links to Web documents from other Web documents.
  - The type of categories that you want to use for this collection. You can select either none, rule-based categories, or model-based categories.  
If you select rule-based categories, the taxonomy and the rules of the WebSphere Portal collection are migrated to enterprise search.  
You can select model-based categories only if you previously migrated the model-based taxonomy from WebSphere Portal to enterprise search.If errors occur during migration, see the `MigrationWizard.log` file that is in the directory where the migration wizard is installed.  
You can now configure additional settings for the migrated enterprise search collections and start the collection processes by using the enterprise search administration console.
10. Start the crawling, parsing, and indexing processes for the migrated collection from the enterprise search administration console.
11. After you determine that the migrated collection is searchable in enterprise search, delete the original collection in the Portal Search Engine.
12. Optional: As a WebSphere Portal administrator, take the following steps if you want to enable users to search the migrated collection from a portal in WebSphere Portal.
  - a. Deploy the enterprise search portlet in your WebSphere Portal installation.  
In a WebSphere Portal Server cluster, this should be done on the server where the WebSphere Application Server deployment manager is installed. The deployment manager distributes the enterprise search portlet to the other servers in the WebSphere Portal Server cluster.
  - b. Add the enterprise search portlet to the appropriate portal pages.  
In WebSphere Portal, access control of the search portlet is modeled by accessibility to specific pages and portlets. Although collection settings are migrated, the portlet must be positioned manually by the WebSphere Portal Server administrator.

### Related concepts

Chapter 3, “Enterprise search collections,” on page 19

An enterprise search collection contains the entire set of sources that users can search with a single query.

“Rule-based categories” on page 75

You can configure rules to control which documents are associated with categories in an enterprise search collection.

“Model-based categories” on page 76

A category tree enables you to view all the rule-based categories in a collection. You use the category tree to create categories, delete categories, and edit the rules that associate documents with categories.

“Static ranking” on page 85

For certain types of documents, you can associate a static ranking factor that increases the importance of those documents in the search results.

---

## Migrated collection settings

When you migrate collections from IBM WebSphere Portal, the migration wizard creates default settings for collections and crawlers.

If the same setting exists in Portal Search Engine collections and enterprise search collections, then the wizard uses the Portal Search Engine setting when it migrates the collection to enterprise search. For settings that exist only in enterprise search, the wizard uses the settings that you specify when you migrate the collection or the default settings for collections in enterprise search.

### Settings that exist in Portal Search Engine and enterprise search

The migration wizard migrates the following settings for each collection that you migrate:

- The Portal Search Engine sites within the Portal Search Engine collection
- The collection language
- The taxonomy (or category tree) and the rules for the rule-based categories, if the enterprise search collection uses rule-based categorization

Each Portal Search Engine site in a collection is consolidated into an enterprise search Web crawler. The migration wizard migrates the following crawler settings:

- The start URL
- The number of parallel crawling processes
- The crawling depth
- The timeout (in seconds) for retrieving a document
- The default character set
- Include and exclude rules for crawling

### Settings that exist only in enterprise search

When you migrate a collection, you specify information about the collection. The migration wizard migrates those settings and uses the default settings for collections in enterprise search to configure each collection that you migrate.

You can modify the collection and crawler configurations by using the enterprise search administration console. The values that are shown in parentheses ( ) are the default settings for the migrated data.

- The collection name
- The document ranking strategy, for example, the document date
- The type of categorization that is used, such as rule-based or none
- Whether to use the search cache and how many queries with search results the search cache can hold (yes, 5000)

- Whether to monitor search response times and issue an alert if a limit is exceeded (yes, 5 seconds)
- Whether to use access controls (no)
- A schedule to refresh the index
- A schedule to reorganize the index
- The log detail level (all messages)

The migration wizard also creates the following settings for each crawler:

- The crawler name
- The crawler description
- The maximum page length
- The document security settings
- The document multipurpose Internet mail extensions (MIME) types that need to be crawled, if applicable to the data source type

---

## Migration wizard log file

The migration wizard writes all messages to the WpsMigratorLog.log file in the directory where the migration wizard is installed.

For each migrated collection, the WpsMigratorLog.log log file contains the values of all the settings that were read from the WebSphere Portal Search Engine, and specifies where these settings were imported to enterprise search collections.



---

## Chapter 15. Backup and restore for enterprise search

You can back up and restore the enterprise search system with the backup and restore scripts.

If the system fails because of an irrecoverable error, you must re-install IBM DB2 Information Integrator OmniFind Edition and then run the restore script. You can also use these scripts to restore essential system files to one or more new servers.

The scripts back up and restore the following files:

- Configuration files from the ES\_NODE\_ROOT/master\_config directory
- Database files for the crawlers
- Index build files

When you back up the system, the script creates the following subdirectories under the backup directory (*backup\_directory*) name that you specify:

**master\_config**

Contains the configuration files from the ES\_NODE\_ROOT/master\_config directory

**database**

Contains the database files from the crawler server

**data** Contains the index build files from the index server

You must have enough disk space available to back up the enterprise search system files to another directory. The backup and restore scripts do not check the files. When you start a backup, most system sessions are temporarily unavailable. The search processes will continue to run. You should start the backup after you reorganize the index so that you have the most current index.

On a multiple server installation, back up and restore the system from the enterprise search index server. Because the index server creates a DB2 catalog, the index server can access and back up the DB2 tables on the crawler server.

---

### Backing up the enterprise search system

You can back up the enterprise search system by using the esbackup.sh script for UNIX or the esbackup.bat script for Microsoft Windows.

**Procedure**

To back up the enterprise search system:

1. On the index server, log in as the enterprise search administrator. This user ID was specified during the installation of DB2 II OmniFind.
2. Enter the following command:

UNIX: esbackup.sh -c -d -i *backup\_directory*

Windows: esbackup.bat -c -d -i *backup\_directory*

You can specify the following options:

**-c** Backs up configuration files. This option is the default.

- d Backs up crawled documents in the database.
- i Backs up index files.

---

## Restoring the enterprise search system

You can restore system configuration files after you re-install DB2 II OmniFind by using the `esrestore.sh` script for UNIX or the `esrestore.bat` script for Microsoft Windows.

### Procedure

To restore the enterprise search system from a backup:

1. On the index server, log in as the enterprise search administrator. This user ID was specified during the installation of DB2 II OmniFind.
2. Start the common communications layer (CCL) if it is not already started:

UNIX: `startccl.sh -bg`

Windows: `startccl.bat`

Windows only: to start the CCL in the background, click **Start** → **Programs** → **Administrative Tools** → **Services**, and restart the IBM DB2 Information Integrator OmniFind Edition service.

3. Stop the controller if it is not already stopped:

`esadmin stop`

4. Enter the following command:

UNIX: `esrestore.sh -c -d -i backup_directory`

Windows: `esrestore.bat -c -d -i backup_directory`

You can specify the following options:

- c Restores configuration files
- d Restores crawled documents in the database
- i Restores index files

---

## Restoring enterprise search system files to new servers

You can back up system files from one enterprise search server and restore the system files to one or more new servers instead of restoring them to the current servers.

### Prerequisites

You must install DB2 II OmniFind on the new servers before you run the restore script.

### About this task

The server information that is stored in the `ES_NODE_ROOT/master_config/nodes.ini` file is not included in the backup files.

### Procedure

To restore the enterprise search system files to one or more new servers:

1. On the index server, log in as the enterprise search administrator. This user ID was specified during the installation of DB2 II OmniFind.
2. Run the backup script on the current (old) index server:

UNIX: `esbackup.sh`

Windows: `esbackup.bat`

3. Use an FTP program to send all the files in the *backup\_directory* to the new index server.
4. Run the restore script on the new index server:

UNIX: `esrestore.sh`

Windows: `esrestore.bat`





---

## Chapter 16. Enterprise search log files

You can choose the types of messages that you want to log for a collection and for the system, specify options for creating and viewing log files, and specify options for receiving e-mail about messages.

During normal operations, the enterprise search components write log messages to a common log file. This log file is in the `ES_NODE_ROOT/logs` directory on the index server. You can use the administration console to view this common log data.

If a problem occurs, such as a network communication failure, the components write log messages to a logs directory on the server where the component is installed. To view these local log files, use a file viewer on that computer, such as the `tail` utility on a UNIX system. You cannot use the administration console to view these types of log files.

When you configure log files, you can choose the types of messages that you want to log (such as error or warning messages), specify how often old log files are to be discarded to make room for new log files, specify a maximum size for the log files, and select the language of the messages. You can also specify options for receiving e-mail whenever certain messages or certain types of messages are logged.

When you monitor log files, you can choose which log file you want to open. You can filter the content of the log file so that you view only messages of a specific severity level (such as error messages only) or messages that were produced by a specific component. When you view a log file, you can view details about individual messages. For example, you might want to see the name of the function that produced the message and other information that can help you take corrective action, if necessary.

---

### Configuring log files

You can specify the types of messages that you want to log and specify options for creating log files.

#### Prerequisites

To configure collection-level log files, you must be a member of the enterprise search administrator role or be a collection administrator for the collection. To configure system-level log files, you must be an enterprise search administrator.


#### About this task

To prevent log files from consuming too much disk space, the system rotates log files, and always starts a new log file whenever the current date changes. If one log file grows to its maximum allowable size, and the date did not change, the system creates a new log file. When the maximum number of log files is reached, the oldest log file is discarded so that a new one can be created.

To receive e-mail about logged messages, you first specify information about how the e-mail is to be delivered. You then specify which messages you want to receive e-mail for.

## Procedure

To configure enterprise search log files:

1. If you want to configure options for creating and rotating system-level log files:
  - a. Click **System** to open the System view.
  - b. Select the Log page.
2. If you want to configure options for creating and rotating collection-level log files:
  - a. Click **Collections** to open the Collections view.
  - b. In the list of collections, locate the collection that you want to specify options for and click  **Edit**.
3. On the Log page, in the **Type of information to log** field, select the types of messages that you want to log:

### Error messages only

Error messages indicate that an undesirable situation or unexpected behavior occurred and that the process cannot continue. You must take action to correct the problem.

### Error and warning messages

Warning messages indicate a possible conflict or inconsistency, but they do not cause a process to stop. This option is the default.

### All messages

Information messages provide general information about the system or current task and do not require any corrective action.

4. In the **Maximum size of each log file** field, type the maximum number of megabytes for each log file. The default value is 5 MB.

When the log file grows to this size, a new log file is created, up to the maximum number of log files that you allow. By keeping log files relatively small, you can view them more efficiently.
5. In the **Maximum number of log files** field, type the maximum number of log files that you want to create. The default value is 10.

If you want to ensure that older log messages are available for review, increase this value. If you are more interested in recent messages and do not need to maintain a long history of activity, decrease this value.
6. In the **Default locale** field, select the language that you want to use to log messages. The default value is English.
7. If you want to receive e-mail automatically whenever certain messages or certain types of messages are logged, click **Configure e-mail options for messages** so that you can specify options for receiving e-mail.
8. Click **Apply**.

---

## Receiving e-mail about logged messages

You can specify options to receive e-mail whenever certain messages, or certain types of messages, are logged.

### Prerequisites

To configure e-mail options for collection-level messages, you must be a member of the enterprise search administrator role or be a collection administrator for the collection. To configure options for your Simple Mail Transfer Protocol (SMTP)

server, or to configure e-mail options for system-level messages, you must be a member of the enterprise search administrator role.

### About this task

When you configure alerts, you can choose an option to log messages whenever certain events occur. If you enable those options, you can then configure options to receive e-mail automatically whenever those messages are logged. You can also specify options to receive e-mail when other messages are logged, not just messages that are triggered by events.


To receive e-mail about logged messages, you first specify information about how the e-mail is to be delivered. You then specify which messages you want to receive e-mail for.

### Procedure

To configure e-mail options for messages:

1. Configure e-mail delivery options:
  - a. Click **System** to open the System view.
  - b. On the Log page, click **Configure e-mail options for messages**.
  - c. On the Configure e-mail options for system messages page, in the **SMTP mail server to use for delivering e-mail** field, type the fully qualified host name or IP address of the SMTP server that you want to use.  
Enterprise search uses this server to send e-mail to the addresses that you specify.
  - d. In the **Frequency to check for e-mail** field, specify how often you want the system to check for eligible messages and send e-mail about them.  
The system combines all the messages for a specific e-mail address into one message, and sends that message at the frequency that you specify.
  - e. Click **OK**.
2. Configure options for receiving e-mail about system messages:
  - a. Click **System** to open the System view.
  - b. On the Log page, click **Configure e-mail options for messages**.
  - c. On the Configure e-mail options for system messages page, select the **Enable alerts for system-level messages** check box.
  - d. In the **E-mail address for receiving alerts** field, type one or more e-mail addresses. Typically, an enterprise search administrator should receive information about system messages.  
Separate each address with a comma. For example:  
steinbeck@us.ibm.com, yeats@ireland.ibm.com, dante@it.ibm.com.
  - e. If you want to receive e-mail about all error messages that are logged, select the **All error messages** check box.
  - f. If you want to receive e-mail only when certain system-level messages are logged, type the message IDs for those messages. Type one message ID per line. For example:  
FFQC4819E  
FFQ00005E

If you enabled system-level alerts, this field already includes the message IDs for the messages that are logged when a monitored event occurs.

- g. Click **OK**.
- 3. Configure options for receiving e-mail about messages for a specific collection:
  - a. Click **Collections** to open the Collections view.
  - b. In the list of collections, locate the collection that you want to specify options for and click  **Edit**.
  - c. On the Log page, click **Configure e-mail options for messages**.
  - d. On the Configure e-mail options for collection messages page, select the **Enable alerts for collection-level messages** check box.
  - e. In the **E-mail address for receiving alerts** field, type one or more e-mail addresses. Typically, a collection administrator should receive information about collection-level messages.  
Separate each address with a comma. For example:  
steinbeck@us.ibm.com, yeats@ireland.ibm.com, dante@it.ibm.com.
  - f. If you want to receive e-mail about all error messages that are logged, select the **All error messages** check box.
  - g. If you want to receive e-mail only when certain system-level messages are logged, type the message IDs for those messages. Type one message ID per line. For example:  
FFQC4819E  
FFQ00005E  
  
If you enabled collection-level alerts, this field already includes the message IDs for the messages that are logged when a monitored event occurs.
  - h. Click **OK**.

#### Related tasks

“Configuring collection-level alerts” on page 94

By configuring alerts, you can ensure that messages are written to the log file whenever certain collection-level events occur. You can also receive e-mail whenever messages about these events are logged.

“Configuring system-level alerts” on page 95

By configuring alerts, you can ensure that messages are written to the log file whenever certain system-level events occur. You can also receive e-mail whenever messages about these events are logged.

---


## Viewing log files


You can view log messages that the system and collection components write to a common log file. You can also specify filters to view messages of a specific severity level and messages from specific components.


### Prerequisites

All enterprise search administrative users can view log files for the collections that they are authorized to administer. To view system-level log files, you must be a member of the enterprise search administrator role.

### Procedure

1. To view the log files for a single collection:
  - a. Click **Collections** to open the Collections view.
  - b. In the list of collections, locate the collection that you want to view, click  **Monitor**, and open the Log page.

**Tip:** If you are editing a collection and are already on the Log page, you can click  **Monitor** to change to the view for monitoring the collection.

2. To view system-level log files:
  - a. Click **System** to open the System view.
  - b. Click  **Monitor** to change to the view for monitoring the system, and open the Log page

3. In the **Log file** field, select the log file that you want to view.

The name of each log file includes the log file type (such as a system or collection identifier), the date that the file was created, and a numeric suffix that indicates the order in which the file was created on that date. For example:

```
log_file_type_2004-05-26_1.log  
log_file_type_2004-05-26_2.log  
log_file_type_2004-05-25_1.log  
log_file_type_2004-05-25_2.log  
log_file_type_2004-05-25_3.log
```

4. If you want to view only messages of specific severity levels, select the appropriate check boxes in the **Severity** field.
5. If you want to view only messages from specific components, select the appropriate check boxes in the **Component** field.

6. Click **View log**.

The log file is displayed. For each message, you see the date that the message was issued, the message severity level, the name of the component that issued the message, and the message ID and error text.

7. If you want to see more detailed information about a specific message, click  **Details**.

Details about the specific message are displayed. You see the host name of the enterprise search server where the message occurred, the name of the file that produced the error, the function and line number where the error occurred, the process ID, and the thread ID.



---

## Chapter 17. URI formats in enterprise search

The uniform resource identifier (URI) of each document in an enterprise search collection indicates the type of crawler that added the document to the collection.

You can specify URIs or URI patterns when you configure categories, scopes, and quick links for a collection. You also specify the URI when you need to remove documents from the index, or to view detailed status information about a specific URI.

Search the collection to determine the URIs or URI patterns for a document. You can click the URIs in the search results to retrieve documents that you are interested in. You can copy the URI from the search results to use the URI in the enterprise search administration console. For example, you can specify a URI or a pattern based on a URI to automatically associate documents that match that URI with an enterprise search category.

### Content Manager crawlers

The URI format for documents that are crawled by a Content Manager crawler is:

*cm://Server Name/Item Type Name/PID*

#### Parameters

*Server Name*

The name of the IBM DB2 Content Manager library server.

*Item Type Name*

The name of the target item type.

*PID*

The DB2 Content Manager persistent identifier.

### DB2 crawlers

The URI format for documents that are crawled by a DB2 crawler is:

*db2://Database Name/Table Name  
/Unique Identifier Column Name1/Unique Identifier Value1  
[/Unique Identifier Column Name2/Unique Identifier Value2/...  
/Unique Identifier Column NameN/Unique Identifier ValueN]*

#### Parameters:

URL encoding is applied to all of the fields.

*Database Name*

The internal name of the database or the alias for the database.

*Table Name*

The name of the target table, including the name of the schema.

*Unique Identifier Column Name1*

The name of the first Unique Identifier column in the table.

*Unique Identifier Value1*

The value of the first Unique Identifier column.

*Unique Identifier Column NameN*

The name of the *n*th Unique Identifier column in the table.

### *Unique Identifier ValueN*

The value of the *n*th Unique Identifier column.

## **Exchange Server crawlers**

The URI format for documents that are crawled by an Exchange Server crawler is:

`exchange://OWA Path[?useSSL=true]`

### **Parameters**

#### *OWA Path*

The Outlook Web Access (OWA) path, without the protocol.

#### **[?useSSL=true]**

Added when the protocol of the original OWA path is HTTPS.

## **Notes crawlers**

The URI format for documents that are crawled by a Notes crawler is:

`domino://Server Name[:Port Number]/Database Replica ID/Database Path and Name  
/[View Universal ID]/Document Universal ID  
[?AttNo=Attachment No AttName=Attachment File Name]`

### **Parameters**

URL encoding is applied to all of the fields.

#### *Server Name*

The name of the Lotus Notes server.

#### *Port Number*

The port number for the Lotus Notes server. The port number is optional.

#### *Database Replica ID*

The identifier for the database replica.

#### *Database Path and Name*

The path and file name for the NSF database on the target Lotus Notes server.

#### *View Universal ID*

The View Universal ID that is defined on the target database. This ID is specified only when the document is selected from a view or folder. If you do not designate a view or folder to crawl (for example if you specify that you want to crawl all documents in a database), the View Universal ID is not specified.

#### *Document Universal ID*

The Document Universal ID that is defined in the document that is crawled by the crawler.

#### *Attachment No*

A consecutive number, starting from zero, for each attachment. The attachment number is optional.

#### *Attachment File Name*

The original name of the attachment file. The attachment file name is optional.



## UNIX file system crawlers

The URI format for documents that are crawled by a UNIX file system crawler is:

`file:///Directory Name/Filename`

### Parameters

URL encoding is applied to all of the fields.

*Directory Name*

The absolute path name for the directory.

*Filename*

The name of the file.

## VeniceBridge crawlers

The URI format for documents that are crawled by a VeniceBridge crawler is:

`vbr://Server Name/Repository System ID/Repository Persistent ID  
/Item ID/Version ID  
/Item Type/?[Page=Page Number&]JNDI properties`

### Parameters

URL encoding is applied to all of the fields.

*Server Name*

The name of the VeniceBridge server.

*Repository System ID*

The system ID for the repository.

*Repository Persistent ID*

The persistent ID for the repository.

*Item ID*

The ID for the item.

*Version ID*

The ID for the version. If the version ID is blank, this value indicates the latest version of the document.

*Item Type*

The type of the item (CONTENT).

*Page Number*

The page number.

*JNDI properties*

The JNDI properties for the J2EE application client. There are two types of properties:

#### **java.naming.factory.initial**

The name of the class for the application server that is used to create the EJB handle.

#### **java.naming.provider.url**

The URL to the naming service for the application server that is used to request the EJB handle.

## Windows file system crawlers

The URI formats for documents that are crawled by a Windows file system crawler are:

file:///Directory Name/Filename  
file:///Network Folder Name/Directory Name/Filename

### Parameters

URL encoding is applied to all of the fields.

#### *Directory Name*

The absolute path name for the directory.

#### *Filename*

The name of the file.

#### *Network Folder Name*

For documents on remote servers only, the name of the shared folder on Windows network.

### Related tasks

“Removing URIs from the index” on page 57

To prevent users from searching documents in a collection, you can remove the URIs for those documents from the index.

“Configuring quick links” on page 74

To create a quick link for an enterprise search collection, you associate the URI of a document with the keywords that trigger its inclusion in the search results.

“Configuring categories” on page 79

You can create any number of categories for a collection, and each category can contain any number of rules. The rules determine which documents are associated automatically with the category.

“Configuring scopes” on page 81

When you configure a scope for an enterprise search collection, you specify the URIs, or URI patterns, for a range of documents in the index that users are allowed to search.

“Viewing details about a URI” on page 96

You can view detailed information about a URI. You can see current and historical information about how the document that is represented by this URI is crawled, indexed, and searched.

---

## Chapter 18. Enterprise Search messages

---

### General system messages

#### **FFQO0164E: Unable to load properties file *file\_name*.**

The configuration file *file\_name* could not be loaded.

Check the log for additional messages. Verify that the file exists and is readable. Check the access permissions of the file and for all files that are included by this configuration file.

#### **FFQO0165E: Unable to load included properties file *file\_name*.**

The configuration file *file\_name* that was included by another configuration file could not be loaded.

Check the log for additional messages. Verify that the file exists and is readable. Check the access permissions of the file and for all files that are included by this configuration file.

#### **FFQO0168E: Unable to save properties to file *file\_name*.**

A configuration could not be saved to the file *file\_name*.

Verify that the directory or file system into which the configuration file was supposed to be saved exists and is writable. Check whether a write-protected file with the same name already exists.

---

### Migration wizard messages

#### **FFQW0001E: No message was found for the *message\_key* key. Contact the enterprise search administrator.**

The message file that is used by the wizard was not found.

Verify that enterprise search components of DB2 Information Integrator OmniFind Edition was installed correctly.

#### **FFQW0002E: File *file\_name* cannot be read. Change your selection.**

The selected file or directory can not be read by the user that started the wizard.

If the file is correct, contact an administrator to change the access rights of the file.

**FFQW0003E: No collection name is set for file *file\_name*.**

Each collection that you want to migrate must have a collection name that does not already exist in enterprise search.

Specify a collection name that does not exist in enterprise search.

**FFQW0008E: The user ID or password is invalid.**

To log in to the migration wizard, you must enter the user ID and password of the enterprise search administrator that was entered during installation.

Enter the user ID and password of the enterprise search administrator and try again.

**FFQW0011E: An error occurred while importing the WebSpherePortal Search Engine collection in file *file\_name*. At line *line\_number* and column *column\_number*, the following error occurred: *error\_message*. The file could not be imported.**

The document contains unexpected content. It might not be a valid collection configuration file, or the file was corrupted.

Check whether the file is the one that you exported from WebSphere Portal. If this is the case, export the configuration settings again from the Portal Search engine collection into a new file and run the wizard again.

**FFQW0012E: An error occurred while creating collection *collection\_name*. The following error was returned: *error\_code*.**

Enterprise search administration returned an error when creating the enterprise search collection.

Correct the problem and try again.

**FFQW0013E: An error occurred while creating a Web crawler for collection *collection\_name*. The following error occurred: *error\_code*.**

Enterprise search administration returned an error when creating the Web crawler.

Correct the problem and try again.

**FFQW0014E: An error occurred while creating collection *collection\_name*. No session could be found for the ID *user\_name*.**

The wizard could not connect to enterprise search.

Ensure that enterprise search is running.

**FFQW0016E: An error occurred while creating collection *collection\_name*: *error\_message*.**

Enterprise search administration returned an error when creating the enterprise search collection.

Check the wizard log file for more information. Correct the problem and try again.

**FFQW0019E: An error occurred while creating a crawler configuration file - Exception Message: *error\_message***

Enterprise search administration returned an error when creating a Web crawler configuration file.

More information can be found in the wizard log file. Correct the problem and try again.

**FFQW0120I: The migration completed successfully. See the log file for more information.**

This message is for your information only. No action is required.

**FFQW0140I: The collection was imported from the WebSphere Portal configuration file *file*.**

This message is for your information only. No action is required.

**FFQW0141I: Collection *collection\_name* was created successfully.**

This message is for your information only. No action is required.

**FFQW0142I: The Web crawler for collection *collection\_name* was created successfully.**

This message is for your information only. No action is required.

**FFQW0154I: Creating the crawler data for collection *collection\_name*.**

This message is for your information only. No action is required.

**FFQW0155I:** The following settings for the maximum number of crawl threads were found: *list\_thread\_numbers*. Imported value is *maximum\_threads*.

This message is for your information only. No action is required.

**FFQW0156I:** The following settings for the maximum crawl path depths were found: *list\_path\_depths*. Imported value is *maximum\_path\_depth*.

This message is for your information only. No action is required.

**FFQW0157I:** The following settings for the crawl document timeout were found: *list\_timeout\_limits*. Imported value is *maximum\_timeout\_limit*.

This message is for your information only. No action is required.

**FFQW0158I:** The following settings for default character sets were found: *list\_charsets*. Imported value is *charset*.

This message is for your information only. No action is required.

**FFQW0159I:** The configuration file *file* was created. Return code: *return\_code*.

This message is for your information only. No action is required.

---

## Configuration manager messages

**FFQC1000E:** Collection name *collection\_name* already exists.

The collection name *collection\_name* that was entered already exists as a collection name in enterprise search for another collection.

Enter a different collection name.

**FFQC1003E:** Session name *session\_name* already exists.

The session name *session\_name* that was entered already exists as a session name in enterprise search for another session.

Enter a different session name.

**FFQC1008E:** Session *session\_id* is already configured. Unable to add session.

The session *session\_id* that you are trying to add is already configured.

You can remove the existing session and re-create a new session by using the session name *session\_id*, or if you are satisfied with the existing session configurations, you can leave the current session and take no further action.

### **FFQC2034E: No index available to do content refresh.**

Content refresh can only be done on an existing index. No index was found for the specified collection.

Ensure that a valid index was built before you refresh the index.

### **FFQC5217E: An error occurred while loading property file *file\_name*.**

An error occurred while attempting to load the property file contents for *file\_name*.

Browse to the file *file\_name* and ensure that the contents of the file do not contain malformed data.

### **FFQC5228W: File name *file\_name* has file size of *file\_size*. This exceeds file replacement size limit of *file\_limit*.**

Text replacement for temporary variables was not done for the configuration file *file\_name*. For each configuration file that the system creates, a scan of temporary variables replaces values that are determined during run-time with real-time values. This file has a size of *file\_size*, which exceeds the maximum size on which the file scans are done, *file\_limit*.

The file probably does not contain temporary variables, so no scan is needed and no user action is required.

---

## **Crawler messages**

### **FFQD1003E: ERROR: User *user* does not exist. Check *file*.**

The user that was specified in the es.cfg file does not exist.

Verify that DB2 Information Integrator OmniFind Edition (enterprise search) was installed correctly.

### **FFQD1011E: ERROR: *file* was not found.**

The specified file was not found.

Verify that DB2 Information Integrator OmniFind Edition (enterprise search) was installed correctly.

### **FFQD1031E: ERROR: This script is not running on the crawler server. \nRun this script on a crawler server.**

This script is not running on the crawler server.

For a multiple server configuration, run this script on a crawler server. For a single server configuration, verify that the value of InstalledComponents in the configuration file contains "crawler".

**FFQD1034E: ERROR: The environment variable CMBROOT is not defined. \nVerify that DB2 Information Integrator for Content was installed correctly.**

The environment variable CMBROOT was not defined. DB2 Information Integrator for Content might not be installed.

Verify that DB2 Information Integrator for Content was installed correctly.

**FFQD1035E: ERROR: The environment variable CMCOMMON is not defined. \nVerify that DB2 Information Integrator for Content was installed correctly.**

The environment variable CMCOMMON was not defined. DB2 Information Integrator for Content might not be installed.

Verify that DB2 Information Integrator for Content was installed correctly.

**FFQD1036E: ERROR: The environment variable ES\_NODE\_ROOT is not defined. \nVerify that DB2 Information Integrator OmniFind Edition was installed correctly and that this script was run by the enterprise search administrator user.**

The environment variable ES\_NODE\_ROOT was not defined.

Verify that DB2 Information Integrator OmniFind Edition was installed correctly and that this script was run by the enterprise search administrator user.

**FFQD1037E: ERROR: The environment variable ES\_INSTALL\_ROOT is not defined. \nVerify that DB2 Information Integrator OmniFind Edition was installed correctly and this script was run by the enterprise search administrator user.**

The environment variable ES\_INSTALL\_ROOT was not defined.

Verify that DB2 Information Integrator OmniFind Edition was installed correctly and that this script was run by the enterprise search administrator user.

**FFQD1101E: ERROR: Undefined platform.**

The setup script was executed on unsupported platform.

Ensure that the setup script is run on a supported platform.



**FFQD1106E: ERROR: Run this setup shell as enterprise search administrator user, not as the root user.**

The setup script must be run by the enterprise search administrator user.

Change the current user to the enterprise search user and run the setup script again.

**FFQD1107E: ERROR: ES\_CFG environment variable is not defined. Check that DB2 Information Integrator OmniFind Edition is correctly installed on this computer.**

ES\_CFG environment variable was not detected by setup script.

Confirm that DB2 Information Integrator OmniFind Edition is installed correctly and check that the ES\_CFG environment variable is set properly.

**FFQD1108E: ERROR: The enterprise search configuration file does not exist.**

The enterprise search configuration file es.cfg does not exist in node root (ES\_NODE\_ROOT) directory.

Check that DB2 Information Integrator OmniFind Edition is installed correctly and check that the enterprise search configuration file es.cfg exists in the node root directory.

**FFQD1123E: ERROR: Lotus Notes directory was not found.**

The Lotus Notes directory that was specified was not found in the crawler server.

Check that Lotus Notes (for Windows) or Lotus Domino Server (for UNIX) was properly installed on the crawler server.

**FFQD1128E: ERROR: Lotus Notes data directory was not found.**

The Lotus Notes data directory that was specified was not found on the crawler server.

Check that Lotus Notes directory was deployed on the crawler server properly.

**FFQD1138E: ERROR: Failed to remove *filename*. Remove it and run this script again.**

This script cannot delete the file *filename*.

Remove the file manually and run this script again.

**FFQD1139E: ERROR: Failed to create *filename*.**

This script cannot create the file *filename*.

Check file and directory privilege so that the file can be created.

**FFQD1142E: ERROR: *filename* was not found.**

This script cannot find the file *filename*.

Check that the file exists.

**FFQD1407E: ERROR: Venetica VeniceBridge directory was not found.**

The Venetica VeniceBridge directory that was specified was not found on the crawler server.

Check that the Venetica VeniceBridge was installed on the crawler server properly.

**FFQD1420E: ERROR: WebSphere Application Server directory was not found.**

WebSphere Application Server directory specified was not found on the crawler server.

Check that WebSphere Application Server was installed on the crawler server properly.

**FFQD2005E: The crawler server is not configured for the specified crawler type. Confirm that the crawler server is configured by the appropriate setup script.**

The crawler server is not configured for the crawler type you want to use.

Check that the crawler server is configured by the appropriate setup script properly and run the setup script.

**FFQD2007E: An error occurred while registering the *api\_name* API for operation *operation\_class*.**

The *api\_name* discovery API could not be registered because of an internal error.

Check that enterprise search and prerequisite software are correctly installed on the crawler server.

**FFQD2100E: The SQL exception occurred while accessing the database. Error code: *error\_message*. Error message: *error\_code*.**

The error occurred while accessing the DB2 database server.

See the IBM DB2 Universal Database Message Reference.

**FFQD2102E: User name or password is incorrect.**

The user name, password, or both that were used to establish a connection to the DB2 server were incorrect.

Try the operation again with correct user name and password.

**FFQD2103E: The following error occurred while executing the DB2 discovery API. Return code: *return\_code*. Error message: *error\_message*.**

The error occurred while discovering the DB2 database server.

See IBM DB2 Universal Database Message Reference.

**FFQD2104E: A connection cannot be established.**

A connection to the specified DB2 database server failed.

Try the operation again. Verify that the DB2 server is available.

**FFQD2105E: The table or view *table\_or\_view\_name* was not found for schema name *schema\_name*.**

The discovery process could not discover the columns because the specified table or view does not exist in the DB2 database server.

Specify the table or view that exists in the DB2 database server.

**FFQD2106E: The DB2 instance *server\_name:port\_number* was not found.**

The specified DB2 instance was not found.

Check that the server name and port number to identify the DB2 instance are correct.

**FFQD2130E: A Notes exception occurred while accessing the server. ID: *error\_id*. Error message: *error\_message***

The error occurred while accessing the Domino server.

See the error message for more information.

**FFQD2131E: The specified ID file does not exist or cannot be read.**

The specified user ID file does not exist or you do not have permission to read the file.

Verify that the user ID file path that you entered is correct and that you can read the file.

**FFQD2132E: A Notes session cannot be created.**

The discovery process could not create a Notes session to the Domino server.

See the log files for more information about where the problem occurred.

**FFQD2133E: Other Notes crawlers are running. They must be stopped because they are holding the lock *lock\_name*. Path: *lock\_filepath*. Name: *lock\_filename*.**

The discovery process cannot coexist with other Notes crawler processes. One of them was holding the lock *lock\_name*.

Stop the Notes crawler and try again.

**FFQD2135E: The crawler server is not configured for NRPC. Check that the crawler server is configured by the appropriate setup script.**

The correct Notes Java class for NRPC (Notes Remote Procedure Call) could not be loaded. A setup procedure was required for the Notes crawler.

Perform the setup procedure for the Notes crawler.

**FFQD2160E: The directory *directory name* is not valid.**

The specified directory name was not found.

Enter an existing directory name.

**FFQD2161E: There are more than *levels of subdirectories to discover* available subdirectories. To reduce this number, specify a Main directory name or pattern or specify fewer levels of subdirectories to discover.**

There are more than *levels of subdirectories to discover* available subdirectories.

To reduce this number, specify a main directory name or pattern or specify fewer levels of subdirectories to discover.

**FFQD2162E: The specified name *directory name* is not a directory name but a file name.**

The specified name was a file name.

Enter an existing directory name.

**FFQD2163E: The specified pattern for directory *directory name or pattern* is not valid.**

The specified pattern to search is not correct.

Enter the correct pattern to search directories. To type a pattern, use an asterisk (\*) as a wildcard character to replace one or more characters in the name.

**FFQD2164E: An internal error occurred while discovering file directories.**

The specified directory on the crawler server was not found.

Verify that enterprise search was successfully installed and try again.

**FFQD2190E: The network path for directory *network path name* was not found (directory=*network path name*). Error message: *detail error message*.**

The specified network path was not found.

Enter the correct pattern to search directories. To type a pattern, use an asterisk (\*) as a wildcard character to replace one or more characters in the name.

**FFQD2191E: The specified drive letter is not valid.**

The specified drive letter or the specified file separator is not valid.

Enter the correct drive letter to search directories.

**FFQD2210E: The user ID, password, or both are incorrect. The specified user ID is *user\_id*.**

The user ID, password, or both that were entered to connect to the Content Manager server were incorrect.

Try the operation again with the correct user ID and password.

**FFQD2211E: The connection to the Content Manager server *server* failed.**

A connection to the specified Content Manager server failed.

Try the operation again. Verify that the Content Manager server is available.

**FFQD2212I: Connected to Content Manager server *server*.**

The crawler server successfully connected to the Content Manager server.

This message is for your information only. No action is required.

**FFQD2213E: CMCOMMON keyword is not found in discovery.properties file.**

The discovery process could not discover the Content Manager servers because there was no CMCOMMON keyword in discovery.properties file under ES\_NODE\_ROOT/master\_config/discovery. The Content Manager crawler might not be configured by the appropriate setup script.

Configure the Content Manager crawler with the appropriate setup script and try the operation again.

**FFQD2214E: CM INI file *file* was not found.**

The discovery process could not discover the Content Manager servers because the CM INI file was not found. DB2 Information Integrator for Content might not be configured correctly.

Verify that the DB2 Information Integrator for Content is installed and configured correctly.

**FFQD2215E: Cannot read the CM INI file, or the CM INI file *file* was not found.**

The discovery process could not discover the Content Manager servers because CM INI file was not found. DB2 Information Integrator for Content might not be configured correctly.

Verify that the DB2 Information Integrator for Content is installed and configured correctly.

**FFQD2216E: The item type *item\_type* does not exist.**

The discovery process could not discover the Content Manager attributes because the specified item type does not exist on the Content Manager server.

Specify an item type that exists on the Content Manager server.

**FFQD2217E: CM DKException occurred.**

An exception occurred while discovering the Content Manager server.

See IBM DB2 Content Manager for Multiplatforms/IBM DB2 Information Integrator for Content for Multiplatforms, Messages and Codes for more information.

### **FFQD2218E: CM Exception occurred.**

An exception occurred while discovering the Content Manager server.

See IBM DB2 Content Manager for Multiplatforms/IBM DB2 Information Integrator for Content for Multiplatforms, Messages and Codes for more information.

### **FFQD2240E: A truststore file must be specified to use SSL.**

You must specify a truststore file to use SSL.

Prepare a truststore file on the crawler server and enter its path.

### **FFQD2241E: A truststore password must be specified to use SSL.**

You must specify a truststore password for the file to use SSL.

Enter a password for the truststore file.

### **FFQD2242E: The truststore file *truststore\_file* does not exist.**

The specified truststore file did not exist or you did not have permission to read the file.

Verify that the inputted truststore file path is correct and you can read the file.

### **FFQD2243E: An error occurred while validating the truststore file *truststore\_file*.**

The specified truststore file did not exist or you did not have permission to read the file.

Verify that the inputted truststore file path is correct and you can read the file.

### **FFQD2244E: The truststore file format or password is incorrect. The specified truststore is *truststore\_file*.**

The truststore file format is incorrect. The supported formats are JKS, JCEKS and PKCS#12.

Verify that the truststore file format.

### **FFQD2245E: An unknown host name *host\_name* was detected.**

The host name in the specified public folder URL was not found.

Verify that the host name is correct.

**FFQD2246E: The URL *public\_folder\_URL* is incorrect.**

The specified public folder URL was not found.

Verify that the public folder URL is correct.

**FFQD2247E: An I/O error occurred while connecting to *public\_folder\_URL*.**

One or more errors occurred while connecting to the public folder server.

Verify that network and the public folder server are working.

**FFQD2248E: A protocol exception occurred while sending a request to *public\_folder\_URL*.**

One or more errors occurred while sending a request to the public folder server.

See the log files for more information about where the problem occurred.

**FFQD2249E: An I/O error occurred while sending a request to *public\_folder\_URL*.**

One or more errors occurred while sending a request to the public folder server.

See the log files for more information about where the problem occurred.

**FFQD2250E: The number of folders (*actual\_folder\_count*) exceeds the limit *folder\_count\_limit*.**

The number of found subfolders must be the limit or less.

Change the pattern for finding subfolders to reduce the number of folders.

**FFQD2251E: The user ID, password, or both are incorrect. The specified user name is *user\_name*.**

An incorrect user ID, an incorrect password or both were specified.

Verify the user name and the password and try again.

**FFQD2252E: The specified folder *public\_folder\_URL* does not exist.**

The specified public folder server existed, but the folder did not exist on the server.

Verify the public folder URL is correct and the public folder server is working correctly.



**FFQD2253E: An unexpected HTTP response was received. HTTP State Code: *HTTP\_state\_code*. HTTP Response Message: *HTTP\_response\_message*.**

The public folder server returned an unexpected HTTP response.

Verify the public folder server is working correctly and see the log files for more information where the problem occurred.

**FFQD2254E: An I/O error occurred while receiving a response from *public\_folder\_URL*.**

One or more errors occurred while receiving a response from the public folder server.

See the log files for more information about where the problem occurred.

**FFQD2255E: The server response is incorrect XML.**

The public folder server returned a XML document which might contain incorrect elements and characters.

Verify the public folder does not contain broken documents.

**FFQD2256E: An I/O error occurred while parsing the received XML.**

One or more errors occurred while receiving a response from the public folder server.

See the log files for more information about where the problem occurred.

**FFQD2257E: An XML parser error occurred.**

Could not create a XML parser for parsing responses from the public folder server.

Verify that DB2 Information Integrator OmniFind Edition (enterprise search) was successfully installed and try again.

**FFQD2258E: The specified URL *public\_folder\_URL* is valid, but it is not a public folder.**

The specified URL can be accessed from the crawler, but it is not a public folder.

Verify that the public folder server is working correctly and Outlook Web Access is also working on the server.

**FFQD2270E: The specified JNDI property is not valid.**

The discovery process could not connect to the VeniceBridge server with the specified JNDI properties.

Check the specified JNDI properties are correct and the VeniceBridge server is available.

**FFQD2271E: The specified repository name is not valid.**

The discovery process could not discover the item classes since the specified repository did not exist in the VeniceBridge server.

Specify the repository that exists on the VeniceBridge server.

**FFQD2272E: The specified user name or password is not valid.**

The user name, password, or both to logon to the repository were incorrect.

Try the operation again with correct user name and password.

**FFQD2273E: The specified item class name is not valid.**

The discovery process could not discover the properties because the specified item class did not exist in the VeniceBridge server.

Specify the item class that exists on the VeniceBridge server.

**FFQD2274E: A VeniceBridge API error occurred. Error message: *error\_message*.**

The error occurred while accessing to the VeniceBridge server.

See the error message from the VeniceBridge server for more information.

**FFQD2275E: The VeniceBridge server instance cannot be created. Error message: *error\_message*.**

The error occurred while accessing to the VeniceBridge server.

See the error message from the VeniceBridge server for more information.

**FFQD2276E: The VeniceBridge server cannot be located. Error message: *error\_message*.**

The error occurred while accessing to the VeniceBridge server.

See the error message from the VeniceBridge server for more information.

**FFQD2277E: A communication error occurred. Error message: *error\_message*.**

The error occurred while accessing to the VeniceBridge server.

See the error message from the VeniceBridge server for more information.

**FFQD2278E: The specified repository *repository\_name* was not found. Error message: *error\_message*.**

The specified repository did not exist in the VeniceBridge server.

Specify the repository which exists in the VeniceBridge server.

**FFQD2279E: The VeniceBridge user instance was not initialized. Error message: *error\_message*.**

The error occurred while accessing to the VeniceBridge server.

See the error message from the VeniceBridge server for more information.

**FFQD2280E: Logon to *repository\_name* repository failed. Error message: *error\_message*.**

The discovery process could not log on to the *repository\_name* repository.

See the error message from the VeniceBridge server for more information.

**FFQD2282E: The specified item class *item\_class\_name* was not found. Repository: *repository\_name***

The *item\_class\_name* item class did not exist in *repository\_name* repository in the VeniceBridge server.

Specify the item class which exists in the repository.

**FFQD3000I: *crawler\_name* crawler started successfully.**

This message is for your information only. No action is required.

**FFQD3001E: *crawler\_name* crawler failed to start.**

The crawler process failed to be started.

See the log files for more information about where the problem occurred.

**FFQD3002I: *crawler\_name* crawler stopped successfully.**

This message is for your information only. No action is required.

**FFQD3003E: *crawler\_name* crawler did not stop successfully.**

The crawler process failed to be stopped successfully.

See the log files for more information about where the problem occurred.

**FFQD3004I: *crawler\_name* crawler was paused successfully..**

This message is for your information only. No action is required.

**FFQD3005E: *crawler\_name* crawler failed to pause.**

The crawler failed to be paused.

See the log files for more information about where the problem occurred.

**FFQD3006I: *crawler\_name* crawler resumed successfully.**

This message is for your information only. No action is required.

**FFQD3007E: *crawler\_name* crawler failed to resume.**

The crawler failed to be resumed.

See the log files for more information about where the problem occurred.

**FFQD3008E: Factory exception occurred. Verify that the modules are installed properly.**

The installed jar files were not found or were not set correctly.

Verify that enterprise search was successfully installed and try again.

**FFQD3009I: Start the crawl for *crawler\_name* on *server\_name*.**

This message is for your information only. No action is required.

**FFQD3010I: Finished the crawl for *crawler\_name* on *server\_name*.**

This message is for your information only. No action is required.

**FFQD3011E: Failed to stop for *target\_host\_name* on *target\_name*.**

The crawler failed to stop crawling for the specified target.

See the log files for more information about where the problem occurred.

**FFQD3012I: The crawler is already crawling on server *server\_name*.**

This message is for your information only. No action is required.

**FFQD3013W: The crawled data was truncated. URI: *URI*. Server: *target\_host\_name*. Target: *target\_name*.**

The crawled data was truncated.

Change the maximum page size for a crawler on the administration console if all the data needs to be crawled.

**FFQD3014E: An exception occurred while handling crawled history.**

The internal error occurred in the crawler.

See the log files for more information about where the problem occurred.

**FFQD3015E: An exception occurred while handling crawled history. URI: *uniform\_resource\_identifier*. Server: *server\_name*. Target: *crawler\_name*.**

The internal error occurred in the crawler.

See the log files for more information about where the problem occurred.

**FFQD3016E: An exception occurred while accessing the database.**

The internal error occurred in the crawler.

See the log files for more information about where the problem occurred.

**FFQD3017E: An exception occurred while accessing the database. URI: *uniform\_resource\_identifier*. Server: *server\_name*. Target: *crawler\_name*.**

The internal error occurred in the crawler.

See the log files for more information about where the problem occurred.

**FFQD3018E: an exception occurred while handling the crawler configuration.**

The internal error occurred in the crawler.

See the log files for more information about where the problem occurred.

**FFQD3019I: The server schedule is set for *server\_name* server.**

This message is for your information only. No action is required.

**FFQD3020I: The schedule is disabled for *server\_name* server.**

This message is for your information only. No action is required.

**FFQD3021W: The schedule for specified server *server\_name* is already set.**

The specified scheduled crawl has already been enabled.

This message is for your information only. No action is required.

**FFQD3022W: The schedule for specified server *server\_name* is already disabled.**

The specified scheduled crawl has already been disabled.

This message is for your information only. No action is required.

**FFQD3023W: The server *server\_name* is not scheduled.**

The specified crawl was not scheduled.

To enable or disable the crawler schedule, the crawl need to be scheduled in advance.

**FFQD3024E: The return code of fetch is incorrect. Return code: *return\_code*. State: *state***

The internal error occurred in the crawler.

See the log files for more information about where the problem occurred.

**FFQD3025E: An exception occurred while fetching data.**

The internal error occurred in the crawler.

See the log files for more information about where the problem occurred.

**FFQD3026E: An exception occurred while processing thread *thread\_id*.**

The internal error occurred in the crawler.

See the log files for more information about where the problem occurred.

**FFQD3027E: An exception occurred for the crawler.**

The internal error occurred in the crawler.

See the log files for more information about where the problem occurred.

**FFQD3028E: The specified server *server\_name* is not found.**

The internal error occurred in the crawler.

See the log files for more information about where the problem occurred.

**FFQD3029E: an exception occurred while initializing the connection manager.**

The internal error occurred during initialization of the crawler.

See the log files for more information about where the problem occurred.

**FFQD3030E: An exception occurred while initializing the state manager.**

The internal error occurred during initialization of the crawler.

See the log files for more information about where the problem occurred.

**FFQD3031E: An exception occurred while crawler status was being requested.**

The internal error occurred in the crawler. The monitor function did not work correctly.

See the log files for more information about where the problem occurred.

**FFQD3032E: An exception occurred while the crawler space status was being requested.**

The internal error occurred in the crawler. The monitor function did not work correctly.

See the log files for more information about where the problem occurred.

**FFQD3033E: An exception occurred while the crawl space status was being requested.**

The internal error occurred in the crawler. The monitor function did not work correctly.

See the log files for more information about where the problem occurred.

**FFQD3034I: Deleting obsolete documents.**

This message is for your information only. No action is required.

**FFQD3035W: Interrupted to delete obsolete documents. Will try again next time.**

This message is for your information only. No action is required.

**FFQD3036I: Finished deleting obsolete documents for *crawler\_name* on *server\_name*.**

This message is for your information only. No action is required.

**FFQD3037E: An exception occurred while updating internal status.**

The internal error occurred in the crawler.

See the log files for more information about where the problem occurred.

**FFQD3038E: An error occurred because of inconsistency of status. Information: *additional\_data***

The internal error occurred in the crawler.

See the log files for more information about where the problem occurred.

**FFQD3039E: An error occurred while accessing internal status. Information: *additional\_data***

The internal error occurred in the crawler.

See the log files for more information about where the problem occurred.

**FFQD3040E: An exception occurred while deleting obsolete records.**

The internal error occurred in the crawler.

See the log files for more information about where the problem occurred.

**FFQD3041I: The schedule component started successfully.**

The internal error occurred in the crawler.

See the log files for more information about where the problem occurred.

**FFQD3042E: The schedule component failed to start.**

The internal error occurred in the crawler.

See the log files for more information about where the problem occurred.

**FFQD3043I: The schedule component stopped successfully.**

The internal error occurred in the crawler.

See the log files for more information about where the problem occurred.

**FFQD3044E: The schedule component failed to stop.**

The internal error occurred in the crawler.

See the log files for more information about where the problem occurred.



**FFQD3045I: Starting the scheduled crawl for *crawler\_name*.**

This message is for your information only. No action is required.

**FFQD3046E: An exception occurred in the schedule component.**

The internal error occurred in the crawler.

See the log files for more information about where the problem occurred.

**FFQD3047I: The internal repository was initialized successfully.**

A crawler initialized the crawler internal repository.

This message is for your information only. No action is required.

**FFQD3048E: An unknown internal repository type *internal\_repository\_type* was detected.**

One or more errors occurred while initializing the crawler internal repository.

Verify that enterprise search was successfully installed and try again.

**FFQD3049E: An internal repository initialization error.**

One or more errors occurred while initializing the crawler internal repository.

Verify that enterprise search was successfully installed and try again.

**FFQD3050E: Access violation.**

One or more errors occurred while initializing the crawler internal repository.

Verify that enterprise search was successfully installed and try again.

**FFQD3051E: Another repository is open.**

An another crawler internal repository was already opened and initialized.

Verify that enterprise search was successfully installed and try again.

**FFQD3052E: There is no database connection.**

The crawler internal repository was opened but a database connection for using the repository did not exist.

Verify that the DB2 database for enterprise search is correctly working and try again.

**FFQD3053E: A database error occurred while opening the internal repository.**

A database connection for using the crawler internal repository could not be established.

Verify that the DB2 database for enterprise search is correctly working and try again.

**FFQD3054E: A database error occurred while closing the internal repository.**

One or more errors occurred while closing the crawler internal repository.

See the SQLException messages in the log files for more information.

**FFQD3055E: A database error occurred while writing a record to the internal repository. The record hash is *record\_hash*.**

One or more errors occurred while writing information to the crawler internal repository.

See the SQLException messages in the log files for more information.

**FFQD3056E: A database error occurred while getting the last modified date. The record hash is *record\_hash*.**

One or more errors occurred while accessing to the crawler internal repository.

See the SQLException messages in the log files for more information.

**FFQD3057E: A database error occurred while updating the last modified date. The record hash is *record\_hash*.**

One or more errors occurred while writing information to the crawler internal repository.

See the SQLException messages in the log files for more information.

**FFQD3058E: A database error occurred while detecting a record change. The record hash is *record\_hash*.**

One or more errors occurred while accessing to the crawler internal repository.

See the SQLException messages in the log files for more information.

**FFQD3059E: A database error occurred while getting checksum. The record hash is *record\_hash*.**

One or more errors occurred while accessing to the crawler internal repository.

See the SQLException messages in the log files for more information.

**FFQD3060E: An I/O error occurred while getting checksum. The record hash is *record\_hash*.**

One or more errors occurred while accessing to the crawler internal repository.

See the IOException messages in the log files for more information.

**FFQD3061E: A database error occurred while checking a record. The record hash is *record\_hash*.**

One or more errors occurred while accessing to the crawler internal repository.

See the SQLException messages in the log files for more information.

**FFQD3062E: A database error occurred while writing an excluded record. The record hash is *record\_hash*.**

One or more errors occurred while accessing to the crawler internal repository.

See the SQLException messages in the log files for more information.

**FFQD3063E: A database error occurred while deleting a record from the repository. The record hash is *record\_hash*.**

One or more errors occurred while accessing to the crawler internal repository.

See the SQLException messages in the log files for more information.

**FFQD3064E: A database error occurred while getting obsolete records. The target ID is *record\_hash*, and the threshold time is *threshold\_time*.**

One or more errors occurred while accessing to the crawler internal repository.

See the SQLException messages in the log files for more information.

**FFQD3065I: The internal repository was opened successfully.**

A crawler opened the crawler internal repository.

This message is for your information only. No action is required.

**FFQD3066I: The internal repository was closed successfully.**

A crawler closed the crawler internal repository.

This message is for your information only. No action is required.

**FFQD3067I: Loading crawler configuration files  
*general\_config\_file* and *crawler\_config\_file*.**

A crawler started loading the crawler configuration files.

This message is for your information only. No action is required.

**FFQD3068I: Loading of the configuration files completed  
successfully.**

A crawler finished loading the crawler configuration files.

This message is for your information only. No action is required.

**FFQD3069E: An error occurred while creating the  
configuration factory.**

A crawler finished loading the crawler configuration files.

This message is for your information only. No action is required.

**FFQD3070E: Loading of configuration file  
*general\_config\_file* failed.**

One or more errors occurred while loading the crawler configuration file.

See the exception messages in the log files for more information. typically the exception is the IOException.

**FFQD3071E: No configuration file was loaded.**

A crawler did not load configuration files.

Verify that the crawler which you want to start was configured correctly and try again.

**FFQD3072E: An unknown configuration object was detected.**

One or more errors occurred while loading the crawler configuration files.

Verify that enterprise search was successfully installed and try again.

### **FFQD3073E: Getting document builder failed.**

Could not create a XML parser for loading the crawler configuration files.

Verify that enterprise search was successfully installed and try again.

### **FFQD3074E: A configuration file format error was detected.**

A crawler configuration file is not valid a XML document.

See the log files for more information.

### **FFQD3075E: An I/O error occurred while parsing the configuration.**

One or more errors occurred while loading the crawler configuration files.

See the IOException messages in the log files for more information.

### **FFQD3076E: The crawler configuration file *crawler\_config\_file* was not found.**

There was no crawler configuration file on the expected path.

Verify that the crawler which you want to start was configured correctly and try again.

### **FFQD3077E: The expected the root element name in the crawler configuration file is *expected\_root\_element*, but *detected\_root\_element* was detected.**

The root element of the crawler configuration file had unexpected name.

Verify that the crawler which you want to run was configured correctly and try again.

### **FFQD3078E: Attribute *attribute\_name* is required in *element\_name* element.**

A required attribute was not found in the crawler configuration file.

Verify that the crawler which you want to run was configured correctly and try again.

### **FFQD3079E: Duplicate name *value\_of\_name\_attribute* was detected in element *element\_name*.**

The same target name was contained in the crawler configuration file.

Verify that the crawler which you want to run was configured correctly and try again.

**FFQD3080E: Incorrect date string format was detected in *element\_name* element.**

The crawler configuration file contained an incorrect date string.

Verify that the crawler which you want to run was configured correctly and try again.

**FFQD3081E: Incorrect interval string format was detected in *element\_name* element.**

The crawler configuration file contained an incorrect crawl interval time.

Verify that the crawler which you want to run was configured correctly and try again.

**FFQD3082E: *element\_name* element's value *actual\_value* is out of bounds.**

The crawler configuration file contained one or more incorrect values.

Verify that the crawler which you want to run was configured correctly and try again.

**FFQD3083W: The XML parser reported a warning.**

The crawler configuration file parser reported a XML warning.

No action is required.

**FFQD3084E: The XML parser reported an error.**

The crawler configuration file parser reported a XML error.

Verify that the crawler which you want to run was configured correctly and try again.

**FFQD3085E: The XML parser reported an unrecoverable error.**

The crawler configuration file parser reported an unrecoverable XML error.

Verify that the crawler that you want to run was configured correctly and try again.

**FFQD3086E: The formatter is not initialized.**

The formatter to generate XML metadata for search engine was not created.

See the log files for more information about where the problem occurred.

**FFQD3087E: Cannot convert to XML.**

The formatter does not generate XML metadata for search engine..

See the log files for more information about where the problem occurred.

**FFQD3088W: Parametric search is set to be enabled for a nonparametric field. The parametric search flag was disabled.**

Crawled field is not a number though crawler is specified this field as parametric searchable. The parametric search will be disabled for this field.

Specify this field as non-parametric field from the administrator console. See the online help for enterprise search to perform this operation.

**FFQD3089W: Metadata was truncated.**

Metadata was truncated due to the limitation of the max document size setting. It is not searchable by truncated metadata.

Increase "Maximum Page Length size" of the crawler from administration console to crawl larger metadata. See the help for enterprise search to perform this operation.

**FFQD3090W: Temporary file *filename* is not deleted.**

Temporary file *filename* is not deleted.

Check that temporary directory for the crawler has appropriate privilege to delete.

**FFQD3091W: The formatted data record is null. No action is required.**

No metadata or binary content fetched by crawler will write to enterprise search repository.

No action is required.

**FFQD3102E: Cannot open internal data store. DBName: *database\_name*. DBSchema: *schema\_name*. DBUser: *database\_user*. TableName: *table\_name*.**

Publisher cannot open internal database *database\_name* to store crawled data.

See the log files for more information about where the problem occurred.

**FFQD3103E: Cannot close internal data store. DBName:** *database\_name*. **DBSchema:** *schema\_name*. **DBUser:** *database\_user*. **TableName:** *table\_name*.

Publisher cannot close internal database *database\_name* to store crawled data.

See the log files for more information about where the problem occurred.

**FFQD3104E: Cannot write metadata to file *file\_name* because of FileNotFoundException.**

Publisher cannot open the temporary file that stored crawled data.

See the log files for more information about where the problem occurred.

**FFQD3105E: Cannot write to internal data store. DBName:** *database\_name*. **DBSchema:** *schema\_name*. **DBUser:** *database\_user*. **TableName:** *table\_name*.

Publisher cannot write the information crawled by crawler into internal database *database\_name*.

See the log files for more information about where the problem occurred.

**FFQD3106E: Cannot write to internal data store. DBName:** *database\_name*. **DBSchema:** *schema\_name*. **DBUser:** *database\_user*. **TableName:** *table\_name*.

Publisher cannot write the information crawled by crawler into internal database *database\_name*.

See the log files for more information about where the problem occurred.

**FFQD3107E: Cannot update by deletion flag to internal data store. DBName:** *database\_name*. **DBSchema:** *schema\_name*. **DBUser:** *database\_user*. **TableName:** *table\_name*.

Publisher cannot mark deletion flag to internal table to delete information from search index. Information about nonexistent document in target datasource will remain in search index.

See the log files for more information about where the problem occurred.

**FFQD3108E: Cannot update by deletion flag to internal data store. DBName:** *database\_name*. **DBSchema:** *schema\_name*. **DBUser:** *database\_user*. **TableName:** *table\_name*.

Publisher cannot mark deletion flag to internal table to delete information from search index. Information about nonexistent document in target datasource will remain in search index.



See the log files for more information about where the problem occurred.

**FFQD3109W: The crawled file *file\_name* is updated by other programs during updating internal data store.**

Crawled file *file\_name* is modified because the file crawled by crawler.

No action is required.

**FFQD3110E: Cannot close file stream.**

The filestream for crawled data stored in temporary file is not closed due to an error.

See the log files for more information about where the problem occurred.

**FFQD3111E: Password decryption failed.**

The password decryption module reported one or more errors.

See the log files for more information about where the problem occurred.

**FFQD3500E: JDBC class *jdbc\_driver\_class* is not found.**

JDBC driver class *jdbc\_driver\_class* was not found in the classpath.

Verify that enterprise search was successfully installed and try again.

**FFQD3501W: Exceptions occurred while altering *table\_name* table.**

One or more exceptions occurred while trying to alter *table\_name* table.

See the exceptions (typically SQLException from DB2) to determine the actual cause of this error.

**FFQD3502W: Exceptions occurred while creating *table\_name* table.**

One or more exceptions occurred while trying to create *table\_name* table.

See the exceptions (typically SQLException from DB2) to determine the actual cause of this error.

**FFQD3503W: Transaction rollback failed.**

One or more exceptions occurred while trying to rollback the transaction.

See the exceptions (typically SQLException from DB2) to determine the actual cause of this error.

### **FFQD3504W: Transaction commit failed.**

One or more exceptions occurred while trying to commit the transaction.

See the exceptions (typically SQLException from DB2) to determine the actual cause of this error.

### **FFQD3505W: Renaming table *orig\_table\_name* to table *dest\_table\_name* failed. Crawler will try to copy and drop it instead.**

One or more exceptions occurred which trying to rename *orig\_table\_name* table to *dest\_table\_name* table. Crawler will try to create a new table, copy all the data into it, and drop the original one instead.

No action is required.

### **FFQD3506W: Exceptions occurred while deleting all data from *table\_name* table.**

One or more exceptions occurred while trying to delete all the data from *table\_name* table.

See the exceptions (typically SQLException from DB2) to determine the actual cause of this error.

### **FFQD3507W: Exceptions occurred while inserting data into *table\_name* table.**

One or more exceptions occurred while trying to insert data into *table\_name* table.

See the exceptions (typically SQLException from DB2) to determine the actual cause of this error.

### **FFQD3508W: Exceptions occurred while dropping *table\_name* table.**

One or more exceptions occurred while trying to drop *table\_name* table.

See the exceptions (typically SQLException from DB2) to determine the actual cause of this error.

### **FFQD3509W: This document is rejected because the URI *URI* for *table\_name* table (*uri\_length* byte) is too long. Maximum length is *maximum\_length* byte.**

The crawled document was not stored into internal table because of too long URI *URI*.

No action is required.

**FFQD3510W: This document is rejected because it is too large for *table\_name* table (URI : *URI*, *document\_size* byte). Maximum length is *maximum\_size* byte.**

The crawled document was not stored into internal table because of too large document.

No action is required.

**FFQD3511W: This document is rejected because it has too much metadata for *table\_name* table (URI : *URI*, *metadata\_size* byte). Maximum length is *maximum\_size* byte.**

The crawled document was not stored into an internal table because the metadata was too extensive.

No action is required.

**FFQD3512W: An unexpected row count was returned while inserting into *table\_name* table. Expected count was *expected\_count*, but it was *actual\_count*.**

While inserting data into *table\_name* table, it could not be inserted because of expected count of inserted rows was not equal to actual count.

No action is required. If this message appears again, verify that enterprise search was successfully installed and try again.

**FFQD3513W: An unexpected row count was returned while updating *table\_name* table. Expected count was *expected\_count*, but it was *actual\_count*.**

While updating data into *table\_name* table, it could not be updated because of expected count of updated rows was not equal to actual count.

If this message appears again, verify that enterprise search was successfully installed and try again.

**FFQD3514W: An unexpected row count was returned while deleting from *table\_name* table. Expected count was *expected\_count*, but it was *actual\_count*.**

While deleting data from *table\_name* table, it could not be deleted because of expected count of deleted rows was not equal to actual count.

If this message appears again, verify that enterprise search was successfully installed and try again.

**FFQD3515W: Exceptions occurred while querying *table\_name* table.**

One or more exceptions occurred while trying to execute select statement against *table\_name* table.

See the exceptions (typically SQLException from DB2) to determine the actual cause of this error.

**FFQD3516W: Exceptions occurred while inserting into *table\_name* table.**

One or more exceptions occurred while trying to executing insert statement against *table\_name* table.

See the exceptions (typically SQLException from DB2) to determine the actual cause of this error.

**FFQD3517W: Exceptions occurred while updating *table\_name* table.**

One or more exceptions occurred while trying to execute update statement against *table\_name* table.

See the exceptions (typically SQLException from DB2) to determine the actual cause of this error.

**FFQD3518W: Exceptions occurred while deleting from *table\_name* table.**

One or more exceptions occurred while trying to execute delete statement against *table\_name* table.

See the exceptions (typically SQLException from DB2) to determine the actual cause of this error.

**FFQD3519W: Exceptions occurred while preparing query *table\_name* table. SQL is *sql\_statement*.**

One or more exceptions occurred while preparing select statement against *table\_name* table.

See the exceptions (typically SQLException from DB2) to determine the actual cause of this error.

**FFQD3520W: Exceptions occurred while preparing insert into *table\_name* table. SQL is *sql\_statement*.**

One or more exceptions occurred while preparing insert statement against *table\_name* table.

See the exceptions (typically SQLException from DB2) to determine the actual cause of this error.

**FFQD3521W: Exceptions occurred while preparing update *table\_name* table. SQL is *sql\_statement*.**

One or more exceptions occurred while preparing update statement against *table\_name* table.

See the exceptions (typically SQLException from DB2) to determine the actual cause of this error.

**FFQD3522W: Exceptions occurred while preparing delete from *table\_name* table. SQL is *sql\_statement*.**

One or more exceptions occurred while preparing delete statement for *table\_name* table.

See the exceptions (typically SQLException from DB2) to determine the actual cause of this error.

**FFQD3523W: The connection manager was already initialized.**

One of the internal manager which managing database connection has already been initialized.

No action is required.

**FFQD3524E: JNDI look-up failed for *jndi\_name*.**

JNDI lookup was failed for *jndi\_name*.

See exceptions (if exists) and make sure the JNDI name *jndi\_name* is correct and available.

**FFQD3534E: Exceptions occurred while *component\_name* was changing state from *from\_state* to *to\_state*.**

One or more exceptions occurred while changing the internal state form *from\_state* to *to\_state*.

See other log messages and exceptions to determine the actual cause of this error.

**FFQD3540E: Failed to get file lock *file\_name* because of IOException.**

Failed to get file lock for crawler that uses file lock facility.

See the log files for more information about where the problem occurred.

**FFQD3541E: Failed to get file lock *file\_name* because of FileNotFoundException.**

Failed to get file lock for crawler that uses file lock facility.

See the log files for more information about where the problem occurred.

**FFQD3543E: Failed to release file lock *file\_name*.**

Failed to release file lock for crawler that uses file lock facility.

See the log files for more information about where the problem occurred.

**FFQD3544E: Failed to release file lock *file\_name* because file lock is null.**

Failed to release file lock for crawler because no lock was obtained.

See the log files for more information about where the problem occurred.

**FFQD3545E: Algorithm *algorithm\_name* for loading truststore file was not found.**

The algorithm for loading truststore file was not contained in the JRE.

Verify that DB2 Information Integrator OmniFind Edition (enterprise search) was successfully installed and try again.

**FFQD3546E: No certificate was loaded.**

The specified truststore file did not contain a suitable certificate.

Verify that the truststore file contains a valid certificate and try again.

**FFQD3547E: Unknown keystore type *keystore\_type* is detected.**

An unexpected truststore file type was detected.

Verify that DB2 Information Integrator OmniFind Edition (enterprise search) was successfully installed and try again.

**FFQD3548E: Truststore file format or password is not valid. The specified truststore file is *truststore\_file*.**

The truststore file format or, and the specified password for the truststore file might be incorrect. The supported formats are JKS, JCEKS and PKCS#12.

Verify that the truststore file format and the password are correct and try again.

**FFQD3549E: An I/O error occurred while closing the truststore file *truststore\_file*.**

One or more I/O errors occurred while closing the truststore file.

Verify that the truststore file can be accessed and try again.

**FFQD3550E: Failed to initialize javax.net.ssl.TrustManager.**

One or more errors occurred while initializing the javax.net.ssl.TrustManager object.

See the exception messages in the log files for more information.

**FFQD3551E: Failed to initialize javax.net.ssl.SSLContext.**

One or more errors occurred while initializing the javax.net.ssl.SSLContext object.

See the exception messages in the log files for more information.

**FFQD3553E: Exceptions occurred while creating connection to the database *database\_name*.**

One or more exceptions occurred while creating connection to the database *database\_name*.

See the exceptions (typically SQLException from DB2) to determine the actual cause of this error.

**FFQD4000E: Notes class is not found. Setup for Notes Crawler is required.**

Necessary Notes Java class cannot be loaded. Setup procedure is required for Notes crawler.

Perform setup procedure to work Notes crawler.

**FFQD4001E: Notes NRPC class (Notes.jar) is not found. Setup for Notes NRPC crawler is required.**

Necessary Notes Java class for NRPC (Notes Remote Procedure Call) crawl cannot be loaded. Setup procedure is required for Notes crawler.

Perform setup procedure to work Notes crawler.

**FFQD4002E: Notes native library *library\_name* for NRPC access was not found. Setup for Notes NRPC crawler is required.**

Necessary Notes library for NRPC( Notes Remote Procedure Call) crawl cannot be loaded. Setup procedure is required for Notes crawler.

Perform setup procedure to work Notes crawler.

**FFQD4003E: Notes DIIOP class (NCS0.jar) is not found. Setup for Notes DIIOP crawler is required.**

Necessary Notes Java class for DIIOP (Domino Internet Inter-ORB Protocol) crawl cannot be loaded. Setup procedure is required for Notes crawler.

Perform setup procedure to work Notes crawler.

**FFQD4004E: Cannot determine Notes library name from detected platform *platform*.**

Necessary Notes library name could not be determined for NRPC (Notes Remote Procedure Call) crawl.

Ensure that the Notes crawler runs on the supported platform.

**FFQD4005E: Information for server, target, or both were not specified.**

Domino server or target was not set properly.

Configure the Notes crawler setting from administration console. See Administering Enterprise Search.

**FFQD4006E: Server name *server\_string* in the configuration file does not represent the Lotus Notes server and database.**

Server string *server\_string* does not conform the format like:  
"server\_name!!database\_name."

Configure the Notes crawler setting from administration console. See Administering Enterprise Search.

**FFQD4007E: Server name *server\_string* in the configuration file does not represent Lotus Notes database *database\_name*. Database name must be specified.**

Server string *server\_string* does not contain database name.

Configure the Notes crawler setting from administration console. See Administering Enterprise Search



**FFQD4008E: Connection to the database *database\_name* on the server *server\_name* is already active.**

The connection to the server *server\_name* is already established and the crawler tried to connect to the server without disconnecting.

See the log files for more information about where the problem occurred.

**FFQD4009E: Cannot connect to the Lotus Notes database *database\_name* on server *server\_name*.**

the Notes crawler cannot connect to the server because of the invalid crawler setting or server setting.

See the log files for more information about where the problem occurred. Check that the target Domino server and target database can be accessed by the crawler server. For example, check whether the Domino server is unavailable or whether the database access control list was changed and set not to crawl.

**FFQD4010E: Cannot change the ID to *id\_file\_name*.**

The Notes crawler cannot change the ID to *id\_file\_name*.

See the log files for more information about where the problem occurred.

**FFQD4011E: Cannot disconnect from Lotus Notes database *database\_name* on the server *server\_name*.**

The Domino server or the database could not disconnect.

See the log files for more information about where the problem occurred.

**FFQD4012E: Cannot change the ID to *id\_file\_name*.**

The Notes crawler cannot change the ID to *id\_file\_name*.

See the log files for more information about where the problem occurred.

**FFQD4013E: Cannot verify next record existence. Server: *server\_name*. Database: *database\_name*. View/Folder: *view\_folder\_name*.**

The Notes crawler cannot verify whether the Notes document exists while the crawler fetches documents from the specified Notes database.

See the log files for more information about where the problem occurred.

**FFQD4014E: Cannot change the ID to *id\_file\_name*.**

Notes crawler cannot change the ID to *id\_file\_name*.

See the log files for more information about where the problem occurred.

### **FFQD4015E: Cannot create the URI.**

An error occurred while creating a uniform resource identifier (URI).

See the log files for more information about where the problem occurred.

### **FFQD4016E: Cannot gather the unchanged attachment names since the last crawl.**

An error occurred while gathering file names from unchanged documents in the Notes database.

See the log files for more information about where the problem occurred.

### **FFQD4017E: Cannot fetch the target content. Server: *server\_name*. Database: *database\_name*. View/Folder: *view\_folder\_name*.**

An error occurred while fetching the target content from database *database\_name* on the server *server\_name*.

See the log files for more information about where the problem occurred.

### **FFQD4018E: Undefined record status *status\_code* was found in the internal repository.**

An unexpected status code *status\_code* was returned from an enterprise search internal table.

See the log files for more information about where the problem occurred.

### **FFQD4019E: Cannot obtain the attachment file name.**

An error occurred while obtaining the attachment file name.

See the log files for more information about where the problem occurred.

### **FFQD4020E: Cannot write to the temporary file *filename*.**

An error occurred while writing content to the temporary file.

Check that the appropriate privilege is given to temporary directory where the temporary file will be generated.

### **FFQD4021E: Cannot fetch the attachment file.**

An error occurred while fetching attachment file from Notes document.

See the log files for more information about where the problem occurred. Check Domino server availability.

**FFQD4022E: An I/O error occurred while reading or writing binary content.**

An I/O error occurred while reading or writing binary content.

See the log files for more information about where the problem occurred. Check the directory capacity of temporary file.

**FFQD4023E: Undefined record status *status\_code*.**

An unexpected status code *status\_code* was returned from enterprise search internal table.

See the log files for more information about where the problem occurred.

**FFQD4024E: Cannot evaluate the string length because of an I/O error.**

Notes crawler cannot evaluate string length because of an I/O error.

See the log files for more information about where the problem occurred.

**FFQD4025W: Cannot obtain last modified date from the Lotus Notes document. 0 will be used instead.**

Fetches Date/Time field does not have a valid value. 0 is used as this field value.

No action is required.

**FFQD4026W: Cannot obtain the creation date from the Lotus Notes document. 0 will be used instead.**

Fetches Date/Time field does not have a valid value. 0 is used as this field value.

No action is required.

**FFQD4027W: Cannot obtain the security field value.**

Specified field as security value does not exist or does not have any value.

No action is required.

**FFQD4028W: Cannot obtain file name for the data source metadata.**

Specified field as security value does not exist or does not have any value.

No action is required.

**FFQD4029W: Cannot obtain last modified date from the Lotus Notes document. 0 will be used instead.**

Fetches Date/Time field does not have a valid value. 0 is used as this field value.

No action is required.

**FFQD4030W: Cannot obtain creation date from the Lotus Notes document. 0 will be used instead.**

Fetches Date/Time field does not have a valid value. 0 is used as this field value.

No action is required.

**FFQD4031E: The connection is already active for server *server\_name* using user *user\_name*.**

The connection is already active for the crawler.

See the log files for more information about where the problem occurred.

**FFQD4032E: Access to the server *server\_name* is denied. Check the server setting that allows Java/CORBA access for user *user\_name*.**

Server refuses the access from crawler user.

Configure Domino server so that crawler can have access to specified target database. See Administering Enterprise Search or Lotus Domino Administrator Help.

**FFQD4033E: User ID *user\_name*, password, or both are not valid for server *server\_name*.**

Server refuses accessing crawler due to invalid user name or password.

Configure a valid user ID and password that is specified in Notes crawler setting from administration console.

**FFQD4034E: Cannot create a Notes session for server *server\_name* by user *user\_name*.**

The Notes crawler cannot access the Domino server and its database due to the invalid setting of the crawler or the Domino server is unavailable.

Check the Notes crawler setting (especially server name, user ID and its password) from the administration console. Check the target Domino server; HTTP and DIIOP task, target server and database privilege to the DIIOP crawler user, and so on. See Administering Enterprise Search.

**FFQD4035E: An error occurred while connecting to database *database\_name* on server *server\_name*. Check that you can access the database and that the crawl target is available to user *user\_name*.**

The Notes crawler cannot access the Domino server and its database due to the invalid setting of the crawler or Domino server.

Check the Notes crawler setting (especially the server name, user ID, and its password) from the administration console. Check the target Domino server condition; HTTP and DIIOP task, target server and database privilege to the DIIOP crawler user, so that the Notes crawler can access the target database. See *Administering Enterprise Search*, *Lotus Domino Designer Help* or *Lotus Domino Administrator Help*.

**FFQD4036E: Cannot connect to database *database\_name* on server *server\_name*. Check that you can access the database and that the crawl target is available to user *user\_name*.**

Check the Notes crawler setting (especially the server name, user ID, and its password) from the administration console. Check the target Domino the server; HTTP and DIIOP task, target server, and database privilege to the DIIOP crawler user so that the Notes crawler can access the target database. See *Administering Enterprise Search*, *Lotus Domino Designer Help* or *Lotus Domino Administrator Help*.

**FFQD4037E: Cannot select formula-selected documents. Check that you can access the documents and that the formula is based on a Notes formula. Server: *server\_name*. User: *user\_name*. Database: *database\_name*. Formula: *formula*.**

The specified formula is not valid as a Notes formula. Or the Domino server or database was denied access to the crawler.

Check and modify the specified Notes formula from the administration console. Check the crawler, Domino server, and target database so that crawler can access the target database properly. See *Administering Enterprise Search* or *Lotus Domino Designer Help*.

**FFQD4038E: An error occurred while getting View or DocumentCollection *view\_or\_folder\_name* from database *database\_name* on server *server\_name*. Check that you can access the database and that the crawl target is available to specified user *user\_name*.**

Specified target (Notes View or Folder) is not valid (target view or folder does not exist) or appropriate privilege is not given to the target to read document from crawler.

Specify the appropriate view or folder as a crawling target from administration console to read documents. Change the privilege of the Notes View or Folder so that the crawler can read it.

**FFQD4039E: Cannot get View or DocumentCollection from database *database\_name* on server *server\_name*. Check that you can access the database and that the crawl target is available to user *user\_name*. View/Folder: *view\_or\_folder\_name*.**

Specified target (Notes View or Folder) is not valid (target view or folder does not exist) or appropriate privilege is not given to the target to read document from crawler.

Specify appropriate view or folder as a crawling target from Administration console to read documents. Change privilege of the Notes View or Folder so that the crawler can read.

**FFQD4040E: Cannot get View or DocumentCollection from database *database\_name* on server *server\_name*. Check that you can access the database and that the crawl target is available to user *user\_name*. View/Folder: *view\_or\_folder\_name*.**

Specified target (Notes View or Folder) is not valid (target view or folder does not exist) or appropriate privilege is not given to the target to read document from crawler.

Specify appropriate view or folder as a crawling target from administration console to read documents. Change privilege of the Notes View or Folder so that the crawler can read.

**FFQD4041E: Cannot initialize Notes thread.**

Notes thread is not started due to an error.

See the log files for more information about where the problem occurred.

**FFQD4042E: Cannot obtain current Notes user ID *id\_filename* for server *server\_name*.**

Notes crawler cannot obtain Notes user ID string created session to the server.

See the log files for more information about where the problem occurred.

**FFQD4043E: Connection for server *server\_name* is already active. ID file: *id\_filename*.**

The connection is already active for the crawler.

See the log files for more information about where the problem occurred.

**FFQD4044E: An error occurred while connecting to database *database\_name* on server *server\_name*. Check that you can access the database and that the crawl target is available to user *id\_filename*.**

Notes crawler cannot have access to the Domino server and its database due to the invalid setting of the crawler or Domino server.

Check Notes crawler setting (especially server name, user ID and its password) from administration console. Check the target Domino server condition; server and database privilege to the crawler user ID so that the Notes crawler can have access to the target database. See Administering Enterprise Search, Lotus Domino Designer Help or Lotus Domino Administrator Help.

**FFQD4045E: Cannot connect to database *database\_name* on server *server\_name*. Check that you can access the database and that the crawl target is available to user *id\_filename*.**

Notes crawler cannot have access to the Domino server and its database due to the invalid setting of the crawler or Domino server.

Check Notes crawler setting (especially server name, user ID and its password) from administration console. Check the target Domino server condition; server and database privilege to the crawler user ID so that the Notes crawler can have access to the target database. See Administering Enterprise Search, Lotus Domino Designer Help or Lotus Domino Administrator Help.

**FFQD4046E: Cannot select formula-selected documents. Check that you can access the documents and that the formula is based on a Notes formula. Server: *server\_name*. ID file:*id\_filename*. Database:*database\_name*. Formula: *formula*.**

Specified formula is not valid as Notes Formula. Or Domino server or database is prohibited to have access from crawler.

Check and modify specified Notes Formula from administration console. Check crawler, Domino server and target database so that crawler can have access to target database properly. See Administering Enterprise Search or Lotus Domino Designer Help.

**FFQD4047E: An error occurred while getting View or DocumentCollection from database *database\_name* on server *server\_name*. Check that you can access the database and that the crawl target is available to user *id\_filename*. View/Folder: *view\_or\_folder\_name*.**

Specified target (Notes View or Folder) is not valid or appropriate privilege is not given to the target to read document from crawler.

Specify appropriate view or folder as a crawling target from administration console to read documents. Change privilege of the Notes View or Folder so that the crawler can read.

**FFQD4048E: Cannot get View or DocumentCollection from database *database\_name* on server *server\_name*. Check that you can access the database and that the crawl target is available to user *id\_filename*. View/Folder: *view\_or\_folder\_name*.**

Specified target (Notes View or Folder) is not valid or appropriate privilege is not given to the target to read document from crawler.

Specify appropriate view or folder as a crawling target from administration console to read documents. Change privilege of the Notes View or Folder so that the crawler can read.

**FFQD4049E: Cannot get View or DocumentCollection from database *database\_name* on server *server\_name*. Check that you can access the database and that the crawl target is available to user *id\_filename*. View/Folder: *view\_or\_folder\_name*.**

Specified target (Notes View or Folder) is not valid or appropriate privilege is not given to the target to read document from crawler.

Specify appropriate view or folder as a crawling target from administration console to read documents. Change privilege of the Notes View or Folder so that the crawler can read.

**FFQD4050E: View or DocumentCollection is null.**

An error occurred while fetching document from specified target. specified target is null.

See the log files for more information about where the problem occurred.

**FFQD4051E: Cannot fetch Document from View.**

An error occurred while fetching document from specified target.

See the log files for more information about where the problem occurred.

**FFQD4052E: Cannot fetch the Document from the View or Folder while comparing with formula-selected Document universal ID.**

An error occurred while fetching document from specified target.

See the log files for more information about where the problem occurred.



**FFQD4053E: Cannot fetch Document from DocumentCollection.**

An error occurred while fetching document from specified target.

See the log files for more information about where the problem occurred.

**FFQD4054E: Cannot fetch the Document from DocumentCollection while comparing with formula-selected Document universal ID.**

An error occurred while fetching document from specified target.

See the log files for more information about where the problem occurred.

**FFQD4055E: Current Document is null.**

An error occurred while getting created date of the Notes document.

See the log files for more information about where the problem occurred.

**FFQD4056E: Cannot obtain created date from current Document.**

An error occurred while getting created date of the Notes document.

See the log files for more information about where the problem occurred.

**FFQD4057E: Current document is null.**

An error occurred while getting last modified date of the Notes document.

See the log files for more information about where the problem occurred.

**FFQD4058E: Cannot obtain last modified date from current document.**

An error occurred while getting last modified date of the Notes document.

See the log files for more information about where the problem occurred.

**FFQD4059E: Cannot obtain current document size.**

An error occurred while getting the size of the Notes document.

See the log files for more information about where the problem occurred.

**FFQD4060E: Cannot fetch field from current document.**

An error occurred while getting document field values of the Notes document.

See the log files for more information about where the problem occurred.

**FFQD4061E: Cannot recycle Item.**

An error occurred while releasing memory of the Notes object.

See the log files for more information about where the problem occurred.

**FFQD4062E: Cannot fetch field from current document.**

An error occurred while getting attachment attribute from the Notes document.

See the log files for more information about where the problem occurred.

**FFQD4063E: Cannot recycle item.**

An error occurred while releasing memory of the Notes object.

See the log files for more information about where the problem occurred.

**FFQD4064E: Current EmbeddedObject is null.**

An error occurred while getting input stream from Notes document attachment file.

See the log files for more information about where the problem occurred.

**FFQD4065E: Cannot get InputStream from current EmbeddedObject.**

An error occurred while getting input stream from Notes document attachment file.

See the log files for more information about where the problem occurred.

**FFQD4066E: Current EmbeddedObject is null.**

An error occurred while getting attachment file name from Notes document.

See the log files for more information about where the problem occurred.

**FFQD4067E: Cannot get file name from current EmbeddedObject.**

An error occurred while getting attachment file name from Notes document.

See the log files for more information about where the problem occurred.

**FFQD4068E: Current EmbeddedObject is null.**

An error occurred while getting attachment file size from Notes document.

See the log files for more information about where the problem occurred.

**FFQD4069E: Cannot get the file size from current EmbeddedObject.**

An error occurred while getting attachment file size from Notes document.

See the log files for more information about where the problem occurred.

**FFQD4070E: Current document is null.**

An error occurred while getting form name defined in the Notes document.

See the log files for more information about where the problem occurred.

**FFQD4071E: Cannot get the form name from current document.**

An error occurred while getting form name defined in the Notes document.

See the log files for more information about where the problem occurred.

**FFQD4072E: Current document is null.**

An error occurred while getting Universal Notes ID (UNID) defined in the Notes document.

See the log files for more information about where the problem occurred.

**FFQD4073E: Cannot get document universal Notes ID.**

An error occurred while getting Universal Notes ID (UNID) defined in the Notes document.

See the log files for more information about where the problem occurred.

**FFQD4074E: Current view or DocumentCollection is null.**

An error occurred while getting Universal Notes ID (UNID) defined in the Notes view or folder.

See the log files for more information about where the problem occurred.

**FFQD4075E: Cannot get the universal ID for the View.**

An error occurred while getting Universal Notes ID (UNID) defined in the Notes view or folder.

See the log files for more information about where the problem occurred.

**FFQD4076E: Current database is null.**

An error occurred while getting Notes replica ID defined in the Notes database.

See the log files for more information about where the problem occurred.

**FFQD4077E: Cannot get database replica ID.**

An error occurred while getting Notes replica ID defined in the Notes database.

See the log files for more information about where the problem occurred.

**FFQD4078E: Cannot gather items to be crawled.**

An error occurred while getting field value(s) stored in the Notes document.

See the log files for more information about where the problem occurred.

**FFQD4079E: ID file *id\_filename* does not exist.**

Specified ID file does not exist.

Specify a valid Notes ID file from administration console.

**FFQD4080E: ID authorization failed for the ID file *id\_filename*.**

Specified password is not valid for given Notes user ID file.

Specify valid password for given Notes ID file from administration console.

**FFQD4081E: Cannot create Notes session.**

An error occurred while creating Notes session.

See the log files for more information about where the problem occurred.

**FFQD4082E: Cannot recycle registration.**

An error occurred while releasing memory of the Notes object.

See the log files for more information about where the problem occurred.

**FFQD4083E: Cannot change ID to *id\_filename*.**

Notes crawler cannot switch to specified ID.

See the log files for more information about where the problem occurred.

**FFQD4084E: This session is on DIIOP.**

Notes crawler tried to change ID though the crawler is for DIIOP crawling.

See the log files for more information about where the problem occurred.

**FFQD4085E: Cannot recycle session.**

An error occurred while releasing memory of the Notes object.

See the log files for more information about where the problem occurred.

**FFQD4086E: Cannot recycle current document.**

An error occurred while releasing memory of the Notes object.

See the log files for more information about where the problem occurred.

**FFQD4087E: Cannot recycle View/DocumentCollection.**

An error occurred while releasing memory of the Notes object.

See the log files for more information about where the problem occurred.

**FFQD4088E: Cannot recycle current document.**

An error occurred while releasing memory of the Notes object.

See the log files for more information about where the problem occurred.

**FFQD4089E: Cannot recycle current EmbeddedObject.**

An error occurred while releasing memory of the Notes object.

See the log files for more information about where the problem occurred.

**FFQD4090E: Cannot recycle current EmbeddedObject.**

An error occurred while releasing memory of the Notes object.

See the log files for more information about where the problem occurred.

**FFQD4091W: A URI that was too long was generated. This row will be ignored. Generated URI: *uri*.**

The generated URI is too long to store internal table. The document which this URI designates is not stored into internal table and is not used by search.

No action is required. To reduce the URI length, reduce database relative path from Domino server data directory and reduce attachment file name.

**FFQD4200I: Connected to *database\_URL* using user *user\_name*.**

The DB2 crawler connected to the specified database.

This message is for your information only. No action is required.

**FFQD4201I: Disconnected from *database\_URL*.**

The DB2 crawler disconnected from the specified database.

This message is for your information only. No action is required.

**FFQD4202W: The metadata is truncated because it exceeds the size limit *data\_size\_limit*.**

The DB2 crawler truncated a metadata data because its size was larger than the size limit.

No action is required.

**FFQD4203W: The content is ignored because it exceeds the size limit *data\_size\_limit*.**

The DB2 crawler ignored content because total record size was larger than the size limit.

No action is required.

**FFQD4204W: A row in table *table\_name* on database *database\_URL* was ignored because null key is detected in the column *column\_name*.**

The DB2 crawler ignored a row because null data is detected in the column specified as a unique key column.

No action is required.

**FFQD4205W: A row was ignored because the URI is too long. URI: *URI*.**

The DB2 crawler ignored a row because its URI is too long.

No action is required.

**FFQD4207E: The fields specification is not valid.**

The DB2 crawler configuration file contained incorrect column definition.

Verify that the DB2 crawler configuration file corresponds with the current target table.

**FFQD4208E: User ID, password, or both are incorrect. The connection to *database\_URL* cannot be established.**

The DB2 crawler configuration file contained incorrect user id or password.

Verify that the user name and the password is up-to-date.

**FFQD4209E: Database *database\_URL* was not found.**

The DB2 crawler could not find the database.

Verify that the database exists.

**FFQD4210E: Table or view *database\_URL* was not found.**

The DB2 crawler could not find the table or view.

Verify that the database exists.

**FFQD4211E: Column was not found.**

The DB2 crawler configuration file contained incorrect column definition.

Verify that the DB2 crawler configuration file corresponds with the current target table.

**FFQD4212E: SQL syntax error was detected. Database: *database\_URL*. SQL: *SQL\_statement*.**

The SQL statement for selecting rows was incorrect. The DB2 crawler configuration file might contain an incorrect WHERE clause.

Verify that the condition for selecting rows is correct.

**FFQD4213E: User *user\_name* does not have CONNECT or SELECT privilege. Database: *database\_URL*. Table: *table\_name*.**

The user did not have privileges to access the table.

Grant CONNECT and SELECT privileges to the user.

**FFQD4214E: Database connection to *database\_URL* is full.**

The DB2 crawler could not establish the connection because the number of connection to the database had reached to the maximum.

Increase the maximum number of connection or try again after a while.

**FFQD4215E: Connection was terminated from the database *database\_URL*.**

The database connection was terminated from the database unexpectedly.

Verify the database is working correctly and try again.

**FFQD4216E: Connection to *database\_URL* failed.**

One or more errors occurred while connecting to the database.

See the log files for more information about where the problem occurred.

**FFQD4217E: Disconnection from *database\_URL* failed.**

One or more errors occurred while disconnecting from the database.

See the log files for more information about where the problem occurred.

**FFQD4218E: An I/O error occurred. Database: *database\_URL*.  
Table: *table\_name*.**

One or more errors occurred while transferring data from the database.

See the log files for more information about where the problem occurred.

**FFQD4219E: Database error occurred. Database:  
*database\_URL*. SQLCODE: *SQLCODE*. SQLSTATE: *SQLSTATE*.**

One or more errors occurred while accessing to the database.

See the log files for more information about where the problem occurred.

**FFQD4220E: There is no database connection.**

The DB2 crawler had started successfully, but the database connection to crawl data was null unexpectedly.

See the log files for more information about where the problem occurred.

**FFQD4221E: Columns to be crawled are not specified.  
Database: *database\_URL*. Table: *table\_name*.**

The DB2 crawler configuration file did not contain any column names.

Verify that the crawler which you want to start was configured correctly and try again.

**FFQD4222E: Security field must be a string column.  
Database: *database\_URL*. Table: *table\_name*. Column:  
*column\_name*. Type: *security\_type*.**

The security field was not a string column.

Verify that the DB2 crawler configuration file corresponds with the current target table.



**FFQD4223E: Date field must be a DATE or TIMESTAMP column. Database: *database\_URL*. Table: *table\_name*. Column: *column\_name*. Type: *date\_type*.**

The date field was not a string column.

Verify that the DB2 crawler configuration file corresponds with the current target table.

**FFQD4224E: Unsupported data type was found. Database: *database\_URL*. Table: *table\_name*. Column: *column\_name*. Type: *data\_type*.**

The DB2 crawler detected that the configuration file contained an unsupported data type column.

Verify that the DB2 crawler configuration file corresponds with the current target table.

**FFQD4225E: DATE or TIMESTAMP column is specified as content. Database: *database\_URL*. Table: *table\_name*. Column: *column\_name*.**

The DB2 crawler detected that a not LOB column was specified as content

Verify that the DB2 crawler configuration file corresponds with the current target table.

**FFQD4226E: A nonnumeric column (including BIGINT) is specified as a parametric searchable field. Database: *database\_URL*. Table: *table\_name*. Column: *column\_name*. Type: *data\_type*.**

The DB2 crawler detected that a nonnumeric column was specified as a parametric searchable field. Numeric columns are as follows:

- DATE
- DECIMAL
- DOUBLE
- INTEGER
- REAL
- SMALLINT
- TIME
- TIMESTAMP

Verify that the DB2 crawler configuration file corresponds with the current target table.

**FFQD4227E: A non-binary column is specified as content.**  
**Database:** *database\_URL*. **Table:** *table\_name*. **Column:**  
*column\_name*. **Type:** *data\_type*.

The DB2 crawler detected that a non-binary column was specified as content.  
Binary columns are as follows:

- BLOB
- CLOB
- DBCLOB
- CHARACTER FOR BITDATA
- VARCHAR FOR BITDATA
- LONG VARCHAR FOR BITDATA

Verify that the DB2 crawler configuration file corresponds with the current target table.

**FFQD4228E: Multiple content is not supported. Database:**  
*database\_URL*. **Table:** *table\_name*.

The DB2 crawler detected that two or more content.

Verify that the DB2 crawler configuration file corresponds with the current target table.

**FFQD4300E: An internal error occurred (another connection exists).**

Unexpected error occurred.

Try the operation again. If the same problem occurred again, see your IBM service representative.

**FFQD4301E: The item type *item\_type* does not exist or the item type is a part item type.**

The item type which was specified in configuration file did not exist in the Content Manager server or it was a part item type.

Log on to the administration console and edit the crawl space to remove the item type.

**FFQD4302E: Connection failed. Database:** *database*. **User:**  
*user*.

A connection to the specified Content Manager server failed.

Try the operation again. Make sure that the server is running, and the password is valid.

### **FFQD4303E: Disconnection failed.**

A disconnection from the Content Manager server failed.

See IBM DB2 Content Manager for Multiplatforms/IBM DB2 Information Integrator for Content for Multiplatforms, Messages and Codes for more information.

### **FFQD4304E: An error occurred while fetching content. PID:*pid*.**

An error occurred while fetching content from the Content Manager server.

See IBM DB2 Content Manager for Multiplatforms/IBM DB2 Information Integrator for Content for Multiplatforms, Messages and Codes for more information. Check the specified document is correct.

### **FFQD4305E: An error occurred while creating a temporary file.**

The error occurred when the crawler tried to store a large content as a temporary file.

Refer to the crawler log for more information.

### **FFQD4306E: Database connection does not exist.**

A connection was not established for the Content Manager server. Connection might be destroyed.

Try the operation again. Check the server is running.

### **FFQD4307E: An error occurred while executing a query command *cmd*.**

An error occurred while executing a query command of Content Manager.

See IBM DB2 Content Manager for Multiplatforms/IBM DB2 Information Integrator for Content for Multiplatforms, Messages and Codes for more information.

### **FFQD4308E: An error occurred while fetching a record. PID: *pid*.**

An error occurred while fetching a record from the Content Manager server.

See IBM DB2 Content Manager for Multiplatforms/IBM DB2 Information Integrator for Content for Multiplatforms, Messages and Codes for more information. Check the specified document is correct.

**FFQD4309E: An error occurred while fetching a part. PID: *pid*.**

An error occurred while fetching a part from the Content Manager server.

See IBM DB2 Content Manager for Multiplatforms/IBM DB2 Information Integrator for Content for Multiplatforms, Messages and Codes for more information. Check the specified document is correct.

**FFQD4310E: The attribute *attribute* does not exist.**

The attribute which was specified in the configuration file did not exist in the Content Manager server.

Log on to the administration consol and edit the crawler space. Remove the item type, then add the item type again to configure.

**FFQD4311E: The attribute *attribute* does not exist or the attribute is BLOB type.**

The attribute which was specified in the configuration file did not exist or the attribute was a BLOB type.

Log on to the administration consol and edit the crawler space. Remove the item type, then add the item type again to configure.

**FFQD4312E: The component *component* does not exist.**

The component which was specified in the configuration file did not exist in the Content Manager server.

Log on to the administration consol and edit the crawler space. Remove the item type, then add the item type again to configure.

**FFQD4313E: An error occurred while fetching attributes.**

An error occurred while fetching a value of attribute from the Content Manager server.

See IBM DB2 Content Manager for Multiplatforms/IBM DB2 Information Integrator for Content for Multiplatforms, Messages and Codes for more information.

**FFQD4314E: An error occurred while retrieving the last modified date.**

An error occurred while retrieving the last modified date from the Content Manager server.

See IBM DB2 Content Manager for Multiplatforms/IBM DB2 Information Integrator for Content for Multiplatforms, Messages and Codes for more information.

**FFQD4315E: An error occurred while retrieving a creation date.**

An error occurred while retrieving a creation date from the Content Manager server.

See IBM DB2 Content Manager for Multiplatforms/IBM DB2 Information Integrator for Content for Multiplatforms, Messages and Codes for more information.

**FFQD4316E: An error occurred while parsing a date *date*.**

An error occurred while parsing a date. This is an unexpected error.

Refer to the crawler log for more information.

**FFQD4317E: An error occurred while retrieving an item type *item\_type*.**

An error occurred while retrieving an item type from the Content Manager server.

See IBM DB2 Content Manager for Multiplatforms/IBM DB2 Information Integrator for Content for Multiplatforms, Messages and Codes for more information.

**FFQD4318E: An error occurred while retrieving an original file name.**

An error occurred while retrieving an original file name from the Content Manager server.

See IBM DB2 Content Manager for Multiplatforms/IBM DB2 Information Integrator for Content for Multiplatforms, Messages and Codes for more information. Check that the Resource Manager is running.

**FFQD4319E: The attribute *attribute* specified as the field to control access does not exist.**

The attribute which was specified as the field to control access in the configuration file did not exist in the Content Manager server.

Log on to the administration consol and edit the crawler space. Remove the item type, then add the item type again to configure.

**FFQD4320E: The attribute *attribute* specified as the document date is incorrect.**

The attribute which was specified as the document date in the configuration file was neither a `TIMESTAMP` type nor a `DATE` type.

Log on to the administration consol and edit the crawler space. Remove the item type, then add the item type again to configure.

**FFQD4321E: The attribute *attribute* specified as the document date does not exist.**

The attribute which was specified as the document date in the configuration file did not exist in the Content Manager server.

Log on to the administration console and edit the crawler space. Remove the item type, then add the item type again to configure.

**FFQD4322E: An error occurred while retrieving a document date. Attribute: *attribute*.**

An error occurred while retrieving a document date from the Content Manager server.

See IBM DB2 Content Manager for Multiplatforms/IBM DB2 Information Integrator for Content for Multiplatforms, Messages and Codes for more information.

**FFQD4323W: A URI that was too long was generated. This row will be ignored. Generated URI: *uri*.**

The generated URI was too long. This row will be ignored.

To crawl the data, shorten the URI. URI becomes shorter if the server name or item type name becomes short.

**FFQD4324I: Connected to server. Database: *database*. User: *user*.**

The crawler server successfully connected to the Content Manager server.

This message is for your information only. No action is required.

**FFQD4325I: Disconnected from server.**

The crawler server successfully disconnected from the Content Manager server.

This message is for your information only. No action is required.

**FFQD4326E: Resource Manager is not online or available.**

The Resource Manager was not online or available.

Start the Resource Manager and try the operation again.

**FFQD4327W: An error occurred while fetching content. Skip this document. PID:*pid*.**

An error occurred while fetching a document from the Content Manager server. The document might not be stored in the Content Manager correctly. This document was not crawled.

See IBM DB2 Content Manager for Multiplatforms/IBM DB2 Information Integrator for Content for Multiplatforms, Messages and Codes for more information. Check the specified document is correct.

**FFQD4402W: Property *property\_name* cannot be changed now. Restart the crawler.**

This property *property\_name* cannot be changed while the crawler is running.

Stop the crawler, and restart it again.

**FFQD4409W: NNTP crawler will be paused because the number of articles exceeds the limit *max\_num\_of\_articles*.**

NNTP crawler went into paused state automatically because the crawler reached to the number of articles specified as maximum number of articles.

If more articles are needed, increase the maximum number of articles and restart the crawler. If not, just stop the crawler.

**FFQD4413W: Crawler will sleep *sleep\_interval* seconds because of an error.**

NNTP crawler will sleep for a while because a network error was detected.

If this is because of temporary network unavailability, just wait for a while, and crawler will start crawling again. If not, make sure the network is available on this crawler node.

**FFQD4414E: Updating the server status failed. Crawler will try again later.**

NNTP crawler failed to update server status on the internal repository. Crawler will sleep for a while and try again.

See other log messages to determine the actual cause of this error.

**FFQD4415E: Updating the newsgroup status failed. Crawler will try again later.**

NNTP crawler failed to update newsgroup status on the internal repository. Crawler will sleep for a while and try again.

See other log messages to determine the actual cause of this error.

**FFQD4416E: Inserting the message ID failed. Crawler will try again later.**

NNTP crawler failed to insert message-id of an article to the internal repository. Crawler will sleep for a while and try again.

See other log messages to determine the actual cause of this error.

**FFQD4417E: Inserting the news article failed. The crawler will try again later.**

NNTP crawler failed to insert an article to the internal repository. Crawler will sleep for a while and try again.

See other log messages to determine the actual cause of this error.

**FFQD4418E: Inserting deleted news article failed. Crawler will try again later.**

NNTP crawler failed to insert information of a deleted article to the internal repository. Crawler will sleep for a while and try again.

See other log messages to determine the actual cause of this error.

**FFQD4419W: Crawler will be stopped despite failure to update the server status.**

NNTP crawler could not update server status. This could cause to have obsolete or inconsistent server information which would be shown on admin GUI. This information will be updated when the crawler is started next time.

No action is required.

**FFQD4420W: Crawler will be stopped despite failure to update newsgroup status. This might cause to recrawl some articles.**

NNTP crawler could not update newsgroup status. This could cause to have obsolete or inconsistent server information which would be shown on admin GUI. This information will be updated when the crawler is started next time.

No action is required.

**FFQD4421W: Crawler will be stopped despite failure to update message id. This might cause to recrawl some articles.**

NNTP crawler could not record message-id into internal repository. This could let the crawler to crawl the same article next time, but it will not affect to search result because it will be eliminated as a duplicate.

No action is required.

**FFQD4422W: Crawler will be stopped despite failure to insert article.**

NNTP crawler could not store the article into internal repository, however it will be crawled when the crawler is started next time.



No action is required.

**FFQD4423W: Crawler will be stopped despite failure to insert deleted article.**

NNTP crawler could not store the information of deleted article into internal repository, however it will be done when the crawler is started next time.

No action is required.

**FFQD4427E: The newsgroup list could not be retrieved because of an unknown host *host\_name*.**

NNTP crawler cannot access to the specified server *host\_name* when it tries to retrieve the list of newsgroups on the server because it is unknown host.

Verify the host name is correct.

**FFQD4428E: The newsgroup list could not be retrieved because of a network error or server error on *host\_name***

NNTP crawler could not retrieve the list of newsgroups on the server because of a network error occurred.

See the exceptions in the log, verify the network availability, and confirm the news server is active.

**FFQD4433W: Cleaning was interrupted. This process will be restarted next time.**

NNTP crawler was stopped while writing information of deleted documents into internal repository. This task will be resumed when the crawler is started next time.

No action is required.

**FFQD4434W: Message ID was not found in this article *message\_id*. This article will be ignored.**

NNTP crawler detected an article which has an incorrect format. This article will be ignored.

No action is required.

**FFQD4435W: Message ID is inconsistent (*message\_id1* and *message\_id2* must be identical). This article will be ignored.**

NNTP crawler detected an article which has an incorrect format. This article will be ignored.

No action is required.

**FFQD4436W: Unexpected date format *date*.**

NNTP crawler detected unexpected date format in an article.

No action is required.

**FFQD4437E: I/O Error occurred while parsing article.**

A network I/O error occurred while parsing an article retrieved from news server.

See the exceptions in the log, verify the network availability, and confirm the news server is active.

**FFQD4449E: Errors occurred in *api\_name*.**

An error occurred while executing *api\_name* API.

See the log files for more information about where the problem occurred.

**FFQD4450E: File *file\_name* was not found.**

One of the configuration file *file\_name* was not found. NNTP crawler could not start crawling.

Verify that enterprise search was successfully installed and try again.

**FFQD4451E: Network I/O error occurred.**

Network I/O error occurred while accessing to the server.

See the exceptions in the log, verify the network availability, and confirm the news server is active.

**FFQD4452E: Unexpected response *server\_response* was returned for command *nntp\_command*.**

NNTP crawler received unexpected response *server\_response* from server for NNTP command *nntp\_command*.

Check the news server is serving accordingly.

**FFQD4453W: The status of article pointer could not be retrieved. Return code was *return\_code*.**

NNTP crawler received a return code *return\_code* other than successful return code. This might happen when some of the articles in a newsgroup are deleted from server.

No action is required if the server is active. If not, remove the server from crawl space.

**FFQD4454W: The article could not be retrieved. Return code was *return\_code*.**

NNTP crawler received a return code *return\_code* other than successful return code. In most cases, this is because server is about to stop or unavailable.

No action is required if the server is active. If not, remove the server from crawl space.

**FFQD4455W: The crawler could not move to the next article. Return code was *return\_code*.**

NNTP crawler received a return code *return\_code* other than successful return code. In most cases, this is because server is about to stop or unavailable.

No action is required if the server is active. If not, remove the server from crawl space.

**FFQD4456W: The crawler could not switch to newsgroup *newsgroup\_name* on server *server\_name*. Return code was *return\_code*.**

NNTP crawler received a return code *return\_code* other than successful return code. In most cases, this is because server is about to stop or unavailable.

No action is required if the server is active. If not, remove the server from crawl space.

**FFQD4457W: The newsgroup list could not be retrieved. Return code was *return\_code*.**

NNTP crawler received a return code *return\_code* other than successful return code. In most cases, this is because server is about to stop or unavailable.

No action is required if the server is active. If not, remove the server from crawl space.

**FFQD4458E: Server name *server\_name* could not be resolved.**

NNTP crawler could not resolve the server address for server *server\_name*.

Remove the server from crawl space if the server does not exist. If it exists, confirm the network environment on the crawler node.

**FFQD4461E: The crawler received service disconnected message from server.**

NNTP crawler received service disconnected message from server. The connection to the server will be disconnected.

Verify the availability of the server.

**FFQD4500E: The *directory\_name* directory was not found.**

The specified directory was not found.

Verify the directory on the crawler node and configure crawl space page on administration console.

**FFQD4501E: The *directory\_name* directory was not found.  
Error message : *detail\_message***

The specified directory was not found.

Verify the directory on the crawler node and configure crawl space page on administration console.

**FFQD4502E: Failed to open *filename* file. The file might be used by another program.**

The file might be used by other programs. The file was not crawled.

Stop the other program using the crawled file and restart the crawler.

**FFQD4503E: An exception occurred while reading *filename* file.**

The file might be used by other programs. The file was not crawled.

See the log files for more information about where the problem occurred.

**FFQD4504E: The subdirectory level *levels of subdirectories* is incorrect.**

The specified levels of subdirectories was not valid.

See the log files for more information about where the problem occurred.

**FFQD4505W: The length of directory name is too long to be stored. Directory: *directory\_name*.**

The specified directory name was too long to be stored.

Move the directory to be crawled and configure crawl space page on the administration console.

**FFQD4506W: The length of the URI name is too long to be stored. URI: *directory\_name*.**

The specified URI was too long to be stored. The length of URI name is calculated as UTF-8 byte string.

Move the directory or rename the file name to be crawled and configure crawl space page on administration console.

**FFQD4507E: The network drive cannot be used. Error message : *detail\_message***

The network drive was not available.

Verify the network and the directory on the crawler node.

**FFQD4508E: The user ID *user\_name* or the password is incorrect. Network folder:*directory\_name*.**

An incorrect user ID, an incorrect password or both were specified.

Verify the user name and the password and configure crawl space page on Administration console.

**FFQD4509E: The temporary file *temporary\_file* was not created.**

The File System crawler could not create the temporary file.

Verify the free disk spaces on the crawler node and restart the crawler.

**FFQD4510E: The temporary file *temporary\_file* failed to be loaded.**

The File System crawler could not read the temporary file.

Verify the free disk spaces on the crawler node and restart the crawler.

**FFQD4511E: The temporary file *temporary\_file* failed to be stored.**

The File System crawler could not write the temporary file.

Verify the free disk spaces on the crawler node and restart the crawler.

**FFQD4600E: An I/O error occurred while receiving a response from the URL *public\_folder\_URL*.**

One or more errors occurred while receiving a response from the public folder server.

See the log files for more information about where the problem occurred.

**FFQD4601E: An XML parser configuration error occurred.**

Could not create a XML parser for parsing responses from the public folder server.

Verify that enterprise search was successfully installed and try again.

**FFQD4602E: WebDAV response from URL *public\_folder\_URL* is an incorrect XML document.**

The public folder server returned a XML document which might contain incorrect elements and characters.

Verify the public folder does not contain broken documents.

**FFQD4603I: Connected to *public\_folder\_URL* using user *user\_name*.**

The Exchange crawler connected to the specified URL.

This message is for your information only. No action is required.

**FFQD4604I: Connected to *public\_folder\_URL* using user *user\_name* and truststore file *truststore\_file*.**

The Exchange crawler connected to the specified URL.

This message is for your information only. No action is required.

**FFQD4605I: Disconnected from *public\_folder\_URL*.**

The Exchange crawler disconnected from the specified URL.

This message is for your information only. No action is required.

**FFQD4607E: Connecting to *public\_folder\_URL* failed.**

One or more errors occurred while connecting to the specified URL.

See the log files for more information about where the problem occurred.

**FFQD4608E: Checking next record failed.**

One or more errors occurred while getting a document from the specified URL.

See the log files for more information about where the problem occurred.

**FFQD4609E: Error occurred while getting content from URL *uniform\_resource\_locator*.**

One or more errors occurred while transferring data from the specified URL.

See the log files for more information about where the problem occurred.

**FFQD4610E: An I/O error occurred while getting content from the URL *public\_folder\_URL*.**

One or more errors occurred while transferring data from the specified URL.

See the log files for more information about where the problem occurred.

**FFQD4611E: Could not crawl the public folder *public\_folder\_URL*. The HTTP Status is *HTTP\_state\_code*.**

One or more errors occurred while transferring data from the specified URL.

See the log files for more information about where the problem occurred.

**FFQD4612E: WebDAV request to *public\_folder\_URL* failed. The returned status code and message is *HTTP\_state\_code*, *HTTP\_response\_message*.**

The public folder server returned an unexpected HTTP response.

Verify the public folder server is working correctly and see the log files for more information where the problem occurred.

**FFQD4613W: Time format error was detected while parsing the value *string*.**

The public folder server returned a response which contained an incorrect date string.

Verify that the crawler which you want to run was configured correctly and try again.

**FFQD4614W: The specified URL *document\_URL* already exists.**

The Exchange crawler detected duplicated URL.

No action is required.

**FFQD4616E: Temporary file *temporary\_file* does not exist.**

The Exchange crawler could not find the temporary file.

Verify that the crawler machine have free disk space and try again.

**FFQD4617E: An I/O error occurred while writing objects into the temporary file *temporary\_file*.**

The Exchange crawler could not write data to the temporary file.

Verify that the crawler machine have free disk space and try again.

**FFQD4618E: An I/O error occurred while closing object output stream to the file *temporary\_file*.**

One or more I/O errors occurred while closing the temporary file.

Verify that the crawler machine have free disk space and try again.

**FFQD4619E: An I/O error occurred while closing the file *temporary\_file*.**

One or more I/O errors occurred while closing the temporary file.

Verify that the crawler machine have free disk space and try again.

**FFQD4620E: An I/O error occurred while reading objects from the temporary file *temporary\_file*.**

One or more I/O errors occurred while closing the temporary file.

Verify that the crawler machine is working correctly and try again.

**FFQD4621E: An I/O error occurred while closing object input stream from the file *temporary\_file*.**

One or more I/O errors occurred while closing the temporary file.

Verify that the crawler machine is working correctly and try again.

**FFQD4622E: An unknown host name *host\_name* was detected.**

The host name in the specified public folder URL was not found.

Verify that the host name is correct.

**FFQD4623E: The URL *public\_folder\_URL* is incorrect.**

The specified public folder URL was not found.

Verify that the public folder URL is correct.

**FFQD4624E: Could not refresh connection to *public\_folder\_URL*.**

The Exchange crawler could not reset the connection to the public folder server.

Verify the public folder server is working correctly and see the log files for more information where the problem occurred.



**FFQD4625E: An I/O error occurred while sending a request to *public\_folder\_URL*.**

One or more errors occurred while sending a request to the public folder server.

See the log files for more information about where the problem occurred.

**FFQD4626E: Connecting to *public\_folder\_URL* failed.**

One or more errors occurred while connecting to the specified URL.

See the log files for more information about where the problem occurred.

**FFQD4627E: A truststore file must be specified to use SSL.**

To specify a truststore file is required for using SSL.

Verify that the crawler which you want to run was configured correctly and try again.

**FFQD4628E: A truststore password must be specified to use SSL.**

To specify a password for the truststore file is required for using SSL.

Verify that the crawler which you want to run was configured correctly and try again.

**FFQD4629E: The truststore file *truststore\_file* does not exist.**

The specified truststore file did not exist or you did not have permission to read the file.

Verify that the inputted truststore file path is correct and you can read the file.

**FFQD4631E: Truststore file format or password is incorrect. The specified truststore is *truststore\_file*.**

The truststore file format or, and the specified password for the truststore file might be incorrect. The supported formats are JKS, JCEKS and PKCS#12.

Verify that the truststore file format and the password are correct and try again.

**FFQD4632W: A document was ignored because the URI is too long. URI: *URI*.**

The Exchange crawler ignored a document because its URI is too long.

No action is required.

**FFQD4633W: A folder was ignored because the URI is too long. URI: *URI*.**

The Exchange crawler ignored a folder because its URI is too long.

No action is required.

**FFQD4707E: Server name *server\_string* in configuration file does not represent VeniceBridge server and repository name.**

The server string *server\_string* was not correct format.

Configure VeniceBridge crawler setting from the administration console.

**FFQD4708E: VeniceBridge API error occurred.**

The error occurred while accessing to the VeniceBridge server.

See the log files for more information about where the problem occurred.

**FFQD4709E: VeniceBridge API is not licensed.**

The VeniceBridge crawler could not access the VeniceBridge server because some of the VeniceBridge component was not licensed.

See the log files for more information about where the problem occurred.

**FFQD4710E: VeniceBridge server instance cannot be created.**

The VeniceBridge crawler could not access the VeniceBridge server because the specified JNDI properties were incorrect.

See the log files for more information about where the problem occurred.

**FFQD4711E: VeniceBridge API configuration error occurred.**

The error occurred while accessing to the VeniceBridge server.

See the log files for more information about where the problem occurred.

**FFQD4712E: VeniceBridge server cannot be located.**

The error occurred while accessing to the VeniceBridge server.

See the log files for more information about where the problem occurred.

**FFQD4713E: Communication related error occurred.**

The error occurred while accessing to the VeniceBridge server.

See the log files for more information about where the problem occurred.

**FFQD4714E: The repository *repository\_name* is not found.**

The *repository\_name* repository did not exist in the VeniceBridge server.

Verify that the VeniceBridge crawler configuration corresponds with the current target repository.

**FFQD4715E: The VeniceBridge user instance was not initialized.**

The error occurred while accessing to the VeniceBridge server.

See the log files for more information about where the problem occurred.

**FFQD4716E: Logon to *repository\_name* repository failed.**

The VeniceBridge crawler could not logon to the *repository\_name* repository.

Check a correct user name and password are specified.

**FFQD4718E: The item class *item\_class* was not found for repository *repository\_name*.**

The item class that was specified in configuration file did not exist in the VeniceBridge server.

Log on to the administration console and edit the crawl space to remove the item class.

**FFQD4719E: The property *property* was not found.  
Repository: *repository\_name*. Item class: *item\_class*.**

The property that was specified in configuration file did not exist in the VeniceBridge server.

Verify that the VeniceBridge crawler configuration corresponds with the current target item class.

**FFQD4720E: The property *property* for ranking date was not found. Repository: *repository\_name*. Item class: *item\_class*.**

The property that was specified as the document data in configuration file did not exist in the VeniceBridge server.

Verify that the VeniceBridge crawler configuration corresponds with the current target item class.

**FFQD4721E: The property *property* for security field is not found. Repository: *repository\_name*. Item class: *item\_class*.**

The property that was specified as the security field in configuration file did not exist in the VeniceBridge server.

Verify that the VeniceBridge crawler configuration corresponds with the current target item class.

**FFQD4722E: The specified selection criteria *condition\_statement* has a problem.**

The condition statement that was specified in configuration file was incorrect syntax.

Verify that the syntax of the condition statement is correct.

**FFQD4723E: Ranking date field must be DATE or DATETIME data type. Repository: *repository\_name*. Item class: *item\_class*. Property: *property*. Type: *data\_type*.**

The property that was specified as the document data in configuration file was neither a DATE type nor a DATETIME type.

Verify that the VeniceBridge crawler configuration corresponds with the current target item class.

**FFQD4724E: An unsupported data type *data\_type* was found. Repository: *repository\_name*. Item class: *item\_class*. Property: *property*.**

The property that was specified in configuration file was an unsupported data type.

Verify that the VeniceBridge crawler configuration corresponds with the current target item class.

**FFQD4725E: The executed query has a problem.**

The error occurred while executing a query command to the VeniceBridge server.

See the log files for more information about where the problem occurred.

**FFQD4726E: No order key is found. Key: *order\_key\_property*.**

The property that was specified as order key in configuration file does not exist in the VeniceBridge server.

Verify that the VeniceBridge crawler configuration corresponds with the current target item class.

**FFQD4728W: A URI that was too long was generated. This record will be ignored. URI: *uri*.**

The generated URI is too long to store internal table. The document that this URI designates is not stored into internal table and is not used by search.

No action is required.

**FFQD4729W: The property *property* was not found.**

The property that was specified in configuration file was not found.

Verify that the VeniceBridge crawler configuration corresponds with the current target item class.

**FFQD4730E: An I/O error occurred while writing content.**

The error occurred while transferring data from the VeniceBridge server.

See the log files for more information about where the problem occurred.



---

## Chapter 19. Glossary

This glossary defines terms that are used in the enterprise search interface and documentation.

**administrative role**

A classification of a user that determines the functions that a user can do in the enterprise search administration console. The role also determines which collections the user can administer.

**alert** A message or other indication that identifies a problem or an impending problem.

**anchor text**

The text that is associated with a hyperlink.

**Boolean search**

A search in which one or more search terms are combined by using operators such as AND, NOT, and OR.

**category**

A group of documents that have similar properties.

**category tree**

In enterprise search, a hierarchy of categories that is displayed in the enterprise search administration console.

**collection**

A set of data sources and options for crawling, parsing, indexing, and searching those data sources. Users can search the data sources in a collection with a single query. Enterprise search administrators specify the options that control how the data in a collection is crawled, parsed, indexed, and searched.

**concept extraction**

A search function that identifies significant vocabulary items (such as people, places, or products) in text documents and produces a list of those items. See also theme extraction.

**crawl space**

A set of sources that match specified patterns (such as database names, file system paths, domain names, IP addresses, and URLs) that a crawler reads from to retrieve items for indexing. A crawl space might contain, for example, a set of relational database tables, the item types on one or more DB2 Content Manager servers, or several Notes databases.

**crawler**

A software program that retrieves documents from data sources and gathers information that can be used to create search indexes.

**data source**

Any repository of data from which documents can be retrieved, such as the Web, relational and nonrelational databases, and content management systems.

**data source type**

A grouping of data sources according to the protocol that is used to access the data.

**dequeue**

To remove items from a queue.

**discoverer**

A function of a crawler that determines which data sources are available for the crawler to retrieve information from.

**dynamic ranking**

A type of ranking in which the terms in the query are analyzed with respect to the documents being searched to determine the rank of results. See also text-based scoring. Contrast with static ranking.

**dynamic summarization**

A type of summarization in which the search terms are highlighted and the search results contain phrases that best represent the concepts of the document that the user is searching for. Contrast with static summarization.

**enqueue**

To place items in a queue.

**enterprise search administrator**

An administrative role that enables a user to administer the entire enterprise search.

**escape character**

A character that suppresses or selects a special meaning for one or more characters that follow.

**extensible markup language (XML)**

A text-based tag language used for document processing and for publishing information on the Web.

**fielded search**

A query that is restricted to a particular field.

**free text search**

A search in which the search term is expressed as free-form text [endash] a phrase or a sentence describing in natural language the subject to be searched for.

**fuzzy search**

A search that returns words with spelling that is similar to that of the search term.

**hybrid search**

A combined Boolean search and free text search.

**hyperlink**

An area on a Web page, such as a text string or an image, that is associated with a URI. When clicked, a hyperlink opens the URI.

**hypertext anchor**

In an HTML document, the code that is used to define a hypertext link to another page or to a location elsewhere in the document.

**index** A data structure that references data items to enable the search to quickly find documents that contain the query terms.

**index queue**

A list of index reorganization requests or index refreshes to be processed.



**index refresh**

The process of adding new information to an existing index in an enterprise search system. Contrast with index reorganization.

**index reorganization**

The process of building the index in an enterprise search system. Contrast with index refresh.

**information extraction**

A type of concept extraction that automatically recognizes significant vocabulary items in text documents, such as, names, terms, and expressions.

**language identification**

An enterprise search feature that determines the language of a document.

**lexical affinity**

The relationship of search words to each other in the document. Lexical affinity is used to calculate the relevancy of a result.

**linguistic search**

A search type that browses, retrieves, and indexes a document with terms that are reduced to their base form. For example, mice is indexed as mouse.

**link analysis**

A method that is used to determine what pages in the collection are important to users.

**masking character**

A character that is used to represent optional characters at the front, middle, and end of a search term. Masking characters are normally used for finding variations of a term in a precise index. See also wildcard.

**model-based category**

A taxonomy of predefined terms that is used to determine the subject of a document so that the document can be indexed and searched with documents that have similar content.

**monitor**

In enterprise search, a user who has the authority to observe collection-level processes.

**natural language query**

A type of search that analyzes written expressions (such as "Who runs the finance department?") instead of a simple collection of keywords.

**operator**

In enterprise search, a user who has the authority to observe, start, and stop collection-level processes.

**parametric search**

A type of search that looks for objects that contain a numeric value or attribute, such as dates, integers, or other numeric data types.

**parse**

In enterprise search, to extract the useful text and metadata from a document. This process includes running algorithms to help rank documents and performing linguistic analysis to decompose compound words and enable synonyms.

**parser**

A program that interprets documents that are added to the enterprise search data store. The parser extracts information from the documents and prepares them for indexing, search, and retrieval.

**popular ranking**

A ranking type that adds to a document's existing ranking based on the document's popularity.

**proximity search**

A search type that looks for certain words in the same sentence, paragraph, or document.

**ranking**

The process of assigning an integer value to each document in the search results from a query. The order of the documents in the search results is based on the relevance to the query. A higher rank signifies a closer match. See also dynamic ranking and static ranking.

**rule-based category**

Categories created by defining rules. The rules determine which documents are associated with categories. For example, you can define rules to associate documents that contain or exclude certain words, or that match a URI pattern, with specific categories.

**scope** A group of related URIs that are used to define the range of a search request.

**score** A value (usually between 0 and 1) that indicates how well a document meets the search criteria relative to the other found documents.

**search application**

A program that processes queries, searches the index, returns the search results, and retrieves the source documents for collections in an enterprise search system.

**search cache**

A buffer that holds the data and results of previous search requests.

**search engine**

A program that accepts a search request and returns a list of documents to the user.

**search index file**

The file in which an index is stored in the search engine.

**search results**

A list of documents that match the search request. Search results are often sorted in relevance order as determined by the search engine's measure of how closely a document satisfies the query. Most result listings include the URL for retrieving the document, the title of the document and a short summary. Some also include the date, file size, and other document attributes, such as the categories that the document belongs to.

**security token**

Information about identity and security that is used to authorize access to documents in a collection. Different data source types support different types of security tokens. Examples include user roles, user IDs, group IDs, and other information that can be used to control access to content.

**static ranking**

A type of ranking in which factors about the documents being ranked, such as date, the number of links that point to the document, and so on, augment the rank. Contrast with dynamic ranking.

**static summarization**

A type of summarization in which the search results contain a specified, stored summary from the document. Contrast with dynamic summarization.

**stemming**

See word stemming.

**stop word**

A word that is commonly used, such as "the," "an," or "and" that is ignored by a search application when responding to a query.

**stop word removal**

The process of removing stop words from the index to reduce the size of the text index.

**summarization**

The process of including sentences in search results to briefly describe the content of a document. See also dynamic summarization and static summarization.

**taxonomy**

A classification of objects into groups based on similarities. In enterprise search, a taxonomy organizes data into categories and subcategories. See also category tree.

**text analysis**

The process of extracting semantics and other information from text to enhance the retrieveability of data in a collection.

**text-based scoring**

The process of assigning an integer value to a document that signifies the relevance of the document with respect to the terms in a query. A higher integer value signifies a closer match to the query. See also dynamic ranking.

**theme extraction**

A type of concept extraction that automatically recognizes significant vocabulary items in text documents to extract the theme or topic of a document. See also concept extraction.

**token** The basic syntactic unit of a computing language. A token consists of one or more characters, excluding the blank character and excluding characters within a string constant or delimited identifier.

**tokenizer**

A parsing program that scans text and determines when and if a series of characters can be recognized as a token.

**user agent**

An application that browses the Web and leaves information about itself at the sites that it visits. In enterprise search, the Web crawler is a user agent.

**Uniform Resource Identifier (URI)**

A compact string of characters for identifying an abstract or physical resource. The most common form of URI is the Web page address, which is a particular form or subset of URI called a Uniform Resource Locator (URL). A URI typically describes how to access the resource, the computer that contains the resource, and the name of the resource (a file name) on the computer.

**Uniform Resource Locator (URL)**

A sequence of characters that represents information resources on a computer or in a network such as the Internet. This sequence of characters includes the abbreviated name of the protocol that is used to access the information resource and the information that is used by the protocol to locate the information resource.

**user ID**

A name that is used to associate the user profile with a user when a user signs on to a system.

**weighted term search**

A query in which certain terms are given more importance.

**word stemming**

A process of linguistic normalization in which the variant forms of a word are reduced to a common form. For example, words like "connections," "connective," and "connected" are reduced to "connect."

**wildcard character**

A character that is used to represent optional characters at the front, middle, or end of a search term.

**XML** See extensible markup language.

---

## DB2 Information Integrator documentation

This topic provides information about the documentation that is available for DB2 Information Integrator.

The tables in the following topics provide the official document title, form number, and location of each PDF book. To order a printed book, you must know either the official book title or the document form number. Titles, file names, and the locations of the DB2 Information Integrator release notes and installation requirements are also provided in the following topics.

---

### Documentation about event publishing function for DB2 Universal Database on z/OS

Documentation about event publishing function for DB2 Universal Database on z/OS

#### Purpose

Documentation about event publishing function for DB2 Universal Database on z/OS.

*Table 2. DB2 Information Integrator documentation about event publishing function for DB2 Universal Database on z/OS*

Name	Form number	Location
<i>ASNCLP Program Reference for Replication and Event Publishing</i>	N/A	DB2 Information Integrator Support Web site
<i>Introduction to Replication and Event Publishing</i>	GC18-7567	<ul style="list-style-type: none"><li>• DB2 PDF Documentation CD</li><li>• DB2 Information Integrator Support Web site</li></ul>
<i>Replication and Event Publishing Guide and Reference</i>	SC18-7568	<ul style="list-style-type: none"><li>• DB2 PDF Documentation CD</li><li>• DB2 Information Integrator Support Web site</li></ul>
<i>Tuning for Replication and Event Publishing Performance</i>	N/A	DB2 Information Integrator Support Web site
<i>Release Notes for IBM DB2 Information Integrator Standard Edition, Advanced Edition, and Replication for z/OS</i>	N/A	<ul style="list-style-type: none"><li>• In the DB2 Information Center, <b>Product Overviews &gt; Information Integration &gt; DB2 Information Integrator overview &gt; Problems, workarounds, and documentation updates</b></li><li>• DB2 Information Integrator Installation launchpad</li><li>• DB2 Information Integrator Support Web site</li><li>• The <i>DB2 Information Integrator</i> product CD</li></ul>

---

## Documentation about event publishing function for IMS and VSAM on z/OS

Documentation about event publishing function for IMS and VSAM on z/OS

### Purpose

Documentation about event publishing function for IMS and VSAM on z/OS.

*Table 3. DB2 Information Integrator documentation about event publishing function for IMS and VSAM on z/OS*

<b>Name</b>	<b>Form number</b>	<b>Location</b>
<i>Client Guide for Classic Federation and Event Publisher for z/OS</i>	SC18-9160	DB2 Information Integrator Support Web site
<i>Data Mapper Guide for Classic Federation and Event Publisher for z/OS</i>	SC18-9163	DB2 Information Integrator Support Web site
<i>Getting Started with Event Publisher for z/OS</i>	GC18-9186	DB2 Information Integrator Support Web site
<i>Installation Guide for Classic Federation and Event Publisher for z/OS</i>	GC18-9301	DB2 Information Integrator Support Web site
<i>Operations Guide for Event Publisher for z/OS</i>	SC18-9157	DB2 Information Integrator Support Web site
<i>Planning Guide for Event Publisher for z/OS</i>	SC18-9158	DB2 Information Integrator Support Web site
<i>Reference for Classic Federation and Event Publisher for z/OS</i>	SC18-9156	DB2 Information Integrator Support Web site
<i>System Messages for Classic Federation and Event Publisher for z/OS</i>	SC18-9162	DB2 Information Integrator Support Web site
<i>Release Notes for IBM DB2 Information Integrator Event Publisher for IMS for z/OS</i>	N/A	DB2 Information Integrator Support Web site
<i>Release Notes for IBM DB2 Information Integrator Event Publisher for VSAM for z/OS</i>	N/A	DB2 Information Integrator Support Web site

---

## Documentation about event publishing and replication function on Linux, UNIX, and Windows

Documentation about event publishing and replication function on Linux, UNIX, and Windows

### Purpose

Documentation about event publishing and replication function on Linux, UNIX, and Windows.

*Table 4. DB2 Information Integrator documentation about event publishing and replication function on Linux, UNIX, and Windows*

<b>Name</b>	<b>Form number</b>	<b>Location</b>
<i>ASNCLP Program Reference for Replication and Event Publishing</i>	N/A	DB2 Information Integrator Support Web site

Table 4. DB2 Information Integrator documentation about event publishing and replication function on Linux, UNIX, and Windows (continued)

Name	Form number	Location
<i>Installation Guide for Linux, UNIX, and Windows</i>	GC18-7036	<ul style="list-style-type: none"> <li>• DB2 PDF Documentation CD</li> <li>• DB2 Information Integrator Support Web site</li> </ul>
<i>Introduction to Replication and Event Publishing</i>	GC18-7567	<ul style="list-style-type: none"> <li>• DB2 PDF Documentation CD</li> <li>• DB2 Information Integrator Support Web site</li> </ul>
<i>Migrating to SQL Replication</i>	N/A	DB2 Information Integrator Support Web site
<i>Replication and Event Publishing Guide and Reference</i>	SC18-7568	<ul style="list-style-type: none"> <li>• DB2 PDF Documentation CD</li> <li>• DB2 Information Integrator Support Web site</li> </ul>
<i>SQL Replication Guide and Reference</i>	SC27-1121	DB2 Information Integrator Support Web site
<i>Tuning for Replication and Event Publishing Performance</i>	N/A	DB2 Information Integrator Support Web site
<i>Tuning for SQL Replication Performance</i>	N/A	DB2 Information Integrator Support Web site
<i>Release Notes for IBM DB2 Information Integrator Standard Edition, Advanced Edition, and Replication for z/OS</i>	N/A	<ul style="list-style-type: none"> <li>• In the DB2 Information Center, <b>Product Overviews &gt; Information Integration &gt; DB2 Information Integrator overview &gt; Problems, workarounds, and documentation updates</b></li> <li>• DB2 Information Integrator Installation launchpad</li> <li>• DB2 Information Integrator Support Web site</li> <li>• The <i>DB2 Information Integrator</i> product CD</li> </ul>

## Documentation about federated function on Linux, UNIX, and Windows

Documentation about federated function on Linux, UNIX, and Windows

### Purpose

Documentation about federated function on Linux, UNIX, and Windows.

Table 5. DB2 Information Integrator documentation about federated function on Linux, UNIX, and Windows

Name	Form number	Location
<i>Application Developer's Guide</i>	SC18-7359	<ul style="list-style-type: none"> <li>• DB2 PDF Documentation CD</li> <li>• DB2 Information Integrator Support Web site</li> </ul>

Table 5. DB2 Information Integrator documentation about federated function on Linux, UNIX, and Windows (continued)

Name	Form number	Location
<i>C++ API Reference for Developing Wrappers</i>	SC18-9172	<ul style="list-style-type: none"> <li>• DB2 PDF Documentation CD</li> <li>• DB2 Information Integrator Support Web site</li> </ul>
<i>Data Source Configuration Guide</i>	N/A	<ul style="list-style-type: none"> <li>• DB2 PDF Documentation CD</li> <li>• DB2 Information Integrator Support Web site</li> </ul>
<i>Federated Systems Guide</i>	SC18-7364	<ul style="list-style-type: none"> <li>• DB2 PDF Documentation CD</li> <li>• DB2 Information Integrator Support Web site</li> </ul>
<i>Guide to Configuring the Content Connector for VeniceBridge</i>	N/A	DB2 Information Integrator Support Web site
<i>Installation Guide for Linux, UNIX, and Windows</i>	GC18-7036	<ul style="list-style-type: none"> <li>• DB2 PDF Documentation CD</li> <li>• DB2 Information Integrator Support Web site</li> </ul>
<i>Java API Reference for Developing Wrappers</i>	SC18-9173	<ul style="list-style-type: none"> <li>• DB2 PDF Documentation CD</li> <li>• DB2 Information Integrator Support Web site</li> </ul>
<i>Migration Guide</i>	SC18-7360	<ul style="list-style-type: none"> <li>• DB2 PDF Documentation CD</li> <li>• DB2 Information Integrator Support Web site</li> </ul>
<i>Wrapper Developer's Guide</i>	SC18-9174	<ul style="list-style-type: none"> <li>• DB2 PDF Documentation CD</li> <li>• DB2 Information Integrator Support Web site</li> </ul>
<i>Release Notes for IBM DB2 Information Integrator Standard Edition, Advanced Edition, and Replication for z/OS</i>	N/A	<ul style="list-style-type: none"> <li>• In the DB2 Information Center, <b>Product Overviews &gt; Information Integration &gt; DB2 Information Integrator overview &gt; Problems, workarounds, and documentation updates</b></li> <li>• DB2 Information Integrator Installation launchpad</li> <li>• DB2 Information Integrator Support Web site</li> <li>• The <i>DB2 Information Integrator</i> product CD</li> </ul>

## Documentation about federated function on z/OS

Documentation about federated function on z/OS



## Purpose

Documentation about federated function on z/OS.

Table 6. DB2 Information Integrator documentation about federated function on z/OS

Name	Form number	Location
<i>Client Guide for Classic Federation and Event Publisher for z/OS</i>	SC18-9160	DB2 Information Integrator Support Web site
<i>Data Mapper Guide for Classic Federation and Event Publisher for z/OS</i>	SC18-9163	DB2 Information Integrator Support Web site
<i>Getting Started with Classic Federation for z/OS</i>	GC18-9155	DB2 Information Integrator Support Web site
<i>Installation Guide for Classic Federation and Event Publisher for z/OS</i>	GC18-9301	DB2 Information Integrator Support Web site
<i>Reference for Classic Federation and Event Publisher for z/OS</i>	SC18-9156	DB2 Information Integrator Support Web site
<i>System Messages for Classic Federation and Event Publisher for z/OS</i>	SC18-9162	DB2 Information Integrator Support Web site
<i>Transaction Services Guide for Classic Federation for z/OS</i>	SC18-9161	DB2 Information Integrator Support Web site
<i>Release Notes for IBM DB2 Information Integrator Classic Federation for z/OS</i>	N/A	DB2 Information Integrator Support Web site

## Documentation about replication function on z/OS

Documentation about replication function on z/OS

## Purpose

Documentation about replication function on z/OS.

Table 7. DB2 Information Integrator documentation about replication function on z/OS

Name	Form number	Location
<i>ASNCLP Program Reference for Replication and Event Publishing</i>	N/A	DB2 Information Integrator Support Web site
<i>Introduction to Replication and Event Publishing</i>	GC18-7567	DB2 Information Integrator Support Web site
<i>Migrating to SQL Replication</i>	N/A	DB2 Information Integrator Support Web site
<i>Replication and Event Publishing Guide and Reference</i>	SC18-7568	<ul style="list-style-type: none"><li>• DB2 PDF Documentation CD</li><li>• DB2 Information Integrator Support Web site</li></ul>
<i>Replication Installation and Customization Guide for z/OS</i>	SC18-9127	DB2 Information Integrator Support Web site
<i>SQL Replication Guide and Reference</i>	SC27-1121	<ul style="list-style-type: none"><li>• DB2 PDF Documentation CD</li><li>• DB2 Information Integrator Support Web site</li></ul>
<i>Tuning for Replication and Event Publishing Performance</i>	N/A	DB2 Information Integrator Support Web site

Table 7. DB2 Information Integrator documentation about replication function on z/OS (continued)

Name	Form number	Location
<i>Tuning for SQL Replication Performance</i>	N/A	DB2 Information Integrator Support Web site
<i>Release Notes for IBM DB2 Information Integrator Standard Edition, Advanced Edition, and Replication for z/OS</i>	N/A	<ul style="list-style-type: none"> <li>• In the DB2 Information Center, <b>Product Overviews &gt; Information Integration &gt; DB2 Information Integrator overview &gt; Problems, workarounds, and documentation updates</b></li> <li>• DB2 Information Integrator Installation launchpad</li> <li>• DB2 Information Integrator Support Web site</li> <li>• The <i>DB2 Information Integrator</i> product CD</li> </ul>

## Documentation about enterprise search function on Linux, UNIX, and Windows

Documentation about enterprise search function on Linux, UNIX, and Windows

### Purpose

Documentation about enterprise search function on Linux, UNIX, and Windows.

Table 8. DB2 Information Integrator documentation about enterprise search function on Linux, UNIX, and Windows

Name	Form number	Location
<i>Administering Enterprise Search</i>	SC18-9283	DB2 Information Integrator Support Web site
<i>Installation Guide for Enterprise Search</i>	GC18-9282	DB2 Information Integrator Support Web site
<i>Programming Guide and API Reference for Enterprise Search</i>	SC18-9284	DB2 Information Integrator Support Web site
<i>Release Notes for Enterprise Search</i>	N/A	DB2 Information Integrator Support Web site

## Release notes and installation requirements

Release notes provide information that is specific to the release and fix pack level for your product and include the latest corrections to the documentation for each release. Installation requirements provide information that is specific to the release of your product.

Table 9. DB2 Information Integrator Release Notes and Installation Requirements

Name	File name	Location
<i>Installation Requirements for IBM DB2 Information Integrator Event Publishing Edition, Replication Edition, Standard Edition, Advanced Edition, Advanced Edition Unlimited, Developer Edition, and Replication for z/OS</i>	Prereqs	<ul style="list-style-type: none"> <li>The <i>DB2 Information Integrator</i> product CD</li> <li>DB2 Information Integrator Installation Launchpad</li> </ul>
<i>Release Notes for IBM DB2 Information Integrator Standard Edition, Advanced Edition, and Replication for z/OS</i>	ReleaseNotes	<ul style="list-style-type: none"> <li>In the DB2 Information Center, <b>Product Overviews &gt; Information Integration &gt; DB2 Information Integrator overview &gt; Problems, workarounds, and documentation updates</b></li> <li>DB2 Information Integrator Installation launchpad</li> <li>DB2 Information Integrator Support Web site</li> <li>The <i>DB2 Information Integrator</i> product CD</li> </ul>
<i>Release Notes for IBM DB2 Information Integrator Event Publisher for IMS for z/OS</i>	N/A	DB2 Information Integrator Support Web site
<i>Release Notes for IBM DB2 Information Integrator Event Publisher for VSAM for z/OS</i>	N/A	DB2 Information Integrator Support Web site
<i>Release Notes for IBM DB2 Information Integrator Classic Federation for z/OS</i>	N/A	DB2 Information Integrator Support Web site
<i>Release Notes for Enterprise Search</i>	N/A	DB2 Information Integrator Support Web site

## Viewing release notes and installation requirements

Viewing release notes and installation requirements

### Purpose

To view release notes and installation requirements from the CD on Windows operating systems, enter:

```
x\doc\%L
```

### Parameters

*x* The Windows CD drive letter

*%L*

The locale of the documentation that you want to use, for example, en\_US.

### Purpose

To view release notes and installation requirements from the CD on UNIX operating systems, enter:

*/cdrom/doc/%L*

#### **Parameters**

*cdrom*

The UNIX mount point of the CD

*%L*

The locale of the documentation that you want to use, for example, en\_US.

---

## **Viewing and printing PDF documentation**

Viewing and printing PDF documentation

To view and print the DB2 Information Integrator PDF books from the *DB2 PDF Documentation CD*

1. From the root directory of the *DB2 PDF Documentation CD*, open the *index.htm* file.
2. Click the language that you want to use.
3. Click the link for the document that you want to view.

---

## **Accessing DB2 Information Integrator documentation**

Accessing DB2 Information Integrator documentation

All DB2 Information Integrator books and release notes are available in PDF files from the DB2 Information Integrator Support Web site at [www.ibm.com/software/data/integration/db2ii/support.html](http://www.ibm.com/software/data/integration/db2ii/support.html).

To access the latest DB2 Information Integrator product documentation, from the DB2 Information Integrator Support Web site, click on the Product Information link, as shown in Figure 2 on page 233.

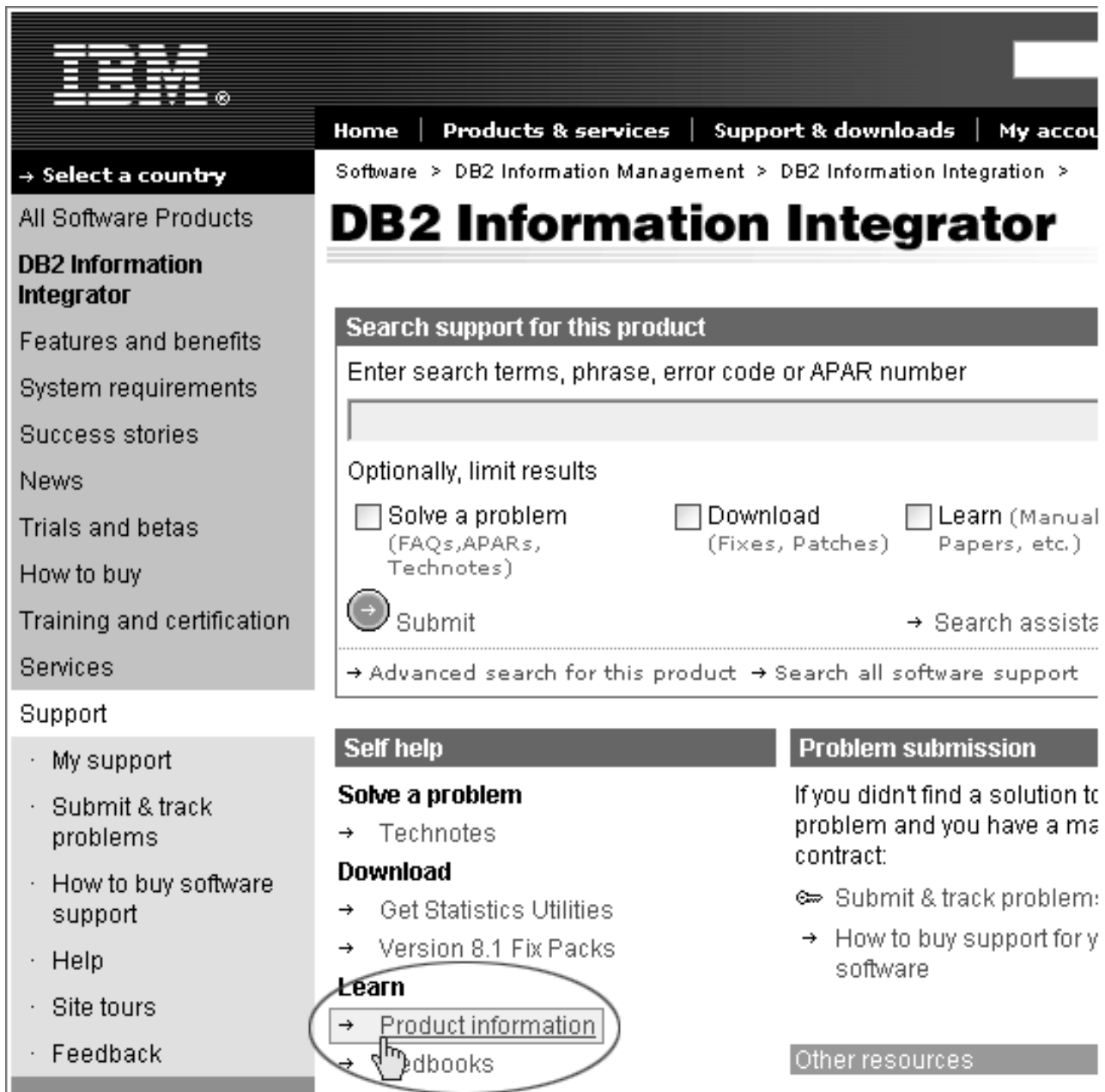


Figure 2. Product information link on the DB2 Information Integrator Support Web site

You can access the latest DB2 Information Integrator documentation, in all supported languages, from the Product Information link:

- DB2 Information Integrator product documentation in PDF files
- Fix pack product documentation, including release notes
- Instructions for downloading and installing the DB2 Information Center for Linux, UNIX, and Windows
- Links to the DB2 Information Center online

The DB2 Information Integrator Support Web site also provides support documentation, IBM Redbooks, white papers, product downloads, links to user groups, and news about DB2 Information Integrator.



---

## Accessibility

Accessibility features help users with physical disabilities, such as restricted mobility or limited vision, to use software products successfully. The following list specifies the major accessibility features in DB2<sup>®</sup> Version 8 products:

- All DB2 functionality is available using the keyboard for navigation instead of the mouse. For more information, see “Keyboard input and navigation.”
- You can customize the size and color of the fonts on DB2 user interfaces. For more information, see “Accessible display.”
- DB2 products support accessibility applications that use the Java<sup>™</sup> Accessibility API. For more information, see “Compatibility with assistive technologies” on page 236.
- DB2 documentation is provided in an accessible format. For more information, see “Accessible documentation” on page 236.

---

## Keyboard input and navigation

### Keyboard focus

Keyboard focus

In UNIX<sup>®</sup> operating systems, the area of the active window where your keystrokes will have an effect is highlighted.

### Keyboard input

Keyboard input

You can operate the DB2 tools using only the keyboard. You can use keys or key combinations to perform operations that can also be done using a mouse. Standard operating system keystrokes are used for standard operating system operations.

For more information about using keys or key combinations to perform operations, see Keyboard shortcuts and accelerators: Common GUI help.

### Keyboard navigation

Keyboard navigation

You can navigate the DB2 tools user interface using keys or key combinations.

For more information about using keys or key combinations to navigate the DB2 Tools, see Keyboard shortcuts and accelerators: Common GUI help.

---

## Accessible display

Accessible display

### Purpose

Accessible display

## Font settings

### Font settings

You can select the color, size, and font for the text in menus and dialog windows, using the Tools Settings notebook.

For more information about specifying font settings, see [Changing the fonts for menus and text: Common GUI help](#).

## Non-dependence on color

### Non-dependence on color

You do not need to distinguish between colors to use any of the functions in this product.

---

## Compatibility with assistive technologies

### Compatibility with assistive technologies

The DB2 tools interfaces support the Java Accessibility API, which enables you to use screen readers and other assistive technologies with DB2 products.

---

## Accessible documentation

### Accessible documentation

Documentation for DB2 is provided in XHTML 1.0 format, which is viewable in most Web browsers. XHTML allows you to view documentation according to the display preferences set in your browser. It also allows you to use screen readers and other assistive technologies.

Syntax diagrams are provided in dotted decimal format. This format is available only if you are accessing the online documentation using a screen-reader.



---

## Notices

This information was developed for products and services offered in the U.S.A. IBM may not offer the products, services, or features discussed in this document in all countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not give you any license to these patents. You can send license inquiries, in writing, to: IBM Director of Licensing IBM Corporation North Castle Drive Armonk, NY 10504-1785 U.S.A.

For license inquiries regarding double-byte (DBCS) information, contact the IBM Intellectual Property Department in your country/region or send inquiries, in writing, to: IBM World Trade Asia Corporation Licensing 2-31 Roppongi 3-chome, Minato-ku Tokyo 106-0032, Japan

The following paragraph does not apply to the United Kingdom or any other country/region where such provisions are inconsistent with local law: INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY, OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions; therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this IBM product, and use of those Web sites is at your own risk.

IBM may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Licensees of this program who wish to have information about it for the purpose of enabling: (i) the exchange of information between independently created programs and other programs (including this one) and (ii) the mutual use of the information that has been exchanged, should contact:

IBM Corporation J46A/G4  
555 Bailey Avenue  
San Jose, CA 95141-1003 U.S.A.

Such information may be available, subject to appropriate terms and conditions, including in some cases payment of a fee.

The licensed program described in this document and all licensed material available for it are provided by IBM under terms of the IBM Customer Agreement, IBM International Program License Agreement, or any equivalent agreement between us.

Any performance data contained herein was determined in a controlled environment. Therefore, the results obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems, and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurements may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements, or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility, or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

All statements regarding IBM's future direction or intent are subject to change or withdrawal without notice, and represent goals and objectives only.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious, and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

#### COPYRIGHT LICENSE:

This information contains sample application programs, in source language, which illustrate programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to IBM for the purposes of developing, using, marketing, or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. IBM, therefore, cannot guarantee or imply reliability, serviceability, or function of these programs. You may copy, modify, and distribute these sample programs in any form without payment to IBM for the purposes of developing, using, marketing, or distributing application programs conforming to IBM's application programming interfaces.

Each copy or any portion of these sample programs or any derivative work must include a copyright notice as follows:

Outside In (®) Viewer Technology, © 1992-2004 Stellent, Chicago, IL., Inc. All Rights Reserved.

## Trademarks

This topic lists IBM trademarks and certain non-IBM trademarks.

The following terms are trademarks of International Business Machines Corporation in the United States, other countries, or both:

IBM  
AIX  
AIX 5L  
DB2  
DB2 Universal Database  
Domino  
Informix  
Lotus  
Lotus Notes  
Notes  
OmniFind  
WebSphere  
z/OS

The following terms are trademarks or registered trademarks of other companies:

Java and all Java-based trademarks and logos are trademarks or registered trademarks of Sun Microsystems, Inc. in the United States, other countries, or both.

Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

Intel, Intel Inside (logos), MMX and Pentium are trademarks of Intel Corporation in the United States, other countries, or both.

UNIX is a registered trademark of The Open Group in the United States and other countries.

Linux is a trademark of Linus Torvalds in the United States, other countries, or both.

Other company, product or service names may be trademarks or service marks of others.



---

## Contacting IBM

To contact IBM customer service in the United States or Canada, call 1-800-IBM-SERV (1-800-426-7378).

To learn about available service options, call one of the following numbers:

- In the United States: 1-888-426-4343
- In Canada: 1-800-465-9600

To locate an IBM office in your country or region, see the IBM Directory of Worldwide Contacts on the Web at [www.ibm.com/planetwide](http://www.ibm.com/planetwide).



---

## Obtaining product information

Information about DB2 Information Integrator is available by telephone or on the Web.

Information about DB2 Information Integrator is available by telephone or on the Web. The phone numbers provided here are valid in the United States.

1. To order products or to obtain general information: 1-800-IBM-CALL (1-800-426-2255)
2. To order publications: 1-800-879-2755
3. Visit the Web at [www.ibm.com/software/data/integration/db2ii/support.html](http://www.ibm.com/software/data/integration/db2ii/support.html).

This site contains the latest information about:

- The technical library
- Ordering books
- Client downloads
- Newsgroups
- Fix packs
- News
- Links to Web resources





---

## Providing comments on the documentation

Please send any comments that you have about this book or other DB2 Information Integrator documentation.

Your feedback helps IBM to provide quality information. Please send any comments that you have about this book or other DB2 Information Integrator documentation. You can use any of the following methods to provide comments:

1. Send your comments using the online readers' comment form at [www.ibm.com/software/data/rcf](http://www.ibm.com/software/data/rcf).
2. Send your comments by e-mail to [comments@us.ibm.com](mailto:comments@us.ibm.com). Include the name of the product, the version number of the product, and the name and part number of the book (if applicable). If you are commenting on specific text, please include the location of the text (for example, a title, a table number, or a page number).



---

# Index

## A

- access controls
  - description 114
  - disabling for a collection 118
  - document-level security 114, 121
  - Portal Search Engine support 121
- active Web sites, monitoring 98, 100
- address rules for Web crawlers 45
- administration console
  - description 7
  - interface 13
  - logging in 15
  - task summary 13
- administrative roles
  - collection administrator 112, 113
  - configuring 113
  - description 112
  - enterprise search administrator 112, 113
  - monitor 112, 113
  - operator 112, 113
- administrator password, changing 16
- AIX
  - configuring Content Manager crawlers 28
  - configuring Notes crawlers 36
  - configuring VeniceBridge crawlers 42
- alerts
  - collection-level 94
  - description 93
  - documents crawled 94
  - documents indexed 94
  - e-mail options 94, 95
  - free space on servers 95
  - receiving e-mail for 134
  - search response times 94
  - system-level 95
- anchor text analysis
  - anchor text-only documents 116
  - description 111
  - enabling collection security 115
  - global analysis 115
- anchor text-only documents
  - description 115
  - enabling 116
  - robots.txt file 115
- APIs
  - data listener 1, 9
  - description 9
  - search and index 9, 59
- authentication, description 114

## B

- backing up enterprise search 129
- backup scripts
  - description 129
  - running 129, 130

## C

- categories
  - categorization type 78
  - category trees 77
  - creating 79
  - description 75
  - migrating from WebSphere Portal 123, 124
  - model-based 77, 78
  - nesting subcategories 77
  - rule-based 75, 78
  - searching 75
  - URI formats 139
- categorization type
  - model-based 77
  - rule-based 75
  - selecting 20, 78
- category rules
  - configuring 79
  - document content 75, 79
  - URI patterns 75, 79
- category trees
  - description 77
  - migrating from WebSphere Portal 123, 124
- character normalization 66
- clitics 67
- collapsing search results 86
- collection administrator
  - description 112
  - role configuration 113
- Collection wizard 19
- collection-level security
  - anchor text analysis 115
  - application IDs 117
  - description 111, 115
  - duplicate document analysis 117
  - enabling 20
- collections
  - allowing all users to access 118
  - anchor text security 115
  - application ID security 117
  - associating with search applications 63
  - creating with Collection wizard 19
  - creating with Collections view 20
  - default migration settings 126
  - deleting 22
  - description 2
  - draft 19
  - duplicate document security 117
  - editing 21
  - migrating from WebSphere Portal 123, 124
  - MigrationWizard.log file 124
  - monitoring 96
  - searching 71
  - security 115
  - system status 96
  - ways to create 19
- Collections view
  - creating collections 20
  - description 13
- com.ibm.es.util.EncryptProperty command 16
- concurrent index builds 56
- config.properties file 61
- Content Manager crawlers
  - configuring 27
  - setting up on AIX 28
  - setting up on Linux 30
  - setting up on Windows 31
- Crawl page, description 13
- crawl rate, monitoring 101
- crawl space
  - alerts about 94
  - description 3
  - editing 26
  - Web crawler configuration 45
- crawler history reports
  - creating 101
  - description 98
  - HTTP return code report 101
  - Site report 101
- crawler properties
  - description 3
  - editing 26
- crawler types
  - combining in a collection 23
  - selecting base values for 23
- crawlers
  - base values for 23
  - combining crawler types 23
  - configuration overview 23
  - Content Manager 27
  - creating 25
  - data listener configuration 50
  - DB2 32
  - default migration settings 126
  - deleting 27
  - description 3
  - editing crawl spaces 26
  - editing crawler properties 26
  - enabling document-level security 23
  - Exchange Server 33
  - initial values for 25
  - monitoring 97
  - NNTP 34
  - Notes 35
  - scheduling 23
  - starting 89, 97
  - stopping 90, 97
  - support for external 9, 50
  - system status 97
  - UNIX file system 40
  - URI formats 139
  - VeniceBridge 41
  - Web 44
  - Windows file system 50
- creating
  - collections 19, 20

- creating (*continued*)
  - crawlers 25
  - quick links 74
  - rule-based categories 79
  - scopes 81
  - Web crawler reports 101
  - XML search fields 82

## D

- data flow, enterprise search system 10
- data listener API 1, 9
- data listener, configuring for external crawlers 50
- data source types
  - DB2 Content Manager item types 27
  - DB2 databases 32
  - Exchange Server public folders 33
  - NNTP news groups 34
  - Notes databases 35
  - relational databases 32
  - support for external 1, 9
  - supported by enterprise search 1
  - UNIX file systems 40
  - VeniceBridge item classes 41
  - Web sites 44
  - Windows file systems 50
- DB2 crawlers, configuring 32
- DB2 II OmniFind
  - administration console 7
  - APIs 9
  - components 2
  - crawler servers 3
  - data flow diagram 10
  - index servers 5
  - integration with WebSphere Portal 119
  - overview 1
  - parsers 4
  - portlet deployment 121
  - search applications 9
  - search servers 6
- Default search application 61
  - deleting
    - collections 22
    - crawlers 27
    - indexes from the queue 106
  - dictionary-based analysis 65, 67
  - dictionary-based segmentation 67
  - DIIOP protocol, Notes crawler configuration 39
  - disabling index schedules 55
  - discovery 3
  - dlAuthentication.prp file 50
  - dlAuthorization.prp file 50
  - dlConfig.prp file 50
  - document importance
    - enabling for a collection 20
    - in migrated collections 124
  - Document Search portlet 121
  - document-level security
    - access controls 114
    - crawler configuration 23
    - description 111, 114
    - Portal Search Engine support 121
    - security tokens 114
  - domain rules for Web crawlers 45

- duplicate document analysis
  - description 111
  - enabling security 117
  - global analysis 117
- dynamic ranking 85
- dynamic summarization 86

## E

- e-mail notifications
  - for alerts 134
  - for messages 134
- editing
  - collections 21
  - crawl spaces 26
  - crawler properties 26
  - data listener properties 50
  - search application properties 61
- enabling index schedules 55
- EncryptProperty command 16
- enterprise search
  - administration console 7
  - administrative roles 112
  - APIs 9
  - backing up 129
  - components 2
  - configuring Portal Search Engine portlets for 121
  - crawler servers 3
  - data flow diagram 10
  - index servers 5
  - integration with WebSphere Portal 119
  - linguistic support 65
  - log files 133
  - monitoring 93
  - overview 1
  - parsers 4
  - portlet deployment 121
  - portlets 119
  - restoring 129
  - restoring from a backup 130
  - restoring system files 130
  - search applications 9
  - search servers 6
  - security 111
  - starting the components 89
  - stopping the components 89, 90
  - URI formats 139
- enterprise search administrator
  - changing the password for 16
  - description 112
  - role configuration 113
- error messages
  - receiving e-mail for 133, 134
  - viewing details 136
- ES\_INSTALL\_ROOT, description 16
- ES\_NODE\_ROOT, description 16
- es.cfg file 16
- ESAdmin application 15
- esbackup.bat script 129, 130
- esbackup.sh script 129, 130
- esrcm.sh script 28, 30
- esrcm.vbs script 31
- escrnote.sh script 36
- escrnte.vbs script 38
- escrvbr.sh script 42

- escrvbr.vbs script 43
- esrestore.bat script 130
- esrestore.sh script 130
- ESSearchApplication application 61, 62
- Exchange Server crawlers, configuring 33
- external crawlers
  - configuring 50
  - data listener API 9
  - data listener configuration 50

## F

- fielded queries 71
- file extensions, excluding from Web crawl space 45
- free space alerts 95
- free-text queries 71

## G

- global analysis
  - anchor text analysis 111, 115
  - description 5
  - duplicate document analysis 111, 117

## H

- HTTP return codes
  - received by Web crawlers 101
  - robots.txt file restrictions 115
  - Web crawler report 101

## I

- index builds
  - starting 106
  - stopping 106
  - system status 106
- Index page, description 13
- index queue 106
- index refresh
  - description 5, 53
  - scheduling 54, 55
- index reorganization
  - description 5, 53
  - scheduling 54, 55
- indexes
  - alerts about 94
  - changing the schedule 55
  - concurrent builds 56
  - deleting from the queue 106
  - description 5, 53
  - disabling the schedule 55, 106
  - enabling the schedule 55, 106
  - monitoring 106
  - removing URIs 57
  - scheduling 54
  - starting 89
  - stopping 90
- IP address rules for Web crawlers 45

## J

Java connector for Content Manager 28, 30, 31

## K

keywords in quick links 73, 74

## L

language detection 66

lemmas 67

lemmatization 65, 67

limiting the Web crawl space 45

linguistic support

character normalization 66

clitics 67

description 65

dictionary-based analysis 65

dictionary-based segmentation 67

language detection 66

lemmas 67

lemmatization 65, 67

n-gram segmentation 68

nondictionary-based analysis 65

nondictionary-based segmentation 68

Okurigana variants 69

orthographic variants in Japanese 69

stop word removal 66

supported languages 67

Unicode normalization 66

Unicode-based white space segmentation 68

word segmentation in Japanese 69

Linux

configuring Content Manager

crawlers 30

configuring Notes crawlers 36

configuring VeniceBridge crawlers 42

log files

default location 133

description 133

e-mail options 134

filtering 136

maximum size 133

migration wizard 127

monitoring 136

rotating 133

severity levels 133

viewing 136

Log page, description 13

logging in to the administration console 15

## M

mapping XML search fields 82

maximum crawl depth 45

maximum recrawl interval 49

migrating

collections 124

model-based taxonomy 123

rule-based taxonomy 124

migration wizard

collections 124

migration wizard (*continued*)

default collection settings 126

default crawler settings 126

description 123

log file 127

model-based taxonomies 123

rule-based taxonomies 124

starting 123, 124

MIME types, including in Web crawl

space 45

minimum recrawl interval 49

model-based categories

description 77

selecting the categorization type 78

model-based taxonomy, migrating from

WebSphere Portal 123

monitor

description 112

role configuration 113

Monitor views, description 13

monitoring

collections 96

crawlers 97

enterprise search 93

log files 136

parsers 105

popular queries 107

recent queries 107

response time history 107

search servers 107

Web crawler active sites 100

Web crawler crawl rate 101

Web crawler thread details 99

Web crawlers 98

## N

n-gram segmentation 68

NNTP crawlers, configuring 34

nondictionary-based analysis 65, 68

nondictionary-based segmentation 68

Notes crawlers

configuring 35

DIIOP protocol configuration 39

NRPC protocol 36, 38

setting up on AIX 36

setting up on Linux 36

setting up on Windows 38

NRPC protocol, Notes crawler

configuration 36, 38

## O

Okurigana variants 69

operator

description 112

role configuration 113

orthographic variants in Japanese 69

## P

parametric queries 71

Parse page, description 13

parsers

data analysis tasks 4

description 4

parsers (*continued*)

monitoring 105

starting 89, 105

stopping 90, 105

system status 105

password, enterprise search

administrator 16

popular queries, monitoring 107

Portal Search Engine

configuring for enterprise search 121

description 119

document-level security 121

passing security tokens 121

portlets

catalog Web site 121

deploying 121

description 119

enterprise search 119

supported versions of WebSphere

Portal 119

Taxonomy Management Portlet 123

prefix rules for Web crawlers 45

## Q

queries

advanced techniques 71

fielded search 71

free-text search 71

parametric 71

spell checking 71

quick links

creating 74

description 73

searching 73

URI formats 139

## R

ranking

dynamic 85

static 85

text-based scoring 85

recent queries, monitoring 107

recently crawled URLs, monitoring 98

recrawl intervals for Web crawlers 49

refreshing indexes 5, 53

removing URIs from an index 57

reorganizing indexes 5, 53

response time history, monitoring 107

restore scripts

description 129

running 130

restoring

from a backup 130

system files to new servers 130

revisiting URLs as soon as possible 49

robots.txt file

anchor text-only documents 115

HTTP return codes 115

rule-based categories

creating 79

description 75

selecting the categorization type 78

rule-based taxonomy, migrating from

WebSphere Portal 123, 124

## S

- sample search application
  - accessing 62
  - default deployment 61
  - description 9
  - property file 61
  - search functions 59
- scheduling
  - crawlers 23
  - index builds 54, 55
- scopes
  - creating 81
  - description 80
  - searching 80
  - URI formats 139
  - URI patterns 80, 81
- scripts
  - esbackup.bat 129, 130
  - esbackup.sh 129, 130
  - escrcm.sh 28, 30
  - escrcm.vbs 31
  - escnote.sh 36
  - escrnte.vbs 38
  - escrvbr.sh 42
  - escrvbr.vbs 43
  - esrestore.bat 130
  - esrestore.sh 130
  - startcl 130
- search and index API 9, 59
- search applications
  - application IDs 117
  - associating with collections 63
  - collection-level security 117
  - custom 59
  - description 9
  - property file 61
  - sample 59
- search cache
  - configuring 109
  - description 109
- Search Center
  - description 119
  - registration portlet 119
- Search page, description 13
- search response time
  - alerts about 94
  - monitoring 107
- search results
  - collapsing 86
  - description 85
  - dynamic ranking 85
  - dynamic summarization 86
  - grouping 86
  - static ranking 85
  - text-based scoring 85
- search servers
  - description 6
  - monitoring 107
  - popular queries 107
  - recent queries 107
  - response time history 107
  - search cache 109
  - starting 89, 107
  - stopping 90, 107
  - system status 107
- searching
  - categories 75

- searching (*continued*)
  - collections 71
  - quick links 73
  - scopes 80
  - XML documents 82
- security
  - access control 114
  - administrative roles 113
  - anchor text analysis 115
  - authentication 114
  - collection-level 115, 118
  - description 111
  - document-level 114
  - duplicate document analysis 117
  - enabling for a collection 20, 111
  - search application IDs 117
- security tokens
  - crawler configuration 114
  - disabling for a collection 118
  - document-level security 114, 121
  - Portal Search Engine processing 121
- Security view, description 13
- seed URLs for Web crawlers 45, 49
- segmentation
  - dictionary-based 67
  - nondictionary-based 68
  - Unicode-based white space 68
- SIAPI (search and index API) 9, 59
- site history reports
  - creating 101
  - description 98
- Site report, creating for Web crawler 101
- SiteDefs.txt file 86
- spell checking in queries 71
- startcl script 130
- starting
  - crawlers 97
  - enterprise search components 89
  - index builds 106
  - migration wizard 123, 124
  - parsers 105
  - search servers 107
- static ranking
  - description 85
  - enabling for a collection 20
  - in migrated collections 124
- stop word removal 66
- stop words 66
- stopping
  - crawlers 97
  - enterprise search components 89, 90
  - index builds 106
  - parsers 105
  - search servers 107
- summarization, dynamic 86
- supported languages
  - dictionary-based linguistic processing 67
  - language detection 66
- synonyms.xml file 123
- system backup 129
- system restore 129, 130
- system status
  - collections 96
  - crawlers 97
  - index builds 106
  - parsers 105

- system status (*continued*)
  - search servers 107
  - Web crawlers 98
- System view, description 13

## T

- task summary, administration
  - console 13
- taxonomies, migrating from WebSphere Portal 123, 124
- Taxonomy Management Portlet 123
- text-based scoring 85
- thread details, monitoring 98, 99
- titles.xml file 123
- treenodes.xml file 123

## U

- Unicode normalization 66
- Unicode-based white space segmentation 68
- UNIX file system crawlers, configuring 40
- URI
  - category rules 75, 79
  - formats in enterprise search 139
  - quick links 73, 74
  - removing from an index 57
  - scopes 80, 81
  - viewing details about 96

## V

- vbr\_access\_services.jar file 42, 43
- VeniceBridge crawlers
  - configuring 41
  - setting up on AIX 42
  - setting up on Linux 42
  - setting up on Windows 43
- viewing
  - log files 136
  - URI details 96
- visiting URLs as soon as possible 49

## W

- Web crawlers
  - active sites 98, 100
  - configuring 44
  - crawl rate 101
  - crawler history 98
  - crawling rules 45
  - creating reports about 101
  - limiting the crawl space 45
  - monitoring 98
  - recently crawled URLs 98
  - recrawl intervals 49
  - seed URLs 45, 49
  - site history 98
  - system status 98
  - thread details 98, 99
  - URL status 98
  - visiting URLs as soon as possible 49

- WebSphere Portal
  - catalog Web site 121
  - category tree migration 123
  - collection migration 123
  - default migration settings 126
  - integration with enterprise search 119
  - model-based taxonomies 123
  - portlet deployment 121
  - Taxonomy Management Portlet 123
  - taxonomy migration 123
- WebSphere Portal Search Center 119
  - description 119
  - registration portlet 119
- Windows
  - configuring Content Manager crawlers 31
  - configuring Notes crawlers 38
  - configuring VeniceBridge crawlers 43
- Windows file system crawlers, configuring 50
- word segmentation, Japanese 69
- WpsMigratorLog.log file 127

## X

- XML documents, searching 82
- XML search fields
  - creating 82
  - description 82
  - mapping elements to 82









Printed in USA

SC18-9283-00



Spine information:



IBM DB2 Information Integrator  
OmniFind Edition

Administering Enterprise Search

Version 8.2