

IBM DB2 Information Integrator  
OmniFind Edition



# Administering Enterprise Search

*Version 8.2.2*



IBM DB2 Information Integrator  
OmniFind Edition



# Administering Enterprise Search

*Version 8.2.2*

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, 2005. All rights reserved.

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

# Contents

## About these topics . . . . . **xiii**

Who should read these topics . . . . . xiii

## 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 . . . . . 7

Enterprise search administration console . . . . . 8

Monitoring an enterprise search system . . . . . 9

Enterprise search log files . . . . . 9

Customizing enterprise search . . . . . 10

Sample search application for enterprise search . . . . . 11

The enterprise search data flow . . . . . 12

## Enterprise search administration overview . . . . . **15**

Logging in to the administration console . . . . . 18

Changing the enterprise search administrator password in a single server configuration . . . . . 19

Changing the enterprise search administrator password in a multiple server configuration . . . . . 20

## Enterprise search collections . . . . . **25**

Creating a collection by using the Collection wizard . . . . . 25

Creating a collection by using the Collections view . . . . . 26

Editing a collection . . . . . 28

Deleting a collection . . . . . 29

## Enterprise search crawler administration . . . . . **31**

Creating a crawler . . . . . 33

Editing crawler properties . . . . . 34

Editing a crawl space . . . . . 34

Deleting a crawler . . . . . 35

Content Manager crawlers . . . . . 36

Configuring the crawler server on AIX for DB2 Content Manager . . . . . 37

Configuring the crawler server on Linux for DB2 Content Manager . . . . . 38

Configuring the crawler server on Windows for DB2 Content Manager . . . . . 39

DB2 crawlers . . . . . 41

Exchange Server crawlers . . . . . 42

NNTP crawlers . . . . . 43

Notes crawlers . . . . . 44

Configuring the crawler server on UNIX for Lotus Notes . . . . . 45

Configuring the crawler server on Windows for Lotus Notes . . . . . 47

Configuring the DIIOP protocol for Notes crawlers . . . . . 49

UNIX file system crawlers . . . . . 49

VeniceBridge crawlers . . . . . 50

Configuring the crawler server on UNIX for VeniceBridge . . . . . 52

Configuring the crawler server on Windows for VeniceBridge . . . . . 53

Web crawlers . . . . . 54

User agent configuration . . . . . 55

Support for JavaScript . . . . . 56

Rules to limit the Web crawl space . . . . . 56

Recrawl interval settings in the Web crawler . . . . . 60

Options for visiting URLs with the Web crawler . . . . . 60

How the Web crawler handles soft error pages . . . . . 60

Web sites served by proxy servers . . . . . 62

Support for crawling secure Web sites . . . . . 63

Cookie administration . . . . . 65

Global Web crawl space configuration . . . . . 66

No-follow and no-index directives . . . . . 68

Overriding no-follow and no-index directives in Web pages . . . . . 68

Windows file system crawlers . . . . . 69

Configuring support for Data Listener applications . . . . . 70

File types supported by enterprise search crawlers . . . . . 71

URI formats in an enterprise search index . . . . . 71

## Parsing documents for enterprise search . . . . . **77**

Working with categories . . . . . 77

Rule-based categories . . . . . 79

Model-based categories . . . . . 80

Category trees . . . . . 81

Selecting the categorization type . . . . . 82

Configuring categories . . . . . 83

Working with XML search fields . . . . . 84

XML search fields . . . . . 84

Mapping XML elements to search fields . . . . . 85

Working with HTML search fields . . . . . 87

HTML search fields . . . . . 87

Mapping HTML metadata elements to search fields . . . . . 87

Custom text analysis . . . . . 88

Associating analysis engines with the system . . . . . 89

Associating an analysis engine with a collection . . . . . 90

Mapping XML elements to a common analysis structure . . . . . 91

Indexing the analysis results . . . . . 92

Enabling support for native XML search . . . . . 94

Linguistic analysis of Chinese, Japanese, and Korean documents . . . . . 94

N-gram segmentation . . . . . 95

Removing new line characters from white space . . . . . 95

Document types associated with collection parsers and Stellent sessions . . . . . 96

Associating document types with a collection parser . . . . . 96

	Default collection parser service rules . . . . .	98
	Associating document types with a Stellent session . . . . .	98
	Default parsing rules for Stellent sessions . . . . .	101

**Enterprise search index**

**administration . . . . . 103**

Scheduling index builds . . . . .	104
Changing the index schedule . . . . .	105
Enabling and disabling the index schedules . . . . .	105
Configuring concurrent index builds . . . . .	106
Options that influence the searchable view of the index . . . . .	107
Wildcard characters in queries. . . . .	108
Configuring support for wildcard characters in queries . . . . .	109
Scopes. . . . .	110
Configuring scopes . . . . .	110
Collapsed URIs. . . . .	111
Collapsing URIs in the search results . . . . .	112
Removing URIs from the index . . . . .	113

**Search server administration. . . . . 115**

Search cache. . . . .	115
Configuring a search cache . . . . .	116
Custom synonym dictionaries . . . . .	116
Synonym support in search applications . . . . .	118
Creating an XML file for synonyms . . . . .	119
Creating a synonym dictionary . . . . .	120
Associating synonym dictionaries with the system . . . . .	120
Associating a synonym dictionary with a collection. . . . .	121
Working with quick links . . . . .	121
Quick links . . . . .	121
Configuring quick links . . . . .	122

**Search applications for enterprise search . . . . . 125**

Associating search applications with collections	126
Sample search application functions . . . . .	126
Editing sample search application properties . . . . .	128
Accessing the sample search application . . . . .	129
Enabling security for the sample search application	130

**Search results from enterprise search 133**

Query techniques . . . . .	133
Text-based scoring. . . . .	135
Static ranking . . . . .	136
Dynamic summarization . . . . .	136
Customizing summaries in search results . . . . .	137

**Enterprise search security. . . . . 139**

Administrative roles . . . . .	140
Configuring administrative users . . . . .	141
Authentication versus access control. . . . .	142
Disabling security for an enterprise application in WebSphere Application Server. . . . .	143
Collection-level security . . . . .	144

Duplicate document analysis . . . . .	144
Anchor text analysis . . . . .	145
Indexing the anchor text in links to forbidden documents . . . . .	145
Security with search application IDs. . . . .	146
Document-level security. . . . .	147
Validation by stored security tokens . . . . .	147
Validation of current credentials during query processing . . . . .	148
User profiles. . . . .	149
Configuring support for identity management	150
Configuring LDAP server properties . . . . .	151
Disabling document-level security . . . . .	153

**Enterprise search integration with WebSphere Portal . . . . . 155**

Deploying the enterprise search portlets in WebSphere Portal . . . . .	157
Configuring the Portal Search Engine to search enterprise search collections . . . . .	157
Document-level security with the Portal Search Engine . . . . .	157

**Migration from WebSphere Portal to enterprise search. . . . . 159**

Migrating a model-based taxonomy from WebSphere Portal . . . . .	159
Migrating a collection from WebSphere Portal . . . . .	160
Migrated collection settings. . . . .	162
Migration wizard log file . . . . .	163

**Starting and stopping enterprise search . . . . . 165**

Starting the enterprise search components . . . . .	165
Stopping the enterprise search components . . . . .	167

**Monitoring enterprise search activity 169**

Estimating the number of documents in a collection. . . . .	169
Checking the availability of system resources. . . . .	170
Monitoring a collection . . . . .	171
Viewing details about a URI . . . . .	172
Monitoring crawlers . . . . .	173
Viewing details about Web crawler activity . . . . .	174
Web crawler thread details . . . . .	175
Web crawler active sites . . . . .	175
Web crawler crawl rate . . . . .	176
Creating Web crawler reports . . . . .	177
Web crawler HTTP return codes . . . . .	178
Monitoring the parser . . . . .	181
Monitoring index activity for a collection . . . . .	182
Monitoring the enterprise search index queue . . . . .	183
Monitoring the Search servers . . . . .	184
Monitoring the Data Listener . . . . .	185

**Enterprise search log file administration . . . . . 187**

Alerts . . . . .	187
Configuring collection-level alerts . . . . .	188

Configuring system-level alerts . . . . .	189
Configuring log files . . . . .	190
Configuring SMTP server information . . . . .	191
Receiving e-mail about logged messages . . . . .	192
Viewing log files . . . . .	194

**Backup and restore for enterprise search . . . . . 197**

Backing up the enterprise search system . . . . .	197
Restoring the enterprise search system . . . . .	198
Restoring enterprise search system files to new servers . . . . .	198

**Messages for enterprise search . . . . . 201**

Controller messages (FFQC) . . . . .	201
FFQC0008E . . . . .	201
FFQC0020I . . . . .	201
FFQC0023E . . . . .	201
FFQC0027E . . . . .	202
FFQC0032E . . . . .	202
FFQC0641E . . . . .	202
FFQC1000E . . . . .	202
FFQC1002E . . . . .	202
FFQC1003E . . . . .	203
FFQC1008E . . . . .	203
FFQC1010E . . . . .	203
FFQC1018E . . . . .	203
FFQC1023E . . . . .	204
FFQC1041W . . . . .	204
FFQC1042E . . . . .	204
FFQC1043E . . . . .	204
FFQC1044E . . . . .	204
FFQC1046W . . . . .	205
FFQC1072W . . . . .	205
FFQC2034E . . . . .	205
FFQC2605E . . . . .	205
FFQC2607W . . . . .	206
FFQC2611W . . . . .	206
FFQC2613W . . . . .	206
FFQC2616W . . . . .	206
FFQC3025E . . . . .	207
FFQC3026E . . . . .	207
FFQC4815E . . . . .	207
FFQC4838E . . . . .	207
FFQC5201E . . . . .	208
FFQC5217E . . . . .	208
FFQC5228W . . . . .	208
FFQC5233E . . . . .	208
FFQC5234W . . . . .	208
FFQC5235E . . . . .	209
FFQC5238E . . . . .	209
FFQC5800E . . . . .	209
FFQC5801E . . . . .	209
FFQC5802E . . . . .	209
FFQC5803E . . . . .	210
FFQC5804E . . . . .	210
FFQC5805E . . . . .	210
FFQC5806E . . . . .	210
FFQC5807E . . . . .	211
FFQC5808E . . . . .	211

FFQC5809E . . . . .	211
FFQC5810W . . . . .	211
FFQC5811W . . . . .	212
FFQC5812E . . . . .	212
FFQC5813E . . . . .	212
FFQC5814E . . . . .	213
FFQC5815E . . . . .	213
FFQC5816E . . . . .	213
FFQC5817E . . . . .	213
FFQC5818E . . . . .	213
FFQC5819E . . . . .	214
FFQC5820E . . . . .	214
FFQC5821E . . . . .	214
FFQC5822E . . . . .	214
FFQC5823E . . . . .	214
Crawler messages (FFQD) . . . . .	215
FFQD1003E . . . . .	215
FFQD1011E . . . . .	215
FFQD1031E . . . . .	215
FFQD1034E . . . . .	215
FFQD1035E . . . . .	216
FFQD1036E . . . . .	216
FFQD1037E . . . . .	216
FFQD1101E . . . . .	216
FFQD1106E . . . . .	216
FFQD1107E . . . . .	217
FFQD1108E . . . . .	217
FFQD1123E . . . . .	217
FFQD1128E . . . . .	217
FFQD1138E . . . . .	217
FFQD1139E . . . . .	218
FFQD1142E . . . . .	218
FFQD1407E . . . . .	218
FFQD1420E . . . . .	218
FFQD2005E . . . . .	218
FFQD2007E . . . . .	219
FFQD2100E . . . . .	219
FFQD2102E . . . . .	219
FFQD2103E . . . . .	219
FFQD2104E . . . . .	219
FFQD2105E . . . . .	220
FFQD2106E . . . . .	220
FFQD2107E . . . . .	220
FFQD2108E . . . . .	220
FFQD2109E . . . . .	220
FFQD2130E . . . . .	221
FFQD2131E . . . . .	221
FFQD2132E . . . . .	221
FFQD2133E . . . . .	221
FFQD2135E . . . . .	221
FFQD2136E . . . . .	222
FFQD2137E . . . . .	222
FFQD2160E . . . . .	222
FFQD2161E . . . . .	222
FFQD2162E . . . . .	222
FFQD2163E . . . . .	223
FFQD2164E . . . . .	223
FFQD2190E . . . . .	223
FFQD2191E . . . . .	223
FFQD2210E . . . . .	223
FFQD2211E . . . . .	224

FFQD2212I . . . . .	224	FFQD2909E . . . . .	235
FFQD2213E . . . . .	224	FFQD2910I . . . . .	235
FFQD2214E . . . . .	224	FFQD2911E . . . . .	235
FFQD2215E . . . . .	224	FFQD2912W . . . . .	236
FFQD2216E . . . . .	225	FFQD2913I . . . . .	236
FFQD2217E . . . . .	225	FFQD2914I . . . . .	236
FFQD2218E . . . . .	225	FFQD2915I . . . . .	236
FFQD2240E . . . . .	225	FFQD2916I . . . . .	236
FFQD2241E . . . . .	225	FFQD2917E . . . . .	237
FFQD2242E . . . . .	226	FFQD2997E . . . . .	237
FFQD2243E . . . . .	226	FFQD2998E . . . . .	237
FFQD2244E . . . . .	226	FFQD2999E . . . . .	237
FFQD2245E . . . . .	226	FFQD3000I . . . . .	237
FFQD2246E . . . . .	226	FFQD3001E . . . . .	237
FFQD2247E . . . . .	226	FFQD3002I . . . . .	238
FFQD2248E . . . . .	227	FFQD3003E . . . . .	238
FFQD2249E . . . . .	227	FFQD3004I . . . . .	238
FFQD2250E . . . . .	227	FFQD3005E . . . . .	238
FFQD2251E . . . . .	227	FFQD3006I . . . . .	238
FFQD2252E . . . . .	227	FFQD3007E . . . . .	238
FFQD2253E . . . . .	228	FFQD3008E . . . . .	238
FFQD2254E . . . . .	228	FFQD3009I . . . . .	239
FFQD2255E . . . . .	228	FFQD3010I . . . . .	239
FFQD2256E . . . . .	228	FFQD3011E . . . . .	239
FFQD2257E . . . . .	228	FFQD3012I . . . . .	239
FFQD2258E . . . . .	229	FFQD3013W . . . . .	239
FFQD2270E . . . . .	229	FFQD3014E . . . . .	239
FFQD2271E . . . . .	229	FFQD3015E . . . . .	240
FFQD2272E . . . . .	229	FFQD3016E . . . . .	240
FFQD2273E . . . . .	229	FFQD3017E . . . . .	240
FFQD2274E . . . . .	230	FFQD3018E . . . . .	240
FFQD2275E . . . . .	230	FFQD3019I . . . . .	240
FFQD2276E . . . . .	230	FFQD3020I . . . . .	241
FFQD2277E . . . . .	230	FFQD3021W . . . . .	241
FFQD2278E . . . . .	230	FFQD3022W . . . . .	241
FFQD2279E . . . . .	230	FFQD3023W . . . . .	241
FFQD2280E . . . . .	231	FFQD3024E . . . . .	241
FFQD2282E . . . . .	231	FFQD3025E . . . . .	241
FFQD2800E . . . . .	231	FFQD3026E . . . . .	242
FFQD2801W . . . . .	231	FFQD3027E . . . . .	242
FFQD2802I . . . . .	231	FFQD3028E . . . . .	242
FFQD2803I . . . . .	232	FFQD3029E . . . . .	242
FFQD2804I . . . . .	232	FFQD3030E . . . . .	242
FFQD2805I . . . . .	232	FFQD3031E . . . . .	242
FFQD2806E . . . . .	232	FFQD3032E . . . . .	243
FFQD2807E . . . . .	232	FFQD3033E . . . . .	243
FFQD2808E . . . . .	232	FFQD3034I . . . . .	243
FFQD2809E . . . . .	232	FFQD3035W . . . . .	243
FFQD2810E . . . . .	233	FFQD3036I . . . . .	243
FFQD2811E . . . . .	233	FFQD3037E . . . . .	243
FFQD2812E . . . . .	233	FFQD3038E . . . . .	244
FFQD2813E . . . . .	233	FFQD3039E . . . . .	244
FFQD2900E . . . . .	233	FFQD3040E . . . . .	244
FFQD2901I . . . . .	234	FFQD3041I . . . . .	244
FFQD2902W . . . . .	234	FFQD3042E . . . . .	244
FFQD2903I . . . . .	234	FFQD3043I . . . . .	244
FFQD2904E . . . . .	234	FFQD3044E . . . . .	245
FFQD2905E . . . . .	234	FFQD3045I . . . . .	245
FFQD2906I . . . . .	235	FFQD3046E . . . . .	245
FFQD2907E . . . . .	235	FFQD3047I . . . . .	245
FFQD2908E . . . . .	235	FFQD3048E . . . . .	245



FFQD3049E . . . . .	245	FFQD3120E . . . . .	257
FFQD3050E . . . . .	246	FFQD3121W . . . . .	257
FFQD3051E . . . . .	246	FFQD3122E . . . . .	258
FFQD3052E . . . . .	246	FFQD3123E . . . . .	258
FFQD3053E . . . . .	246	FFQD3124E . . . . .	258
FFQD3054E . . . . .	246	FFQD3125E . . . . .	258
FFQD3055E . . . . .	247	FFQD3126E . . . . .	258
FFQD3056E . . . . .	247	FFQD3127I . . . . .	259
FFQD3057E . . . . .	247	FFQD3128E . . . . .	259
FFQD3058E . . . . .	247	FFQD3129E . . . . .	259
FFQD3059E . . . . .	247	FFQD3130E . . . . .	259
FFQD3060E . . . . .	248	FFQD3131W . . . . .	259
FFQD3061E . . . . .	248	FFQD3132W . . . . .	260
FFQD3062E . . . . .	248	FFQD3133E . . . . .	260
FFQD3063E . . . . .	248	FFQD3134W . . . . .	260
FFQD3064E . . . . .	248	FFQD3135E . . . . .	260
FFQD3065I . . . . .	249	FFQD3136E . . . . .	260
FFQD3066I . . . . .	249	FFQD3137E . . . . .	260
FFQD3067I . . . . .	249	FFQD3138E . . . . .	261
FFQD3068I . . . . .	249	FFQD3500E . . . . .	261
FFQD3069E . . . . .	249	FFQD3501W . . . . .	261
FFQD3070E . . . . .	249	FFQD3502W . . . . .	261
FFQD3071E . . . . .	250	FFQD3503W . . . . .	261
FFQD3072E . . . . .	250	FFQD3504W . . . . .	262
FFQD3073E . . . . .	250	FFQD3505W . . . . .	262
FFQD3074E . . . . .	250	FFQD3506W . . . . .	262
FFQD3075E . . . . .	250	FFQD3507W . . . . .	262
FFQD3076E . . . . .	251	FFQD3508W . . . . .	262
FFQD3077E . . . . .	251	FFQD3509W . . . . .	263
FFQD3078E . . . . .	251	FFQD3510W . . . . .	263
FFQD3079E . . . . .	251	FFQD3511W . . . . .	263
FFQD3080E . . . . .	251	FFQD3512W . . . . .	263
FFQD3081E . . . . .	252	FFQD3513W . . . . .	263
FFQD3082E . . . . .	252	FFQD3514W . . . . .	264
FFQD3083W . . . . .	252	FFQD3515W . . . . .	264
FFQD3084E . . . . .	252	FFQD3516W . . . . .	264
FFQD3085E . . . . .	252	FFQD3517W . . . . .	264
FFQD3086E . . . . .	252	FFQD3518W . . . . .	264
FFQD3087E . . . . .	253	FFQD3519W . . . . .	265
FFQD3088W . . . . .	253	FFQD3520W . . . . .	265
FFQD3089W . . . . .	253	FFQD3521W . . . . .	265
FFQD3090W . . . . .	253	FFQD3522W . . . . .	265
FFQD3091W . . . . .	253	FFQD3523W . . . . .	266
FFQD3102E . . . . .	254	FFQD3524E . . . . .	266
FFQD3103E . . . . .	254	FFQD3534E . . . . .	266
FFQD3104E . . . . .	254	FFQD3540E . . . . .	266
FFQD3105E . . . . .	254	FFQD3541E . . . . .	266
FFQD3106E . . . . .	254	FFQD3543E . . . . .	266
FFQD3107E . . . . .	255	FFQD3544E . . . . .	267
FFQD3108E . . . . .	255	FFQD3545E . . . . .	267
FFQD3109W . . . . .	255	FFQD3546E . . . . .	267
FFQD3110E . . . . .	255	FFQD3547E . . . . .	267
FFQD3111E . . . . .	256	FFQD3548E . . . . .	267
FFQD3112E . . . . .	256	FFQD3549E . . . . .	268
FFQD3113E . . . . .	256	FFQD3550E . . . . .	268
FFQD3114E . . . . .	256	FFQD3551E . . . . .	268
FFQD3115E . . . . .	256	FFQD3553E . . . . .	268
FFQD3116E . . . . .	257	FFQD4000E . . . . .	268
FFQD3117E . . . . .	257	FFQD4001E . . . . .	269
FFQD3118E . . . . .	257	FFQD4002E . . . . .	269
FFQD3119E . . . . .	257	FFQD4003E . . . . .	269

FFQD4004E . . . . .	269	FFQD4065E . . . . .	282
FFQD4005E . . . . .	269	FFQD4066E . . . . .	282
FFQD4006E . . . . .	270	FFQD4067E . . . . .	282
FFQD4007E . . . . .	270	FFQD4068E . . . . .	282
FFQD4008E . . . . .	270	FFQD4069E . . . . .	282
FFQD4009E . . . . .	270	FFQD4070E . . . . .	283
FFQD4010E . . . . .	270	FFQD4071E . . . . .	283
FFQD4011E . . . . .	271	FFQD4072E . . . . .	283
FFQD4012E . . . . .	271	FFQD4073E . . . . .	283
FFQD4013E . . . . .	271	FFQD4074E . . . . .	283
FFQD4014E . . . . .	271	FFQD4075E . . . . .	284
FFQD4015E . . . . .	271	FFQD4076E . . . . .	284
FFQD4016E . . . . .	271	FFQD4077E . . . . .	284
FFQD4017E . . . . .	272	FFQD4078E . . . . .	284
FFQD4018E . . . . .	272	FFQD4079E . . . . .	284
FFQD4019E . . . . .	272	FFQD4080E . . . . .	284
FFQD4020E . . . . .	272	FFQD4081E . . . . .	285
FFQD4021E . . . . .	272	FFQD4082E . . . . .	285
FFQD4022E . . . . .	273	FFQD4083E . . . . .	285
FFQD4023E . . . . .	273	FFQD4084E . . . . .	285
FFQD4024E . . . . .	273	FFQD4085E . . . . .	285
FFQD4025W . . . . .	273	FFQD4086E . . . . .	286
FFQD4026W . . . . .	273	FFQD4087E . . . . .	286
FFQD4027W . . . . .	274	FFQD4088E . . . . .	286
FFQD4028W . . . . .	274	FFQD4089E . . . . .	286
FFQD4029W . . . . .	274	FFQD4090E . . . . .	286
FFQD4030W . . . . .	274	FFQD4091W . . . . .	286
FFQD4031E . . . . .	274	FFQD4092E . . . . .	287
FFQD4032E . . . . .	274	FFQD4093E . . . . .	287
FFQD4033E . . . . .	275	FFQD4094E . . . . .	287
FFQD4034E . . . . .	275	FFQD4095E . . . . .	287
FFQD4035E . . . . .	275	FFQD4096E . . . . .	287
FFQD4036E . . . . .	275	FFQD4097E . . . . .	288
FFQD4037E . . . . .	276	FFQD4098E . . . . .	288
FFQD4038E . . . . .	276	FFQD4099E . . . . .	288
FFQD4039E . . . . .	276	FFQD4100E . . . . .	288
FFQD4040E . . . . .	277	FFQD4101W . . . . .	288
FFQD4041E . . . . .	277	FFQD4102E . . . . .	289
FFQD4042E . . . . .	277	FFQD4200I . . . . .	289
FFQD4043E . . . . .	277	FFQD4201I . . . . .	289
FFQD4044E . . . . .	277	FFQD4202W . . . . .	289
FFQD4045E . . . . .	278	FFQD4203W . . . . .	289
FFQD4046E . . . . .	278	FFQD4204W . . . . .	290
FFQD4047E . . . . .	278	FFQD4205W . . . . .	290
FFQD4048E . . . . .	279	FFQD4207E . . . . .	290
FFQD4049E . . . . .	279	FFQD4208E . . . . .	290
FFQD4050E . . . . .	279	FFQD4209E . . . . .	290
FFQD4051E . . . . .	279	FFQD4210E . . . . .	290
FFQD4052E . . . . .	279	FFQD4211E . . . . .	291
FFQD4053E . . . . .	280	FFQD4212E . . . . .	291
FFQD4054E . . . . .	280	FFQD4213E . . . . .	291
FFQD4055E . . . . .	280	FFQD4214E . . . . .	291
FFQD4056E . . . . .	280	FFQD4215E . . . . .	291
FFQD4057E . . . . .	280	FFQD4216E . . . . .	292
FFQD4058E . . . . .	280	FFQD4217E . . . . .	292
FFQD4059E . . . . .	281	FFQD4218E . . . . .	292
FFQD4060E . . . . .	281	FFQD4219E . . . . .	292
FFQD4061E . . . . .	281	FFQD4220E . . . . .	292
FFQD4062E . . . . .	281	FFQD4221E . . . . .	292
FFQD4063E . . . . .	281	FFQD4222E . . . . .	293
FFQD4064E . . . . .	282	FFQD4223E . . . . .	293

FFQD4224E . . . . .	293	FFQD4453W . . . . .	306
FFQD4225E . . . . .	293	FFQD4454W . . . . .	306
FFQD4226E . . . . .	293	FFQD4455W . . . . .	306
FFQD4227E . . . . .	294	FFQD4456W . . . . .	307
FFQD4228E . . . . .	294	FFQD4457W . . . . .	307
FFQD4230E . . . . .	294	FFQD4458E . . . . .	307
FFQD4231E . . . . .	295	FFQD4461E . . . . .	307
FFQD4232E . . . . .	295	FFQD4500E . . . . .	308
FFQD4300E . . . . .	295	FFQD4501E . . . . .	308
FFQD4301E . . . . .	295	FFQD4502E . . . . .	308
FFQD4302E . . . . .	296	FFQD4503E . . . . .	308
FFQD4303E . . . . .	296	FFQD4504E . . . . .	308
FFQD4304E . . . . .	296	FFQD4505W . . . . .	308
FFQD4305E . . . . .	296	FFQD4506W . . . . .	309
FFQD4306E . . . . .	296	FFQD4507E . . . . .	309
FFQD4307E . . . . .	297	FFQD4508E . . . . .	309
FFQD4308E . . . . .	297	FFQD4509E . . . . .	309
FFQD4309E . . . . .	297	FFQD4510E . . . . .	309
FFQD4310E . . . . .	297	FFQD4511E . . . . .	310
FFQD4311E . . . . .	297	FFQD4600E . . . . .	310
FFQD4312E . . . . .	298	FFQD4601E . . . . .	310
FFQD4313E . . . . .	298	FFQD4602E . . . . .	310
FFQD4314E . . . . .	298	FFQD4603I . . . . .	310
FFQD4315E . . . . .	298	FFQD4604I . . . . .	310
FFQD4316E . . . . .	298	FFQD4605I . . . . .	311
FFQD4317E . . . . .	299	FFQD4607E . . . . .	311
FFQD4318E . . . . .	299	FFQD4608E . . . . .	311
FFQD4319E . . . . .	299	FFQD4609E . . . . .	311
FFQD4320E . . . . .	299	FFQD4610E . . . . .	311
FFQD4321E . . . . .	300	FFQD4611E . . . . .	312
FFQD4322E . . . . .	300	FFQD4612E . . . . .	312
FFQD4323W . . . . .	300	FFQD4613W . . . . .	312
FFQD4324I . . . . .	300	FFQD4614W . . . . .	312
FFQD4325I . . . . .	300	FFQD4616E . . . . .	312
FFQD4326E . . . . .	301	FFQD4617E . . . . .	313
FFQD4327W . . . . .	301	FFQD4618E . . . . .	313
FFQD4328E . . . . .	301	FFQD4619E . . . . .	313
FFQD4402W . . . . .	301	FFQD4620E . . . . .	313
FFQD4409W . . . . .	301	FFQD4621E . . . . .	313
FFQD4413W . . . . .	302	FFQD4622E . . . . .	313
FFQD4414E . . . . .	302	FFQD4623E . . . . .	314
FFQD4415E . . . . .	302	FFQD4624E . . . . .	314
FFQD4416E . . . . .	302	FFQD4625E . . . . .	314
FFQD4417E . . . . .	302	FFQD4626E . . . . .	314
FFQD4418E . . . . .	303	FFQD4627E . . . . .	314
FFQD4419W . . . . .	303	FFQD4628E . . . . .	315
FFQD4420W . . . . .	303	FFQD4629E . . . . .	315
FFQD4421W . . . . .	303	FFQD4631E . . . . .	315
FFQD4422W . . . . .	303	FFQD4632W . . . . .	315
FFQD4423W . . . . .	304	FFQD4633W . . . . .	315
FFQD4427E . . . . .	304	FFQD4707E . . . . .	316
FFQD4428E . . . . .	304	FFQD4708E . . . . .	316
FFQD4433W . . . . .	304	FFQD4709E . . . . .	316
FFQD4434W . . . . .	304	FFQD4710E . . . . .	316
FFQD4435W . . . . .	305	FFQD4711E . . . . .	316
FFQD4436W . . . . .	305	FFQD4712E . . . . .	316
FFQD4437E . . . . .	305	FFQD4713E . . . . .	317
FFQD4449E . . . . .	305	FFQD4714E . . . . .	317
FFQD4450E . . . . .	305	FFQD4715E . . . . .	317
FFQD4451E . . . . .	306	FFQD4716E . . . . .	317
FFQD4452E . . . . .	306	FFQD4718E . . . . .	317

FFQD4719E . . . . .	318		FFQV0005E . . . . .	330
FFQD4720E . . . . .	318		FFQV0006E . . . . .	330
FFQD4721E . . . . .	318		FFQV0007E . . . . .	330
FFQD4722E . . . . .	318		FFQV0008E . . . . .	330
FFQD4723E . . . . .	318		FFQV0009E . . . . .	331
FFQD4724E . . . . .	319		FFQV0010E . . . . .	331
FFQD4725E . . . . .	319		FFQV0011E . . . . .	331
FFQD4726E . . . . .	319		FFQV0012E . . . . .	331
FFQD4728W . . . . .	319		FFQV0021E . . . . .	331
FFQD4729W . . . . .	319		Migration wizard messages (FFQW).	332
FFQD4730E . . . . .	320		FFQW0001E . . . . .	332
FFQD4731E . . . . .	320		FFQW0002E . . . . .	332
FFQD4732E . . . . .	320		FFQW0003E . . . . .	332
FFQD4733E . . . . .	320		FFQW0008E . . . . .	332
FFQD4734E . . . . .	321		FFQW0011E . . . . .	332
FFQD4735E . . . . .	321		FFQW0012E . . . . .	333
FFQD4736E . . . . .	321		FFQW0013E . . . . .	333
FFQD4737W . . . . .	321		FFQW0014E . . . . .	333
FFQD4738W . . . . .	321		FFQW0016E . . . . .	333
FFQD4739E . . . . .	322		FFQW0019E . . . . .	333
FFQD4741W . . . . .	322		FFQW0120I . . . . .	334
General system messages (FFQG).	322		FFQW0140I . . . . .	334
FFQG0026E . . . . .	322		FFQW0141I . . . . .	334
Operating system subsystem (OSS) messages			FFQW0142I . . . . .	334
(FFQO) . . . . .	322		FFQW0154I . . . . .	334
FFQO0034E . . . . .	322		FFQW0155I . . . . .	334
FFQO0086E . . . . .	323		FFQW0156I . . . . .	335
FFQO0141E . . . . .	323		FFQW0157I . . . . .	335
FFQO0164E . . . . .	323		FFQW0158I . . . . .	335
FFQO0165E . . . . .	323		FFQW0159I . . . . .	335
FFQO0168E . . . . .	323			
FFQO0193W . . . . .	324			
FFQO0223E . . . . .	324			
FFQO0258E . . . . .	324			
FFQO0259E . . . . .	324			
FFQO0273W . . . . .	324			
Parser messages (FFQT) . . . . .	325			
FFQT0059E . . . . .	325			
FFQT0060E . . . . .	325			
FFQT0061E . . . . .	325			
FFQT0062E . . . . .	325			
FFQT0065I . . . . .	326			
FFQT0066I . . . . .	326			
FFQT0067I . . . . .	326			
FFQT0072I . . . . .	326			
FFQT0078W . . . . .	326			
FFQT0084E . . . . .	327			
FFQT0085E . . . . .	327			
FFQT0086W . . . . .	327			
FFQT0138E . . . . .	327			
Utility messages (FFQU). . . . .	328			
FFQU0269E . . . . .	328			
FFQU0270E . . . . .	328			
FFQU0271E . . . . .	328			
FFQU0272E . . . . .	328			
FFQU0273E . . . . .	329			
Synonym dictionary messages (FFQV) . . . . .	329			
FFQV0001E . . . . .	329			
FFQV0002E . . . . .	329			
FFQV0003E . . . . .	329			
FFQV0004E . . . . .	330			

**Glossary of terms for enterprise search . . . . . 337**

**DB2 Information Integrator documentation . . . . . 343**

Documentation about event publishing function for DB2 Universal Database on z/OS . . . . .	343
Documentation about event publishing function for IMS and VSAM on z/OS . . . . .	344
Documentation about event publishing and replication function on Linux, UNIX, and Windows . . . . .	344
Documentation about federated function on Linux, UNIX, and Windows . . . . .	345
Documentation about federated function on z/OS . . . . .	346
Documentation about replication function on z/OS . . . . .	347
Documentation about enterprise search function on Linux, UNIX, and Windows . . . . .	348
Release notes and installation requirements . . . . .	348
Viewing release notes and installation requirements . . . . .	349
Viewing and printing PDF documentation. . . . .	350
Accessing DB2 Information Integrator documentation . . . . .	350

**Accessibility . . . . . 353**

Keyboard input and navigation . . . . .	353
Keyboard focus. . . . .	353
Keyboard input. . . . .	353
Keyboard navigation . . . . .	353
Accessible display . . . . .	353

Font settings . . . . . 354  
Non-dependence on color . . . . . 354  
Compatibility with assistive technologies . . . . . 354  
Accessible documentation . . . . . 354

**Contacting IBM . . . . . 355**  
Obtaining product information . . . . . 355

Providing comments on the documentation . . . . . 355

**Notices . . . . . 357**  
Trademarks . . . . . 359

**Index . . . . . 361**



---

## About these topics

Use this information when you administer an IBM® DB2® Information Integrator OmniFind™ Edition Version 8.2.2 system.

DB2 Information Integrator OmniFind Edition (DB2 II OmniFind Edition) provides a technology called *enterprise search*. The components for enterprise search are installed when you install the DB2 II OmniFind Edition product. The term *enterprise search* is used throughout the DB2 II OmniFind Edition documentation except for references to installation paths and product packaging labels.

The administration documentation for enterprise search covers the following topics:

- An introduction to enterprise search concepts, components, and capabilities
- Instructions on how to create, monitor, and administer collections
- Information about how data is parsed, and the ways that you can customize parsing activities to optimize search and retrieval
- Information about how parsed data is indexed, and the ways that you can administer indexing activity
- Information about how the Search servers search for data and return search results
- Information about enabling custom search applications to 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 taxonomies and collections to enterprise search category trees and collections
- Instructions for backing up and restoring the system
- Instructions for creating and viewing log files
- Explanations of enterprise search messages

---

## Who should read these topics

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

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

To use this information effectively, you must be familiar with Web applications and have experience with the data sources that you want to search.





---

## 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 easily locate the information that you need.

The enterprise search components, which are installed with IBM DB2 Information Integrator OmniFind Edition (DB2 II OmniFind Edition), 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, an enterprise search system handles 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 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 as 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**

“Enterprise search security” on page 139

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®, Microsoft® SQL Server, 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 (Documentum, FileNet Panagon Content Services, and Hummingbird<sup>®</sup> item classes). VeniceBridge is being renamed to IBM WebSphere Information Integrator Content Edition.
- Web sites (on the Internet or within your intranet)

With the Data Listener 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" in "Programming Guide and API Reference for Enterprise Search"

---

## **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.

Support for federated searching enables users to search more than one collection with a single query. The search results potentially include documents from all data sources and all collections in your enterprise search system.

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.

Other DB2 Information Integrator OmniFind Edition components enable you to specify security preferences, 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**

“Enterprise search crawler administration” on page 31

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.

“Parsing documents for enterprise search” on page 77

To enhance the retrievability of documents, you can specify options for how documents and metadata are to be parsed, analyzed, and categorized before they are added to the enterprise search index.

“Enterprise search index administration” on page 103

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.

“Search server administration” on page 115

Options that you can specify for the Search servers include using cache space for returning search results, controlling the maximum length of document summaries in the search results, associating a custom synonym dictionary so that users can find documents that contain synonyms of query terms, and returning predefined URIs in the search results whenever certain terms appear in the query.

## **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 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 Network News Transfer Protocol (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.

## 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 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 and HTML documents

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

### Categories

This option 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.

### Custom text analysis

Application developers can create custom analysis programs to perform complex linguistic analysis of the data that you need to search. You can

plug these programs into the enterprise search system and use them to annotate the content of your collections. By indexing the annotations, you enable collections for semantic search.

For example, users can search for query terms that occur in proximity to each other or that occur in the same sentence, or they can search for relationships between query terms (such as documents that discuss an IBM salesperson named Smith, not an IBM engineer named Smith).

#### **Support for n-gram segmentation**

To enhance the retrievability of documents that were written in Chinese, Japanese, or Korean, you can enable the n-gram segmentation method of lexical analysis. This form of analysis does not use white space to delimit words. (You cannot change the segmentation method after you create a collection.)

#### **Support for searching XML documents with native XML search**

A native XML search can provide more precise search results by searching XML markup. For example, a query might specify that a word must occur in a particular XML element.

- 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**

*“Working with categories” on page 77*

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 84*

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

*“HTML search fields” on page 87*

HTML search fields enable users to query attributes of HTML documents.

*“Linguistic support” in “Programming Guide and API Reference for Enterprise Search”*

*“Text analysis included in enterprise search” in “Programming Guide and API Reference for Enterprise Search”*

#### **Related tasks**

*“Monitoring the parser” on page 181*

Monitor the parser when you need to view information about documents that are analyzed by the parser before they are added to the enterprise search index.

Options enable you to review statistics and administer parser activity.

## **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 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.

You can specify options for the following indexing activities:

- To enable users to specify trailing wildcard characters, you can build support for expanding the query terms into the index, or you can specify that the query terms are to be expanded during query processing. The decision that you make involves a trade-off between resource usage and query response time.
- 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 scopes, users can search and retrieve documents from only those subsets of the collection.
- You can specify options for collapsing search result documents that have the same URI prefix. You can also specify a group name so that documents with different URI prefixes can be collapsed together in the search results.
- After an index is built, you can remove URIs that you want to prevent users from searching.

### Related concepts



“Enterprise search index administration” on page 103

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.

“Wildcard characters in queries” on page 108

You can enable users to include a wildcard character in query terms and search for words that start with a specified pattern.

“Scopes” on page 110

Configure a scope when you want to present users with a limited view of a collection.

“Collapsed URIs” on page 111

Enterprise search can organize the search results so that result documents that have the same URI prefix in the index are grouped together and collapsed in the search results.

### Related tasks

“Scheduling index builds” on page 104

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

“Configuring concurrent index builds” on page 106

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.

“Removing URIs from the index” on page 113

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

“Monitoring index activity for a collection” on page 182

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 183

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 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.
- You can specify a default language for searching documents in the collection.
- If your application developers create a custom synonym dictionary, you can associate it with specific collections. When users query the collection, documents that contain synonyms of the query terms are included in the search results.
- 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.

In a multiple server configuration, failure protection is available at the collection level, not just at the server level. If a collection on one Search server becomes unavailable for any reason, then the queries for that collection are routed automatically to the other Search server.

#### **Related concepts**

“Search applications for enterprise search” on page 125

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 115

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

“Custom synonym dictionaries” on page 116

To improve the quality of the search results, you can enable users to search for synonyms of their query terms when they search a collection.

“Quick links” on page 121

Quick links enable you to provide users with links to documents that are predetermined to be relevant to the query terms.

#### **Related tasks**

“Monitoring the Search servers” on page 184

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.

## **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 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 select the Index page to change the index schedule or select the Parse page to modify a rule for parsing XML documents.

#### **Related concepts**

“Enterprise search administration overview” on page 15

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, associate search applications with collections, and specify information to enforce security.

“Administrative roles” on page 140

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 18

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 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.

You can also configure alerts, which enable you to receive e-mail about certain monitored activities whenever a monitored event occurs. For example, you can receive an alert if the search response time exceeds a specified threshold.

### Related concepts

“Monitoring enterprise search activity” on page 169

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.

“Starting and stopping enterprise search” on page 165

After you create a collection, you must start the components for crawling, parsing, indexing, and searching data. Stop and restart the components after you make changes to the collection.

## Enterprise search log files

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

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 (including alerts), 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 enterprise search session.

### Related concepts

“Enterprise search log file administration” on page 187

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, receiving alerts, and receiving e-mail about messages.

“Alerts” on page 187

You can configure enterprise search to write messages to the log file whenever it detects that certain events occurred.

“Messages for enterprise search” on page 201

Explanations and suggestions for corrective action are provided for most enterprise search messages.

#### **Related tasks**

“Configuring log files” on page 190

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

“Configuring SMTP server information” on page 191

Before you can receive e-mail about enterprise search activities, you must configure information about your Simple Mail Transfer Protocol (SMTP) server.

“Receiving e-mail about logged messages” on page 192

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

“Viewing log files” on page 194

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 enterprise search sessions.

## **Customizing enterprise search**

The application programming interfaces for enterprise search enable you to create custom search applications, custom applications to update the content of collections, custom programs for text analysis, and custom synonym dictionaries.

After installing 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, then add data to a collection or remove data from a collection.

You can enhance the retrievability of information by integrating custom programs for linguistic analysis with your enterprise search collections. After you add custom text annotators to the system, you can associate them with collections. When users query a collection, they benefit from the word associations that your custom programs build into the index. For example, users can search for relationships between terms, not just on the terms themselves.

You can also enhance the retrievability of information by integrating custom synonym dictionaries that reflect, for example, acronyms, abbreviations, and vocabulary terms that are specific to your industry. After you add custom synonym dictionaries to the system, you can associate them with collections. When users query a collection, they benefit by seeing documents that contain synonyms of their query terms in the search results.

#### **Related concepts**

“Search applications for enterprise search” on page 125

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)" in "Programming Guide and API Reference for Enterprise Search"

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

#### **Related tasks**

"Configuring support for Data Listener applications" on page 70

You can extend enterprise search by using the Data Listener API to create an external crawler. Your custom Data Listener applications can add data to a collection, remove data from a collection, or instruct a Web crawler to visit and revisit URLs.

## **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 DB2 Information Integrator OmniFind Edition. The sample search application demonstrates many of the search and retrieval functions that are available for enterprise search. The application is also a working example that enables you to search all active collections in your enterprise search system. You can use the sample application to test new collections before you make the collections available to users.

The sample search application demonstrates support for federated search by enabling you to search one or more collections at a time. The application also demonstrates the following functions:

- Quick links
- Collapsed search results
- Ability to browse documents by the categories or scopes that they belong to
- Spell checker for query terms
- Synonym search

If you enable global security in WebSphere Application Server and your collections, and if you specify that you want to validate current user credentials when you configure options for Lotus Notes documents, you can use the sample search application to configure a user profile. Credentials that you specify in the user profile are encrypted and stored so that you can search the collection without being prompted to present credentials for authentication.

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

#### **Related concepts**

"Search applications for enterprise search" on page 125

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)" in "Programming Guide and API Reference for Enterprise Search"

"Sample search application functions" on page 126

The sample search application for enterprise search demonstrates most of the

search functions that you can build into your custom search applications. You can use this application to search one, several, or all collections at a time.

#### **Related tasks**

“Editing sample search application properties” on page 128

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 use a different search application as the default application.

“Accessing the sample search application” on page 129

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

“Enabling security for the sample search application” on page 130

If you enable global security in WebSphere Application Server and want to use the sample search application to search secure collections, you must change configuration settings in the sample application and in WebSphere Application Server.

---

## **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 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.

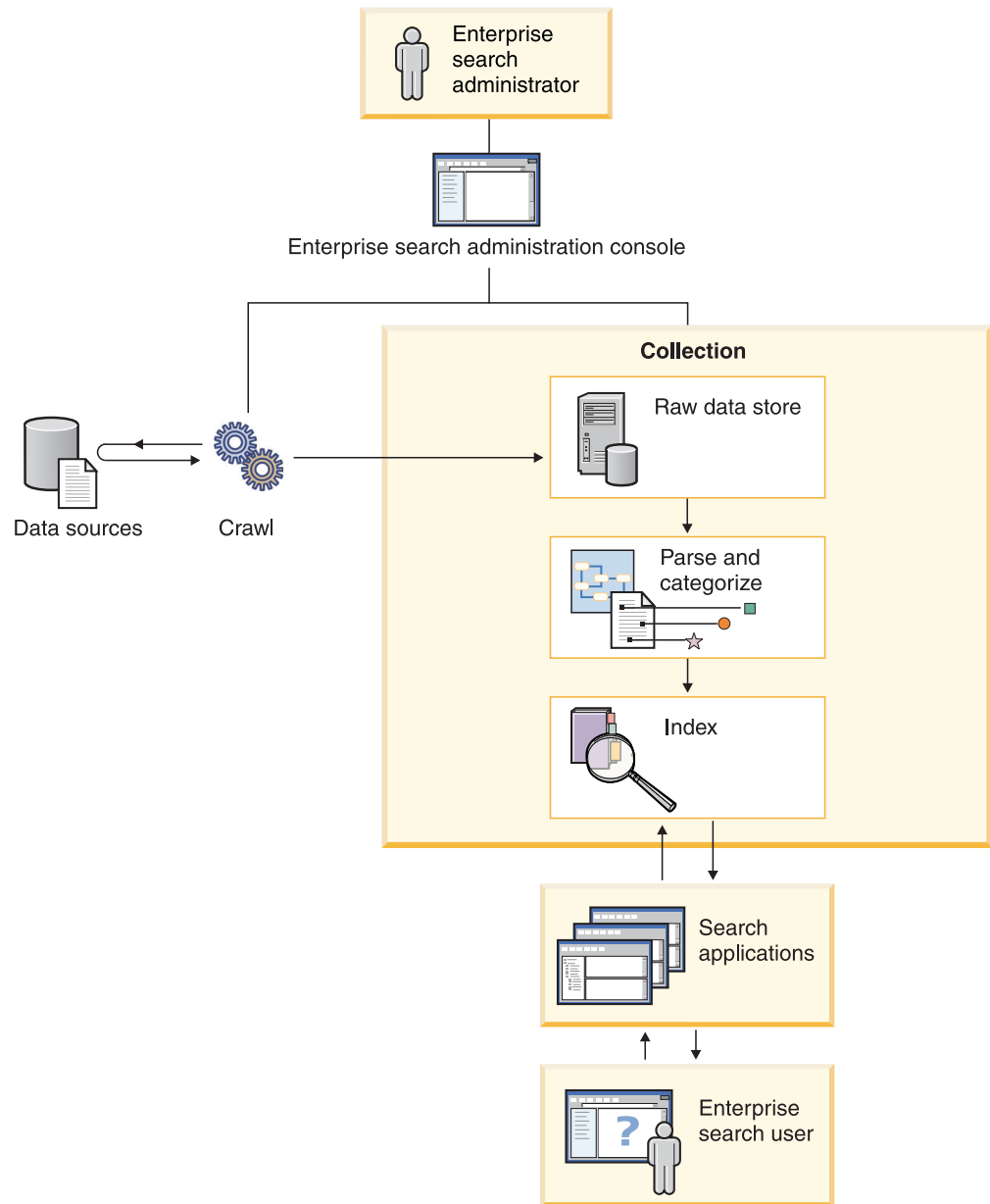


Figure 1. How data flows through an enterprise search system



---

## 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, associate search applications with collections, and specify information to enforce security.

### 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 and to specify options for how you want to crawl data and make the data searchable.

4. Click  **Edit** for your new collection, then select a page to specify options for the collection.

**Attention:** If you click the Back or Refresh button in the Web browser, inconsistent results and a potential loss of data can occur. When you configure collections and need to go to the previous page or to refresh information in the administration console, click the **Previous** and **Refresh** buttons in the enterprise search administration console instead of the Back and Refresh buttons in the Web browser.

- Optional: On the General page, you can specify options that apply to the entire collection:

- You can edit general options to change the name or description of the collection, or change the estimated size of the collection.
- You can view information about the collection that you cannot change, such as the collection ID or the static ranking method for ranking documents in the search results.
- If security was enabled for the collection when it was created, you can enable or disable document-level security controls.

- 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:

- You can specify whether XML documents are to be parsed so that they can be searched with native XML search.
- You can associate documents with categories, which enables users to search a subset of the collection or browse search result documents by the categories that they belong to.

- You can map XML elements and HTML metadata elements to search fields in the index, which enables users to specify the field names in queries and search specific parts of documents.
- If custom analysis engines were added to the enterprise search system, you can choose one to use with the collection, and then specify text analysis options to enhance the retrievability of information.
- On the Index page, configure 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. You can also do the following optional activities:
  - Enable users to specify wildcard characters in query terms.
  - Configure scopes, so that users search a limited part of the collection instead of all documents in the index.
  - Collapse search results, so that all documents from the same source are grouped and collapsed together in the search results.
  - Remove URIs from the index. For example, you might need to prevent users from seeing certain documents after the collection is created.
- Optional: On the Search page, specify options for searching documents in the collection:
  - You can set aside cache space for search results and change the default language of the collection.
  - If custom synonym dictionaries were added to the enterprise search system, you can select one to use for searching the collection.
  - You can specify a display length for document summaries in the search results.
  - If you want specific URIs to appear automatically in the search results whenever a query includes particular keywords or phrases, you can configure quick links.
- Optional: On the Log page, you can do the following activities:
  - Specify options for the types of messages that you want to log and how often you want the log files to be recycled.
  - 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.
  - Specify options for receiving e-mail whenever certain messages or certain types of messages are logged.

## Start the components

After you specify the data sources to crawl and options for collecting and searching the data, you can start the processes for building the collection. The order in which you start components is critical. Crawlers must crawl data before it can be parsed, parsers must analyze the crawled data before it can be indexed, and the index must be refreshed or reorganized before the Search servers can start to process search requests.

## System view

If you are a member of the enterprise search administrator role, you can click **System** on the toolbar to do the following activities (collection administrators, operators, and monitors cannot access this view):

- Check the availability of system resources



- Configure client Data Listener applications
- Add custom text analysis engines and synonym dictionaries to the system
- Specify how many indexes can be built concurrently
- Configure alerts for system-level events
- Specify options for logging messages that are produced by system-level sessions
- Specify information about your mail server so that you can receive e-mail about enterprise search activities

## Security view


If you are a member of the enterprise search administrator role, you can click **Security** to specify security options. Collection administrators, operators, and monitors cannot access this view.

If you enable security in IBM WebSphere Application Server, you can 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.

With security enabled in WebSphere Application Server, you can also configure information about your lightweight directory access protocol (LDAP) server. When users search collections, the Search servers can use the LDAP server data to enforce restrictions on which sources an individual user can search.

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

## Monitor view

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 concepts

“Enterprise search administration console” on page 8

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.

### Related tasks

“Starting the enterprise search components” on page 165

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 167

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 Collection wizard” on page 25

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 collection by using the Collections view” on page 26

Use the Collections view to create an empty collection. You can then edit the collection to specify options for adding data to the collection and making the collection searchable.

“Creating a crawler” on page 33

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 171

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.

### Before you begin

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 Edition 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. 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.

After you finish administering collections, you can click **Logout** to log out of the console. You can then log in with a different ID and password, or you can close the Web browser to exit the administration console.

#### **Related concepts**

“Enterprise search administration console” on page 8

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.

“Administrative roles” on page 140

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

#### **Related tasks**

“Starting the enterprise search components” on page 165

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 in a single server configuration**

The password for the enterprise search administrator is stored in an encrypted format. To change the password, use the `eschangepw` script.

#### **Before you begin**

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 password for the initial enterprise search administrator ID is specified when DB2 II OmniFind Edition is installed.

To change the password, you must run the `eschangepw` script to disseminate the change throughout the enterprise search system. The installation program creates two environment variables that you can use with the `eschangepw` script:

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

The enterprise search installation directory.

#### **ES\_NODE\_ROOT**

The enterprise search data directory. The password for the enterprise search administrator ID is stored in the `es.cfg` file in this directory.

Because the `eschangepw` script is installed in the `ES_INSTALL_ROOT/bin` directory, you can run it from anywhere in the system.

#### **Procedure**

To change the enterprise search administrator password in a single server configuration:

1. Log in as the enterprise search administrator.
2. Stop the server by entering this command: `esadmin stop`

3. Open the WebSphere Application Server administrative console and stop the server1 and ESsearchServer enterprise applications.
4. Change the system password for the enterprise search administrator user ID by using UNIX operating system commands or the Microsoft Windows change password facility.
5. Run the following script, where *newValue* is the password that you specified in step 4:

Operating system	Command
UNIX	eschangepw.sh <i>newValue</i>
Windows	eschangepw <i>newValue</i>

6. In the WebSphere Application Server administrative console, start the server1 and ESsearchServer applications.
7. Recycle the DB2 II OmniFind Edition common communications layer (CCL) by entering the following commands:

Operating system	Commands
UNIX	stopccl.sh, then startccl.sh -bg
Windows command prompt	stopccl, then startccl
Windows Services administrative tool	<ol style="list-style-type: none"> <li>1. Launch Windows Services.</li> <li>2. Right-click <b>DB2 Information Integrator OmniFind Edition</b> and select <b>Stop</b>.</li> <li>3. Right-click <b>DB2 Information Integrator OmniFind Edition</b> again and select <b>Properties</b>.</li> <li>4. Click the <b>Log On</b> tab.</li> <li>5. Change the password by specifying the <i>newValue</i>, and click <b>OK</b>.</li> <li>6. Right-click <b>DB2 Information Integrator OmniFind Edition</b> again, and select <b>Start</b>.</li> </ol>

8. Restart enterprise search by entering this command: esadmin start.

---

## Changing the enterprise search administrator password in a multiple server configuration

The password for the enterprise search administrator is stored in an encrypted format. To change the password, use the eschangepw script to change it on all computers in your enterprise search system.

### Before you begin

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.

The password for the enterprise search administrator must be the same on all computers that belong to this installation of DB2 II OmniFind Edition.

### About this task

The password for the initial enterprise search administrator ID is specified when DB2 II OmniFind Edition is installed.

To change the password, and to disseminate the change throughout the enterprise search system, you must run the `eschangepw` script on each computer in your multiple server configuration. The installation program creates two environment variables that you can use with the `eschangepw` script:

**ES\_INSTALL\_ROOT**

The enterprise search installation directory.

**ES\_NODE\_ROOT**

The enterprise search data directory. The password for the enterprise search administrator ID is stored in the `es.cfg` file in this directory.

Because the `eschangepw` script is installed in the `ES_INSTALL_ROOT/bin` directory, you can run it from anywhere in the system.

**Procedure**

To change the enterprise search administrator password in a multiple server configuration:

1. On the computer where you installed the index server, log in as the enterprise search administrator.
2. Start the system maintenance mode by entering this command: `esadmin system maintenance`
3. Change the system password for the enterprise search administrator user ID by using UNIX operating system commands or the Microsoft Windows change password facility.
4. Run the following script, where *newValue* is the password that you specified in step 3:

Operating system	Command
UNIX	<code>eschangepw.sh newValue</code>
Windows	<code>eschangepw newValue</code>

5. Recycle the DB2 II OmniFind Edition common communications layer (CCL) by entering the following commands on the index server:

Operating system	Commands
UNIX	<code>stopccl.sh</code> , then <code>startccl.sh -bg</code>
Windows command prompt	<code>stopccl</code> , then <code>startccl</code>
Windows Services administrative tool	<ol style="list-style-type: none"> <li>1. Launch Windows Services.</li> <li>2. Right-click <b>DB2 Information Integrator OmniFind Edition</b> and select <b>Stop</b>.</li> <li>3. Right-click <b>DB2 Information Integrator OmniFind Edition</b> again and select <b>Properties</b>.</li> <li>4. Click the <b>Log On</b> tab.</li> <li>5. Change the password by specifying the <i>newValue</i>, and click <b>OK</b>.</li> <li>6. Right-click <b>DB2 Information Integrator OmniFind Edition</b> again, and select <b>Start</b>.</li> </ol>

6. On the computer where you installed the crawler server, log in as the enterprise search administrator.
7. Change the system password for the enterprise search administrator user ID by using UNIX operating system commands or the Windows change password facility. This password must match the password that you specified in step 3 on page 21.
8. Run the following script, where *newValue* is the password that you specified in step 3 on page 21:

Operating system	Command
UNIX	exchangepw.sh <i>newValue</i>
Windows	exchangepw <i>newValue</i>

9. Recycle CCL on the crawler server by entering the following commands:

Operating system	Commands
UNIX	stopccl.sh, then startccl.sh -bg
Windows command prompt	stopccl, then startccl
Windows Services administrative tool	<ol style="list-style-type: none"> <li>1. Launch Windows Services.</li> <li>2. Right-click <b>DB2 Information Integrator OmniFind Edition</b> and select <b>Stop</b>.</li> <li>3. Right-click <b>DB2 Information Integrator OmniFind Edition</b> again and select <b>Properties</b>.</li> <li>4. Click the <b>Log On</b> tab.</li> <li>5. Change the password by specifying the <i>newValue</i>, and click <b>OK</b>.</li> <li>6. Right-click <b>DB2 Information Integrator OmniFind Edition</b> again, and select <b>Start</b>.</li> </ol>

10. On the computer where you installed the index server, log in as the enterprise search administrator, and enter this command to start enterprise search:  
esadmin start
11. On one of the computers where you installed a Search server, log in as the enterprise search administrator.
12. Stop the IBM HTTP Web Server. (This step prevents the Network Dispatcher from routing queries to this server.)
13. Stop CCL on the Search server by entering the following commands:

Operating system	Commands
UNIX	stopccl.sh
Windows command prompt	stopccl
Windows Services administrative tool	<ol style="list-style-type: none"> <li>1. Launch Windows Services.</li> <li>2. Right-click <b>DB2 Information Integrator OmniFind Edition</b> and select <b>Stop</b>.</li> </ol>

14. Open the WebSphere Application Server administrative console and stop the server1 and ESSearchServer applications.

15. Change the system password for the enterprise search administrator user ID by using UNIX operating system commands or the Windows change password facility. This password must match the password that you specified in step 3 on page 21.
16. Run the following script, where *newValue* is the password that you specified in step 3 on page 21:

Operating system	Command
UNIX	<code>eschangepw.sh newValue</code>
Windows	<code>eschangepw newValue</code>

17. Restart CCL on the Search server by entering the following commands:

Operating system	Commands
UNIX	<code>startccl.sh -bg</code>
Windows command prompt	<code>startccl</code>
Windows Services administrative tool	<ol style="list-style-type: none"> <li>1. Launch Windows Services.</li> <li>2. Right-click <b>DB2 Information Integrator OmniFind Edition</b> and select <b>Properties</b>.</li> <li>3. Click the <b>Log On</b> tab.</li> <li>4. Change the password by specifying the <i>newValue</i>, and click <b>OK</b>.</li> <li>5. Right-click <b>DB2 Information Integrator OmniFind Edition</b> again, and select <b>Start</b>.</li> </ol>

18. On the second search server, repeat steps 11 on page 22 through 17.
19. Start the IBM HTTP Web Server.
20. In the WebSphere Application Server administrative console, start the `server1` and `ESSearchServer` enterprise applications.
21. On the computer where you installed the index server, log in as the enterprise search administrator, and enter this command to start enterprise search:  
`esadmin start`
22. Log in to the enterprise search administration console, monitor a collection, click the Search page, and then click **Stop** and **Start** to restart the Search servers. Repeat this step for each collection in your enterprise search system.





---

## Enterprise search collections

An enterprise search collection contains the entire set of sources that users can search with a single query. Through federation, users can search multiple collections 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 and monitor the collection, the enterprise search system, and security options.

### Collection federation

If support for federation is built into the search application, users can search multiple collections at the same time. Federation also enables you to scale beyond the size limitation for a collection (which is 20 000 000 documents per collection). For example, users can search two collections that each contain 20 000 000 documents.

Search quality is dependent on the scores that are generated by individual collections, which are then merged to produce the final result set. The results are the same as submitting two separate searches and then merging and ranking the results.

#### Related tasks

“Monitoring a collection” on page 171

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 160

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.

### Before you begin

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. If you enabled security for the collection, the  **Collection security is enabled** icon is displayed next to the collection name.
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 (named Default) to search your new collection.

#### Related concepts

“Enterprise search administration overview” on page 15

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, associate search applications with collections, and specify information to enforce security.

---

## Creating a collection by using the Collections view

Use the Collections view to create an empty collection. You can then edit the collection to specify options for adding data to the collection and making the collection searchable.

### Before you begin

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.** Specify a descriptive name for the content or purpose of the collection.
  - **Collection security.** Specify whether you want to enable security for the collection. After you create the collection, you cannot change this setting. If collection security is enabled, you can later specify options for enforcing document-level access controls.
  - **Document importance (static ranking model).** Specify a strategy for assigning a static ranking factor that will be used to rank documents in the search results. After you create the collection, you cannot change this value.
  - **Categorization type.** Specify whether you want to be able to search for documents by the categories that they belong to.
  - **Default language.** Specify the default language for searching documents in the collection.
3. Accept the default values for the following fields, or specify options that you want to use with this collection:
  - **Description.** By default, no description is created.
  - **Estimated number of documents.** The default estimated size of the collection is 1 000 000 documents. The system uses this value to estimate the memory and disk resources for the collection, not to limit the size of the collection.
  - **Location for collection data.** The default location for collection-related files is on the index server. After you create the collection, you cannot change this value.
  - **Collection ID.** The default collection ID is based on the collection name. After you create the collection, you cannot change this value. (If you specify a custom collection ID, your search applications call the collection with this identifier instead of the potentially cryptic identifier that the system creates.)
  - **N-gram segmentation.** The default segmentation method is Unicode-based, white space segmentation. Select the option to use n-gram segmentation only if your collection includes Chinese, Japanese, or Korean documents and you want the parser to use n-gram segmentation to delimit words instead. After you create the collection, you cannot change this value.
4. Click **OK**.

The Collections view lists your new collection with other collections in your enterprise search system. If you enabled security for the collection, the  **Collection security is enabled** icon is displayed next to the collection name.

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 applications to search your new collection until you are ready to make it available for users to search with your custom search applications.

#### **Related concepts**

“Enterprise search administration overview” on page 15

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, associate search applications with collections, and specify information to enforce security.

---

## **Editing a collection**

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

### **Before you begin**


To edit 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 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.

### **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 and view settings that you cannot change. If security was enabled for the collection when it was created, you can configure document-level security options.

**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 types of 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, which enable users  
| to search subsets of the collection, and you can configure rules that  
| enable users to search specific parts of XML and HTML documents. If  
| custom analysis engines were added to the enterprise search system,  
| you can select one to use for analyzing and annotating content in this  
| collection, and then specify options for analyzing documents and  
| adding the results of analysis to the index.

| **Index** Specify schedules for reorganizing the entire index and refreshing the  
| index with new and changed content. You can also configure options  
| for using wildcard characters in queries, limiting the view of the  
| collection to a range of URIs, collapsing search results from the same  
| Web site, and removing URIs from the index.

#### **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 function that ensures the return of  
| predetermined URIs whenever a user includes specific words or  
| phrases in a query. If custom synonym dictionaries were added to the  
| enterprise search system, you can select one to use for searching this  
| collection.

| **Log** Specify the types of messages that you want to log and options for  
| creating and rotating log files. You can also configure alerts so that you  
| can be notified when certain events occur, and specify options for  
| receiving e-mail whenever certain messages, or certain types of  
| messages, are logged.

---

## **Deleting a collection**

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

### **Before you begin**

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.

| **Tip:** You might see a message about the requested operation timing out even  
| though the process is still running in the background. To determine whether  
| the task has completed, click **Refresh** in the administration console (do not  
| click **Refresh** in the Web browser). The delete process is finished when the  
| collection name no longer appears in the list of collections.

### **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**.

---

## 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

If you enable security for a collection when you create it, you can configure document-level security options. Each crawler can associate security tokens with the documents that it crawls. If you specify that you want to use document-level security when you configure the crawler, the crawler associates the security tokens that you specify with each document, and these tokens are added to the index with the documents.

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 capability 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.

For Lotus Notes sources, you can specify additional security controls. For example, you can specify that you want the user's credentials to be validated against access controls that are maintained by the original data source during query processing. This validation of current credentials can be done instead of or in addition to the validation that occurs when the user's credentials are compared to security tokens that are stored in the enterprise search index.

## 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 173

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

“URI formats in an enterprise search index” on page 71

The uniform resource identifier (URI) of each document in an enterprise search index 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.

### Before you begin


To create a crawler, you must 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.

#### **Related concepts**

“Enterprise search administration overview” on page 15

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, associate search applications with collections, and specify information to enforce security.

---

## **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.

### **Before you begin**



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
4.  **Crawler properties**.
5. Change the crawler properties, then click **OK**.
5. For the changes to become effective, stop and restart the crawler. (If you change only the crawler name or description, you do not need to 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.

### **Before you begin**



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.

### Before you begin

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.



### About this task

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

**Tip:** Because this task takes time to complete, you might see a message about the requested operation timing out even though the process is still running in the background. To determine whether the task has completed, intermittently click **Refresh** in the administration console (do not click **Refresh** in the Web browser). The delete process is finished when the crawler name no longer appears in the list of crawlers.

### 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 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 37.
- “Configuring the crawler server on Linux for DB2 Content Manager” on page 38.
- “Configuring the crawler server on Windows for DB2 Content Manager” on page 39.

### 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 and specify which attributes can be returned in the search results.
- Configure document-level security options. If security was enabled for the collection when it was created, the crawler can associate security data with documents in the index, which enables search applications to enforce access controls based on the stored security data.

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 options for crawling documents in an item type.
- Change document-level security options.
- 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 Edition on an IBM AIX<sup>®</sup> 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<sup>™</sup> 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 username and password for the DB2 Content Manager library server 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 CAdmin_ID CAdmin_password
```

5. Run the setup script for the Content Manager crawler on the crawler server:
  - a. Change to the ES\_INSTALL\_ROOT/bin directory:

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

```
escrcm.sh
```
6. Stop and restart the enterprise search system, including 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 Edition 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 Edition 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 username and password for the DB2 Content Manager library server 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 CMadmin_ID CMadmin_password
```
  5. Run the setup script for the Content Manager crawler on the crawler server:
    - a. Change to the ES\_INSTALL\_ROOT/bin directory:
 

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

```
./escrcm.sh
```
  6. Stop and restart the enterprise search system, including 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 Edition 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 Edition 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 username and password for the DB2 Content Manager library server 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 CMadmin_ID CMadmin_password
```

4. Run the setup script for the Content Manager crawler on the crawler server:
  - a. Change to the ES\_INSTALL\_ROOT\bin directory:  
cd %ES\_INSTALL\_ROOT%\bin
  - b. Start the following script and answer the prompts:  
escrcm.vbs

5. Stop and restart the enterprise search system, including all sessions on the enterprise search common communications layer (CCL):

- a. At a command prompt, stop the enterprise search system:  
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 enterprise search system:  
`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 Microsoft SQL Server 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.  
**Attention:** To optimize the performance of the discovery processes (and to prevent the crawler configuration process from timing out), 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 which columns can be returned in the search results.
- Configure document-level security options. If security was enabled for the collection when it was created, the crawler can associate security data with documents in the index, which enables search applications to enforce access controls based on the stored security data.

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 options for crawling the columns in a table.
- Change document-level security options.
- 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.
- Configure document-level security options. If security was enabled for the collection when it was created, the crawler can associate security data with documents in the index, which enables search applications to enforce access controls based on the stored security data.

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 options for crawling and searching documents in a subfolder.
- Change document-level security options.
- 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 design, 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.
- Specify patterns to include news groups, and specify patterns to exclude certain news groups from the crawl space.
- Configure document-level security options. If security was enabled for the collection when it was created, the crawler can associate security data with documents in the index, which enables search applications to enforce access controls based on the stored security data.

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.
- Change document-level security options.

- 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 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 45.
- “Configuring the crawler server on Windows for Lotus Notes” on page 47.
- “Configuring the DIIOP protocol for Notes crawlers” on page 49.

**Requirement:** 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). When you create the crawler, you select the databases or directories to crawl from a single Lotus Notes server. Later, when you edit the crawl space, you can add documents from another Lotus Notes server that you want to include in the same crawl space. When you create or edit the crawler, you can specify whether you want to crawl all databases or directories on the server, or whether you want to crawl specific databases, 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 documents in the crawl space.
- Identify the Lotus Notes server host name, port, and communications protocol.
- Select the databases or directories that you want to crawl.
- Set up a schedule for crawling the databases or directories.

- Select the documents that you want to crawl. You can crawl all documents in a directory, all documents in a database, or documents from selected views and folders of a database.
- Specify options for making the fields in various databases, views, and folders searchable. For example, you can exclude certain fields from the crawl space and specify options for searching attachments.
- Configure document-level security options. If security was enabled for the collection when it was created, the crawler can associate security data with documents in the index, which enables search applications to enforce access controls based on the stored access control lists or security tokens. You can also select an option to validate the user's credentials at the time the user submits a query. In this case, instead of comparing the user's credentials to indexed security data, the credentials are compared to current access control lists that are maintained by the original data source.

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 directories, databases, views, and folders to the crawl space. You can add documents from the same Lotus Notes server or specify information about a new Lotus Notes server that you want to include in the crawl space.
- Change the password that the crawler uses to access a Lotus Notes server.
- Change the schedule for crawling databases or directories.
- Change the options for crawling documents in an individual directory, database, view, or folder.
- Change security options for searching documents in an individual directory, database, view, or folder.
- Edit a set of default options for document-level security and for how fields are to be crawled and searched. The Notes crawler uses these options with all directories, databases, views, and folders that you do not configure individual options for.
- If a directory, database, view, or folder uses default options that were defined for the entire crawl space, specify that you want to configure options for the documents in that resource. If you no longer want to use the options that you configured explicitly for a directory, database, view, or folder, specify that you want to start using the default options for the crawl space.
- Remove Lotus Notes directories, databases, views, and folders from the crawl space.

## Configuring the crawler server on UNIX for Lotus Notes

If you install DB2 II OmniFind Edition 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® 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 Edition 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 when DB2 II OmniFind Edition was installed).
  - b. Start the following script, which 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?.

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 system, including 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 Edition 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 Edition 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 with the enterprise search administrator ID (this user ID was specified when DB2 II OmniFind Edition was installed).

- b. Start the following script, which is installed in the %ES\_INSTALL\_ROOT%\bin directory:

escrnote.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 system, including all sessions on the enterprise search common communications layer (CCL):

- a. At a command prompt, stop the enterprise search system:

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 enterprise search system:

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.

### Before you begin

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 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 configure options for crawling this Lotus Notes, specify this user ID and password on the Specify the 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 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 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.
- Configure document-level security options. If security was enabled for the collection when it was created, the crawler can associate security data with documents in the index, which enables search applications to enforce access controls based on the stored security data.

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 options for crawling documents in a subdirectory.
- Change document-level security options.
- Remove subdirectories from the crawl space.

---

## VeniceBridge crawlers

To include Venetica VeniceBridge repositories in an enterprise search collection, you must configure a VeniceBridge crawler. (VeniceBridge is being renamed to IBM WebSphere Information Integrator Content Edition.)

### Setting up the crawler server

Before you can crawl VeniceBridge sources, you must run a script on the crawler server. This script, which is provided with DB2 Information Integrator OmniFind Edition, enables the VeniceBridge crawler to communicate with the target 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 52.
- “Configuring the crawler server on Windows for VeniceBridge” on page 53.

## Configuring a VeniceBridge crawler

You can use the VeniceBridge crawler to crawl Documentum, FileNet Panagon Content Services and Hummingbird repositories. You must configure a separate crawler for each 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 server.
- Select the server that you want to crawl.
- Select the Documentum, FileNet, and Hummingbird 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 that you want to crawl a particular version of a repository.
- Configure document-level security options. If security was enabled for the collection when it was created, the crawler can associate security data with documents in the index, which enables search applications to enforce access controls based on the stored security data.

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 options for crawling documents in an item class.
- Change document-level security options.
- Remove item classes from the crawl space.

## Configuring the crawler server on UNIX for VeniceBridge

If you install DB2 II OmniFind Edition 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 Edition 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 Edition 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 when DB2 II OmniFind Edition was installed).
  - b. Start the following script, which is installed in the `$ES_INSTALL_ROOT/bin` directory), and answer the prompts:

```
esrvbr.sh
```
3. Stop and restart the enterprise search system, including 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 Edition 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 Edition 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 ID (this user ID was specified when DB2 II OmniFind Edition was installed).
  - b. Start the following script, which is installed in the `%ES_INSTALL_ROOT%\bin` directory, and answer the prompts:
 

```
escrvbr.vbs
```
3. Stop and restart the enterprise search system, including all sessions on the enterprise search common communications layer (CCL):
  - a. At a command prompt, stop the enterprise search system:
 

```
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 enterprise search system:
 

```
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 then follows links in documents to crawl additional documents. The Web crawler can crawl and extract links from individual pages or *framesets* (pages that are created with HTML frames).

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.
- Specify patterns and rules to allow and forbid visits to certain Web sites or certain types of documents.
- Specify rules for how the Web crawler handles soft error pages.
- Configure document-level security options. If security was enabled for the collection when it was created, the crawler can associate security data with documents in the index, which enables search applications to enforce access controls based on the stored security data.
- Specify options for crawling password-protected Web sites (the Web servers to be crawled must use HTTP basic authentication or HTML forms to prompt for passwords).
- Specify options to crawl Web sites that are served by a proxy server.

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.

- Change the crawl space. For example, you can:
  - 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.
  - Exclude Web sites or certain types of documents from the crawl space.
  - Change the rules for handling soft error pages.
  - Change document-level security options.
  - Change the security credentials that enable the crawler to crawl password-protected Web sites or Web sites that are served by a proxy server.

## User agent configuration

The Web crawler complies with the Robots Exclusion protocol. To crawl a Web site that uses this protocol, ensure that the robots.txt file on the Web site allows the user agent name that you configure for the Web crawler to access the Web site.

When the enterprise search system is started, the Web crawler loads the user agent name that you configure for it. Before the crawler downloads a page from a Web site that it has not previously visited (or that it has not visited for some time), the crawler first tries to download a file called robots.txt. This file is in the root directory of the Web site.

If the robots.txt file does not exist, the Web site is open to unrestricted crawling. If the file does exist, it specifies what areas of the site (directories) are off limits to crawlers. The robots.txt file specifies permissions for crawlers by identifying their user-agent name.

The Robots Exclusion protocol is voluntary, but the enterprise search Web crawler complies with it:

- If a robots.txt file contains an entry for the user agent name that is configured for the Web crawler, then the Web crawler complies with the restrictions on that user agent.
- If the user agent name does not appear in the robots.txt file, but the last entry specifies `User-agent: *` (which means any user agent) and the restriction is `Disallow: /` (which means do not allow any crawling, starting at the root of the Web site), then the Web crawler is barred from crawling that site.
- If the user agent name does not appear in the robots.txt file, but the last entry specifies `User-agent: *` and the restriction is `Allow: /`, then the Web crawler is allowed to crawl that site.

Web site administrators often specify a final entry that bars access to all crawlers that are not explicitly granted access. If you are configuring a new Web crawler and you know that some of the Web sites that you want to crawl use the Robots Exclusion protocol, ask the Web site administrators to add an entry for your crawler to their robots.txt files.

Be sure to specify the same user agent name in the Web crawler's properties and in all robots.txt files that belong to the Web sites of interest.

If none of the Web sites to be crawled use the Robots Exclusion protocol, then the value that you specify for the user agent property typically does not matter. However, some application servers, JSPs, and servlets tailor their responses to the user agent name. For example, different responses exist to handle browser incompatibilities. The user agent name that you specify for the Web crawler might



matter in these situations, regardless of the Robots Exclusion protocol. If you need to crawl these types of sites, consult with the Web site administrators to ensure that the Web crawler is allowed access.

## Support for JavaScript

The Web crawler for enterprise search can find some links (URLs) that are contained in the JavaScript™ portions of Web documents.

The Web crawler can find both relative and absolute links. If an HTML document contains a BASE element, the crawler uses that element to resolve relative links. Otherwise, the crawler uses the document's own URL.

Support for JavaScript is limited to link extraction. The crawler does not parse JavaScript, does not build a DOM (Document Object Model), and does not interpret or execute JavaScript statements. The crawler looks for strings in the document content (including, but not limited to the JavaScript portions) that are likely to be URLs in JavaScript statements. This means two things:

- Some URLs will be found that are ignored by the stricter HTML parser. The crawler will reject anything that is not a syntactically valid URL, but some of the valid URLs returned by the scanning step might be of low interest for searching.
- Document content that is generated by JavaScript, such as when a human user views a page with a browser and the browser executes some JavaScript, will not be seen by the Web crawler, and thus will not be indexed.

## 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`.

A domain rule that does not specify a port number applies to all ports on that domain. In the following example, all ports on the `foo` domain are allowed:

```
allow domain foo.ibm.com
```

If a domain rule specifies a port number, then the rule applies only to that port. In the following example, only port 443 on the `bar` domain is allowed:

```
allow domain bar.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:

```
allow 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.)

```
forbid 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 should end with the following default rule, which eliminates domains that were not allowed by a preceding rule:

```
forbid domain *
```

This rule is critical, because it prevents the crawl space from including the entire Internet.

- 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.

```
forbid 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.

## How the Web crawler handles soft error pages

You can configure the Web crawler to handle custom pages that Web site administrators create when they do not want to return a standard error code in response to requests for certain pages.

If an HTTP server cannot return the page that a client requests, the server normally returns a response that consists of a header with a return code. The return code indicates what the problem is (such as error 404, which indicates that the file could not be found). Some Web site administrators create special pages that explain the problem in more detail and configure the HTTP server to return these pages instead. These custom pages are called soft error pages.

Soft error pages can distort the Web crawler's results. For example, instead of receiving a header that indicates a problem, the crawler receives a soft error page and the return code 200, which indicates the successful download of a valid HTML page. But this downloaded soft error page is not related to the requested URL, and its content is nearly identical each time it is returned in place of a requested page. These irrelevant and near-duplicate pages distort the index and search results.

To handle this situation, you can specify options for handling soft error pages when you configure the Web crawler. The Web crawler needs the following information about each Web site that returns soft error pages:

- A URL pattern for a site that uses soft error pages. This URL pattern consists of the protocol (HTTP or HTTPS), the host name, port number (if non standard), and path name. You can use an asterisk (\*) as a wildcard character to match one or more characters up to the next occurrence of a non-wildcard character in the pattern. The pattern that you specify is case sensitive.
- A title pattern for text that corresponds to the <TITLE> tag of an HTML document. You can use the asterisk (\*) as a wildcard character to specify this pattern. This pattern that you specify is case sensitive.
- A content pattern for text that corresponds to the content of an HTML document. The content is not just the content of the <BODY> tag, if a <BODY> tag is present. The content is everything that comes after the HTTP header in the file. You can use the asterisk (\*) as a wildcard character to specify this pattern. This pattern that you specify is case sensitive.
- An integer that represents the return code to use for documents that match the URL, title, and content patterns that you specified.

## Example

This following configuration tells the Web crawler to compare all valid HTML pages (return code 200) that are returned from the `http://www.mysite.com/hr/*` Web site to the specified title and content patterns. If the <TITLE> tag of a page begins with "Sorry, the page" and the content of the document contains anything (\*), then the crawler handles the page the same way it would a return code 404 (the page was not found).

*Table 1. Soft error page example*

URL Pattern	Title pattern	Content pattern	HTTP return code
<code>http://www.mysite.com/hr/*</code>	Sorry, the page*	*	404

You can create multiple entries for the same Web site to handle different return codes. Each return code from the same Web site requires its own entry in the Web crawler's configuration.

## Using wildcards

The URL, title, and content patterns are not regular expressions. The asterisk character matches any characters up to the next occurrence of any non-wildcard character. For example:

```
*404 matches any characters404
404: * matches 404: any characters
http://*.mysite.com/* matches http://any host.mysite.com/any file
* matches any characters
```

## Performance impact

When you configure options for handling soft error pages, you increase the amount of crawler processing time because all successfully crawled pages must be checked. More processing time is required to check for pattern matches and determine whether a page or a replacement return code should be returned.

## Web sites served by proxy servers

If the Web crawler is not permitted direct access to a network, you can configure the crawler to use an HTTP proxy server to access the content that you want to crawl.

If access to a TCP/IP network is not available on the computer where the Web crawler is to run, or if access is restricted to privileged processes, you can configure the Web crawler to use an HTTP proxy server. An HTTP proxy is a process that listens at a specified port on a specified host for HTTP requests. The proxy server relays requests to the Web server, and relays responses from the Web server to the requesting client (the Web crawler). A proxy server can run on the same computer with the Web crawler, or run on a different computer.

In non-proxy crawling, a request for a URL is sent directly to the host. With proxy crawling, the request is sent to the proxy server.

When you create a Web crawler or edit the crawl space, specify information about the proxy servers that the Web crawler is to use when crawling pages the proxy server's domain. Obtain the following information before you add a proxy server to the crawl space:

### Proxy server domains

The domains that are served by the proxy server. You can use an asterisk (\*) as a wildcard character. For example, \* matches all domains that are served by this proxy server, and \*.resource.com matches all domains that end in resource.com.

### Proxy server host name or IP address

The DNS host name or dotted IP address of the proxy server.

### Proxy server port number

The TCP/IP port number where the proxy server listens for HTTP proxy requests.

After you add a proxy server, you must select it and position it in the order that you want the crawler to process it. List the more specific domain names first, and put the more generic domain names lower in the list. When the Web crawler evaluates a candidate URL, it uses the proxy server data that is specified for the

first domain in the list that matches the candidate URL. (URLs that do not match any proxy rule are assumed to be directly accessible to the crawler.)

## Support for crawling secure Web sites

By specifying credentials in the enterprise search administration console, you can enable the Web crawler to access restricted content, such as documents that require a password for access.

If a Web server uses HTTP basic authentication or HTML form-based authentication to restrict access to Web sites, you can specify credentials in the Web crawler's configuration that enable pages on the password-protected Web sites to be crawled. You can also specify options for manually configuring cookie files.

### Web sites protected by HTTP basic authentication

If a Web server uses HTTP basic authentication to restrict access to Web sites, you can specify authentication credentials that enable the Web crawler to access password-protected pages.

To determine whether a user (or client application) has permission to access pages on a Web site, many Web servers use a client authentication scheme called HTTP basic authentication to establish the user's identity. Typically, this interaction is interactive:

- When an HTTP user agent (such as a Web browser) requests a page that is protected by HTTP basic authentication, the Web server responds with a 401 return code, which indicates that the requestor is not authorized to access the requested page.
  - The Web server also challenges the requestor to present credentials that can be used to verify whether the user is allowed to access the restricted content.
  - The Web browser presents the user with a dialog that requests a user name, password, and any other information that is required to constitute the user's credentials.
  - The Web browser encodes the credentials, then includes them when it repeats the request for the protected page.
  - If the credentials are valid, the Web server responds with a 200 return code and the contents of the requested page.
  - Subsequent requests for pages from the same Web server typically include the same credentials, enabling the authorized user to access additional restricted content without being challenged to specify credentials with each request.
- Once a user's identity is established, the Web server and HTTP user agent typically exchange tokens, called *cookies*, that enable knowledge of the user's login status to be maintained between HTTP requests.

Because the Web crawler does not run interactively, the credentials that enable it to crawl password-protected pages must be specified before the crawler begins crawling. When you create a Web crawler or edit the crawl space, specify information about each secure Web site that needs to be crawled.

To specify this information, you must work closely with the administrators for the Web sites or Web servers that are protected by HTTP basic authentication. They must provide you with the security requirements for the Web sites to be crawled, including all information that is used to authenticate the Web crawler's identity and determine that the crawler has permission to crawl the restricted pages.



If security was enabled for the collection when the collection was created, you can specify security tokens, such as user IDs, group IDs, or user roles, to control access to documents when you configure the crawler. The Web crawler associates these security tokens with every document that it crawls in the file system tree for the specified root URL. The tokens are used in addition to any document-level security tokens that you configure for the entire Web crawl space.

The order of the URLs is important. After you add information about a password-protected Web site, you must position it in the order that you want the crawler to process it. List the more specific URLs first, and put the more generic URLs lower in the list. When the Web crawler evaluates a candidate URL, it uses the authentication data that is specified for the first URL in the list that matches the candidate URL.

### **Web sites protected by form-based authentication**

If a Web server uses HTML forms to restrict access to Web sites, you can specify authentication credentials that enable the Web crawler to access password-protected pages.

To determine whether a user (or client application) has permission to access pages on a Web site, many Web servers use HTML forms to establish the user's identity. Typically, this interaction is interactive:

- When an HTTP user agent (such as a Web browser) requests a page that is protected by form-based authentication, the Web server checks to see whether the request includes a cookie that establishes the user's identity.
- If the cookie is not present, the Web server prompts the user to enter security data into a form. When the user submits the form, the Web server returns the required cookies, and the request for the password-protected page is allowed to proceed.
- Future requests that include the required cookies are also allowed to proceed. The authorized user is able to access additional restricted content without being challenged to fill in a form and specify credentials with each request.

Because the Web crawler does not run interactively, the credentials that enable it to crawl password-protected pages must be specified before the crawler begins crawling. When you create a Web crawler or edit the crawl space, specify information about each secure Web site that needs to be crawled.

The fields that you specify correspond to the fields that an interactive user fills in when prompted by the Web browser, and any hidden or static fields that are required for a successful login.

To specify this information, you must work closely with the administrators for the Web sites or Web servers that are protected by form-based authentication. They must provide you with the security requirements for the Web sites to be crawled, including all information that is used to authenticate the Web crawler's identity and determine that the crawler has permission to crawl the restricted pages.

The order of the URL patterns is important. After you add information about a password-protected Web site, you must position it in the order that you want the crawler to process it. List the more specific URL patterns first, and put the more generic URL patterns lower in the list. When the Web crawler evaluates a candidate URL, it uses the form data that is specified for the first URL pattern in the list that matches the candidate URL.



## Cookie administration

Typically, cookie administration occurs automatically, with no action required from an enterprise search administrator. If necessary, you can manually specify cookies for a Web crawling session.

Cookies are opaque tokens that a Web server returns to a user agent as part of an HTTP response header. They are meaningful only to the Web server that issued them, and they are used to maintain state between HTTP requests. For example, during client authentication, the Web server might return a cookie that enables the server to determine that an authenticated user is already logged in. The presence of the cookie enables the user to issue additional requests for pages on that Web server without being prompted to log in again.

The Web crawler retains cookies that are received from Web servers and uses them for the duration of the crawler instance. It stores the cookies in a `cookies.ini` file, which is rewritten by the crawler at the end of every crawler session. When the Web crawler stops, it saves all unexpired cookies, then reloads them at the start of the next session.

If you manually specify cookies, store them in a separate file, and then merge them with the cookies in the `cookies.ini` file when needed. The crawler does not discard unexpired cookies, but if a problem prevents the writing of the entire cookie collection, you do not want to lose the cookies that you manually specified. You must merge your cookies with the cookies that the crawler automatically maintains before the start of a crawling session.

### Cookie format

Cookies that you plan to merge with the enterprise search `cookies.ini` file must be in a particular format.

- Each cookie must be on a single line. Blank lines and comments are permitted, but they will not be preserved in the `cookies.ini` file.
- Each cookie must have the following format:

```
CookieN(cookie_length,URL_length)cookie_text,validation_URL
```

#### **Cookie**

A required keyword that indicates the start of a cookie entry.

The Cookie keyword cannot contain blanks and it must have a single digit appended to it, either 0, 1, or 2. The digit indicates the cookie type: version-0 (Netscape), version-1 (RFC2109), or version-2 (RFC2965). Port lists are not supported in RFC2965 cookies.

*cookie\_length*

The length in characters of the associated cookie text.

*URL\_length*

The length in characters of the associated validation URL.

*cookie\_text*

The content of the cookie that is to be sent to the originating Web server. This string (which represents the right side of the Set-Cookie directive in an HTTP response header) specifies the cookie's name and value pair and any other content (such as a path, security setting, and so on) to be sent with the cookie. This string is followed by a comma (,) separator.

*validation\_URL*

The URL at which this cookie was discovered. This URL is used to determine where to send the cookie (for example, by supplying a domain

name and path name). The validation URL must satisfy the originating Web server's security and privacy restrictions on cookies.

The following example is shown on two lines for readability; cookies that you specify must be on a single line:

```
Cookie0(53,40)ASPSESSIONIDQSQTACSD=SLNSIDFNLSIDNLSIDNLSNL;path=/  
https://www.ibm.com:443/help/solutions/
```

## Configuring cookies for the Web crawler

You can manually specify cookies for a Web crawling session, and merge them with cookies that the Web crawler maintains.

### Before you begin

To manually configure cookies for the Web crawler to use, you must be an enterprise search administrator.

### Procedure

To manually configure cookies for a Web crawler:

1. From the enterprise search administration console, monitor the collection that you want to specify cookies for, and stop the Web crawler.
2. Log in as the enterprise search administrator on the crawler server (this user ID was specified when DB2 II OmniFind Edition was installed).
3. Change to the data directory for the crawler that you want to configure, where *crawler\_session\_ID* is an ID that was assigned to the crawler session by the enterprise search system. For example:  
`ES_NODE_ROOT/data/col_56092.WEB_88534`
4. Edit the `cookies.ini` file, append the cookie entries that you manually specified to the ones that are already listed, then save and exit the file. Ensure that your cookies do not override any that are already present.
5. From the enterprise search administration console, restart the Web crawler that you stopped.

## Global Web crawl space configuration

You can configure a global crawl space for Web crawlers, which enables you to better control the removal of URLs from the index.

Each Web crawler is configured with a crawl space that defines the URLs that are to be crawled or not crawled. Discovered URLs that are in the crawl space are retained (in a database) for later crawling; URLs that are not in the crawl space are discarded. If the crawler starts with an empty database, the crawl space definition and database remain consistent while the crawler runs.

Sometimes a crawler is stopped, and its crawl space is reduced (for example, by new rules that forbid pages to be crawled). When the crawler is restarted, its crawl space definition and database become inconsistent. The database contains URLs (some crawled and some not crawled) which are not in the new, smaller crawl space.

If a collection has only one Web crawler, the Web crawler can restore consistency by changing the HTTP return codes for these URLs to 760 (which specifies that they are to be excluded) and requesting the removal of the now-excluded pages from the index.

If you divide the crawl space between two or more Web crawlers (for example, to ensure some pages are crawled more often than the rest), each Web crawler maintains independent database tables (initially empty), and they each crawl a different part of the Web crawl space. The original crawler's crawl space is then reduced to whatever is left after the parts to be crawled by other crawlers are removed. Problems arise when the original crawler attempts to restore consistency by removing the moved pages from the index. Because the moved pages are now being crawled by other crawlers, the pages should remain in the index.

By configuring a higher level, global crawl space you can identify URLs that are not to be crawled by the original crawler, but are not to be removed from the index, either. URLs that are no longer in any crawler's crawl space continue to be marked for exclusion by the discovery processes, and are removed from the index when they are recrawled.

The global crawl space is defined by a configuration file named `global.rules`, which must exist in the crawler configuration directory (the presence of a `global.rules` file enables the global crawl space function). If this file exists, it is read during crawler initialization. If this file does not exist, the crawler operates with a single-level crawl space, and removes documents from the index as necessary to maintain consistency between its crawl space definition and database.

If a global crawl space exists, the crawler rules URLs in or out as before, but will request the removal of a URL from the index only if the URL is not in any Web crawl space.

The `global.rules` file has the same syntax as the local `crawl.rules` file, except that it can contain only domain name rules. This restriction enables a crawl space to be partitioned between crawlers only on the basis of DNS host names, not IP addresses or HTTP prefix patterns. URLs that are excluded by URL prefix or IP address rules in the local crawl space (as defined in the `crawl.rules` file) are unaffected by the global crawl space; such URLs are still excluded.

The global crawl space is used only to prevent the removal of URLs, which are excluded from one crawler's crawl space by a local domain rule, from the index. The following rules apply in the following order:

1. If a URL from the crawler's database is excluded by a local prefix rule or address rule, the URL is assigned return code 760 and it is removed from the index. The URL will not be crawled again.
2. If a URL from the crawler's database is excluded by a local domain rule, and there is no global crawl space, the URL is assigned return code 760, and it is removed from the index. The URL will not be crawled again.
3. If a URL from the crawler's database is excluded by a local domain rule, but explicitly allowed by a rule in the global crawl space, the URL is assigned return code 761. The crawler will not crawl the URL again, but it is not removed from the index (it is assumed to be in some other crawler's local crawl space).
4. If a URL from the crawler's database is excluded by a local domain rule, and not explicitly allowed by a rule in the global crawl space, the URL is assigned return code 760, and removed from the index.

Because the global crawl space is consulted only to prevent the deletion of URLs that have already been excluded by the local crawl space, the default result from the global crawl space, if no rule applies to a candidate URL, is to forbid it from being crawled.

The `global.rules` file must exist in the `master_config` directory of every crawler that shares the global crawl space. You must carefully edit all copies of the `global.rules` file and the individual local `crawl.rules` files to ensure that they remain mutually consistent.

## No-follow and no-index directives

You can improve search quality by specifying directives for the Web crawler that control whether links on pages are followed and whether pages are indexed.

Some Web pages have no-follow or no-index directives, which instruct robots (such as the Web crawler) to not follow links found in those pages, to not include the contents of those pages in the index, or to not do either of these actions.

Controlling these settings can improve the quality of the crawl. For example, some directory pages can contain thousands of links but no other useful content; those pages should be crawled, and their links followed, but there is no benefit to indexing the directory pages themselves.

There might also be times when you want the crawler to go no lower in a hierarchy, but the desired leaf pages contain links and do not contain no-follow directives. Because some of these pages are automatically generated, they have no owners who might insert the required directives.

To specify rules for crawling such pages, you create or edit a configuration file named `followindex.rules`. Use the following guidelines when you specify rules in this file:

- The rules that you configure must specify URL prefixes (you cannot identify Web sites by IP address or DNS host name).
- The URL prefixes can include asterisks (\*) as a wildcard character to allow or forbid multiple sites with similar URLs.
- Order is significant (the crawler applies the first rule that matches a candidate URL).
- The rules, which explicitly allow and forbid following or indexing, override other settings, including those in the target document.

## Overriding no-follow and no-index directives in Web pages

You can specify rules in a configuration file to control whether the Web crawler follows links to pages or indexes pages that contain no-follow or no-index directives.

### Before you begin

To specify no-follow and no-index directives for the Web crawler, you must be an enterprise search administrator. The directives that you specify override directives that exist in the pages to be crawled.

### Procedure

To override no-follow and no-index directives:

1. From the enterprise search administration console, monitor the collection that you want to configure rules for, and stop the Web crawler.
2. Log in as the enterprise search administrator on the crawler server (this user ID was specified when DB2 II OmniFind Edition was installed).

3. Change to the configuration directory for the crawler that you want to configure, where *crawler\_session\_ID* is an ID that was assigned to the crawler session by the enterprise search system. For example:  
`ES_NODE_ROOT/master_config/col_56092.WEB_88534`
4. Create or edit a file named `followindex.rules`.
5. Type rules for the crawler in the following format, where *URLprefix* is the beginning characters for the Web sites that you want to allow or forbid to be followed or indexed:  
`forbid follow URLprefix`  
`allow follow URLprefix`  
`forbid index URLprefix`  
`allow index URLprefix`
6. Save, then exit the file.
7. From the enterprise search administration console, restart the Web crawler that you stopped.

---

## 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 UNIX computer, you cannot use that server to crawl Windows file system sources (the Windows 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 a user ID and password that enables the crawler to access files in a particular subdirectory.
- Configure document-level security options. If security was enabled for the collection when it was created, the crawler can associate security data with documents in the index, which enables search applications to enforce access controls based on the stored security data.

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 options for crawling and searching documents in a subdirectory.
- Change document-level security options.
- Remove subdirectories from the crawl space.

---

## Configuring support for Data Listener applications

You can extend enterprise search by using the Data Listener API to create an external crawler. Your custom Data Listener applications can add data to a collection, remove data from a collection, or instruct a Web crawler to visit and revisit URLs.

### Before you begin

To configure Data Listener applications, you must be a member of the enterprise search administrator role.

### About this task


A client Data Listener application enables the crawling of data source types that cannot be crawled by the default crawlers for enterprise search. Before you can use a Data Listener application, you must configure credentials that enable the application to access and update collections.

When your client Data Listener application connects to the Data Listener, it must pass in the client application ID and password and the ID of the collection to be updated. This information must match the information that you configure for the application in the administration console.



The Data Listener is started automatically when the enterprise search system is started. If you change the port number after you configure the application in the administration console, you must restart the Data Listener.

### Procedure

To configure support for your Data Listener applications:

1. Click **System** to open the System view.
2. Click  **Edit** to change to the system editing view.
3. On the Data Listener page, click **Configure support for Data Listener applications**.
4. On the Data Listener Applications page, specify authentication information that enables your client Data Listener applications to access enterprise search collections.

The Data Listener client IDs must be unique within the enterprise search system, and each Data Listener application that you add can update a single collection only.

5. Click **OK**.
6. If you changed the Data Listener port number, take the following steps to restart the Data Listener:
  - a. Click  **Monitor** to change to the system monitoring view.
  - b. On the Data Listener page, click  **Restart**.

#### **Related concepts**

“Customizing enterprise search” on page 10

The application programming interfaces for enterprise search enable you to create custom search applications, custom applications to update the content of collections, custom programs for text analysis, and custom synonym dictionaries.

“Data listener” in “Programming Guide and API Reference for Enterprise Search”

#### **Related tasks**

“Monitoring the Data Listener” on page 185

Monitor the Data Listener to see its status and to view details about client Data Listener application activity.

---

## **File types supported by enterprise search crawlers**

Enterprise search collections support a wide variety of document types and content types.

When you configure the crawl space for a Web crawler, you specify which types of the documents the crawler is to include or exclude. The Web crawler can crawl any Web-accessible document that meets the following conditions:

- The document must have a MIME type that matches one of the MIME types configured to be acceptable on the Types of Documents to Include page.
- The file extension of the document cannot match any of the file extensions configured to be excluded in the File Extensions to Exclude page.

For many of the other crawler types, you can specify which MIME types and file extensions you want to include or exclude when you configure options for an individual data source (such as a subdirectory, a subfolder, or an item class). A number of file types are excluded from crawling by default, but you can add and delete entries as needed to ensure that the crawler includes the types documents that want to crawl.

---

## **URI formats in an enterprise search index**

The uniform resource identifier (URI) of each document in an enterprise search index 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.

### Example

```
cm://cmsrvctg/ITEMTYPE1/92+3+ICM8+icmnlsdb12+ITEMTYPE159+26+A1001001A  
03F27B94411D1831718+A03F27B+94411D183171+14+1018
```

## 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.

### Examples

Local, cataloged database:

```
db2://LOCALDB/SCHEMA1.TABLE1/MODEL/ThinkPadA20
```

Remote, uncataloged database:

```
db2://myserver.mycompany.com:50001/REMOTEDB/SCHEMA2.TABLE2/NAME/DAVID
```



## 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.

### Examples

Document body:

`exchange://exchangesvr.ibm.com/public/RootFolder1/Folder1/Document.EML`

Document attachment:

`exchange://exchangesvr.ibm.com/public/RootFolder1/Folder1/Document.EML/  
AttachedFile.doc`

Enabled for SSL:

`exchange://exchangesvr.ibm.com/public/TeamRoom/Folder1/Document.EML?  
useSSL=true`

## 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 Number&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 Number*

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.

### **Examples**

A document that was selected for crawling by view or folder:

```
domino:///dominosvr.ibm.com/49256D3A000A20DE/Database.nsf/  
8178B1C14B1E9B6B8525624F0062FE9F/0205F44FA3F45A9049256DB20042D226
```

A document that was not selected for crawling by view or folder:

```
domino:///dominosvr.ibm.com/49256D3A000A20DE/Database.nsf//  
0205F44FA3F45A9049256DB20042D226
```

A document attachment:

```
domino:///dominosvr.ibm.com/49256D3A000A20DE/Database.nsf//  
0205F44FA3F45A9049256DB20042D226?AttNo=0&AttName=AttachedFile.doc
```

### **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.

#### **Example**

```
file:///home/user/test.doc
```

### **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.

#### **Examples**

Documentum:

```
vbr://vbrsrv.ibm.com/Documentum/c06b/094e827780000302//CONTENT/?  
java.naming.provider.url=iiop%3A%2F%2Fmyvbr.ibm.com%3A2809&  
java.naming.factory.initial=com.ibm.websphere.naming.WsnInitContextFactory
```

FileNet PanagonCS:

```
vbr://vbrsrv.ibm.com/PanagonCS/4a4c/003671066//CONTENT/?Page=1&  
java.naming.provider.url=iiop%3A%2F%2Fmyvbr.ibm.com%3A2809&  
java.naming.factory.initial=com.ibm.websphere.naming.WsnInitContextFactory
```

#### **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.

#### **Examples**

Local file system:

```
file:///d:/directory/test.doc
```

Network file system:

```
file:///filesrv.ibm.com/directory/file.doc
```

#### **Related concepts**

*“Enterprise search crawler administration” on page 31*

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 tasks**

*“Removing URIs from the index” on page 113*

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 122*

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 83*

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 110*

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 172*

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.

---

## Parsing documents for enterprise search

To enhance the retrievability of documents, you can specify options for how documents and metadata are to be parsed, analyzed, and categorized before they are added to the enterprise search index.

The options that you can specify for parsing document content and optimizing the retrievability of information include the following:

### 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 documents that belong to specific categories.

### Configuring search fields

You can map elements in XML documents to search fields in the index. You can also map metadata elements in HTML documents to search fields. By creating search fields in the enterprise search index, you enable users to query specific parts of XML and HTML documents and improve the precision of the search results.

### Configuring text analysis options

If custom text analysis engines were added to the enterprise search system, you can select one to use with a collection. After you associate an analysis engine with a collection, you can specify options for mapping content so that it can be linguistically analyzed and annotated. You can also specify how the results of the analysis are to be mapped to search fields and searchable attributes in the enterprise search index.

### Configuring options for parsing Chinese, Japanese, and Korean documents

You can specify options for using n-gram segmentation to parse documents that are written in the Chinese, Japanese, and Korean languages. You can also remove new line characters from the white space in Chinese and Japanese documents.

### Enabling native XML search

If your collection includes XML documents, you can enable them to be searched with native XML query syntax, such as XPath and XML fragments. A native XML search enables users to specify queries based on the relationships between various XML elements.

### Related concepts

"Linguistic support" in "Programming Guide and API Reference for Enterprise Search"

"Text analysis included in enterprise search" in "Programming Guide and API Reference for Enterprise Search"

"Unstructured information management architecture (UIMA)" in "Programming Guide and API Reference for Enterprise Search"

---

## 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 capability, 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 based on document content (for example, 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.

### **Applying categorization changes**

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 all documents, reparse 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. If you change categories or category rules after documents are crawled and indexed, search quality will be degraded until you recrawl all documents, reparse the documents, and reorganize the index.

### **Document content fields**

To create a collection with model-based categories or with rule-based categories that use document content rules (as opposed to URI pattern rules), ensure that the documents to be categorized contain content fields.

Model-based categories and category rules that are based on document content operate on the content fields of documents and do not take document metadata into account. Only documents that contain content fields, or that contain fields that can be configured to be content fields when the crawler is configured, can be categorized with these categorization options.

You can configure model-based categories or categories based on content rules only for documents that are crawled by the following crawlers:

- Web and NNTP crawlers.
- UNIX file system and Windows file system crawlers.
- DB2 crawlers. When you create the crawler, specify which fields constitute content fields when you configure crawling options for individual tables (the

final step in the crawler wizard). If the crawler already exists, edit the crawl space, select a table to edit, then specify the which fields constitute content fields in that table.

- Content Manager crawlers. When you create the crawler, specify which fields constitute content fields when you configure crawling options for individual item types (the final step in the crawler wizard). If the crawler already exists, edit the crawl space, select an item type to edit, then specify the which fields constitute content fields in that item type.

#### **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.

## **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`

Because all URIs are treated as patterns, the system ignores any asterisks that you specify as wildcard characters at the start or end of the pattern. For example, \*/hr/\* and /hr/ match the same set of URIs.

URI pattern rules are not case sensitive. If a URI contains spaces, the URI pattern must adhere to the enterprise search rules for encoding URIs. The following example shows correct and incorrect ways to specify a URI for a Windows file system path:

Incorrect URI: file:///c:/program files/  
Correct URI: file:///c:/program+files/

### 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
```

Content rules undergo the same linguistic normalizations as Search and Index API (SI-API) queries. However, the syntax for content rules supports a subset of the operations available in the SI-API query syntax. Only the following special characters are allowed:

- + Precede a term with a plus sign to indicate that the term must occur in the document.
- Precede a term with a minus sign to indicate that the term must not occur in the document.
- " Enclose two or more terms in quotation marks to indicate that the exact phrase must occur in the document.

### Related tasks

"Migrating a collection from WebSphere Portal" on page 160

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 159  
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 160

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.

### Before you begin

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, after the documents are parsed, 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

Select this option if you want to use a taxonomy that contains category 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

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.

### Before you begin

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.

**Important:** If you change categories or category rules 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, reparse the documents, and reorganize the index.

### 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 Category Tree page, select the location in the tree where you want to add a category and click **Create a category**.

If you select the root, the new category is created at the root level. If you select a 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: `file:///c:/program+files/finance`

- If you want enterprise search to analyze words in the content fields of documents 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 +"fiscal year"`

e. Click **Finish**.

Your new category is listed on the Category Tree page with the other categories that belong to this collection.

**Related concepts**

“Migration from WebSphere Portal to enterprise search” on page 159  
Enterprise search provides a migration wizard that you can use to migrate taxonomies and collections from IBM WebSphere Portal to enterprise search.

**Related tasks**

“Migrating a model-based taxonomy from WebSphere Portal” on page 159  
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.

**Related reference**

“URI formats in an enterprise search index” on page 71  
The uniform resource identifier (URI) of each document in an enterprise search index 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.

## 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.

### Before you begin

To map XML elements to search fields, you must be a member of the enterprise search administrator role or be a 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:

---

## Working with HTML search fields

Map HTML metadata elements to search fields in the index if you want to enable users to search specific metadata sections of HTML documents.

You use the enterprise search administration console to map HTML metadata elements to search fields.

### HTML search fields

HTML search fields enable users to query attributes of HTML documents.

Metadata elements in HTML documents are similar to document attributes in that they provide information about the document, how it is formatted, and how it is allowed to be accessed on the Web. For example:

```
<meta http-equiv="Content-Type" content="text/html; charset=utf-8" /> ]
<meta name="copyright" content="(C) Copyright IBM Corporation 2005" />
<meta name="content.owner" content="(C) Copyright IBM Corporation 2005" />
<meta name="security" content="public" />
<meta name="abstract" content="This topic describes an IBM product." />
<meta name="format" content="XHTML" />
```

Enterprise search can associate, or map, the names of HTML metadata elements with search field names. When you configure parsing options for a collection, you specify which HTML metadata elements are to be mapped to which search field names. By mapping HTML metadata elements to search fields, you enable users to find documents with those elements by specifying the search 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 HTML documents, and you specify that the copyright and abstract metadata elements are to be indexed as search fields, users can query these specific elements. A search for `copyright:IBM` finds HTML documents that have IBM in the copyright metadata.

#### 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.

### Mapping HTML metadata elements to search fields

When you map an HTML metadata element to a search field, you specify which HTML metadata elements users can search by specifying a field name in a query.

#### Before you begin

To map HTML metadata elements to search fields, you must be a member of the enterprise search administrator role or be a collection administrator for the collection that the HTML documents belong to.

#### About this task


When you create an HTML field mapping, or add, change, or delete fields in an existing HTML field mapping, the changes become effective after you stop and restart the parser. The changes apply only to data that is parsed after you restart



the parser. To update documents that have already been parsed and indexed, you must crawl and parse the documents again, and then reorganize the index.

### Procedure

To map HTML metadata elements to search fields:

1. Click **Collections** to open the Collections view.
2. In the list of collections, locate the collection that you want to configure and click  **Edit**.
3. On the Parse page, click **Map HTML metadata to fields**.
4. On the HTML Field Mappings page, click **Add field** to add a blank line to the list of field mapping rules.
5. In the **Field name** field, type a name that you want to associate with the HTML metadata element that you are mapping. Users can specify this field name when they query HTML documents in this collection.
6. In the **HTML metadata field** field, type the name of the HTML metadata element that you want to map.
7. To enable users to query this field and view it in the search results, select the **Search by field name** and **Show in search results** check boxes.
8. If the data type of this field is **DECIMAL**, **DOUBLE**, **INTEGER**, **SHORT**, **TIME**, or **TIMESTAMP**, and you want to enable users to specify parametric queries when searching this field, select the **Parametric search** check box.
9. Click **OK**.

### Examples:

Users can now query the mapped field names to find HTML documents with specific metadata. For example, if you mapped an HTML metadata element named `description` to a search field named `abstract`, users might enter a query similar to the following to find documents about Thinkpads:

```
abstract:thinkpad
```

---

## Custom text analysis

You can improve the quality and precision of search results by integrating custom text analysis algorithms with enterprise search collections.

DB2 Information Integrator OmniFind Edition supports the IBM Unstructured Information Management Architecture (UIMA), which is a framework for creating, discovering, composing, and deploying text analysis functions. Application developers create and test analysis algorithms for the content to be searched, then create a processing engine archive (.pear file) that includes all of the resources required to use the archive for enterprise search. To be able to search collections with your custom analysis algorithms, you must add the archive (which contains the analysis engine) to the enterprise search system.

The analysis logic component in an analysis engine is called an *annotator*. Each annotator performs specific linguistic analysis tasks. An analysis engine can contain any number of annotators, or it can be a composite of several analysis engines, each of which contain their own custom annotators.



The information produced by the annotators is referred to as the *analysis results*. Analysis results, which correspond to the information that you want to search for, are written to a data structure called a *common analysis structure*.

When you configure text analysis options for a collection, you do the following tasks:

- Select the analysis engine that you want to use for annotating documents in the collection.
- If your collection contains XML documents with meaningful markup, and you want to use this markup in your custom text analysis, associate XML mapping files with the collection, and map the output of the XML mapping to a common analysis structure.

For example, you can map the content of <addressee> and <customer> elements to Person annotations in the common analysis structure. These annotations can then be accessed by your custom annotators, which might detect additional information (for example, they might detect the gender of the Person). You can also map Person annotations to the enterprise search index, which allows users to search for Persons without having to know the original XML elements.

If you want to allow users to specify the original XML elements in queries, then you do not need to define any XML mappings. Instead, you can configure parsing options and enable native XML search for the collection.

- Map a common analysis structure to the enterprise search index, which enables the annotated documents to be searched by semantic search.

For example, depending on the entities and relationships that are detected by the annotators, users can search for concepts that occur in the same sentence (such as a specific person and any competitor name), or a keyword and a concept (such the term Alex and a phone number).

#### **Related concepts**

"Custom text analysis integration" in "Programming Guide and API Reference for Enterprise Search"

"Unstructured information management architecture (UIMA)" in "Programming Guide and API Reference for Enterprise Search"

"Workflow for custom analysis integration" in "Programming Guide and API Reference for Enterprise Search"

"Text analysis algorithms" in "Programming Guide and API Reference for Enterprise Search"

"Semantic search applications" in "Programming Guide and API Reference for Enterprise Search"

"Semantic search query" in "Programming Guide and API Reference for Enterprise Search"

## **Associating analysis engines with the system**

If you create a custom analysis engine, you must add it to the system before you can use it for enterprise search. Collections can use the analysis engine to analyze and annotate documents and improve the precision of search results.

### **Before you begin**





To add analysis engines to the system, you must be a member of the enterprise search administrator role.

### **About this task**

Application developers can create a processing engine archive (.pear file) that adheres to the UIMA framework for text analysis and includes all of the resources required to search enterprise search collections. To be able to search collections with your custom analysis algorithms, you must add the archive (which contains the *analysis engine*) to the enterprise search system.

### Procedure

To add a custom analysis engine to the enterprise search system:

1. Click **System** to open the System view.
2. Click  **Edit** to change to the system editing view.
3. On the Parse page, click **Add Analysis Engine**.
4. Type a descriptive name for the new analysis engine. The system uses this display name to identify the analysis engine throughout the administration console.
5. Either type the fully qualified path for the custom analysis engine (.pear file) that you want to add, or click **Browse** to select the file from your system.
6. Click **OK**. Your analysis engine is listed on the Parse page.
7. To change the display name for the analysis engine, or to see a list of the collections that use this analysis engine, click  **Edit**.
8. To view the XML source for the analysis engine, click  **Details**. By viewing the XML source, you can see the descriptor for the analysis engine. The descriptor shows what information is produced by this engine.
9. To remove the analysis engine, click  **Delete**. You cannot remove an analysis engine from the system if any collections are still using it.

#### Related concepts

"Workflow for custom analysis integration" in "Programming Guide and API Reference for Enterprise Search"

"Custom text analysis integration" in "Programming Guide and API Reference for Enterprise Search"

"Unstructured information management architecture (UIMA)" in "Programming Guide and API Reference for Enterprise Search"

## Associating an analysis engine with a collection

If custom analysis engines are associated with the enterprise search system, you can select one to use with a collection. Users can then specify semantic queries when searching the collection, and improve the quality and precision of the search results.

### Before you begin

To associate an analysis engine with a collection, you must be a member of the enterprise search administrator role or be a collection administrator for the collection.

### About this task


If an analysis engine is already associated with this collection, the following actions occur when you associate a different analysis engine:

- If you select **No custom analysis** when you select an analysis engine, then all text analysis mappings that you previously defined for the collection are reset. The collection begins using the system default values.

- If you select the name of a different custom analysis engine, then all text analysis mappings that you previously defined for the collection are retained. For example, if you change from engine\_1 to engine\_2, then engine\_2 inherits the XML mapping files that you configured for engine\_1.

### Procedure

To associate an analysis engine with a collection:

1. Click **Collections** to open the Collections view.
2. In the list of collections, locate the collection that you want to configure and click  **Edit**.
3. On the Parse page, click **Configure text analysis options**.
4. Click **Select a different analysis engine**. If no custom analysis engines were added to the enterprise search system, or if the collection is currently using the default analysis algorithms, the analysis engine name is **Default**.
5. On the Select an Analysis Engine for This Collection page, select the name of the analysis engine that you want to use with this collection. If no analysis engines are available, or if you select **No custom analysis**, then the parser applies default text analysis rules as it annotates documents and prepares documents for the index.
6. Click **OK**.

#### Related concepts

"Workflow for custom analysis integration" in "Programming Guide and API Reference for Enterprise Search"

"Custom text analysis integration" in "Programming Guide and API Reference for Enterprise Search"

"Unstructured information management architecture (UIMA)" in "Programming Guide and API Reference for Enterprise Search"

## Mapping XML elements to a common analysis structure

If custom analysis engines are associated with the enterprise search system, and if an analysis engine is associated with a collection, you can specify which custom XML mapping files you want to use for parsing documents in the collection.

### Before you begin

To map XML elements to a common analysis structure, you must be a member of the enterprise search administrator role or be a collection administrator for the collection.

### About this task

To enable custom text analysis to access specific elements in XML documents, or to map several XML elements to a common Type for use in semantic search, you can create custom XML mapping files.

The XML mapping files must adhere to the UIMA framework for text analysis.

When you add XML mapping files to a collection that uses a custom analysis engine, you enable XML elements in source documents to be mapped to annotations in the common analysis structure. These annotations can then be used by your custom analysis engine. You can then map the analysis results (in the




common analysis structure) to the index, and enable users to query specific annotations when the search the collection by using semantic search

For example, you can map the content of addressee and customer elements to Person annotations in the common analysis structure. These annotations can then be accessed by your custom annotators, which might detect additional information (for example, they might detect the gender of the Person). You can also map Person annotations to the enterprise search index, which allows users to search for Persons without having to know the original XML elements.

If you want to allow users to specify the original XML elements in queries, then you do not need to define any XML mappings. Instead, you can configure parsing options and enable native XML search for the collection.

### Procedure

To map XML elements to a common analysis structure:

1. Click **Collections** to open the Collections view.
2. In the list of collections, locate the collection that you want to configure and click  **Edit**.
3. On the Parse page, click **Configure text analysis options**.
4. In the **Map XML elements to common analysis structures** area, click **Add Mapping**.
5. On the Map XML Elements to a Common Analysis Structure page, type a descriptive name for the XML mapping file.
6. Either type the fully qualified path for the XML mapping file that you want to add, or click **Browse** to select the file from your system.
7. Click **OK**. Your XML mapping file is listed on the Text Analysis Options page.
8. To view the XML source for the XML mapping file, click  **View XML source**. By viewing the XML source, you can see the XML structure and learn more about how XML elements are being mapped for retrievability.
9. To delete the XML mapping file, click  **Delete**.

#### Related concepts

"Workflow for custom analysis integration" in "Programming Guide and API Reference for Enterprise Search"

"Custom text analysis integration" in "Programming Guide and API Reference for Enterprise Search"

"Unstructured information management architecture (UIMA)" in "Programming Guide and API Reference for Enterprise Search"

"Approaches for mapping XML document structures to a common analysis structure" in "Programming Guide and API Reference for Enterprise Search"

"XML mapping configuration file" in "Programming Guide and API Reference for Enterprise Search"

"XML mapping sample and the output results" in "Programming Guide and API Reference for Enterprise Search"

## Indexing the analysis results

If custom analysis engines are associated with the enterprise search system, and if an analysis engine is associated with a collection, you can specify which analysis results you want to use in semantic search queries.

## Before you begin

To map a common analysis structure to the index, you must be a member of the enterprise search administrator role or be a collection administrator for the collection.

## About this task



By mapping the results of custom text analysis (in a common analysis structure) to the enterprise search index, you enable users to specify semantically precise queries and improve the quality of the search results. When you index the analysis results, you enable users to query annotations that were added by your custom analysis engine or annotations that were created when XML elements were mapped to the common analysis structure.

For example, depending on the entities and relationships detected by the annotators, users can search for concepts that occur in the same sentence (such as a specific person and any competitor name), or a keyword and a concept (such as the term Alex and a phone number).

You can index a common analysis structure only if you associated a custom analysis engine with the collection. If you use the default analysis engine, you must use the default common analysis structure. Enterprise search uses a default mapping from the common analysis structure to the enterprise search index.

## Procedure

To index the results of analysis:

1. Click **Collections** to open the Collections view.
2. In the list of collections, locate the collection that you want to configure and click  **Edit**.
3. On the Parse page, click **Configure text analysis options**.
4. In the **Map one common analysis structure to searchable content in the index** area, click **Select a different set of analysis results**.
5. On the Select a Common Analysis Structure for This Collection page, select the mapping that you want to use with the enterprise search index:
  - Select **Default** if you want to use the default system rules for mapping analysis results to the enterprise search index. If you did not associate an analysis engine with this collection, this option is the only one available.
  - If you associated an analysis engine with this collection, you can select **Custom** to specify which analysis results you want to use. Either type the fully qualified path for the mapping that you want to use, or click **Browse** to select the file from your system.
6. Click **OK**. The mapping file that you selected is listed in the **Analysis results** field on the Text Analysis Options page.
7. To view the XML source for the mapping file, click  **View XML source**. By viewing the XML source, you can see which analysis results users can access when they search the collection. You can also see how users access and use the analysis results, such as by using semantic search or by using a field that is returned in the search results.

## Related concepts

"Workflow for custom analysis integration" in "Programming Guide and API Reference for Enterprise Search"

"Custom text analysis integration" in "Programming Guide and API Reference for Enterprise Search"

"Unstructured information management architecture (UIMA)" in "Programming Guide and API Reference for Enterprise Search"

"Approaches for indexing custom analysis results" in "Programming Guide and API Reference for Enterprise Search"

"Definition of a feature path" in "Programming Guide and API Reference for Enterprise Search"

"Writing the index build configuration file" in "Programming Guide and API Reference for Enterprise Search"

---

## Enabling support for native XML search

If a collection includes XML documents, you can enable users to use the XML markup when searching for documents by enabling native XML search for the collection.

### Before you begin

To enable support for searching XML documents with native XML search, you must be a member of the enterprise search administrator role or be a collection administrator for the collection.


### About this task

A native XML search, such as XPath or XML fragments, can provide more precise search results by exploiting the XML markup of the documents within the query. For example, users can specify that a query term must occur within a certain XML element or attribute.

For example, invoices from a computer retailer that are in XML format might contain <order> entries that include <company> and <computertype> elements. To retrieve invoices that contain orders for IBM notebooks, a keyword search for IBM and notebook might retrieve documents that include Dell notebook computers and IBM desktop models. By using XML search, you can specify that IBM must appear within the <company> element, that notebook must appear in the <computertype> element, and that both elements must be under the same <order> element. This way, you retrieve invoices that specifically discuss IBM notebooks.

### Procedure

To enable users to search a collection with native XML search:

1. Click **Collections** to open the Collections view.
2. In the list of collections, locate the collection that you want to configure and click  **Edit**.
3. On the Parse page, click **Configure parsing options**.
4. Select **Enable users to search XML documents with native XML search**.
5. Click **OK**.

---

## Linguistic analysis of Chinese, Japanese, and Korean documents

To enhance the retrievability of documents written in the Chinese, Japanese, and Korean languages, you can specify linguistic analysis options.

For Chinese, Japanese, and Korean documents, you can specify that the parser is to use the n-gram segmentation method for lexical analysis. For Chinese and Japanese documents, you can also configure the parser to remove new line characters from white space.

## N-gram segmentation

When you create a collection, you select the type of lexical analysis that you want to use for parsing Chinese, Japanese, and Korean documents.

Unicode-based white space segmentation uses blank space as the delimiter between words. N-gram segmentation considers overlapping sequences of any number of characters as a single word. For languages like Chinese, Japanese, and Korean, which do not use blanks as word delimiters, n-gram segmentation can return better search results than Unicode-based white space segmentation.

You choose the segmentation method that you want to use for parsing documents when you create a collection. After the collection is created, you can view the setting by viewing parsing options, but you cannot change it.

## Removing new line characters from white space

In languages where white space is not used to delimit word boundaries, such as Chinese and Japanese, you can configure the parser to remove certain white space characters that cause line breaks.

### Before you begin

To complete this task, you must log in as an enterprise search administrator.

### About this task

If this option is enabled for a collection, and a document is written in Chinese or Japanese, then the parser will remove any sequence of new line or tab characters that separate two-letter characters from the non-ASCII character range. The following characters are removed:

- Tab (0x09 )
- LF or line feed (0x0A)
- CR or carriage return (0x0D).

For the change to become effective, stop and restart the parser. To apply the change to documents that are already parsed and stored in the index, crawl and parse the documents again, and then reorganize the index.

### Procedure

To remove new line characters from white space:

1. On the index server, log in as the enterprise search administrator. This user ID was specified during the installation of DB2 II OmniFind Edition.
2. Use a text editor to edit the following file, where *collection\_ID* is the ID that was specified for the collection (or that was assigned by the system) when the collection was created:  
`collection_ID.parserdriver/collection.properties`
3. Change the value of the `removeCjNewlineChars` property from `false` to `true`.



---

## Document types associated with collection parsers and Stellent sessions

To ensure that documents in a crawl space are accurately and efficiently parsed, you can create configuration files to specify which types of documents are to be parsed by the collection parser and which are to be parsed by a Stellent document filtering session.

In an enterprise search collection, most document formats are processed by built-in HTML or XML parsers. Certain types of documents are typically not parsed (such as Postscript documents), and other types of documents are handled by Stellent parsing functions (such as Microsoft Word, Microsoft Excel, Microsoft PowerPoint, Lotus Freelance, Lotus 123, PDF, RT, and Ichitaro document types).

Because metadata can be misleading, plain text and HTML documents might be sent to the Stellent session in error, and then sent back to one of the built-in parsers, a situation that can impact performance. To avoid this situation, you can create configuration files to control where and how different types of documents are parsed.

Associating document types with the collection parser and Stellent session involves the following tasks:

1. Configuring document types for the collection parser. This step involves creating a configuration file that maps document types to the parser that is used by a collection. You can create one of these configuration files per collection.
2. Configuring document types for the Stellent session. This step involves creating a configuration file that maps document types to the Stellent document filters that are used by a collection. You can create one of these configuration files per collection.
3. Stopping and restarting the parser. For the changes to become effective, use the enterprise search administration console to monitor the collection for which you configured document types, then stop and restart the parser.

### Associating document types with a collection parser

To associate particular types of documents with a collection parser, you create a configuration file. There is no support for this task in the enterprise search administration console.

#### Before you begin

To complete this task, you must log in as an enterprise search administrator.

#### About this task

If the configuration file does not exist, the collection parser uses the default parser service rules. If the configuration file exists, rules in the file specify:

- Which documents you want to send to the Stellent session, depending on the file extension or the content type.
- How to parse documents whose type is unknown because of incomplete metadata.

The format of the file is a sequence of lines, where each line is one of the following rules:



| **EXTENSION** *extension parser*

| All documents whose URL ends on the specified extension will be  
| processed by the specified parser. Do not include the period in the  
| extension; comparison is not case sensitive.

| **CONTENTTYPE** *type/subtype parser*

| All documents whose content type matches the specified type/subtype will  
| be processed by the specified parser. Given the content type t/s of a  
| document, a match occurs if t equals type, and either s equals subtype or  
| the subtype is a wildcard (the asterisk, \*).

| **UNKNOWN** *parser*

| All documents whose extension and content type are not known (that is,  
| not made available by the crawler), will be processed by the specified  
| parser.

| **DEFAULT** *parser*

| All documents that are not covered by any of the other rules will be  
| processed by the specified parser.

| In all cases, *parser* must specify `html`, `xml`, `stellent`, or `none`, where `none` means  
| that documents of that type are not to be parsed.

| If more than one rule matches a document, then the more specific rule prevails,  
| disregarding the order in which the rules appear:

- An **EXTENSION** rule is more specific than a **CONTENTTYPE** rule.
- A **CONTENTTYPE** rule that includes a subtype is more specific than one with a wildcard. For example, a rule for content type `application/postscript` has priority over a rule for `application/*`.
- There should not be two rules for the same extension or content type. In that case, it is up to the implementation which of the rules is given priority.

| **Procedure**

| To associate document types with the collection parser:

1. On the index server, log in as the enterprise search administrator. This user ID was specified when DB2 II OmniFind Edition was installed.
2. Create the configuration file as follows, where *collection\_ID* identifies the collection that you want to configure:  
ES\_NODE\_ROOT/master\_config/*collection\_ID*.parserdriver/parserTypes.cfg
3. Use a text editor to edit the file and specify parser service rules, then save and exit the file.
4. For the changes to become effective, use the enterprise search administration console to monitor the parser for the collection, and stop and restart the parser.

| **Example**

| In this example, the built-in HTML parser processes all documents with the  
| extension `txt`, `htm` or `html`, with a content type that begins with `text/`, or with an  
| unknown extension and content type. The built-in XML parser processes all  
| documents with extension `xml` or with content type `text/xml`. All other documents,  
| including those with a content type that starts with `application/`, are sent to the  
| Stellent session.

| EXTENSION doc stellent  
| EXTENSION txt html  
| EXTENSION htm html

```

EXTENSION html html
EXTENSION xml xml
EXTENSION ps none
CONTENTTYPE text/xml xml
CONTENTTYPE text/* html
CONTENTTYPE application/* stellent
UNKNOWN html
DEFAULT stellent

```

## Default collection parser service rules

If you do not create a configuration file to map file types and content types to the parser for a collection, default rules are used to parse documents.

The default rules used by the collection parser are as follows:

```

EXTENSION pdf stellent
EXTENSION ppt stellent
EXTENSION prz stellent
EXTENSION lwp stellent
EXTENSION doc stellent
EXTENSION rtf stellent
EXTENSION xls stellent
EXTENSION 123 stellent
EXTENSION vsd stellent
EXTENSION vdx stellent
EXTENSION jxw stellent
EXTENSION jsw stellent
EXTENSION jtw stellent
EXTENSION jaw stellent
EXTENSION juw stellent
EXTENSION jbw stellent
EXTENSION jvw stellent
EXTENSION jfw stellent
EXTENSION jtt stellent
EXTENSION jtd stellent
EXTENSION jttc stellent
EXTENSION jtdc stellent
EXTENSION jtdx stellent
EXTENSION ps none
EXTENSION xml xml
EXTENSION txt text
EXTENSION htm html
EXTENSION html html
EXTENSION shtml html
EXTENSION xhtml html
EXTENSION asp html

CONTENTTYPE application/postscript none
CONTENTTYPE application/* stellent
CONTENTTYPE text/rtf stellent
CONTENTTYPE text/richtext stellent
CONTENTTYPE text/xml xml
CONTENTTYPE text/html html
CONTENTTYPE text/plain text

UNKNOWN html
DEFAULT html

```

## Associating document types with a Stellent session

To specify which types of documents are to be parsed by Stellent document filters, you create a configuration file. There is no support for this task in the enterprise search administration console.

### Before you begin

To complete this task, you must log in as an enterprise search administrator.

### About this task

The configuration file specifies:

- Which file types are to be parsed by the Stellent session. A file type corresponds to one of the file types recognized by the Stellent library.
- Which file types are to be sent back to the collection parser for processing with one of the built-in parsers. (This action is needed because the collection parser might send a document to the Stellent session in error, due to misleading metadata.)
- Which file types are to be rejected because they are not supported in enterprise search.

If the configuration file was specified but does not exist, the parser will fail to start. If no configuration file was specified for the `OutsideInSupportedTypes` property in the `stellent.properties` file, then the default parsing rules for Stellent sessions will be used.

The configuration file lists document types and how they are to be handled. The format of the file is a sequence of lines, where each line is one of the following rules:

```
accept DEFAULT
accept ALL doctype
accept type doctype
native DEFAULT
native type doctype
reject type
```

Where:

*doctype*

Is the value to be used for the doctype query token. Documents can be searched by document type. For example, a user might specify `$doctype::pdf` to search PDF documents.

*type* Is one of the `FI_` values in the Stellent library, and *doctype* is the value to be used for the doctype token if a rule is applied.

#### **DEFAULT**

Means that the list of accepted or native types, depending on the type of the rule, includes the default list. This option enables you to extend the default configuration instead of replacing it.

**A11** Means that all types that are not explicitly listed are accepted with the specified doctype token.

The rules in the configuration file are processed as follows:

- If there is an accept rule for *type*, including the default list if accept DEFAULT was specified, it is accepted.
- Else, if there is a reject rule for *type*, it is not accepted.
- Else, if accept ALL was specified, it is accepted.
- Otherwise, it is not accepted.

If the document type is accepted, then the *doctype* value that was specified in the rule that was applied is used. This value is sent back to the collection parser along with the parsed content. If the document type is not accepted, the following behavior occurs:

- If there is a native rule for *type* (including the default parsing rules if native DEFAULT was specified), the document is sent back to the built-in parser along with the value for the *doctype* token that is specified by this rule. The value of *doctype* must be either *txt*, *htm* or *xml*, indicating plain text, HTML or XML, respectively.
- Otherwise, the document is rejected and will not be parsed.

## Procedure

To associate document types with the Stellent session:

1. On the index server, log in as the enterprise search administrator. This user ID was specified when DB2 II OmniFind Edition was installed.
2. Edit the  
ES\_NODE\_ROOT/master\_config/collection\_ID.stellent/stellent.properties file, where *collection\_ID* identifies the collection that you want to configure.
3. For the `OutsideInSupportedTypes` property, specify the absolute path of the configuration file that you are creating.

For example, you might create the following configuration file for a single collection, and store it with other collection-specific files:

```
ES_NODE_ROOT/master_config/collection_ID.stellent/stellenttypes.cfg
```

As another example, you might create the following configuration file so that you can use the same settings with all collections, and store this file with other system-level files. (If you use this approach, be sure to specify this path in the `stellent.properties` file for each collection, as specified in step 2.)

```
ES_INSTALL_ROOT/default_config/stellent/stellenttypes.cfg
```

4. Use a text editor to create the configuration file and specify Stellent parsing rules, then save and exit the file.
5. For the changes to become effective, use the enterprise search administration console to monitor the parser for the collection, and stop and restart the parser.

## Example

In the following configuration file, the Stellent session accepts, in addition to the default list, the Microsoft Visio format.

```
accept DEFAULT
accept FI_VISI03 visio
accept FI_VISI04 visio
accept FI_VISI05 visio
accept FI_VISI06 visio
```

In the following configuration file, Postscript documents will be supported and searchable with a document type of `ps`; the X pixmap format (XPM) will be sent back to the built-in text parser; the PNG image format will be rejected; and all other file types will be accepted and made searchable with a document type of `other`.

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

## Default parsing rules for Stellent sessions

If you do not create a configuration file to map file types to Stellent session document filters, default rules are used to parse documents.

The default rules used by the Stellent session are as follows:

```
ACCEPT FI_WORD4          doc
ACCEPT FI_WORD5          doc
ACCEPT FI_RTF             rtf
ACCEPT FI_WINWORD1       doc
ACCEPT FI_WINWORD1COMPLEX doc
ACCEPT FI_WINWORD2       doc
ACCEPT FI_WORD6          doc
ACCEPT FI_WINWORD6       doc
ACCEPT FI_ICHITAR03      jxw
ACCEPT FI_ICHITAR04      jsw
ACCEPT FI_WINWORD1J      doc
ACCEPT FI_WINWORD5J      doc
ACCEPT FI_RTFJ           rtf
ACCEPT FI_WINWORD7       doc
ACCEPT FI_WORDPRO        lwp
ACCEPT FI_WINWORD97      doc
ACCEPT FI_ICHITAR08      jtd
ACCEPT FI_WORDPRO97      lwp
ACCEPT FI_WINWORD2000    doc
ACCEPT FI_WINWORD2002    doc
ACCEPT FI_WINWORD2003    doc
ACCEPT FI_123R1          123
ACCEPT FI_123R2          123
ACCEPT FI_123R3          123
ACCEPT FI_EXCEL          xls
ACCEPT FI_EXCEL3         xls
ACCEPT FI_EXCEL4         xls
ACCEPT FI_123R4          123
ACCEPT FI_EXCEL5         xls
ACCEPT FI_123R6          123
ACCEPT FI_EXCEL97        xls
ACCEPT FI_123R9          123
ACCEPT FI_EXCEL2000      xls
ACCEPT FI_EXCEL2002      xls
ACCEPT FI_EXCEL2003      xls
ACCEPT FI_FREELANCE      prz
ACCEPT FI_POWERPOINT4    ppt
ACCEPT FI_POWERPOINT3    ppt
ACCEPT FI_POWERPOINT7    ppt
ACCEPT FI_FREELANCE3     prz
ACCEPT FI_POWERPOINTMAC3 ppt
ACCEPT FI_POWERPOINTMAC4 ppt
ACCEPT FI_PDF            pdf
ACCEPT FI_EXTPOWERPOINT4 ppt
ACCEPT FI_EXTPOWERPOINTMAC4 ppt
ACCEPT FI_POWERPOINTMACB3 ppt
ACCEPT FI_POWERPOINTMACB4 ppt
ACCEPT FI_POWERPOINT97   ppt
ACCEPT FI_PDFMACBIN      pdf
ACCEPT FI_POWERPOINT9597 ppt
ACCEPT FI_POWERPOINT2000 ppt
ACCEPT FI_POWERPOINT2    ppt

NATIVE FI_HTML           htm
NATIVE FI_HTML_LATIN2    htm
NATIVE FI_HTML_JAPANESESJIS htm
NATIVE FI_HTML_JAPANESEEUCL htm
NATIVE FI_HTML_CHINESEBIG5 htm
NATIVE FI_HTML_CHINESEEUCL htm
NATIVE FI_HTML_CHINESEGB htm
```

NATIVE FI_HTML_KOREANHANGUL	htm
NATIVE FI_HTML_CYRILLIC1251	htm
NATIVE FI_HTML_CYRILLICKO18	htm
NATIVE FI_CYRILLIC1251	txt
NATIVE FI_CYRILLICKO18	txt
NATIVE FI_W2KHTML	htm
NATIVE FI_XL2KHTML	htm
NATIVE FI_PP2KHTML	htm
NATIVE FI_XML	xml
NATIVE FI_WML	xml
NATIVE FI_HTML_JAPANESEJIS	htm
NATIVE FI_WML_CHINESEBIG5	xml
NATIVE FI_WML_CHINESEEEUC	xml
NATIVE FI_WML_CHINESEGB	xml
NATIVE FI_WML_CYRILLIC1251	xml
NATIVE FI_WML_CYRILLICKO18	xml
NATIVE FI_WML_JAPANESEJIS	xml
NATIVE FI_WML_JAPANESEJIS	xml
NATIVE FI_WML_JAPANESEEEUC	xml
NATIVE FI_WML_JAPANESEEEUC	xml
NATIVE FI_WML_KOREANHANGUL	xml
NATIVE FI_WML_LATIN2	xml
NATIVE FI_HTMLUNICODE	htm
NATIVE FI_XML_DOCTYPE_HTML	htm
NATIVE FI_XHTML	htm
NATIVE FI_ASCII	txt
NATIVE FI_ANSI	txt
NATIVE FI_UNICODE	txt
NATIVE FI_ASCII8	txt
NATIVE FI_ANSI8	txt
NATIVE FI_MAC	txt
NATIVE FI_MAC8	txt
NATIVE FI_SHIFTJIS	txt
NATIVE FI_CHINESEGB	txt
NATIVE FI_HANGEUL	txt
NATIVE FI_CHINESEBIG5	txt
NATIVE FI_LATIN2	txt
NATIVE FI_JAPANESE_EUC	txt
NATIVE FI_HEBREW_OLDCODE	txt
NATIVE FI_HEBREW_PC8	txt
NATIVE FI_HEBREW_E0	txt
NATIVE FI_HEBREW_WINDOWS	txt
NATIVE FI_ARABIC_710	txt
NATIVE FI_ARABIC_720	txt
NATIVE FI_ARABIC_WINDOWS	txt
NATIVE FI_7BITTEXT	txt
NATIVE FI_JAPANESE_JIS	txt
NATIVE FI_CENTRALEU_1250	txt
NATIVE FI_UTF8	txt
NATIVE FI_EBCDIC_37	txt
NATIVE FI_EBCDIC_273	txt
NATIVE FI_EBCDIC_277	txt
NATIVE FI_EBCDIC_278	txt
NATIVE FI_EBCDIC_280	txt
NATIVE FI_EBCDIC_284	txt
NATIVE FI_EBCDIC_285	txt
NATIVE FI_EBCDIC_297	txt
NATIVE FI_EBCDIC_500	txt
NATIVE FI_EBCDIC_870	txt
NATIVE FI_EBCDIC_871	txt
NATIVE FI_EBCDIC_1026	txt

---

## 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 (file system space, processor speed, and memory)
- How many documents need to be crawled and recrawled
- The type of data that to crawl
- How often you change category rules (the changes 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 several 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

To specify options that influence the user's view of the index, you can also do the following tasks:

- Configure support for wildcard characters in query terms
- Configure scopes to limit the range of documents that users can search
- Collapse documents from the same source in the search results
- Remove URIs from the index

### **Related concepts**

“Enterprise search indexes” on page 5

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

#### **Related tasks**

“Monitoring index activity for a collection” on page 182

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 183

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.

### **Before you begin**

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. On the Index page, click **Schedule index builds**.
4. To specify how often the index is to be refreshed with new content, specify the following options on the Schedule Index Builds page in the **Specify a schedule to refresh the index** area:
  - a. In the **Start on** area, in the **Year, Month, Day, Hour, and Minute** 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).

5. 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 **Year, Month, Day, Hour, and Minute** 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).

6. Click **OK**.

#### **Related concepts**

“Enterprise search indexes” on page 5

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

## **Changing the index schedule**


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

### **Before you begin**

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.

### **Before you begin**






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.

### Before you begin

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. Click  **Edit** to change to the system editing view.
3. On the Index page, click **Configure indexing options**.
4. On the System-Level Indexing Options page, type the number of indexes that can share system resources and be built at the same.
5. Click **OK**.

#### Related concepts

“Enterprise search indexes” on page 5

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

---

## Options that influence the searchable view of the index

After documents are indexed, you can specify options that control how users can search for documents and view documents in the search results.

To specify options that influence the user’s view of the index, you can do the following tasks:

- Configure support for trailing wildcard characters in query terms. You can build support for wildcard queries into the index, or you can specify options to expand the query terms during query processing.
- Configure scopes to limit the range of documents that users can search. When users search the collection, they search only the documents that belong to the scope, not the entire index.
- Collapse documents from the same source in the search results. You can group documents that match a URI or URI pattern in the index, and show only the top result documents in the search results (users can specify options to see the collapsed result documents).
- Remove URIs from the index. You might need to temporarily prevent users from searching particular documents in the index.

## Wildcard characters in queries

You can enable users to include a wildcard character in query terms and search for words that start with a specified pattern.

A wildcard query term is a term that contains a trailing asterisk (\*). When a user submits a query that includes a wildcard character, the search results include all documents in the index that match the query term plus all documents in the index that match the pattern represented by the wildcard character. Pattern matching is done only for trailing characters in the query term (for example, sea\* can match search, season, and seals).

When you configure wildcard settings for an index, you can choose whether you want to enable users to specify wildcard characters in queries and, if so, how this support is to be provided. Wildcard query terms are expanded to all the terms in the index that match. You can either build term expansion into the index, or you expand the terms when queries are submitted.

Any changes that you make to the wildcard settings become effective after the next index reorganization.

### Index expansion

To build support for wildcard characters into the index, you specify how many leading characters in a word must match the pattern characters in a query term in order for a match to occur. For example, if you specify 4, then the first four characters in a word must match the query term that uses a wildcard character.

To illustrate this example, the word technical matches the query term tech\* but does not match the query term te\*.

When the index is refreshed or reorganized, all possible expansions for each term in a document are indexed along with the original terms. An advantage of this approach is that no additional time is required to expand the terms during query processing. However, this approach increases the size of the index, which means you must have sufficient system resources available to accommodate the larger index.

This approach is most useful if the size of the collection is relatively small, or where the space and time to build the index are less important than query response time. For example, you might choose this approach to search a catalog or an employee directory.

### Query expansion

To expand queries and apply pattern matching rules when users submit queries that contain wildcard characters, you specify how many variations of a query term constitute a match. For example, if you specify 50, then up to 50 variations of a query term can qualify as matches of the query term.

To illustrate this example, the query term tech\* matches the words technical, technique, technology, and up to 50 different words that begin with the characters tech.

Although query expansion has only a minor impact on the size of the index, it can degrade query performance. The search processes must iterate over all possible

expansions of the wildcard query term, up to the limit that you specify in the wildcard settings. If there are more variations for a term than the allowed number of expansions, then alphabetical order is used to determine the matching documents. For example, `technical` qualifies as a match of `tech*` before `technique` or `technology`.

This approach is most useful if the size of the collection is relatively large and the space and time to build the index must be minimized. For example, you might choose this approach for e-mail archives, where the index must keep up with the rapidly changing documents, but query response time is less important.

#### Related concepts

“Enterprise search indexes” on page 5

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

## Configuring support for wildcard characters in queries

When you configure indexing options for an enterprise search collection, you can specify whether you want to enable users to include wildcard characters in query terms.

### Before you begin


To configure support for wildcard characters in queries, you must be a member of the enterprise search administrator role or be a collection administrator for the collection that the index belongs to.

### About this task

When you specify wildcard settings, your changes become effective after the next index reorganization.

### Procedure

To configure support for wildcard characters in queries:

1. Click **Collections** to open the Collections view.
2. In the list of collections, locate the collection that you want to configure and click  **Edit**.
3. On the Index page, click **Configure indexing options**.
4. On the Indexing Options page, select the wildcard setting that you want to use with this collection:
  - Select **No wildcard support** if you do not want to enable users to specify wildcard characters in query terms when they search this collection.
  - Select **Build support for wildcard characters into the index** if you prefer to enable support for wildcard characters by increasing the size of the index. If you select this option, also specify how many leading characters in a word must match the pattern characters in a query term in order for a match to occur.
  - Select **Expand query terms when users submit queries** if you prefer to expand the query terms during query processing. This option can impact query performance. If you select this option, also specify how many variations of the query term can be returned in the search results.
5. Click **OK**.

## Scopes

Configure a scope when you want to present users with a limited view of a collection.

A scope is a group of related URIs 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, not the entire index. To use this capability, your search applications must include support for searching scopes.

When you create a scope, you specify a range of URIs in the index that users are able to search. Limiting the documents that users can search helps ensure that documents in the search results are specific to the information that 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 configure scopes, you might need to reorganize the index twice before the changes become effective. If you configure scopes before the first index reorganization for the collection, users will be able to search the collection, but they will not be able to see the scope data in the search results. Reorganize the index again to ensure that search results reflect the range of URIs in the scope.

If you configure scopes after the index has been reorganized at least once, the changes become effective after the next index reorganization.

### **Related concepts**

“Enterprise search indexes” on page 5

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

## 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.

### **Before you begin**

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.

When you configure scopes, you might need to reorganize the index twice before the changes become effective. If you configure scopes before the first index reorganization for the collection, users will be able to search the collection, but they will not be able to see the scope data in the search results. Reorganize the index again to ensure that search results reflect the range of URIs in the scope.

If you configure scopes after the index has been reorganized at least once, the changes become effective after the next index reorganization.

### 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 Index 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

“URI formats in an enterprise search index” on page 71

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

## Collapsed URIs

Enterprise search can organize the search results so that result documents that have the same URI prefix in the index are grouped together and collapsed in the search results.

When results are collapsed, the top result typically appears flush left. One or more lower ranking results are grouped and indented below the top result.

To collapse result documents that have different URI prefixes as a single group, you can associate the URI prefixes with a group name that you create. For example, if you have three servers for managing financial data, you can group documents from all three servers in the search results and collapse the lower ranking results below the top result documents.

Search applications can use the URI prefix or the group name to collapse documents in the search results. In the sample search application for enterprise search, the top two search result documents are displayed. If more than two result documents with the same URI prefix (or that belong to the same URI group) are returned, you can select an option to see the collapsed results.

### How to organize URI prefixes and group names

When you use the administration console to configure rules for collapsing search results, you specify the URI prefixes of the documents that you want to collapse and optionally associate the URI prefixes with a group name.



The order of the URI prefixes that you configure is important. The Index component uses the order of the URI prefixes when it computes the value of each URI in a collection. For each URI:

1. The Index component scans through the URI prefixes in the rules for collapsing search results sequentially.
2. When the Index component finds the first URI prefix that matches a prefix of a document in the index, it associates the group name (or the URI prefix, if the rule does not specify a group name) as an extra search term for the document.
3. If a document cannot be matched to a URI prefix, then:
  - For Web URIs, the Index component uses the host name of the URL as the URI prefix.
  - For NNTP URIs, the Index component uses the first message ID in the value of the reference header as the URI prefix.

After you add URI prefix to the list of those that are to be collapsed in the search results, you must position the URI prefix in the order that you want the Index component to scan it and potentially associate it as an extra search term with documents in the index:

- When you add a URI prefix and do not associate it with a group name, you can select the individual URI prefix and move it up or down in the list.
- When you add a URI prefix and associate it with a group name, you move the entire group of URI prefixes that belong to the same group whenever you move a URI prefix up or down in the list. (The order of URI prefixes within a group does not matter; selecting an individual URI prefix automatically selects the entire group.)

## Query syntax to search by group name

The enterprise search query syntax supports the ability to specify a group name and search only the URI prefixes that are associated with that group name. To retrieve documents that belong to the same group, use the following syntax, where *group* is either the group name or the URI prefix that belongs to the group:

```
samegroupas:group
```

For example:

```
samegroupas:hr
```

```
samegroupas:http://mycompany.server3.com/news/
```

### Related concepts

“Enterprise search indexes” on page 5

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

## Collapsing URIs in the search results

You can specify options for grouping and collapsing result documents that have the same URI prefix in the search results. You can also create a group name that enables result documents with different URI prefixes to be collapsed together.

### Before you begin

To specify options for collapsing search results, you must be a member of the enterprise search administrator role or be a collection administrator for the collection.






## About this task

The changes that you make for collapsing search results do not take effect until the next time that the index is reorganized.

## Procedure

To specify options for collapsing search results:

1. Click **Collections** to open the Collections view.
2. In the list of collections, locate the collection that you want to configure and click  **Edit**.
3. On the Index page, click **Collapse search results**.
4. On the Collapse Search Results page, click **Add URI Prefix**.
5. On the Add a URI Prefix for Collapsing Results page, type the URI prefix for documents that you want to collapse in the search results. For example:  
db2://financial/ROI/  
cm://enterprise/finance/  
http://server1.com/finance/
6. Optionally type a descriptive group name that you want to associate with this URI prefix. To collapse result documents from several sources as a single group, type the same group name when you add each URI prefix.
7. Click **OK**.
8. On the Collapse Search Results page, position the new rule in the order that you want the Index component to scan it:
  - If you added a URI prefix and did not associate it with a group name, the new URI prefix appears at the bottom of the list. Use the arrow keys to move it to the correct position.
  - If you associated the new URI prefix with a group name, the new URI prefix appears at the bottom of the set of URI prefixes that belong to the same group. Use the arrow keys to move the entire group of URI prefixes to the correct position.
9. To change the URI prefix or group name, select the URI prefix and click  **Edit**.
10. To remove a URI prefix from the list, select the URI prefix and click  **Remove**.

## 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.

### Before you begin

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 Index 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 concepts

“Enterprise search indexes” on page 5

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

#### Related reference

“URI formats in an enterprise search index” on page 71

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

---

## Search server administration

Options that you can specify for the Search servers include using cache space for returning search results, controlling the maximum length of document summaries in the search results, associating a custom synonym dictionary so that users can find documents that contain synonyms of query terms, and returning predefined URIs in the search results whenever certain terms appear in the query.

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 URI, title, description, date, data type, and so on.

When you configure the Search servers for a collection, you specify options that influence how queries are processed, including options that can impact query performance:

### **Configuring a search cache**

To optimize query performance, you can specify that result documents are to be stored in a cache, and you can configure the amount of space to allocate for cached documents.

### **Configuring a maximum display length for document summaries**

Most result documents show a summary of the document content to help users decide whether the document is one that they want to retrieve. You can specify how much space is to be used in the search results to display this summary information.

### **Specifying a different default language**

A default language for searching documents in the collection is specified when the collection is created, but you can specify a different language as needed.

### **Associating a synonym dictionary**

If your application developers created a custom synonym dictionary, you can select one to use with a collection. When users search the collection, documents that contain synonyms of the query terms are included in the search results.

### **Configuring quick links**

You can predetermine URIs to be returned for certain keywords and phrases. When users specify the keywords or phrases in a query, 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.

### **Related concepts**

[“Search results from enterprise search” on page 133](#)

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.

---

## 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).

When you make changes to the search cache settings, you must restart the Search servers for the changes to become effective.

#### **Related concepts**

“Search servers for enterprise search” on page 7

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

---

## **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.

### **Before you begin**

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.

### **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, click **Configure search server options**.
4. On the Search Server Options page, select the **Use the search cache** check box.
5. In the **Maximum number of entries in the search cache** field, type the maximum number of search result sets that the search cache can hold.
6. Click **OK**.
7. For your changes to become effective, stop and restart the Search servers.

---

## **Custom synonym dictionaries**

To improve the quality of the search results, you can enable users to search for synonyms of their query terms when they search a collection.

If you create a synonym dictionary, add it to the enterprise search system, and associate it with a collection, users can search for documents that contain synonyms of their query terms when they search the collection. By expanding queries in this manner, users are more likely to find all documents of interest, not just documents that match their precise query terms. Because you define which

words are synonyms of each other when you create the synonym dictionary, you help ensure that users find relevant documents without having to specify all variations of the query term.

For example, your organization might use acronyms and abbreviations to refer to departments, equipment, and so on, or the documents in your collections might contain vocabulary that is specific to your industry. By creating a synonym dictionary, you can ensure that a query that includes an acronym (such as ACL) returns documents that discuss the expansion of that acronym (such as ACLs, access control lists, access controls, and so on).

The enterprise search query language support synonyms by allowing users to prepend a tilde operator to a query term. For example, the query ~WAS might return documents that discuss WebSphere Application Server. Application developers can also make synonym support available through query properties, which do not require special syntax.

Synonym dictionaries contain variants of words and have the following characteristics:

- The words are not specific to a language, but they can be used in different languages. There is only one synonym dictionary per collection.
- The words are not inflected. All possible inflections must be added to the synonym list. For example, an inflection can be the singular and the plural form of the word (such as ACL and ACLs).

Most of the terms that you add to a synonym dictionary are strict semantic equivalents, which means that if term A is a synonym of term B, then B is a synonym of A. Every time A is used in a query, B can be used, and vice versa.

However, you can also add terms that correspond to different uses of a term, including generic or more specific variants of the term. For example, you can have one synonym group that includes both building and house, and another group that includes bank, shore, and credit union.

The less strict that the relationship is between the terms, the larger the search result, although some of the search results might not be relevant to the query. The Search and Index API provides methods that allow users to select the appropriate synonyms when they submit a search request, and methods that show users which query terms were expanded to which synonyms.

When you want to use synonyms in a query term that contains white space characters, the term must appear between double quotation marks. If you enter an acronym such as WAS, the expanded form of the term will be found. However, if you enter WebSphere Application Server without enclosing the phrase between double quotation marks, the acronym WAS will not be found.

To create a synonym dictionary, an expert in the subject matter of the collection needs to create a synonym list in XML format (or work with an application developer to create the XML file). Then, using a tool that is provided with DB2 II OmniFind Edition, the XML file must be converted to a binary (a .dic file). An enterprise search administrator then uploads this binary to the system and assigns it a display name. Collection administrators can select a synonym dictionary to use for searching documents in a collection when they configure Search server options for a collection.

To associate synonym dictionaries with the system and collections, you use the enterprise search administration console.

**Requirement:** After you add a custom synonym dictionary to the system, you cannot edit it. To revise the synonyms that are available to a collection, you must:

1. Update the source XML file.
2. Convert the XML source to a new dictionary file.
3. Remove the synonym dictionary from the collections that use it.
4. Delete the synonym dictionary from the system.
5. Add the new version of the synonym dictionary to the system.

#### **Related concepts**

"Search servers for enterprise search" on page 7

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

"Text analysis included in enterprise search" in "Programming Guide and API Reference for Enterprise Search"

## **Synonym support in search applications**

Users can expand the search results by searching for documents that contain synonyms of the query terms.

The Search and Index API (SIAPI) interface for enterprise search supports several ways for users to search for synonyms of query terms:

- The SIAPI query syntax supports the tilde (~) operator for synonym expansion. If the user prepends this operator to a query term, synonym expansion is performed for the word. For example, the query ~WAS returns documents that discuss WebSphere Application Server and any other synonyms that exist for this abbreviation.
- Synonym expansion can be enabled using the SIAPI synonym expansion interface from within a search application. Query terms can be automatically expanded to include synonyms, or the search application might include options that enable the user to specify whether synonyms of the query terms are to be returned in the search results.

During automatic synonym expansion, synonym lookup is performed on all query words and content fields. The search results include documents that contain either the query terms or synonyms of the query terms. The search results also show which terms were expanded to which synonyms.

In a user-driven scenario, the search application shows the user which synonyms were found for each query word before the query is actually performed. The user then selects which terms to include in the search, or reformulates the search to remove original query terms. In this scenario, the user controls which terms to include in the query — either the strict equivalences, or varying word senses or usages.

#### **Related concepts**

"Text analysis included in enterprise search" in "Programming Guide and API Reference for Enterprise Search"

"Search and Index API (SIAPI)" in "Programming Guide and API Reference for Enterprise Search"

## Creating an XML file for synonyms

To expand queries in enterprise search to include synonyms of the query terms, you must specify which words qualify as synonyms of each other in an XML file.

### About this task

The XML file that lists the synonyms must comply with the schema shown in the following example.

```
<?xml version="1.0" encoding="UTF-8"?>
<synonymgroups xmlns="http://www.ibm.com/of/822/synonym/xml">
  <synonymgroup>
    <synonym>Think Pad</synonym>
    <synonym>Notebook</synonym>
    <synonym>Notebooks</synonym>
  </synonymgroup>
  <synonymgroup>
    <synonym>WebSphere Application Server</synonym>
    <synonym>WAS</synonym>
  </synonymgroup>
</synonymgroups>
```

### Restrictions

You must group words that are synonyms of each other (the <synonym> elements) in a <synonymgroup> element. A synonym can include white-space characters, but it cannot include punctuation characters, such as a comma (,) or vertical bar (|), because these characters might interfere with the enterprise search query syntax.

You must enumerate all possible inflections of the terms that you add as synonyms, such as the singular and plural forms of a word. You do not need to enumerate normalizations of the term, such as the removal of accents or umlauts (enterprise handles normalization automatically). For example, if you want to include the term météo as a synonym, you do not need to include the term METEO, too.

### Procedure

To create a list of synonyms for enterprise search:

1. Create an XML file. To avoid XML syntax errors, use an XML editor or XML authoring tool of your choice.
2. Add a <synonymgroup> element, then insert a <synonym> element for each word that is to be treated as a synonym of other words in the synonym group.  
Be sure to include your mappings in a <synonymgroups xmlns="http://www.ibm.com/of/822/synonym/xml"> element. The namespace (specified in the xmlns attribute) needs to be exactly as shown.
3. Repeat the preceding step until you have specified all of the synonyms that you want to use for searching documents in an enterprise search collection.
4. Save and exit the XML file.

After you create the XML file, you must convert it to a synonym dictionary so that it can be added to the enterprise search system.

### Related concepts

"Character normalization" in "Programming Guide and API Reference for Enterprise Search"



## Creating a synonym dictionary

After you create or update a list of synonyms in an XML file, you must convert the XML file to a synonym dictionary.

### About this task

To create a synonym dictionary, use the command line tool called `essyndictbuilder`, which is supplied with DB2 II OmniFind Edition. The tool is in the `ES_INSTALL_ROOT/bin` directory.

The input to the tool is the XML file that lists your synonyms, and the output from the tool is a synonym dictionary. The dictionary must have the suffix `.dic`. For example, `c:\mydictionaries\products.dic`.

The default location for both files is the directory where the script is invoked. If a dictionary with the same name exists, the script produces an error.

### Procedure

To create a synonym dictionary for enterprise search:

1. On the index server, log in as the enterprise search administrator. This user ID was specified when DB2 II OmniFind Edition was installed.
2. Enter the following command, where *XML\_file* is the fully qualified path to the XML file that contains the list of synonyms and *DIC\_file* is the fully qualified path to the synonym dictionary.

UNIX: `essyndictbuilder.sh XML_file DIC_file`

Windows: `essyndictbuilder.bat XML_file DIC_file`

After you create a synonym dictionary, use the enterprise search administration console to add the dictionary to the enterprise search system and associate it with one or more collections.

Only the generated `.dic` file is uploaded to the enterprise search system. Ensure that the source XML file is kept in an access-controlled environment, with the appropriate backup strategy in place. You need this XML file to update your synonym dictionary.

## Associating synonym dictionaries with the system


If you create custom synonym dictionaries for searching the documents in a collection, you must associate the dictionaries with the enterprise search system. You can later choose which synonym dictionary you want to use for searching a collection.

### Before you begin

To add your custom synonym dictionaries for use with enterprise search queries, you must be a member of the enterprise search administrator role.

### Procedure

To associate synonyms with the enterprise search system:

1. Click **System** to open the System view.
2. Click  **Edit** to change to the system editing view.



3. On the Search page, click **Add Synonym Dictionary**.
4. On the Add a Synonym Dictionary page, type a unique display name for the synonym dictionary and optionally type a description.
5. Either type the fully qualified path for the synonym dictionary (.dic file), or click **Browse** to locate the file on your system.
6. Click **OK**. Your custom synonym dictionary is added to the enterprise search system and becomes available for searching collections.

## Associating a synonym dictionary with a collection


If synonym dictionaries are associated with the enterprise search system, you can select one to use when searching a collection. If a query term matches a term in the dictionary, then result documents that contain synonyms of that term are also returned in the search results.

### Before you begin

To select a synonym dictionary for a collection, you must be a member of the enterprise search administrator role or be a collection administrator for the collection.

### Procedure

To associate a synonym dictionary with a collection:

1. Click **Collections** to open the Collections view.
2. In the list of collections, locate the collection that you want to configure and click  **Edit**.
3. On the Search page, click **Configure search server options**.
4. In the **Synonym dictionary name** field on the Search Server Options page, select the name of the synonym dictionary that you want to use when users query this collection.

The list of available synonym dictionaries includes all synonym dictionaries that were added to the enterprise search system.

5. Click **OK**.

---

## 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 documents that are predetermined to be relevant to the query terms.

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.

#### **Related concepts**

“Search servers for enterprise search” on page 7

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

## **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.

### **Before you begin**

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.

You do not need to restart the Search servers for your changes to become effective.

### **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**

“URI formats in an enterprise search index” on page 71

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



---

## 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 demonstrates many of the search and retrieval functions that are available for enterprise search. The sample application is also a working example that demonstrates how you can use the IBM Search and Index API (SI-API) to build interactive, custom search applications that reflect the goals of your enterprise.

Unless you change properties in the default configuration file, 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 the collections available to users.

You do not need to associate the sample application with specific collections (or all collections) in the enterprise search administration console before you can use it. In a production environment, enterprise search administrators control which search applications are allowed to search various collections.

### 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 applications, work seamlessly in both environments.

#### Related concepts

[“Search servers for enterprise search” on page 7](#)

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

[“Customizing enterprise search” on page 10](#)

The application programming interfaces for enterprise search enable you to create custom search applications, custom applications to update the content of collections, custom programs for text analysis, and custom synonym dictionaries.

[“Sample search application for enterprise search” on page 11](#)

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

[“Query syntax” in “Programming Guide and API Reference for Enterprise Search”](#)

[“Query techniques” on page 133](#)

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

[“Search results from enterprise search” on page 133](#)

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 146

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

“Document-level security” on page 147

If security is enabled for a collection when it is created, you can configure document-level security controls. 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 157

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

“Search and Index API (SI-API)” in “Programming Guide and API Reference for Enterprise Search”

---

## Associating search applications with collections

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

### Before you begin

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 **Configure search applications**.
3. On the Associate Search Applications With Collections page, click **Add Search Application**.
4. Type the name of the search application.
5. 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.

6. Click **OK**.

---

## Sample search application functions

The sample search application for enterprise search demonstrates most of the search functions that you can build into your custom search applications. You can use this application to search one, several, or all collections at a time.

### Collection-level functions

With these functions you can:

- Select the collections that you want to search (in a production environment, 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.
- Control the size of the query area. For example, you might want to maximize the query area to type a complex query, then review the query terms before you submit the query.
- 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.
- Search specific data sources instead of all sources in the collection.
- Specify which language your query terms are in.
- Specify that you want to see a specific field in the search results (typically, an administrator determines which fields are displayed in the search results, but you can request additional fields).
- Be prompted with suggestions for spelling corrections if possibly misspelled words are detected in the query string.
- Expand the query to include synonyms of your query terms. If a synonym dictionary is associated with the collection, documents that contain synonyms of your query terms are returned in the search results.
- 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 named category.

## Search result functions

With these functions you can:

- See the search results that match your query. You can clear the search results from the display at any time.
- Control how many result documents appear on each page, and browse forward and backward through the result set.
- Hide and display details about the result documents, and hide or display summaries of 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 source. For example, if one source returns 100 documents, the two most relevant documents are shown 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, by document date in ascending order, or by document date in descending order.
- 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 Web, NNTP, or Microsoft Exchange Server 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.

#### **Related concepts**

“Sample search application for enterprise search” on page 11

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

---

## **Editing sample 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 use a different search application as the default application.

#### **About this task**

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

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 Edition 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, then save and exit the file:

#### **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 as a default application and want to use a different application instead.

**Tip:** When the application name is `Default`, you can use the sample search application to search all collections with a single query.

#### **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.

**port** 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.

**timeout** Specify a the number of seconds to wait for a response from the Search server before a search request times out. This number must be an integer (such as 60, not 60.5 or sixty). If you do not specify a timeout value, the default value is 30 seconds.

**username** If you enabled global security in WebSphere Application Server, specify a user name. This user name enables enterprise search to enforce user authentication, and authenticate users with WebSphere Application Server, when search requests are received.

**password** Specify the password for the specified username.

3. Use the WebSphere Administrative Console to stop and restart the ESSearchApplication application:
  - a. On the Search server, start the WebSphere Application Server Administrative Console.

You can start this program by using the Windows **Start** menu, or you can open a Web browser and go to `http://hostname:port/admin`, where *hostname* is the host name of the Search server and *port* is the port number for the WebSphere Application Server Administrative Console. Typically, the Administrative Console port is 9090.
  - b. When you are prompted for a user ID and password, enter a user name and password that is registered in the WebSphere Application Server user registry. If you are using the local operating system as your user registry, you can specify the enterprise search administrator ID and password.
  - c. After you log in to the Administrative Console, click **Applications** and then click **Enterprise Applications**.
  - d. Select the check box for ESSearchApplication and click **Stop**.
  - e. After the application stops, select the check box for the application again, and click **Start**.

#### **Related concepts**

“Sample search application for enterprise search” on page 11

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

---

## **Accessing the sample search application**

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

### **Before you begin**

You must configure the sample search application 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 the collections 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.  
If your Web server is not configured to use port 80, you also need to specify the correct port number. For example:  
`http://SearchServer.com:9080/ESSearchApplication/`
2. If security is enabled in WebSphere Application Server, log in to the application with a valid user ID and password.  
If any of the collections that are available to the search application are enabled for security, and if the secure collections include Lotus Notes documents, you can configure a user profile. On the My Profile page, specify credentials for accessing the LDAP directory and various Lotus Notes data sources. You can then search those sources without being challenged to log in to them.
3. On the Search page, select the collections that you want to search and submit a query.

### Related concepts

“Sample search application for enterprise search” on page 11

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

---

## Enabling security for the sample search application

If you enable global security in WebSphere Application Server and want to use the sample search application to search secure collections, you must change configuration settings in the sample application and in WebSphere Application Server.

### Before you begin

- You must be a member of the enterprise search administrator role.
- You must enable global security in WebSphere Application Server. If you installed DB2 II OmniFind Edition as a multiple server configuration, then enable global security on the Search servers. See the WebSphere Application Server documentation for instructions on how to enable global security.
- If you choose Lightweight Directory Access Protocol (LDAP) for the user registry in WebSphere Application Server, then you must add the enterprise search administrator ID and password to your LDAP registry (this ID and password were specified when DB2 II OmniFind Edition was installed).

For example, if the enterprise search administrator ID is `adminUser`, then the user entry in the LDAP registry might be `uid=adminUser,ou=Employees,o=IBM,c=US`. See the LDAP server documentation for instructions.

If a collection includes Lotus Notes data sources, and you want to validate current user credentials during query processing, then you must use LDAP for the WebSphere Application Server user registry.

### Procedure

To enable security for the sample application for enterprise search:

1. Update search application properties in the administration console:
  - a. Log in to the enterprise search administration console as a user with enterprise search administrator privileges.
  - b. Click **Security** on the toolbar.
  - c. On the Search Applications page, click **Configure search applications**.
  - d. Click **Add Search Application** and, in the **Search application name** field, type the enterprise search administrator ID that was specified when DB2 II OmniFind Edition was installed.
  - e. Ensure that **All collections** is selected, and click **OK**.
2. Edit the config.properties file:
  - a. If you are using UNIX, open a console window. If you are using Microsoft Windows, open a command prompt window.
  - b. Change to the WEB-INF directory for the sample search application. The following examples are shown on two lines for readability; specify the command on a single line:  
**UNIX:**

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

  
**Windows:**

```
cd %ES_INSTALL_ROOT%\installedApps\ESSearchApplication.ear\  
ESSearchApplication.war\WEB-INF
```
  - c. Use a text editor to edit the config.properties file.
  - d. Change the username property to the name of a valid WebSphere Application Server user.
  - e. Change the password property to the password for the specified user.
  - f. Save and exit the file.
3. Restart the ESSearchApplication application in WebSphere Application Server:
  - a. On the Search server, start the WebSphere Application Server Administrative Console.  
  
You can start this program by using the Windows **Start** menu, or you can open a Web browser and go to `http://hostname:port/admin`, where *hostname* is the host name of the Search server and *port* is the port number for the WebSphere Application Server Administrative Console. Typically, the Administrative Console port is 9090.
  - b. When you are prompted for a user ID and password, enter the administrator ID and password that were specified when global security was enabled in WebSphere Application Server.
  - c. After you log in to the Administrative Console, click **Applications** and then click **Enterprise Applications**.
  - d. Select the check box for ESSearchApplication and click **Stop**.
  - e. After the application stops, select the check box for ESSearchApplication again, and click **Start**.

### Related concepts

| "Sample search application for enterprise search" on page 11  
| You can use the sample search application for enterprise search as a template  
| for developing custom search applications.

---

## Search results from enterprise search

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 support a rich query syntax and 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

“Search applications for enterprise search” on page 125

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 server administration” on page 115

Options that you can specify for the Search servers include using cache space for returning search results, controlling the maximum length of document summaries in the search results, associating a custom synonym dictionary so that users can find documents that contain synonyms of query terms, and returning predefined URIs in the search results whenever certain terms appear in 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**

“Search applications for enterprise search” on page 125

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.

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

---

**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.

### Related tasks

“Migrating a collection from WebSphere Portal” on page 160

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

---

## 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.

| When configuring Search server options for a collection, an administrator can  
| specify the maximum display length for document summaries in the search results.  
| Because the summary includes highlighting characters, the buffer returned to the  
| search application will be larger than the specified maximum value. The display  
| length, however, will not exceed the specified maximum value, although the  
| summary might be shorter (depending on the summary data extracted from the  
| source document).



---

## Customizing summaries in search results

Each result document for an enterprise search query includes a summary. You can customize the amount of information that each summary contains.

### About this task

You can customize search result descriptions by changing values for the following properties in the `ES_NODE_ROOT/master_config/collection_ID.runtime.node1/runtime-generic.properties` file:

#### **MinWordsPerSentence**

The minimum number of words that a description sentence can contain. The default value is 4.

#### **MaxWordsPerSentence**

The maximum number of words that a description sentence can contain. The default value is 20.

#### **NumberOfReturnedSentences**

The number of sentences that constitute a document's description. The default value is 5.

#### **MaxSentencesPerDocument**

The maximum number of sentences in a document that will be considered as candidates in the process of creating the description. The default value is 1000.

### Procedure

To customize document summaries in the search results:

1. On the Search servers, log in as the enterprise search administrator. This user ID was specified when DB2 II OmniFind Edition was installed.
2. Use a text editor to edit the following file, where *coll\_ID* is the ID that was specified for the collection (or that was assigned by the system) when the collection was created:  
`ES_NODE_ROOT/master_config/coll_ID.runtime.node1/runtime-generic.properties`

**Tip:** To determine the mapping between a collection name and its ID, see the `ES_NODE_ROOT/master_config/collections.ini` file.

3. Change the properties that you want to customize, then save and exit the file.
4. Stop and restart the Search servers to apply the changes.



---

## 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 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 use the search application to search collections.

### Collection-level security

When you create a collection, you can enable security at the collection level. You cannot change this setting after the collection is created. If you do not enable collection-level security, you cannot later specify document-level security controls.

When collection-level security is enabled, the global analysis processes apply the following 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 in the links of forbidden documents is excluded from the index. 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 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 can associate security tokens with each document that it crawls. The security tokens are stored with the documents in the index. For the Notes crawler, you can also specify that the user's credentials are to be validated with current access control data (as configured on the native data source) during query processing.

Your search applications can use the security tokens and user credentials 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.

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 DB2 Information Integrator OmniFind Edition (DB2 II OmniFind Edition), 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 WebSphere Application Server, this user ID is the only user ID that you can use to access the enterprise search administration console.

If you enable global security in 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 Edition 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 concepts**

“Enterprise search administration console” on page 8

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.

#### **Related tasks**

“Logging in to the administration console” on page 18

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

“Starting the enterprise search components” on page 165

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

---

## **Configuring administrative users**

By configuring administrative roles, you can restrict access to collections and control the functions that each administrative user can do.

### **Before you begin**

Before you assign a user to an administrative role, ensure that security is enabled in 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 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.

---

## Disabling security for an enterprise application in WebSphere Application Server

To control which DB2 II OmniFind Edition activities require user authentication, you can disable global security for individual enterprise applications in WebSphere Application Server.

### About this task

The DB2 II OmniFind Edition installation program deploys three enterprise applications to WebSphere Application Server:

- The ESAdmin application contains the interface for the enterprise search administration console.
- The ESSearchApplication application contains the interface for the sample search application.
- The ESSearchServer application provides all remote communication for the DB2 II OmniFind Edition SI-API implementation and enables the SI-API interfaces to communicate with the Search server.

By default, all three enterprise applications support WebSphere Application Server global security. When these applications detect that global security is enabled, they begin authenticating all requests that they receive.

Some organizations might want to enable or disable security for specific DB2 II OmniFind Edition enterprise applications. For example, you might want to authenticate all users who access the enterprise search administration console, but not authenticate users who use the DB2 II OmniFind Edition SI-API interfaces or the sample search application.

### Procedure

To disable security for a particular enterprise application:

1. On the Search server, start the WebSphere Application Server Administrative Console.

You can start this program by using the Windows **Start** menu, or you can open a Web browser and go to `http://hostname:port/admin`, where *hostname* is the host name of the Search server and *port* is the port number for the WebSphere Application Server Administrative Console. Typically, the Administrative Console port is 9090.

2. When you are prompted for a user ID and password, enter the administrator ID and password that were specified when global security was enabled in WebSphere Application Server.

3. After you log in to the Administrative Console, click **Applications** and then click **Enterprise Applications**.
4. Select the check box next to the name of the enterprise application for which you want to disable security.
5. Scroll down and click the **Map security roles to users/groups** link.
6. Locate the **AllAuthenticated** role and select the check box under the **Everyone?** column.
7. Click **OK**.
8. Click the **Save** link to save your changes.
9. If you are using WebSphere Network Deployment, select the **Synchronize changes with Nodes** check box.
10. Click **Save**.
11. Click **Applications** again, and then click **Enterprise Applications**.
12. Select the check box for application that you just modified and click **Stop**.
13. After the application stops, select the check box for application again, and click **Start**.

---

## 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, you can later configure document-level security controls. When collection security is enabled, the enterprise search global analysis processes also apply different rules for indexing duplicate documents and anchor text in 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.

## 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.



## 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. If you do not enable collection security, you can specify whether you want to index the anchor text in links to forbidden documents when you configure individual 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 associate 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.

If you enable collection security when you create a collection, anchor text processing is disabled. The anchor text is no longer indexed with a document unless it actually appears in the document or in the document metadata. This security control ensures that users are not exposed to information in documents that they are not allowed to access because the anchor text to forbidden documents is not associated with the documents that they do have access to.

Enabling collection security can enhance the security of Web documents by enabling 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, you can select an option to index the anchor text in links to forbidden documents when you configure individual Web crawlers. If you specify that the anchor text is to be indexed, the analysis and indexing processes index the anchor text in all pages that are retrieved by the Web crawler. If you specify that anchor text is not to be indexed, the anchor text in links to forbidden documents is excluded from the index.

## Indexing the anchor text in links to forbidden documents

If a document includes links to documents that the Web crawler is forbidden to crawl, and if you did not enable collection security, you can specify whether you want to retain the anchor text for those links in the index when you configure a Web crawler.

### Before you begin

To configure options for indexing anchor text, you must be a member of the enterprise search administrator role or be a collection administrator for the Web crawler that you want to configure.

### About this task

Directives in a robots.txt file or in the metadata of Web documents can prevent the Web crawler from accessing documents on a Web site. If a document that the Web



crawler is allowed to crawl includes links to forbidden documents, you can specify how you want to handle the anchor text for those links.

If you enable security for a collection when you create it, anchor text indexing is automatically disabled to prevent users from learning about pages that the Web crawler is not allowed to crawl. You cannot change this setting.

If you do not enable security for a collection when you create it, you can specify whether you want to index the anchor text to forbidden documents when you configure the Web crawler. For maximum security, specify that you do not want to index the anchor text in links to forbidden documents. By not indexing anchor text, however, the search results might not include all of the documents that are potentially relevant to a query.

### Procedure

To enable or disable the indexing of anchor text in links to forbidden documents:

1. Click **Collections** to open the Collections view.
2. In the list of collections, locate the collection that you want to configure and click  **Edit**.
3. On the Crawl page, locate the Web crawler that you want to configure and click  **Crawler properties**.
4. Click **Edit advanced Web crawler properties**.
5. To index the anchor text in all documents that this crawler crawls, select the **Index the anchor text in links to forbidden documents** check box. Users will be able to learn about pages that the Web crawler is not allowed to crawl by searching for text that is in the anchor text of links that point to those pages.  
To exclude anchor text in links to forbidden documents from the index, clear this check box. Users will not be able to learn about pages that the Web crawler is not allowed to crawl. The anchor text will be excluded from the index along with the forbidden documents.
6. Click **OK** and then, on the Web Crawler Properties page, click **OK** again.
7. For the changes to become effective, stop and restart the crawler.

To apply the changes to documents that were previously indexed, you must recrawl the documents so that they can be indexed again. If a previous crawl added information about forbidden documents to the index, that information will be removed from the index.

## 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 security tokens (such as user IDs, group IDs, or user roles) 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, see the Search and Index API for enterprise search.

#### **Related concepts**

*"Search applications for enterprise search" on page 125*

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)" in "Programming Guide and API Reference for Enterprise Search"*

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

---

## **Document-level security**

If security is enabled for a collection when it is created, you can configure document-level security controls. Document-level security ensures that users who search collections are able to access only the documents that they are allowed to see.

To control access to documents in the collection, crawlers can collect security for the index. For Lotus Notes documents, you can also validate a user's current credentials whenever a query is submitted. To validate a user's current credentials, you can build support for user profiles into your custom search applications. By storing user profiles, you enable users to access documents without being challenged multiple times to specify their credentials.

#### **Related concepts**

*"Search applications for enterprise search" on page 125*

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 157*

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

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

## **Validation by stored security tokens**

If security is enabled for a collection when it is created, you can configure document-level security controls by storing security data in the index.

By default, each document is considered a public document, which means that it can be searched by all users. For most document types, you can achieve document-level security by associating one or more security tokens with documents and storing these tokens with the documents in the index. When you configure a crawler, you 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 value in one of those fields to enforce access controls. If the data source does not have fields, if you do not want to use a field value for security purposes, or if the field that you specify does not include a value 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. If a data source administrator updates the native access control list, the updated security controls become available the next time that the index is refreshed or reorganized.

Security tokens accompany documents as the documents pass through the stages of parsing, analysis, and indexing. If your search applications enable security, 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 security rules, the user cannot search documents that are protected by the security tokens.

## **Validation of current credentials during query processing**

If security is enabled for a collection when it is created, and the collection includes documents that were crawled by a Notes crawler, you can validate a user's current credentials when the user submits a query.

When you configure a Notes crawler, you can select an option to validate the user's credentials by comparing them to current access controls that are managed by the native repository. Before responding to a query, the Search servers interface with the native repositories to validate the user's current permissions, and then remove all documents that the user does not have permission to view from the search results.

This approach for enforcing document-level security provides a high level of security because the user's credentials are compared to current security data as opposed to security data that is stored in the index. It also ensures that access is controlled by the security mechanisms of the native repository, regardless of how complex those mechanisms might be. Because document filtering occurs in real time, the search results reflect the latest access control settings for each document that matches the search criteria.

Another advantage of this approach is that it does not impact the size of the index (extra space is not required to index the security tokens). However, because the validation requires connections to the native repositories, the approach can impact query performance.

For maximum security, and to minimize the impact on query performance, combine the option for storing security tokens in the index with the option to validate current access controls. When a user submits a query, validation occurs in two stages:

- First, the Search servers use the indexed security data to quickly determine whether the user has permission to access the server and database from which a document was crawled (the index is optimized for speed and produces sub-second response times).

- Next, the Search servers create an interim list that contains only documents that reside in databases on servers that the user is allowed to access, and then they use this list to connect to native repositories and determine whether the user is currently allowed to view the requested document.

If a user has access to a server and database, there is a high probability that the user has access to the documents, but this final filtering stage ensures that only documents that match the user's current permission settings are returned in the search results.

## User profiles

By storing user profiles, you enable users to search secure collections and view documents that they have permission to see without being challenged to specify their access credentials.

If security is enabled for a collection when it is created, and the collection includes documents that were crawled by a Notes crawler, you can configure options that enable user credentials to be encrypted and stored in the enterprise search system.

Users can create a user profile and register their credentials with enterprise search while they use a search application. In the sample search application for enterprise search, this capability is provided by the **My Profile** option on the application toolbar (your custom search applications might implement this capability differently).

The user profile stores the user's Lightweight Directory Access Protocol (LDAP) user name and password and the various credentials that enable the user to access documents in the data sources to be searched. Users can create a user profile if all of the following conditions are true:

- Security is enabled in WebSphere Application Server, and WebSphere Application Server is configured to use an LDAP server for the user registry.
- The schema that supports single sign on capabilities was imported to the LDAP server.
- The LDAP user registry contains an entry that specifies each user's LDAP distinguished name (DN) and password. LDAP must be configured to allow the user's DN and password to be used to add entries under the user's subtree in the user registry.
- Security is enabled in at least one of the collections that the search application can search.
- At least one secure collection includes documents that were crawled by the enterprise search Notes crawler.
- The option to validate current user credentials was selected when document-level security was configured for the Notes crawler.

When users create a profile, they choose whether the LDAP password is to be stored for enterprise search. If this password is stored, users can search collections that require validation without being challenged to authenticate themselves each time they access the search application. If this password is not stored, the user must specify the LDAP password to authenticate themselves before they can search secure collections.

The profile lists all data sources that are available to the search application that require the user's credentials to be validated during query processing. Users can

choose which data sources they want to store credentials for. If the user does not specify credentials for a data source, documents from that data source will be excluded from the search results.

## Configuring support for identity management

Before you can use an LDAP user registry to validate user credentials during query processing, you must configure LDAP to support single-sign on capabilities. In enterprise search, this capability is known as *identity management*.

### Before you begin

To import an LDAP schema, you must be a member of the enterprise search administrator role.

### Restrictions

Security must be enabled in WebSphere Application Server, and WebSphere Application Server must use an LDAP server for the user registry. See the WebSphere Application Server documentation for information about enabling security and configuring a user registry.

### About this task

This task has two parts. To enable identity management in enterprise search, you must:

- Configure IBM Tivoli® Directory Server (ITDS). You must import the enterprise search LDAP schema to ITDS, and configure user permissions so that Tivoli Directory users will be able to read and write LDAP entries.
- Configure Microsoft Active Directory Server (ADS). You must install prerequisite software, import the enterprise search LDAP schema to ADS, and configure user permissions so that Active Directory users will be able to read and write LDAP entries.

### Procedure

1. Import the enterprise search LDAP schema into Tivoli Directory Server:
  - a. Ensure that Tivoli Directory Server is running, then copy or FTP the `tds.add.ldif` file from the `ES_INSTALL_ROOT/bin` directory to the ITDS server.
  - b. On the ITDS server, enter the following command to install the custom schema for enterprise search into the ITDS store, where *LDAP ID* and *LDAP password* are the root DN and password for the LDAP server:

```
ldapadd -D "LDAP ID" -w "LDAP password" -f tds.add.ldif
```
  - c. In a command window, enter the following command or use a third party tool to browse the schema:

```
ldapsearch -b "cn=schema" objectclass=*
```
  - d. Confirm that the following custom attributes were created:

```
vbrInfo
vbrRepoUserName
vbrRepoPassword
vbrRepoOptionalToken
vbrSaltValue
```
  - e. Confirm that the following custom object classes were created:

```
vbrVenetica
vbrRepositoryInfo
vbrRepositoryCredential
vbrUserSalt
```

2. Ensure that enterprise search users have the ability to read and write entries in the LDAP user registry. See the access control section of the Tivoli Directory Server administration documentation for instructions on how to grant user permissions.
3. Install the following software prerequisites on Active Directory Server:
  - Windows Support Tools. This software is available on the Windows 2000 installation disk in the *win2kcd/SUPPORT/TOOLS* directory.
  - Schema Management Snap-In. This software is available on the Windows 2000 installation disk at *win2kcd/I386/AdminPak.msi*.
4. Import the enterprise search LDAP schema into the existing Active Directory schema structure:
  - a. Ensure that Active Directory Server is running, then copy or FTP the *ads.add.ldif* file from the *ES\_INSTALL\_ROOT/bin* directory to the ADS server.
  - b. On the ADS server, enter the following command, where *mydomain* is the base distinguished name (DN) of your LDAP server:

```
ldifde -i -f ads.add.ldif -c "DC=X" "dc=mydomain"
```
  - c. To confirm that the custom LDAP schemas for enterprise search were added, run the *schmmgmt.msc* command to open the Schema Management snap-in. To start this snap-in, click Start, then choose Run, and type .
  - d. After the Schema Management snap-in opens, expand the Classes folder, scroll to the bottom, and confirm that the following custom object classes were created:

```
vbrVenetica
vbrRepositoryInfo
vbrRepositoryCredential
vbrUserSalt
```
5. Ensure that enterprise search users have full control access permissions for the following LDAP objects:

```
vbrVenetica
vbrRepositoryInfo
vbrRepositoryCredential
vbrUserSalt
```

To store credentials in an enterprise search system, enterprise search users must have the ability to read and write objects of these types. See the Active Directory Server administration documentation for instructions on how to grant user permissions.

## Configuring LDAP server properties

If security is enabled for a collection and in WebSphere Application Server, you can store information about your Lightweight Directory Access Protocol (LDAP) server and enable users to search multiple data sources without being prompted to authenticate themselves.

### Before you begin

To configure LDAP server properties, you must be a member of the enterprise search administrator role.



## Restrictions

You can specify information about your LDAP server only if security was enabled for the collection when the collection is created, your search application enabled global security in WebSphere Application Server, WebSphere Application Server uses the LDAP server for the user registry, the collection includes documents that were crawled by a Notes crawler, and the option to validate current user credentials was selected in the Notes crawler configuration.

## About this task

When users access a search application that allows them to configure a user profile, they must specify credentials that enable their identity to be authenticated. The Search servers can access your LDAP server to validate the credentials and determine whether the user has permission to access the collection and, if so, whether the user has permission to access Lotus Domino domains in the collection.

To determine whether an entry exists for a user in the LDAP user registry, enterprise search needs information about your LDAP server. For assistance with specifying this information, contact your LDAP administrator.

The system uses the root credentials, base DN, user ID attribute, and object class to identify LDAP entries for users who log in to the search application. The following LDAP search is simulated in the search application:

```
ldapsearch -D "cn=root" -w secretPassword -b "o=IBM,c=US"  
"(&(uid=adminUser)(objectclass=ePerson))"
```

## Procedure

To configure LDAP server properties:

1. Click **Security** to open the Security view.
2. On the Search Applications page, click **Configure LDAP server properties**.
3. On the LDAP Server Properties page, select the **Validate user credentials with this LDAP server** check box. If you leave this check box clear, the properties that you specify for the LDAP server on this page will be ignored. When users search secure Lotus Notes collections, documents that require validation will be excluded from the search results.
4. Specify information about your LDAP server:
  - The LDAP server name and port. The default port number is 389.
  - The root distinguished name (DN) and password. This information enables enterprise search to retrieve information from the LDAP user registry. For example: cn=root.
  - The base DN for the area of the LDAP user registry where user entries are stored. When determining whether a user is registered, the enterprise search will begin searching the user registry at this location. For example: o=IBM,c=US.
  - The attribute in a user entry that identifies user names. For example: uid.
  - The objectclass that identifies the data type of the user entry. For example: ePerson.
5. Click **OK**.



## Disabling document-level security

You can enable users to search a collection regardless of whether any access controls are associated with the documents in the index. For documents crawled by a Notes crawler, you can also enable users to search a collection without validating current access controls during query processing.

### Before you begin

To configure document-level security options, you must be a member of the enterprise search administrator role or be a collection administrator for the collection.

### Restrictions

You can specify document-level security options only if security was enabled for the collection when the collection is created.

### 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, which are stored in the index, to enforce access controls when users search the collection.


If you configure a Notes crawler, you can also specify that you want to validate access controls that are currently associated with documents in their native repositories during query processing.

To remove these security restrictions, you can specify that the Search servers are to ignore any security tokens that are passed with a query. You can also enable users to query Lotus Notes documents without having their credentials compared to current access controls.

You might want to disable document-level security temporarily if you are testing a new collection or if you need to troubleshoot a problem with a search application.

### Procedure

To disable document-level access controls:

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 General page, click **Configure document-level security**.
4. On the Document-Level Security page, select the **Ignore document-level access controls in the index** check box if you do not want the security tokens that crawlers associated with documents to be used when users query the collection. Crawlers continue to add security tokens to documents, but the Search servers ignore the tokens and allow users to search the previously protected documents.
5. Select the **Skip the validation of current credentials during query processing** check box if you do not want to validate the access controls that are currently associated with documents in their native Lotus Notes repositories when users submit queries. This check box is available only for documents that were crawled by a Notes crawler.

| If you select this check box, other document-level security options remain in  
| effect. For example, if you specified options to store access controls in the index  
| when you configured the Notes crawler, those security controls continue to  
| apply unless you also select the **Ignore document-level access controls in the**  
| **index** check box.

| 6. Click **OK**.

---

# 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 Edition) 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 Edition.

### Related concepts

“Migration from WebSphere Portal to enterprise search” on page 159  
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 Edition 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.

### Before you begin

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 Information Integrator OmniFind Edition 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**

“Search applications for enterprise search” on page 125

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 147

If security is enabled for a collection when it is created, you can configure document-level security controls. Document-level security ensures that users who search collections are able to access only the documents that they are allowed to see.

---

## Migration from WebSphere Portal to enterprise search

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

“Enterprise search integration with WebSphere Portal” on page 155

You can expand the search capabilities of IBM WebSphere Portal by deploying enterprise search portlets in WebSphere Portal and the WebSphere Portal Search Center.

### Related tasks

“Configuring categories” on page 83

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.

---

## 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 **Import model-based taxonomy files from WebSphere Portal**, 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 80

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

#### **Related tasks**

“Configuring categories” on page 83

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.

---

## **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 the search settings from the Portal Search Engine in WebSphere Portal**, 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 use the enterprise search administration console to configure additional settings for the migrated collections.

**Requirement:** When you configure Web crawler properties for a collection that you migrated, you must specify an e-mail address for receiving comments about the crawler and a user agent name (for assistance, click **Help** while you configure Web crawler properties).

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**

“Enterprise search collections” on page 25

An enterprise search collection contains the entire set of sources that users can search with a single query. Through federation, users can search multiple collections with a single query.

“Rule-based categories” on page 79

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

“Model-based categories” on page 80

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 136

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

Before you start a newly migrated Web crawler, review all of the crawler properties and crawl space settings and ensure that all required values are specified (required fields are marked with a red asterisk). In particular, ensure that you specify an e-mail address for receiving comments about the crawler and a user agent name for the crawler. For assistance, click **Help** while you configure Web crawler properties.

---

## 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.



---

## 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 after you make changes to 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 need to stop and restart them only when you change the configuration settings (such as updating categories or increasing the size of the search cache).

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.

### Related concepts

“Monitoring an enterprise search system” on page 9

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

---

## 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.

### Before you begin

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. To use the enterprise search administration console or search application:
  - a. If it is not already running, start the IBM HTTP Server.
  - b. If they are not already running, use the WebSphere Application Server Administrative Console to start the ESAdmin and ESSearchApplication enterprise applications.

2. If the enterprise search system is not running, start it:
  - a. On the Index server, log in with the enterprise search administrator user ID that was specified when DB2 II OmniFind Edition was installed.
  - b. Enter the following command:
 

```
esadmin start
```
3. 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.
4. On the Collections view, locate the collection that you want to administer and click  **Monitor**.
5. 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 crawling all updates** icon or the **Start a full crawl** 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.
6. 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 how the data is parsed (such as updating categories or XML field mappings).
7. 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.
8. 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 the search cache settings.

#### **Related concepts**

“Enterprise search administration overview” on page 15

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, associate search applications with collections, and specify information to enforce security.

“Administrative roles” on page 140

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 18

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.

### Before you begin

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.

If you want to stop the enterprise search system instead of stopping individual components, you can log in on the Index server with the enterprise search administrator ID (this ID was specified when DB2 II OmniFind Edition was installed). Then, enter the following command:

```
esadmin stop
```

### Procedure

To stop enterprise search components:


1. Click **Collections** in the administration console 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 crawler properties, stop and restart the crawler to incorporate the changes. If you change the crawl space and want to apply the changes to documents that are already in the collection, you must also recrawl the documents.

**Tip:** You might see a message about the requested operation timing out even though the process is still running in the background. To determine whether the task has completed, click **Refresh** in the administration console (do not click **Refresh** in the Web browser). The process is finished when the status icon for the crawler indicates that it is stopped.

4. On the Parse page, click  **Stop** to stop the parser.

When you change the rules for parsing data, stop and restart the parser to incorporate the changes. The changes apply only to newly crawled documents. If you want to apply the changes to documents that are already in the index, you must start a full crawl to recrawl all of the documents, which enables them to then be parsed and indexed again.

5. On Index page, click  **Stop** to stop an index that is currently being refreshed or reorganized.

You can also stop an index build while you are monitoring the index queue. To do this, select **System** on the toolbar, 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 servers. Typically, you need to stop and restart the Search servers only when you change the search cache settings.

**Related concepts**

“Enterprise search administration overview” on page 15

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, associate search applications with collections, and specify information to enforce security.



---

## 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 concepts

“Monitoring an enterprise search system” on page 9

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

### Related tasks

“Starting the enterprise search components” on page 165

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 167

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.

---

## Estimating the number of documents in a collection

When you create or edit an enterprise search collection, you provide an estimate for how many documents you expect the collection to hold. The Resource Manager uses this number to estimate the memory and the disk resources that are required for the collection, but not to enforce a limit on the size of the collection.

### Before you begin

To change the estimated size of 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 the collection grows to the size that you estimate, the system does not stop adding documents to the index. The Resource Manager issues warnings when your estimate of memory and disk resource requirements is more than what is currently available in the system. The warnings allow you to prevent future out-of-resource problems.

If you configure alerts for the collection and enable the option to be notified when the number of documents in the index exceeds a limit, the default limit matches

the value that you specify for the estimated number of document in the collection. The Monitor uses this number in conjunction with the alert threshold percentage that you specify and sends e-mail when the maximum number of documents configured for the collection is being approached.

### Procedure

To provide an estimate for the potential size of a collection:

1. Click **Collections** to open the Collections view.
2. In the list of collections, locate the collection that you want to configure and click  **Edit**.
3. On the General page, click **Configure general options**.
4. In the **Estimated number of documents** field, type a number that represents how large you expect the collection to grow. The default value is 1 000 000 documents.

---

## Checking the availability of system resources

If you check system resources after you create a collection or crawler but before you start crawling documents, you can determine whether the available system resources are adequate for running DB2 II OmniFind Edition at full capacity based on your current configuration settings.

### Before you begin

To check system resources, you must be a member of the enterprise search administrator role.

### About this task

When you create a collection or crawler, the system automatically checks the availability of resources. You can also select an option to check the availability of system resources at any time.

The system compares how much space you estimated that you would use when you configured collection and crawler properties to the available system space, and then displays a message that informs you about the availability of resources. The message indicates which server might have insufficient resources, how much space your estimate of the size of the collection requires, and how much free space is available.

If resources are insufficient, try one of the following corrective actions:

- Increase the size of the file system identified in the message text.
- Edit general options for the collection and specify a smaller number for the estimated number of documents.
- Edit crawler properties and specify smaller numbers for the maximum number of documents to crawl, the maximum page size, and the maximum number of threads.
- For a Web crawler, edit the crawler properties to specify smaller numbers for the maximum number of active hosts, maximum number of new documents, and maximum number of documents in temporary storage.
- Edit index properties for the system and reduce the number of index builds that are allowed to run concurrently.

- Avoid having too many processes in multiple collections simultaneously active at any given time. For example, resources can be impacted when crawlers and parsers from multiple collections run at the same time.
- Delete crawlers from the system.
- Delete collections from the system.

If no warnings about potential low resources were detected, there are probably sufficient resources to accommodate the data to be crawled and indexed.

### Procedure

To determine whether sufficient resources are available for the current configuration of your enterprise search system:

1. Click **System** to open the System view.
2. On the General page, click **Check system resources**.

---

## 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.


### Before you begin


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

"Enterprise search administration overview" on page 15

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, associate search applications with collections, and specify information to enforce security.

“Enterprise search collections” on page 25

An enterprise search collection contains the entire set of sources that users can search with a single query. Through federation, users can search multiple collections 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.

### Before you begin

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.


Before you can receive a report, you must ensure that information about your mail server has been configured for enterprise search. You specify this information when you configure e-mail options on the Log page of the System view.


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 in the **E-mail address to notify** field.
  - b. Click **Send Report**.

### Related reference

“URI formats in an enterprise search index” on page 71

The uniform resource identifier (URI) of each document in an enterprise search index 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.


### Before you begin


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** or **Resume**.

### 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. The first time the crawler crawls a data source, it runs a

full crawl. When a scheduled crawl repeats, the crawler crawls all updates to the data source (document additions, deletions, and modifications).

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 crawling all updates** icon (or the **Start a full crawl** 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 or resume the crawler.

#### **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.

“Enterprise search crawler administration” on page 31

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.

---

## **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.


### **Before you begin**


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.

### Before you begin

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.

- 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.

- 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 distributed 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.

### Table summary

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 return codes are proprietary to enterprise search and the Web crawler.

*Table 2. HTTP return codes 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)

Table 2. HTTP return codes from the Web crawler (continued)

Code	Description	Code	Description	Code	Description	Code	Description
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	3020	Soft redirect
302	Found	412	Precondition failed	616	FBA anomaly	4044	Excluded by robots.txt file
303	See other	413	Request entity too large	617	Encoding error		
304	Not modified	414	Request URI is too long	680	DNS lookup failure		
305	Use proxy	415	Unsupported media type				
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.

**616** Indicates that the login form still appears in the download after reauthentication.

**617** Indicates the inability to create a String from a document's byte content because the encoding string (charset) is invalid or the document contains invalid bytes.

**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.

**3020** Indicates that a document with return code 200 contains a location header that refers the user agent to another URL.

---

## Monitoring the parser

Monitor the parser when you need to view information about documents that are analyzed by the parser before they are added to the enterprise search index. Options enable you to review statistics and administer parser activity.

### Before you begin

If your administrative role limits you to monitoring collections, you can view the status of the parser, but you cannot start or stop the parser.


### About this task


When you monitor parser details, you see a snapshot of parser activity that provides statistics about parsing activities at a specific moment in time. The statistics show you the number of documents that were crawled and are being parsed or waiting to be parsed, and the number of documents that were parsed and are waiting to be stored in the index.


When the parser is active, messages provide you with additional information about the actual state of the parser. For example, the parser might be actively parsing documents; idle (sleeping until there are more documents available to parse, waiting to be restarted because of an error condition, or waiting for the parser service to restart); or paused (waiting, for example, for an index reorganization be completed).

### Procedure


To monitor the parser 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 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 running and you want to see detailed status information about parsing activity, click  **Details**.

If your administrative role allows you to administer processes for a collection, you can start and stop the parser while you view details about parsing activities.

5. 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. Unless you make changes to parsing rules, you can let the parser run continuously.

6. If the parser is running and you want to stop it, click  **Stop**.  
You need to stop and restart the parser when you make changes to parsing rules. For example, if you change the categorization type, category rules, XML field mapping rules, HTML metadata mapping rules, or text analysis options, you must stop and restart the parser before your changes become effective.

**Attention:** To apply parsing changes to documents that are already in the index, you must also recrawl the documents so that they can be parsed and indexed again.

#### **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.

---

## **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.


### **Before you begin**





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.

#### **Related concepts**

“Enterprise search indexes” on page 5

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

“Enterprise search index administration” on page 103

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.

---

## **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.

### **Before you begin**

To administer the index queue, you must be a member of the enterprise search administrator role.

### **About this task**

Multiple indexes can be built at the same 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 indexing resources concurrently.

### **Procedure**

To monitor the index queue:

1. Click **System** to open the System view.
2. 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 remove an index from the index queue, click  **Remove**.

#### **Related concepts**

“Enterprise search indexes” on page 5

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



“Enterprise search index administration” on page 103

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.

---

## 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.


### Before you begin



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. 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**.

If you enable or disable the search cache, make changes to the search cache size, or make changes to quick links, you must stop and restart the Search servers for your changes to become effective.
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 7

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

---

## **Monitoring the Data Listener**


Monitor the Data Listener to see its status and to view details about client Data Listener application activity.

### **Before you begin**

To monitor the Data Listener, you must be a member of the enterprise search administrator role.

### **Procedure**

To monitor the Data Listener:

1. Click **System** to open the System view.
2. On the Data Listener page, view the status icons to see whether the Data Listener is active or stopped.
3. If the Data Listener is running and you want to see detailed status information about client application activity, click  **Details**.

Status icons on the Data Listener Details page indicate whether the Data Listener is running or stopped. The statistics show how many requests are waiting to be processed, the current state of each thread that is working on client application requests, and how many threads are active for a given thread state.

4. If you change the port number for the Data Listener, click  **Restart**.

The Data Listener is started when the enterprise search system is started. You do not need to restart the Data Listener unless you change its port number.

### **Related tasks**

“Configuring support for Data Listener applications” on page 70

You can extend enterprise search by using the Data Listener API to create an external crawler. Your custom Data Listener applications can add data to a collection, remove data from a collection, or instruct a Web crawler to visit and revisit URLs.



---

## Enterprise search log file administration

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, receiving alerts, and 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 events occur, or whenever certain messages or 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 enterprise search session. 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.

### Related concepts

“Enterprise search log files” on page 9

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

“Messages for enterprise search” on page 201

Explanations and suggestions for corrective action are provided for most enterprise search messages.

---

## 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.

#### Related concepts

“Enterprise search log files” on page 9

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

## 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.

### Before you begin

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 and click  **Edit**.
3. On the Log 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.

Because you can configure different limits for different crawlers, separate messages are logged for each crawler. For example, if you use the default alert threshold, allow a DB2 crawler to crawl 2 000 000 documents, and allow a Notes crawler to crawl 1 000 000 documents, one message will be logged when the DB2 crawler crawls 1 800 000 documents and another message will be logged when the Notes crawler crawls 900 000 documents.

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 collection reaches a percentage of the estimated size** check box.
  - b. In the **Percentage** field, specify when you want a message to be logged. Specify this number as a percentage of the estimated number of documents that you expect the collection to hold. The default value is 85 percent.  
The **Limit** field shows the current estimated size of the collection. To change this value, open the General page of the collection, select the option to configure general options, and specify a new value in the **Estimated number of documents** field.  
**Attention:** This limit, and the estimated number of documents that you configure for a collection, are used only for monitoring the growth of the collection. They do not enforce an absolute limit on how large the index can grow.
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. The message IDs for the alerts that you enabled are automatically added to the list of message IDs for which e-mail is to be sent.

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 192

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.

### **Before you begin**

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. Click  **Edit** to change to the system editing view.
3. On the Log page, click **Configure alerts**.
4. 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.
5. 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.
6. 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 192

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

---

## **Configuring log files**

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

### **Before you begin**

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. Click  **Edit** to change to the system editing view.

- c. On the Log page, click **Configure log file options**. The System-Level Log File Options page is displayed.
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**.
  - c. On the Log page, click **Configure log file options**. The Collection-Level Log File Options page is displayed.
3. 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 5MB.

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. Click **OK**.

**Related concepts**

“Enterprise search log files” on page 9

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

---

## Configuring SMTP server information

Before you can receive e-mail about enterprise search activities, you must configure information about your Simple Mail Transfer Protocol (SMTP) server.

**Before you begin**

To configure information about your SMTP server, you must be a member of the enterprise search administrator role.


**About this task**

Several enterprise search administrative functions enable you to receive e-mail. Before you can receive e-mail from any of these functions, you must specify information about your SMTP server:

- If you configure collection-level alerts or system-level alerts, you can receive e-mail whenever those messages are logged. You can also receive e-mail when other messages are logged, not just messages that are triggered by monitored events.
- If you want to see detailed information about a URI in the index, you can receive the report by e-mail.
- If you monitor a Web crawler, and specify that you want to create Web crawler history reports, you can be notified by e-mail after a report is created.

### Procedure

To configure information about your SMTP server:

1. Click **System** to open the System view.
2. Click  **Edit** to change to the system editing view.
3. On the Log page, click **Configure e-mail options for messages**.
4. On the 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.

The system uses this server to send e-mail to the addresses that you specify.

5. 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.

6. Click **OK**.

#### Related concepts

“Enterprise search log files” on page 9

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

---

## Receiving e-mail about logged messages

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

### Before you begin

To configure e-mail options for system-level messages, you must be a member of the enterprise search administrator role. 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.

Before you can receive e-mail, you must first configure information about your Simple Mail Transfer Protocol (SMTP) server so that e-mail can be delivered.



### 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.



## Procedure

To configure e-mail options for messages:

1. If you want to receive e-mail about system messages:
  - a. Click **System** to open the System view.
  - b. Click  **Edit** to change to the system editing view.
  - c. On the Log page, click **Configure e-mail options for messages**.
  - d. On the E-mail Options for System Messages page, select the **Receive e-mail about system-level messages** check box.
  - e. In the **E-mail address for receiving e-mail** 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.
  - f. If you want to receive e-mail about all error messages that are logged, select the **Receive e-mail about all error message** 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 in the **Receive e-mail about certain messages** area. Type one message ID per line. For example:  
FFQC4819E  
FFQ00005E  
Several message IDs are listed by default (click **Help** for a description of these messages).
  - h. Click **OK**.
2. If you want to receive e-mail about messages for a collection:
  - a. Click **Collections** to open the Collections view.
  - b. In the list of collections, locate the collection that you want to configure and click  **Edit**.
  - c. On the Log page, click **Configure e-mail options for messages**.
  - d. On the E-mail Options for Collection Messages page, select the **Receive e-mail about collection-level messages** check box.
  - e. In the **E-mail address for receiving e-mail** 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 **Receive e-mail about all error message** check box.
  - g. If you want to receive e-mail only when certain collection-level messages are logged, type the message IDs for those messages in the **Receive e-mail about certain messages** area. Type one message ID per line. For example:  
FFQC4819E  
FFQ00005E  
Several message IDs are listed by default (click **Help** for a description of these messages).
  - h. Click **OK**.

### Related concepts

“Enterprise search log files” on page 9

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

“Messages for enterprise search” on page 201

Explanations and suggestions for corrective action are provided for most enterprise search messages.

#### Related tasks

“Configuring collection-level alerts” on page 188

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 189

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 enterprise search sessions.

### Before you begin

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. Select 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 sessions, select the appropriate check boxes in the **Session** field.
6. Click **View log**.

|  
|  
|

For each message on the Log File Contents page, you see the date and time that the message was issued, the message severity level, the name of the session 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.**

On the Log Message Details page, 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.

#### **Related concepts**

“Enterprise search log files” on page 9

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



---

## 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 when DB2 II OmniFind Edition was installed.
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.

### Before you begin

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





---

## Messages for enterprise search

Explanations and suggestions for corrective action are provided for most enterprise search messages.

### Related concepts

“Enterprise search log files” on page 9

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

“Enterprise search log file administration” on page 187

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, receiving alerts, and receiving e-mail about messages.

### Related tasks

“Receiving e-mail about logged messages” on page 192

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

---

## Controller messages (FFQC)

### FFQC0008E

An error occurred while building *Reorganize* or *Refresh Type* index for collection *collection\_id*.

#### Explanation:

An error occurred when the index was being built or when the index was checked for consistency.

#### User Response:

Check the system error log and the error log for that collection for more information.

### FFQC0020I

The index build was discarded for collection *collection\_id*, partition *partition\_number*, type *build\_type*.

#### Explanation:

The index build could not be started. The index is already being built or it is already queued to build.

#### User Response:

No is action necessary.

### FFQC0023E

Cannot clear collection *collection\_name*, Parser or index build is running

#### Explanation:

A clear collection command was issued for the collection. However, a parser or index build is already running for the collection *collection\_name*.

#### User Response:

Check whether the parser or index build is running for the specified

collection. If you want to clear this collection, stop the parser, stop or wait for the index build to complete, then issue the clear collection command again.

## FFQC0027E

Partition number is not specified for partitioned collection *collection\_name*.

**Explanation:**

The command needs a partition number to be specified if it is a partitioned collection.

**User Response:**

Rerun the command with a partition number specified in the arguments.

## FFQC0032E

The specified merge size is too small.

**Explanation:**

The size specified for the target merge partition is too small. The controller is unable to find candidate partitions that can participate in a merge and yield a new partition that will fit the specified size.

**User Response:**

Specify a size value that is at least twice the size of the partition.

## FFQC0641E

Unable to write to the new partition because parsed data store path is empty.

**Explanation:**

The controller session returned an empty store path for the new collection partition.

**User Response:**

Check the system error log and the collection error log for information about why the controller returned an empty parsed data store path for the new partition.

## FFQC1000E

Collection name *collection\_name* already exists.

**Explanation:**

The collection name *collection\_name* that was entered already exists as a collection name in enterprise search for another collection.

**User Response:**

Enter a different collection name.

## FFQC1002E

An error occurred while creating the index data directory *directory*.

**Explanation:**

An error occurred while attempting to create the collection's index data directory.

**User Response:**

Ensure that all the directories specified in the collection data directory path have necessary create permissions.

**FFQC1003E**

Session name *session\_name* already exists.

**Explanation:**

The session name *session\_name* that was entered already exists as a session name in enterprise search for another session.

**User Response:**

Enter a different session name.

**FFQC1008E**

Session *session\_id* is already configured. Unable to add session.

**Explanation:**

Session *session\_id* that you are trying to add is already configured.

**User Response:**

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.

**FFQC1010E**

An error occurred while removing collection *colleciton\_id*. Received message code *message\_id*.

**Explanation:**

The attempt to remove the collection failed.

**User Response:**

See the returned message code for more information.

**FFQC1018E**

An error occurred while adding session *session\_id* to collection *collection\_id*. Received message code *message\_id*.

**Explanation:**

A collection can have some session (processes) configured in it. An error was encountered while trying to configure the specified session to the collection.

**User Response:**

See the returned message code for more information.

## FFQC1023E

An error occurred while removing the crawler instance: *crawler\_instance\_id*.  
*message\_id*

**Explanation:**

The collection is no longer in the system. However, the collection's associated crawler instance manager tables could not be removed.

**User Response:**

Contact IBM Software Support for further assistance.

## FFQC1041W

Freed configuration lock handle *lock\_handle* for lock holder *session\_id* with PID *pid*. Session was either stopped or terminated without having released this lock.

**Explanation:**

Configuration lock *lock\_handle* being held by session *session\_id* was freed because the session does not appear to be running.

**User Response:**

The session normally or abnormally terminated without freeing the configuration lock. No action is required.

## FFQC1042E

Collection ID *collection\_id* already exists.

**Explanation:**

The collection ID *collection\_id* that was entered already exists as a collection ID in enterprise search for another collection.

**User Response:**

Enter a different collection ID.

## FFQC1043E

Collection ID is missing or is empty.

**Explanation:**

A collection ID must be specified when creating a collection.

**User Response:**

Enter a collection ID. Zero length IDs are not allowed.

## FFQC1044E

Collection name is missing or is empty.

**Explanation:**

A collection name must be specified when creating a collection.

**User Response:**

Enter a collection name. Zero length names are not allowed.

## FFQC1046W

The appids.properties file was not successfully migrated.

### Explanation:

An error occurred while migrating the appids.properties configuration file. On the controller server (index server), the migration process attempts to move the appids.properties file from the \$ES\_NODE\_ROOT/master\_config/admin directory to the \$ES\_NODE\_ROOT/master\_config directory. On the search servers, the migration process attempts to remove the appids.properties file from the \$ES\_NODE\_ROOT/config/admin directory

### User Response:

For a multiple server installation, on the controller server (index server), move the appids.properties file from the \$ES\_NODE\_ROOT/master\_config/admin directory to the \$ES\_NODE\_ROOT/master\_config directory. On the search servers, remove the appids.properties file from the \$ES\_NODE\_ROOT/config/admin directory. For a single server installation, move the appids.properties file from the \$ES\_NODE\_ROOT/master\_config/admin directory to the \$ES\_NODE\_ROOT/master\_config directory. Also, remove the appids.properties file from the \$ES\_NODE\_ROOT/config/admin directory.

## FFQC1072W

The nodes.ini file was not successfully migrated.

### Explanation:

An error occurred while migrating the nodes.ini configuration file. The search server information is not migrated over from the \$ES\_NODE\_ROOT/master\_config/searchmanager/setup.ini file.

### User Response:

Make sure all search nodes within the \$ES\_NODE\_ROOT/master\_config/nodes.ini file contains the proper values for searchserverhost (default to same as destination), searchserverport (default to port 80), and searchservertimeout (default to 60 seconds).

## FFQC2034E

No index is available to refresh the contents of the index.

### Explanation:

The content of an index can only be refreshed on an existing index. No index was found for the specified collection.

### User Response:

Ensure that a valid index was built before you refresh the content of the index.

## FFQC2605E

The maximum number of *number* queued requests was reached.

### Explanation:

The number of requests for getting the URL status exceeded the request queue limit of *number*.

**User Response:**

Wait for the queued requests to be serviced. Try your request again later.

**FFQC2607W**

Number of crawled documents for crawler session *crawler\_id* exceeds the threshold value.

**Explanation:**

The maximum number of documents configured for the crawler *crawler\_id* was exceeded.

**User Response:**

The crawler will automatically pause its operations to allow the parser to consume queued documents. The crawler will automatically resume operations. You should monitor the crawler to ensure that it resumes its activities.

**FFQC2611W**

Number of collection documents for collection *collection\_id* exceeds the threshold value.

**Explanation:**

The maximum number of documents configured for collection *collection\_id* was exceeded. Exceeding the collection limit for the maximum number of documents might impact system performance and other collections configured in the system.

**User Response:**

You can reconfigure the threshold limit to a greater value. However, increasing the threshold limit might impact performance and system resources. You can also lower the threshold limits of other collections if specific system performance levels must be maintained.

**FFQC2613W**

The search process response for search session *session\_id* exceeds the threshold value.

**Explanation:**

The search server response time for query requests exceeded the configured threshold value.

**User Response:**

If this behavior continues, you can tune the query load on the search server. Performance might also be affected by the sharing of resources on the search server with other processes. Consider tuning the load on the server.

**FFQC2616W**

The available file system space is lower than the configured threshold value on server *node\_id*. Actual free space percentage:

*file\_system\_space\_used\_percentage* Configured free space percentage:

*file\_system\_space\_used\_threshold\_percentage*

**Explanation:**

The available file system space on the specified server has fallen below the configured threshold value. Processes that try to write to the file system might fail if there is no space available on the configured file system.

**User Response:**

Increase the size of the file system on the specified server. You can reconfigure the collections that are writing to this file system to prevent the file system resources from being completely consumed.

**FFQC3025E**

Caught parser tokenizer exception while processing document *URI*. This document is being dropped.

**Explanation:**

An error occurred while parsing and tokenizing a document with specified URI. The document is not being added to the parsed data store and will not appear in the next index build.

**User Response:**

Check the system error log for information about why this document was not parsed.

**FFQC3026E**

Exception caught in socket action: *socket\_operation*.

**Explanation:**

An error occurred while parsing and tokenizing documents. The parser is being stopped.

**User Response:**

Check the system error log for information about why a socket error occurred in the parser. Try to restart the parser.

**FFQC4815E**

Invalid collection ID *collection\_id*.

**Explanation:**

There is no collection configured in the system with the specified collection ID.

**User Response:**

Ensure that you specify an existing collection.

**FFQC4838E**

Invalid server ID *node\_id*.

**Explanation:**

There is no server configured in the system with the specified server ID.

**User Response:**

Ensure that you specify an existing server.

## FFQC5201E

Unexpected error in *api\_name* API call.

**Explanation:**

An unexpected error occurred when making an API call to a session.

**User Response:**

Check the system error log and the collection error log for information about why the API call failed.

## FFQC5217E

An error occurred while loading property file *file\_name*.

**Explanation:**

An error occurred while attempting to load the property file contents for *file\_name*.

**User Response:**

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*.

**Explanation:**

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*.

**User Response:**

The file probably does not contain temporary variables, so no scan is needed and no action is required.

## FFQC5233E

Server *node\_id* is inactive.

**Explanation:**

The enterprise search system is not running on the specified server ID.

**User Response:**

Ensure that CCL is running on this server. You can also restart enterprise search on the controller server (index server).

## FFQC5234W

Unable to start session *session\_id* on server *node\_id*. Server *node\_id* is inactive.

**Explanation:**

Unable to start session. The enterprise search system is not running on the specified server ID.



**User Response:**

Ensure that CCL is running on this server. You can restart enterprise search on the controller server (index server).

## FFQC5235E

The API call *api\_name* returned a NULL result.

**Explanation:**

A NULL result was returned when making an API call to a session.

**User Response:**

Check the system error log and the collection error log for information about why the API call returned a NULL result.

## FFQC5238E

Invalid partition number *partition\_number* for collection *collection\_id*.

**Explanation:**

The collection does not contain the specified partition, or no partition number was specified.

**User Response:**

Ensure that you specify a valid partition number.

## FFQC5800E

An error occurred while estimating the amount of file system space that will be used by session *session identifier*.

**Explanation:**

An error was encountered while generating an estimate of the amount of file system space that will be used by session *session identifier*.

**User Response:**

Monitor the amount of file system space used by session *session identifier* and ensure that adequate file system space is available for proper system operations.

## FFQC5801E

An error occurred while estimating the amount of virtual memory that will be used by session *session identifier*.

**Explanation:**

An error was encountered while generating an estimate of the amount of virtual memory that will be used by session *session identifier*.

**User Response:**

Monitor the amount of virtual memory used by session *session identifier* and ensure that adequate virtual memory is available for proper system operations.

## FFQC5802E

An error occurred while estimating the amount of file system space that will be used by collection *collection identifier*.

**Explanation:**

An error was encountered while generating an estimate of the amount of file system space that will be used by collection *collection identifier*.

**User Response:**

Monitor the amount of file system space used by collection *collection identifier* and ensure that adequate file system space is available for proper system operations.

**FFQC5803E**

An error occurred while estimating the amount of virtual memory that will be used by collection *collection identifier*.

**Explanation:**

An error was encountered while generating an estimate of the amount of virtual memory that will be used by collection *collection identifier*.

**User Response:**

Monitor the amount of virtual memory used by collection *collection identifier* and ensure that adequate virtual memory is available for proper system operations.

**FFQC5804E**

An error occurred while writing the required size of *amount of file system space in MB* for file system *file system identifier* on server *node identifier* associated with collection *collection identifier* to file.

**Explanation:**

The required size of *amount of file system space in MB* for file system *file system identifier* on server *node identifier* associated with collection *collection identifier* could not be written to file.

**User Response:**

Ensure that the file system has enough available space.

**FFQC5805E**

An error occurred while writing the required virtual memory size of *amount of virtual memory in bytes* for collection *collection identifier* on server *node identifier* to file.

**Explanation:**

The required size of *amount of virtual memory in bytes* for the virtual memory on server *node identifier* could not be written to file.

**User Response:**

Ensure that the file system has enough available space.

**FFQC5806E**

An error occurred while reading the required size for file system *file system identifier* on server *node identifier* associated with collection *collection identifier*.

**Explanation:**

The requires size of file system *file system identifier* on server *node identifier* associated with collection *collection identifier* could not be read from file.

**User Response:**  
Check the file system logs.

## FFQC5807E

An error occurred while reading the required virtual memory on server *node identifier*.

**Explanation:**  
The required size for the virtual memory on server {1} could not be read from file.

**User Response:**  
Check the file system logs.

## FFQC5808E

An error occurred while processing the file system size estimates for server *node identifier*.

**Explanation:**  
The calculated amount of file system space used on server *node identifier* could not be compared against the actual amount of file system space used on the server.

**User Response:**  
Check the log files for more details.

## FFQC5809E

An error occurred while processing the virtual memory estimates for server *node identifier*.

**Explanation:**  
The calculated amount of virtual memory used on server *node identifier* could not be compared to the existing amount of virtual memory used on the server.

**User Response:**  
Check the log files for more details.

## FFQC5810W

The available free space on file system *file system identifier* of server *node identifier* might be insufficient for the proper operation of the system. The estimated additional space required from file system *file system identifier* on server *node identifier* is *size* MB, while the free space available is only *size* MB.

**Explanation:**  
The available free space on file system *file system identifier* of server *node identifier* might be insufficient for the proper operation of the system. The estimated additional space required from file system *file system identifier* on server *node identifier* is *size* MB, while the free space available is only *size* MB.

**User Response:**  
To avoid possible space problems, the following these steps:

1. Increase the size of the specified file system.
2. Reconfigure the system with a smaller value for the maximum number of documents.
3. Remove crawlers from the system.
4. Remove collections from the system.
5. Avoid having too many collections simultaneously active within the system at one time.

## FFQC5811W

The available free virtual memory space on server *node identifier* might be insufficient for the proper operation of the system. The estimated additional virtual memory space required on server *node identifier* is *size* MB, while the free virtual memory space available is only *size* MB.

### Explanation:

The available free virtual memory space on server *node identifier* might be insufficient for the proper operation of the system. The estimated additional virtual memory space required on server *node identifier* is *size* MB, while the free virtual memory space available is only *size* MB.

### User Response:

To avoid possible space problems, the follow these steps:

1. Add more memory to the identified server.
2. Reconfigure the system with smaller values for the maximum numbers of documents to index and to crawl.
3. Remove crawlers from the system.
4. Remove collections from the system.
5. Avoid having too many collections simultaneously active within the system at any given time.

## FFQC5812E

An error occurred while determining the maximum available resources in the system. Resource data file: *filename*

### Explanation:

The maximum file system space and memory available to the system cannot be determined.

### User Response:

If the specified file *filename* exists, ensure that it has adequate read permissions. If the file does not exist, ensure that the specified directory has adequate write permissions. Ensure also that all servers of the system are running.

## FFQC5813E

The Resource Manager cannot be started because the session has not been initialized.

### Explanation:

The session cannot be started because it has not been initialized.

### User Response:

The Resource Manager session must be initialized before it can be started.

## FFQC5814E

An error occurred within the Resource Manager session while loading the input file *filename*.

**Explanation:**

The Resource Manager session encountered an error while loading the file *filename*.

**User Response:**

Check that the file *filename* exists and that the file has read and write permissions.

## FFQC5815E

An error occurred while determining the free space for file system *file system identifier* on server *node identifier* associated with collection *collection identifier*.

**Explanation:**

The free space for file system *file system identifier* on server *node identifier* associated with collection *collection identifier* could not be determined.

**User Response:**

Check the log files for more details.

## FFQC5816E

An error occurred while determining the available memory resources on server *node identifier*.

**Explanation:**

The available memory resources on server *node identifier* could not be determined.

**User Response:**

Check the log files for more details.

## FFQC5817E

An error occurred while comparing available file system resources against estimated file system resources on server *node identifier*.

**Explanation:**

The comparison of estimated file system resources to available file system resources on server *node identifier* cannot be performed at this time.

**User Response:**

Check the log files for more details.

## FFQC5818E

An error occurred while comparing available memory resources against estimated memory resources on server *node identifier*.

**Explanation:**

The comparison of estimated memory resources to available memory resources on server *node identifier* cannot be performed at this time.

**User Response:**  
Check the log files for more details.

## FFQC5819E

An error occurred while trying to determine the current number of documents in collection *collection identifier*.

**Explanation:**  
The current number of documents that are contained in collection *collection identifier* cannot be established.

**User Response:**  
Check the log files for more details.

## FFQC5820E

An error occurred while trying to determine the maximum number of documents configured for collection *collection identifier*.

**Explanation:**  
The maximum number of documents configured for collection *collection identifier* cannot be established.

**User Response:**  
Check the log files for more details.

## FFQC5821E

An error occurred while estimating the amount of file system space that is currently being used by collection *collection identifier*.

**Explanation:**  
An error was encountered while generating an estimate of the amount of file system space that is currently being used by collection *collection identifier*.

**User Response:**  
Monitor the amount of file system space that is used by collection *collection identifier* and ensure that adequate file system space is available for proper system operations.

## FFQC5822E

An error occurred while comparing available system resources against estimated system requirements. Resource report: *file name*

**Explanation:**  
The comparison of estimated system requirements to available system resources cannot be performed at this time.

**User Response:**  
Check the log files for more details.

## FFQC5823E

An error occurred while executing the resource checking function of the Resource Manager.

**Explanation:**

The comparison of estimated system requirements to available system resources cannot be performed.

**User Response:**

Check the log files for more details.

---

## Crawler messages (FFQD)

### FFQD1003E

ERROR: User *user* does not exist. Check *file*.

**Explanation:**

The user that was specified in the es.cfg file does not exist.

**User Response:**

Verify that DB2 Information Integrator OmniFind Edition (enterprise search) was installed correctly.

### FFQD1011E

ERROR: *file* was not found.

**Explanation:**

The specified file was not found.

**User Response:**

Verify that DB2 Information Integrator OmniFind Edition (enterprise search) was installed correctly.

### FFQD1031E

ERROR: This script is not running on the crawler server. Run this script on the crawler server.

**Explanation:**

This script is not running on the crawler server.

**User Response:**

For a multiple server configuration, run this script on the 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. Verify that DB2 Information Integrator for Content was installed correctly.

**Explanation:**

The environment variable CMBROOT was not defined. DB2 Information Integrator for Content might not be installed.

**User Response:**

Verify that DB2 Information Integrator for Content was installed correctly.

## FFQD1035E

ERROR: The environment variable CMCOMMON is not defined. Verify that DB2 Information Integrator for Content was installed correctly.

**Explanation:**

The environment variable CMCOMMON was not defined. DB2 Information Integrator for Content might not be installed.

**User Response:**

Verify that DB2 Information Integrator for Content was installed correctly.

## FFQD1036E

ERROR: The environment variable ES\_NODE\_ROOT is not defined. Verify that DB2 Information Integrator OmniFind Edition was installed correctly and that this script was run by the enterprise search administrator user.

**Explanation:**

The environment variable ES\_NODE\_ROOT was not defined.

**User Response:**

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. Verify that DB2 Information Integrator OmniFind Edition was installed correctly and this script was run by the enterprise search administrator user.

**Explanation:**

The environment variable ES\_INSTALL\_ROOT was not defined.

**User Response:**

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.

**Explanation:**

The setup script was executed on an unsupported platform.

**User Response:**

Ensure that the setup script is run on a supported platform.

## FFQD1106E

ERROR: Run this setup shell as the enterprise search administrator user, not as the root user.

**Explanation:**

The setup script must be run by the enterprise search administrator user.



**User Response:**

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.

**Explanation:**

ES\_CFG environment variable was not detected by setup script.

**User Response:**

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.

**Explanation:**

The enterprise search configuration file es.cfg does not exist in server root (ES\_NODE\_ROOT) directory.

**User Response:**

Check that DB2 Information Integrator OmniFind Edition is installed correctly and check that the enterprise search configuration file es.cfg exists in the server root directory (ES\_NODE\_ROOT).

## FFQD1123E

ERROR: The Lotus Notes directory was not found.

**Explanation:**

The Lotus Notes directory that was specified was not found in the crawler server.

**User Response:**

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.

**Explanation:**

The Lotus Notes data directory that was specified was not found on the crawler server.

**User Response:**

Check that Lotus Notes directory was properly deployed on the crawler server.

## FFQD1138E

ERROR: Failed to remove *filename*. Remove it and run this script again.

**Explanation:**

This script cannot delete the file *filename*.

**User Response:**

Remove the file manually and run this script again.

**FFQD1139E**

ERROR: Failed to create *filename*.

**Explanation:**

This script cannot create the file *filename*.

**User Response:**

Check the file and directory privileges so that the file can be created.

**FFQD1142E**

ERROR: *filename* was not found.

**Explanation:**

This script cannot find the file *filename*.

**User Response:**

Check that the file exists.

**FFQD1407E**

ERROR: Venetica VeniceBridge directory was not found.

**Explanation:**

The Venetica VeniceBridge directory that was specified was not found on the crawler server.

**User Response:**

Check that Venetica VeniceBridge was properly installed on the crawler server.

**FFQD1420E**

ERROR: WebSphere Application Server directory was not found.

**Explanation:**

WebSphere Application Server directory specified was not found on the crawler server.

**User Response:**

Check that WebSphere Application Server was properly installed on the crawler server.

**FFQD2005E**

The crawler server is not configured for the specified crawler type. Confirm that the crawler server is configured by the appropriate setup script.

**Explanation:**

The crawler server is not configured for the crawler type you want to use.

**User Response:**

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*.

**Explanation:**

The *api\_name* discovery API could not be registered because of an internal error.

**User Response:**

Check that enterprise search and prerequisite software are correctly installed on the crawler server.

**FFQD2100E**

The SQL error occurred while accessing the database. Error code: *error\_message*. Error message: *error\_code*.

**Explanation:**

The error occurred while accessing the DB2 database server.

**User Response:**

See the IBM DB2 Universal Database Message Reference.

**FFQD2102E**

User name or password is incorrect.

**Explanation:**

The user name, password, or both that were used to establish a connection to the DB2 server were incorrect.

**User Response:**

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*.

**Explanation:**

The error occurred while discovering the DB2 database server.

**User Response:**

See IBM DB2 Universal Database Message Reference.

**FFQD2104E**

A connection cannot be established.

**Explanation:**

A connection to the specified DB2 database server failed.

**User Response:**

Try the operation again. Verify that the DB2 UDB server is available.

## FFQD2105E

The table or view *table\_or\_view\_name* was not found for schema name *schema\_name*.

**Explanation:**

The discovery process could not discover the columns because the specified table or view does not exist on the DB2 database server.

**User Response:**

Specify the table or view that exists in the DB2 database server.

## FFQD2106E

The DB2 instance *server\_name:port\_number* was not found.

**Explanation:**

The specified DB2 instance was not found.

**User Response:**

Check that the server name and port number that identify the DB2 instance are correct.

## FFQD2107E

The other operation for discovering remote databases is requested. The current operation is for the DB2 instance *server\_name:port\_number*.

**Explanation:**

The other operation for discovering databases is running.

**User Response:**

Wait for the other operation to complete and try again. If it is an unintentional operation, check that DB2 Administration Server (DAS) on the specified server is available.

## FFQD2108E

The other operation for discovering local databases is requested.

**Explanation:**

The other operation for discovering databases is running.

**User Response:**

Wait for the other operation to complete and try again.

## FFQD2109E

The discovery operation was interrupted.

**Explanation:**

The requested operation was interrupted.

**User Response:**

Try the operation again.

## FFQD2130E

A Notes error occurred while accessing the server. ID: *error\_id*. Error message: *error\_message*

**Explanation:**

The error occurred while accessing the Domino server.

**User Response:**

See the error message for more information.

## FFQD2131E

The specified ID file does not exist or cannot be read.

**Explanation:**

The specified user ID file does not exist, or you do not have permission to read the file.

**User Response:**

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.

**Explanation:**

The discovery process could not create a Notes session to the Domino server.

**User Response:**

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*.

**Explanation:**

The discovery process cannot coexist with other Notes crawler processes. One of them was holding the lock *lock\_name*.

**User Response:**

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.

**Explanation:**

The correct Notes Java class for NRPC (Notes Remote Procedure Call) could not be loaded. A setup procedure was required for the Notes crawler.

**User Response:**

Perform the setup procedure for the Notes crawler.

## FFQD2136E

Could not get domain name from Domino Directory. ID: *error\_id*. Error message: *error\_message*

**Explanation:**

Could not get domain name from Domino Directory.

**User Response:**

See the error message for more information. Ensure that crawl user can read Domino Directory.

## FFQD2137E

Server name must be a fully qualified Internet host name. Server name: *server\_name*

**Explanation:**

Specified server name *server\_name* was not fully qualified Internet host name. For example, a canonical server name like *svename/org1/org2* is not valid.

**User Response:**

Specify correct fully qualified Internet host name.

## FFQD2160E

The directory *directory name* is not valid.

**Explanation:**

The specified directory name was not found.

**User Response:**

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.

**Explanation:**

There are more than *levels of subdirectories to discover* available subdirectories.

**User Response:**

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.

**Explanation:**

The specified name was a file name.

**User Response:**

Enter an existing directory name.

## FFQD2163E

The specified pattern for directory *directory name or pattern* is not valid.

**Explanation:**

The specified pattern to search is not correct.

**User Response:**

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.

**Explanation:**

The specified directory on the crawler server was not found.

**User Response:**

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*.

**Explanation:**

The specified network path was not found.

**User Response:**

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.

**Explanation:**

The specified drive letter or the specified file separator is not valid.

**User Response:**

Enter the correct drive letter to search directories.

## FFQD2210E

The user ID, password, or both are incorrect. The specified user ID is *user\_id*.

**Explanation:**

The user ID, password, or both that were entered to connect to the Content Manager server were incorrect.

**User Response:**

Try the operation again with the correct user ID and password.

## FFQD2211E

The connection to the Content Manager server *server* failed.

**Explanation:**

A connection to the specified Content Manager server failed.

**User Response:**

Try the operation again. Verify that the Content Manager server is available.

## FFQD2212I

Connected to Content Manager server *server*.

**Explanation:**

The crawler server successfully connected to the Content Manager server.

**User Response:**

This message is for your information only. No action is required.

## FFQD2213E

CMCOMMON keyword is not found in discovery.properties file.

**Explanation:**

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.

**User Response:**

Configure the Content Manager crawler with the appropriate setup script and try the operation again.

## FFQD2214E

CM INI file *file* was not found.

**Explanation:**

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.

**User Response:**

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.

**Explanation:**

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.



**User Response:**

Verify that the DB2 Information Integrator for Content is installed and configured correctly.

## FFQD2216E

The item type *item\_type* does not exist.

**Explanation:**

The discovery process could not discover the Content Manager attributes because the specified item type does not exist on the Content Manager server.

**User Response:**

Specify an item type that exists on the Content Manager server.

## FFQD2217E

CM DKException occurred.

**Explanation:**

An error occurred while discovering the Content Manager server.

**User Response:**

See IBM DB2 Content Manager for Multiplatforms/IBM DB2 Information Integrator for Content for Multiplatforms, Messages and Codes for more information.

## FFQD2218E

A DB2 Content Manager error occurred.

**Explanation:**

An error occurred while discovering the Content Manager server.

**User Response:**

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.

**Explanation:**

You must specify a truststore file to use SSL.

**User Response:**

Prepare a truststore file on the crawler server and enter its path.

## FFQD2241E

A truststore password must be specified to use SSL.

**Explanation:**

You must specify a truststore password for the file to use SSL.

**User Response:**

Enter a password for the truststore file.

## FFQD2242E

The truststore file *truststore\_file* does not exist.

**Explanation:**

The specified truststore file did not exist or you did not have permission to read the file.

**User Response:**

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*.

**Explanation:**

The specified truststore file did not exist or you did not have permission to read the file.

**User Response:**

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*.

**Explanation:**

The truststore file format is incorrect. The supported formats are JKS, JCEKS and PKCS#12.

**User Response:**

Verify that the truststore file format.

## FFQD2245E

An unknown host name *host\_name* was detected.

**Explanation:**

The host name in the specified public folder URL was not found.

**User Response:**

Verify that the host name is correct.

## FFQD2246E

The URL *public\_folder\_URL* is incorrect.

**Explanation:**

The specified public folder URL was not found.

**User Response:**

Verify that the public folder URL is correct.

## FFQD2247E

An I/O error occurred while connecting to *public\_folder\_URL*.

**Explanation:**

One or more errors occurred while connecting to the public folder server.

**User Response:**

Verify that network and the public folder server are working.

**FFQD2248E**

A protocol error occurred while sending a request to *public\_folder\_URL*.

**Explanation:**

One or more errors occurred while sending a request to the public folder server.

**User Response:**

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*.

**Explanation:**

One or more errors occurred while sending a request to the public folder server.

**User Response:**

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*.

**Explanation:**

The number of found subfolders must be the limit or less.

**User Response:**

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*.

**Explanation:**

An incorrect user ID, an incorrect password or both were specified.

**User Response:**

Verify the user name and the password and try again.

**FFQD2252E**

The specified folder *public\_folder\_URL* does not exist.

**Explanation:**

The specified public folder server existed, but the folder did not exist on the server.

**User Response:**

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*.

**Explanation:**

The public folder server returned an unexpected HTTP response.

**User 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*.

**Explanation:**

One or more errors occurred while receiving a response from the public folder server.

**User Response:**

See the log files for more information about where the problem occurred.

## **FFQD2255E**

The server response is incorrect XML.

**Explanation:**

The public folder server returned a XML document which might contain incorrect elements and characters.

**User Response:**

Verify the public folder does not contain broken documents.

## **FFQD2256E**

An I/O error occurred while parsing the received XML.

**Explanation:**

One or more errors occurred while receiving a response from the public folder server.

**User Response:**

See the log files for more information about where the problem occurred.

## **FFQD2257E**

An XML parser error occurred.

**Explanation:**

Could not create a XML parser for parsing responses from the public folder server.

**User Response:**

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.

**Explanation:**

The specified URL can be accessed from the crawler, but it is not a public folder.

**User Response:**

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.

**Explanation:**

The discovery process could not connect to the VeniceBridge server with the specified JNDI properties.

**User Response:**

Check the specified JNDI properties are correct and the VeniceBridge server is available.

**FFQD2271E**

The specified repository name is not valid.

**Explanation:**

The discovery process could not discover the item classes since the specified repository did not exist in the VeniceBridge server.

**User Response:**

Specify the repository that exists on the VeniceBridge server.

**FFQD2272E**

The specified user name or password is not valid.

**Explanation:**

The user name, password, or both to logon to the repository were incorrect.

**User Response:**

Try the operation again with correct user name and password.

**FFQD2273E**

The specified item class name is not valid.

**Explanation:**

The discovery process could not discover the properties because the specified item class did not exist in the VeniceBridge server.

**User Response:**

Specify the item class that exists on the VeniceBridge server.

## **FFQD2274E**

A VeniceBridge API error occurred. Error message: *error\_message*.

**Explanation:**

The error occurred while accessing to the VeniceBridge server.

**User Response:**

See the error message from the VeniceBridge server for more information.

## **FFQD2275E**

The VeniceBridge server instance cannot be created. Error message: *error\_message*.

**Explanation:**

The error occurred while accessing to the VeniceBridge server.

**User Response:**

See the error message from the VeniceBridge server for more information.

## **FFQD2276E**

The VeniceBridge server cannot be located. Error message: *error\_message*.

**Explanation:**

The error occurred while accessing to the VeniceBridge server.

**User Response:**

See the error message from the VeniceBridge server for more information.

## **FFQD2277E**

A communication error occurred. Error message: *error\_message*.

**Explanation:**

The error occurred while accessing to the VeniceBridge server.

**User Response:**

See the error message from the VeniceBridge server for more information.

## **FFQD2278E**

The specified repository *repository\_name* was not found. Error message: *error\_message*.

**Explanation:**

The specified repository did not exist in the VeniceBridge server.

**User Response:**

Specify the repository which exists in the VeniceBridge server.

## **FFQD2279E**

The VeniceBridge user instance was not initialized. Error message: *error\_message*.

**Explanation:**

The error occurred while accessing to the VeniceBridge server.

**User Response:**

See the error message from the VeniceBridge server for more information.

**FFQD2280E**

Logon to *repository\_name* repository failed. Error message: *error\_message*.

**Explanation:**

The discovery process could not log on to the *repository\_name* repository.

**User Response:**

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*

**Explanation:**

The *item\_class\_name* item class did not exist in *repository\_name* repository in the VeniceBridge server.

**User Response:**

Specify the item class which exists in the repository.

**FFQD2800E**

An error was generated from crawler plug-in. Message: *message*

**Explanation:****User Response:**

An error message *message* was generated from crawler plug-in. Refer to the message from crawler plug-in.

**FFQD2801W**

A warning was generated from crawler plug-in. Message: *message*

**Explanation:****User Response:**

A warning message *message* was generated from crawler plug-in. Refer to the message from crawler plug-in.

**FFQD2802I**

An information was generated from crawler plug-in. Message: *message*

**Explanation:****User Response:**

An information message *message* was generated from crawler plug-in. Refer to the message from crawler plug-in.

## FFQD2803I

The crawler plug-in process started successfully.

**Explanation:**

**User Response:**

This message is for your information only. No action is required.

## FFQD2804I

The crawler plug-in process stopped successfully.

**Explanation:**

**User Response:**

This message is for your information only. No action is required.

## FFQD2805I

The crawler plug-in class *class\_name* is loaded successfully.

**Explanation:**

**User Response:**

This message is for your information only. No action is required.

## FFQD2806E

The crawler plug-in process failed to start

**Explanation:**

The crawler plug-in process failed to start.

**User Response:**

See the log files for more information about where the problem occurred.

## FFQD2807E

The crawler plug-in process failed to stop successfully.

**Explanation:**

The crawler process failed to be stopped successfully.

**User Response:**

See the log files for more information about where the problem occurred.

## FFQD2808E

The user-defined class of crawler plug-in cannot be created.

**Explanation:**

The user-defined class derived from CrawlerPlugin class cannot be created.

**User Response:**

Verify that the class name and the CLASSPATH were set correctly.

## FFQD2809E

The crawler plug-in session *session\_name* is not available.



**Explanation:**

The crawler plug-in session is not available. The user defined class could not be created correctly.

**User Response:**

See the log files for more information about where the problem occurred.

**FFQD2810E**

The error occurred in `init()` method of `CrawlerPlugin` class `user_defined_classname`.

**Explanation:**

The `init()` method of user defined class did not work correctly.

**User Response:**

Verify the program of the user defined class.

**FFQD2811E**

The error occurred in `term()` method of `CrawlerPlugin` class `user_defined_classname`.

**Explanation:**

The `term()` method of user defined class did not work correctly.

**User Response:**

Verify the program of the user defined class.

**FFQD2812E**

The error occurred in `isMetadataUsed()` method of `CrawlerPlugin` class `user_defined_classname`.

**Explanation:**

The `isMetadataUsed()` method of user defined class did not work correctly.

**User Response:**

Verify the program of the user defined class.

**FFQD2813E**

The error occurred in `updateDocument()` method of `CrawlerPlugin` class `user_defined_classname`.

**Explanation:**

The `updateDocument()` method of user defined class did not work correctly.

**User Response:**

Verify the program of the user defined class.

**FFQD2900E**

The crawler plug-in server (`server_url`) is not ready. The client could not connect to the server.

**Explanation:**

The crawler plug-in server (*server\_url*) is not ready. The client can not connect to it, or it can not be terminated.

**User Response:**

Confirm that the crawler plug-in server (*server\_url*) is ready.

**FFQD2901I**

The crawler plug-in client (*client\_url*) initialized. It connected to the server (*server\_url*).

**Explanation:****User Response:**

This message is for your information only. No action is required.

**FFQD2902W**

The crawler plug-in client (*client\_url*) received a warning during termination.

**Explanation:**

The crawler plug-in client encountered problems while terminating. The connection closed abnormally.

**User Response:**

Ensure that the server is running properly.

**FFQD2903I**

The crawler plug-in client (*client\_url*) terminated.

**Explanation:****User Response:**

This message is for your information only. No action is required.

**FFQD2904E**

The crawler plug-in client is not initialized.

**Explanation:**

A crawler plug-in client must be initialized before using it.

**User Response:**

Initialize the crawler plug-in client.

**FFQD2905E**

The crawler plug-in client received unsupported data.

**Explanation:**

A crawler plug-in client can receive SocketData object only.

**User Response:**

Confirm that the crawler plug-in server sends a SocketData object.

## FFQD2906I

The crawler plug-in server was created without a specific port number.

**Explanation:**

The crawler plug-in server uses any free port when it initializes.

**User Response:**

This message is for your information only. No action is required.

## FFQD2907E

The crawler plug-in server was not created.

**Explanation:**

The crawler plug-in server requires a crawler plug-in module.

**User Response:**

Check that a crawler plug-in module was created before you create the crawler plug-in server.

## FFQD2908E

The crawler plug-in server could not be started. The listener port *listener\_port* was already in use.

**Explanation:**

The crawler plug-in server could not start. The listener port *listener\_port* is already in use.

**User Response:**

Specify another port number or specify 0 to use any free port.

## FFQD2909E

The crawler plug-in server could not be started.

**Explanation:**

An error occurred while starting the crawler plug-in server. Typically, an I/O error occurs when opening the socket.

**User Response:**

Start the server again, or fix the I/O problem.

## FFQD2910I

The crawler plug-in server was started. The listener port is *listener\_port*.

**Explanation:**

**User Response:**

This message is for your information only. No action is required.

## FFQD2911E

The crawler plug-in server (*server\_url*) is not ready. The server could not be terminated.

**Explanation:**

The crawler plug-in server (*server\_url*) is not ready. The server could not be terminated.

**User Response:**

Check the listener port number of the crawler plug-in server that you want to terminate.

**FFQD2912W**

The crawler plug-in server (*server\_url*) encountered problems while terminating.

**Explanation:**

The crawler plug-in server encountered problems while terminating. The server terminated abnormally.

**User Response:**

Ensure that the client is running properly.

**FFQD2913I**

The crawler plug-in server (*server\_url*) terminated.

**Explanation:****User Response:**

This message is for your information only. No action is required.

**FFQD2914I**

The crawler plug-in server is listening. The listener port is *listener\_port*.

**Explanation:****User Response:**

This message is for your information only. No action is required.

**FFQD2915I**

The crawler plug-in server accepted a connection request from the client (*client\_url*).

**Explanation:****User Response:**

This message is for your information only. No action is required.

**FFQD2916I**

The crawler plug-in server closed the connection to the client (*client\_url*).

**Explanation:****User Response:**

This message is for your information only. No action is required.

## FFQD2917E

The crawler plug-in server received unsupported class object.

**Explanation:**

The crawler plug-in server received unsupported class object. This object could not be send to the crawler plug-in module.

**User Response:**

See the log files for more information about where the problem occurred.

## FFQD2997E

An error occurred for the crawler plug-in socket client.

**Explanation:**

An internal error occurred in the crawler plug-in socket client.

**User Response:**

See the log files for more information about where the problem occurred.

## FFQD2998E

An error occurred for the crawler plug-in socket server.

**Explanation:**

An internal error occurred in the crawler plug-in socket server.

**User Response:**

See the log files for more information about where the problem occurred.

## FFQD2999E

An error occurred for the crawler plug-in socket.

**Explanation:**

An internal error occurred in the crawler plug-in socket.

**User Response:**

See the log files for more information about where the problem occurred.

## FFQD3000I

*crawler\_name* crawler started successfully.

**Explanation:**

**User Response:**

This message is for your information only. No action is required.

## FFQD3001E

*crawler\_name* crawler failed to start.

**Explanation:**

The crawler process failed to be started.

**User Response:**

See the log files for more information about where the problem occurred.

## FFQD3002I

*crawler\_name* crawler stopped successfully.

**Explanation:**

**User Response:**

This message is for your information only. No action is required.

## FFQD3003E

*crawler\_name* crawler did not stop successfully.

**Explanation:**

The crawler process failed to be stopped successfully.

**User Response:**

See the log files for more information about where the problem occurred.

## FFQD3004I

*crawler\_name* crawler was paused successfully..

**Explanation:**

**User Response:**

This message is for your information only. No action is required.

## FFQD3005E

*crawler\_name* crawler failed to pause.

**Explanation:**

The crawler failed to be paused.

**User Response:**

See the log files for more information about where the problem occurred.

## FFQD3006I

*crawler\_name* crawler resumed successfully.

**Explanation:**

**User Response:**

This message is for your information only. No action is required.

## FFQD3007E

*crawler\_name* crawler failed to resume.

**Explanation:**

The crawler failed to be resumed.

**User Response:**

See the log files for more information about where the problem occurred.

## FFQD3008E

Factory error occurred. Verify that the modules are installed properly.

**Explanation:**

The installed JAR files were not found or were not set correctly.

**User Response:**

Verify that enterprise search was successfully installed and try again.

**FFQD3009I**

Start the crawl for *crawler\_name* on *server\_name*.

**Explanation:****User Response:**

This message is for your information only. No action is required.

**FFQD3010I**

Finished the crawl for *crawler\_name* on *server\_name*.

**Explanation:****User Response:**

This message is for your information only. No action is required.

**FFQD3011E**

Failed to crawl for *target\_host\_name* on *target\_name*.

**Explanation:**

The crawler failed to crawl for the specified target.

**User Response:**

See the log files for more information about where the problem occurred.

**FFQD3012I**

The crawler is already crawling on server *server\_name*.

**Explanation:****User Response:**

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*.

**Explanation:**

The crawled data was truncated.

**User Response:**

Change the maximum page size for a crawler on the administration console if all the data needs to be crawled.

**FFQD3014E**

An error occurred while handling crawled history.

**Explanation:**

The internal error occurred in the crawler.

**User Response:**

See the log files for more information about where the problem occurred.

**FFQD3015E**

An error occurred while handling crawled history. URI: *uniform\_resource\_identifier*. Server: *server\_name*. Target: *crawler\_name*.

**Explanation:**

The internal error occurred in the crawler.

**User Response:**

See the log files for more information about where the problem occurred.

**FFQD3016E**

An error occurred while accessing the database.

**Explanation:**

The internal error occurred in the crawler.

**User Response:**

See the log files for more information about where the problem occurred.

**FFQD3017E**

An error occurred while accessing the database. URI: *uniform\_resource\_identifier*. Server: *server\_name*. Target: *crawler\_name*.

**Explanation:**

The internal error occurred in the crawler.

**User Response:**

See the log files for more information about where the problem occurred.

**FFQD3018E**

an error occurred while handling the crawler configuration.

**Explanation:**

The internal error occurred in the crawler.

**User Response:**

See the log files for more information about where the problem occurred.

**FFQD3019I**

The server schedule is set for *server\_name* server.

**Explanation:****User Response:**

This message is for your information only. No action is required.



## FFQD3020I

The schedule is disabled for *server\_name* server.

**Explanation:**

**User Response:**

This message is for your information only. No action is required.

## FFQD3021W

The schedule for specified server *server\_name* is already set.

**Explanation:**

The specified scheduled crawl has already been enabled.

**User Response:**

This message is for your information only. No action is required.

## FFQD3022W

The schedule for specified server *server\_name* is already disabled.

**Explanation:**

The specified scheduled crawl has already been disabled.

**User Response:**

This message is for your information only. No action is required.

## FFQD3023W

The server *server\_name* is not scheduled.

**Explanation:**

The specified crawl was not scheduled.

**User Response:**

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*

**Explanation:**

The internal error occurred in the crawler.

**User Response:**

See the log files for more information about where the problem occurred.

## FFQD3025E

An error occurred while fetching data.

**Explanation:**

The internal error occurred in the crawler.

**User Response:**

See the log files for more information about where the problem occurred.

## FFQD3026E

An error occurred while processing thread *thread\_id*.

**Explanation:**

The internal error occurred in the crawler.

**User Response:**

See the log files for more information about where the problem occurred.

## FFQD3027E

An error occurred for the crawler.

**Explanation:**

The internal error occurred in the crawler.

**User Response:**

See the log files for more information about where the problem occurred.

## FFQD3028E

The specified server *server\_name* is not found.

**Explanation:**

The internal error occurred in the crawler.

**User Response:**

See the log files for more information about where the problem occurred.

## FFQD3029E

an error occurred while initializing the connection manager.

**Explanation:**

The internal error occurred during initialization of the crawler.

**User Response:**

See the log files for more information about where the problem occurred.

## FFQD3030E

An error occurred while initializing the state manager.

**Explanation:**

The internal error occurred during initialization of the crawler.

**User Response:**

See the log files for more information about where the problem occurred.

## FFQD3031E

An error occurred while crawler status was being requested.

**Explanation:**

The internal error occurred in the crawler. The monitor function did not work correctly.

**User Response:**

See the log files for more information about where the problem occurred.

## FFQD3032E

An error occurred while the status of the crawl space was being requested.

**Explanation:**

The internal error occurred in the crawler. The monitor function did not work correctly.

**User Response:**

See the log files for more information about where the problem occurred.

## FFQD3033E

An error occurred while the crawl space status was being requested.

**Explanation:**

The internal error occurred in the crawler. The monitor function did not work correctly.

**User Response:**

See the log files for more information about where the problem occurred.

## FFQD3034I

Deleting obsolete documents.

**Explanation:**

**User Response:**

This message is for your information only. No action is required.

## FFQD3035W

Interrupted to delete obsolete documents. Will try again next time.

**Explanation:**

**User Response:**

This message is for your information only. No action is required.

## FFQD3036I

Finished deleting obsolete documents for *crawler\_name* on *server\_name*.

**Explanation:**

**User Response:**

This message is for your information only. No action is required.

## FFQD3037E

An error occurred while updating internal status.

**Explanation:**

The internal error occurred in the crawler.

**User Response:**

See the log files for more information about where the problem occurred.

## FFQD3038E

An error occurred because of inconsistency of status. Information:  
*additional\_data*

**Explanation:**

The internal error occurred in the crawler.

**User Response:**

See the log files for more information about where the problem occurred.

## FFQD3039E

An error occurred while accessing internal status. Information:  
*additional\_data*

**Explanation:**

The internal error occurred in the crawler.

**User Response:**

See the log files for more information about where the problem occurred.

## FFQD3040E

An error occurred while deleting obsolete records.

**Explanation:**

The internal error occurred in the crawler.

**User Response:**

See the log files for more information about where the problem occurred.

## FFQD3041I

The schedule component started successfully.

**Explanation:**

The internal error occurred in the crawler.

**User Response:**

See the log files for more information about where the problem occurred.

## FFQD3042E

The schedule component failed to start.

**Explanation:**

The internal error occurred in the crawler.

**User Response:**

See the log files for more information about where the problem occurred.

## FFQD3043I

The schedule component stopped successfully.

**Explanation:**

The internal error occurred in the crawler.

**User Response:**

See the log files for more information about where the problem occurred.

**FFQD3044E**

The schedule component failed to stop.

**Explanation:**

The internal error occurred in the crawler.

**User Response:**

See the log files for more information about where the problem occurred.

**FFQD3045I**

Starting the scheduled crawl for *crawler\_name*.

**Explanation:**

**User Response:**

This message is for your information only. No action is required.

**FFQD3046E**

An error occurred in the schedule component.

**Explanation:**

The internal error occurred in the crawler.

**User Response:**

See the log files for more information about where the problem occurred.

**FFQD3047I**

The internal repository was initialized successfully.

**Explanation:**

A crawler initialized the crawler internal repository.

**User Response:**

This message is for your information only. No action is required.

**FFQD3048E**

An unknown internal repository type *internal\_repository\_type* was detected.

**Explanation:**

One or more errors occurred while initializing the crawler internal repository.

**User Response:**

Verify that enterprise search was successfully installed and try again.

**FFQD3049E**

An internal repository initialization error.

**Explanation:**

One or more errors occurred while initializing the crawler internal repository.

**User Response:**

Verify that enterprise search was successfully installed and try again.

**FFQD3050E**

Access violation.

**Explanation:**

One or more errors occurred while initializing the crawler internal repository.

**User Response:**

Verify that enterprise search was successfully installed and try again.

**FFQD3051E**

Another repository is open.

**Explanation:**

An another crawler internal repository was already opened and initialized.

**User Response:**

Verify that enterprise search was successfully installed and try again.

**FFQD3052E**

There is no database connection.

**Explanation:**

The crawler internal repository was opened but a database connection for using the repository did not exist.

**User Response:**

Verify that the DB2 database for enterprise search is correctly working and try again.

**FFQD3053E**

A database error occurred while opening the internal repository.

**Explanation:**

A database connection for using the crawler internal repository could not be established.

**User Response:**

Verify that the DB2 database for enterprise search is correctly working and try again.

**FFQD3054E**

A database error occurred while closing the internal repository.

**Explanation:**

One or more errors occurred while closing the crawler internal repository.

**User Response:**

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*.

**Explanation:**

One or more errors occurred while writing information to the crawler internal repository.

**User Response:**

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*.

**Explanation:**

One or more errors occurred while accessing to the crawler internal repository.

**User Response:**

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*.

**Explanation:**

One or more errors occurred while writing information to the crawler internal repository.

**User Response:**

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*.

**Explanation:**

One or more errors occurred while accessing to the crawler internal repository.

**User Response:**

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*.

**Explanation:**

One or more errors occurred while accessing to the crawler internal repository.

**User Response:**

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*.

**Explanation:**

One or more errors occurred while accessing to the crawler internal repository.

**User Response:**

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*.

**Explanation:**

One or more errors occurred while accessing to the crawler internal repository.

**User Response:**

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*.

**Explanation:**

One or more errors occurred while accessing to the crawler internal repository.

**User Response:**

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*.

**Explanation:**

One or more errors occurred while accessing to the crawler internal repository.

**User Response:**

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*.

**Explanation:**

One or more errors occurred while accessing to the crawler internal repository.



**User Response:**  
See the SQLException messages in the log files for more information.

### **FFQD3065I**

The internal repository was opened successfully.

**Explanation:**  
A crawler opened the crawler internal repository.

**User Response:**  
This message is for your information only. No action is required.

### **FFQD3066I**

The internal repository was closed successfully.

**Explanation:**  
A crawler closed the crawler internal repository.

**User Response:**  
This message is for your information only. No action is required.

### **FFQD3067I**

Loading crawler configuration files *general\_config\_file* and *crawler\_config\_file*.

**Explanation:**  
A crawler started loading the crawler configuration files.

**User Response:**  
This message is for your information only. No action is required.

### **FFQD3068I**

Loading of the configuration files completed successfully.

**Explanation:**  
A crawler finished loading the crawler configuration files.

**User Response:**  
This message is for your information only. No action is required.

### **FFQD3069E**

An error occurred while creating the configuration factory.

**Explanation:**  
A crawler finished loading the crawler configuration files.

**User Response:**  
This message is for your information only. No action is required.

### **FFQD3070E**

Loading of configuration file *general\_config\_file* failed.

**Explanation:**

One or more errors occurred while loading the crawler configuration file.

**User Response:**

See the error messages in the log files for more information. typically the error is the IOException.

**FFQD3071E**

No configuration file was loaded.

**Explanation:**

A crawler did not load configuration files.

**User Response:**

Verify that the crawler which you want to start was configured correctly and try again.

**FFQD3072E**

An unknown configuration object was detected.

**Explanation:**

One or more errors occurred while loading the crawler configuration files.

**User Response:**

Verify that enterprise search was successfully installed and try again.

**FFQD3073E**

Getting document builder failed.

**Explanation:**

Could not create a XML parser for loading the crawler configuration files.

**User Response:**

Verify that enterprise search was successfully installed and try again.

**FFQD3074E**

A configuration file format error was detected.

**Explanation:**

A crawler configuration file is not valid a XML document.

**User Response:**

See the log files for more information.

**FFQD3075E**

An I/O error occurred while parsing the configuration.

**Explanation:**

One or more errors occurred while loading the crawler configuration files.

**User Response:**

See the IOException messages in the log files for more information.

## FFQD3076E

The crawler configuration file *crawler\_config\_file* was not found.

**Explanation:**

There was no crawler configuration file on the expected path.

**User Response:**

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.

**Explanation:**

The root element of the crawler configuration file had unexpected name.

**User Response:**

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.

**Explanation:**

A required attribute was not found in the crawler configuration file.

**User Response:**

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*.

**Explanation:**

The same target name was contained in the crawler configuration file.

**User Response:**

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.

**Explanation:**

The crawler configuration file contained an incorrect date string.

**User Response:**

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.

**Explanation:**

The crawler configuration file contained an incorrect crawl interval time.

**User Response:**

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.

**Explanation:**

The crawler configuration file contained one or more incorrect values.

**User Response:**

Verify that the crawler which you want to run was configured correctly and try again.

## FFQD3083W

The XML parser reported a warning.

**Explanation:**

The crawler configuration file parser reported a XML warning.

**User Response:**

No action is required.

## FFQD3084E

The XML parser reported an error.

**Explanation:**

The crawler configuration file parser reported a XML error.

**User Response:**

Verify that the crawler which you want to run was configured correctly and try again.

## FFQD3085E

The XML parser reported an unrecoverable error.

**Explanation:**

The crawler configuration file parser reported an unrecoverable XML error.

**User Response:**

Verify that the crawler that you want to run was configured correctly and try again.

## FFQD3086E

The formatter is not initialized.

**Explanation:**

The formatter to generate XML metadata for search engine was not created.

**User Response:**

See the log files for more information about where the problem occurred.

**FFQD3087E**

Cannot convert to XML.

**Explanation:**

The formatter does not generate XML metadata for search engine..

**User Response:**

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.

**Explanation:**

Crawled field is not a number though crawler is specified this field as parametric searchable. The parametric search will be disabled for this field.

**User Response:**

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.

**Explanation:**

Metadata was truncated due to the limitation of the max document size setting. It is not searchable by truncated metadata.

**User Response:**

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.

**Explanation:**

Temporary file *filename* is not deleted.

**User Response:**

Check that temporary directory for the crawler has appropriate privilege to delete.

**FFQD3091W**

The formatted data record is null. No action is required.

**Explanation:**

No metadata or binary content fetched by crawler will write to enterprise search repository.

**User Response:**

No action is required.

**FFQD3102E**

Cannot open internal data store. DBName: *database\_name*. DBSchema: *schema\_name*. DBUser: *database\_user*. TableName: *table\_name*.

**Explanation:**

Publisher cannot open internal database *database\_name* to store crawled data.

**User Response:**

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*.

**Explanation:**

Publisher cannot close internal database *database\_name* to store crawled data.

**User Response:**

See the log files for more information about where the problem occurred.

**FFQD3104E**

Cannot write metadata to file *file\_name* because of `FileNotFoundException`.

**Explanation:**

Publisher cannot open the temporary file that stored crawled data.

**User Response:**

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*.

**Explanation:**

Publisher cannot write the information crawled by crawler into internal database *database\_name*.

**User Response:**

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*.

**Explanation:**

Publisher cannot write the information crawled by crawler into internal database *database\_name*.

**User Response:**

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*.

**Explanation:**

Publisher cannot mark deletion flag to internal table to delete information from search index. Information about nonexistent document in target data source will remain in search index.

**User Response:**

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*.

**Explanation:**

Publisher cannot mark deletion flag to internal table to delete information from search index. Information about nonexistent document in target data source will remain in search index.

**User Response:**

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.

**Explanation:**

Crawled file *file\_name* is modified because the file crawled by crawler.

**User Response:**

No action is required.

**FFQD3110E**

Cannot close file stream.

**Explanation:**

The file stream for crawled data stored in temporary file is not closed due to an error.

**User Response:**

See the log files for more information about where the problem occurred.

## FFQD3111E

Password decryption failed.

**Explanation:**

The password decryption module reported one or more errors.

**User Response:**

See the log files for more information about where the problem occurred.

## FFQD3112E

Security token plug-in cannot be loaded. class : *security\_token\_class\_name*

**Explanation:**

Specified class name for security token plug-in cannot be loaded.

**User Response:**

Ensure that security token plugin module is installed properly and specified class name is correct. See manual and release note.

## FFQD3113E

SecurityTokensGeneratorException was thrown by Security tokens plug-in.

**Explanation:**

SecurityTokensGeneratorException was thrown by security tokens plug-in. Crawler will be stopped.

**User Response:**

No action need be taken.

## FFQD3114E

Unexpected exception was thrown by Security tokens plug-in.

**Explanation:**

Unexpected exception was thrown by security tokens plug-in. Crawler will be stopped.

**User Response:**

Security token plug-in module should throw only SecurityTokensGeneratorException to stop crawler. Ensure that security tokens plug-in does not throw unexpected exception.

## FFQD3115E

Temporary file *filename* is not deleted.

**Explanation:**

Temporary file *filename* is not deleted.

**User Response:**

Check that temporary directory for the crawler has appropriate privilege to delete. See the log files for more information about where the problem occurred.



## FFQD3116E

Cannot close file stream.

**Explanation:**

The file stream was not closed due to an error.

**User Response:**

See the log files for more information about where the problem occurred.

## FFQD3117E

MetadataGeneratorException was thrown by metadata plug-in.

**Explanation:**

MetadataGeneratorException was thrown by metadata plug-in. Crawler will be stopped.

**User Response:**

No action need be taken.

## FFQD3118E

Unexpected exception was thrown by metadata plug-in.

**Explanation:**

Unexpected exception was thrown by metadata plug-in. Crawler will be stopped.

**User Response:**

Metadata plug-in module should throw only MetadataGeneratorException to stop crawler. Ensure that metadata plug-in does not throw unexpected exception.

## FFQD3119E

The *metada plug-in configuration file* file was not valid.

**Explanation:**

The configuration file for metadata plug-in were not valid.

**User Response:**

Enter the correct properties in the configuration file for metadata plug-in.

## FFQD3120E

An error occurred while handling the configuration for metadata plug-in.

**Explanation:**

The metadata plug-in function did not work. Crawler will be stopped.

**User Response:**

See the log files for more information about where the problem occurred.

## FFQD3121W

The CLASSPATH was not valid.

**Explanation:**

The CLASSPATH property defined in configuration file was not valid.

**User Response:**

Enter the correct CLASSPATH.

## FFQD3122E

The user-defined class for metadata plug-in function cannot be created.

**Explanation:**

The user-defined class derived from MetadataGenerator class cannot be created.

**User Response:**

Verify that the class name and the CLASSPATH were set correctly.

## FFQD3123E

The *parameter\_name* parameter was missing or not valid.

**Explanation:**

A crawler configuration file was not valid due to *parameter\_name* parameter.

**User Response:**

See the log files for more information.

## FFQD3124E

The *class\_name* class could not be loaded correctly.

**Explanation:**

The specified class *class\_name* could not be loaded correctly.

**User Response:**

Verify that WebSphere Information Integrator OmniFind Edition (enterprise search) was installed correctly.

## FFQD3125E

Cannot open formatter of crawler publisher.

**Explanation:**

Crawler cannot initialize XML formatter.

**User Response:**

See the log files for more information.

## FFQD3126E

Cannot close formatter of crawler publisher.

**Explanation:**

Crawler cannot terminate XML formatter.

**User Response:**

See the log files for more information.

## FFQD3127I

Crawler plug-in was initialized successfully.

**Explanation:**

Crawler formatter initialized crawler plug-in successfully.

**User Response:**

No action is required.

## FFQD3128E

CrawlerPluginSocketException was thrown by crawler plug-in.

**Explanation:**

CrawlerPluginSocketException was thrown by crawler plug-in client. Crawler will be stopped.

**User Response:**

Ensure that the crawler plug-in and its session works correctly.

## FFQD3129E

CrawlerPluginException was thrown by crawler plug-in.

**Explanation:**

CrawlerPluginException was thrown by crawler plug-in. Crawler will be stopped.

**User Response:**

No action need be taken.

## FFQD3130E

Unexpected exception was thrown by crawler plug-in.

**Explanation:**

Unexpected exception was thrown by crawler plug-in client. Crawler will be stopped.

**User Response:**

Ensure that the crawler plug-in and its session works correctly.

## FFQD3131W

This document will not be indexed because crawler plug-in returned null metadata list.

**Explanation:**

This document will not be indexed because crawler plug-in returns null metadata list.

**User Response:**

Crawler plug-in should return null object as a result of the function updateDocument() in order not to index current document. Confirm the specification of crawler plug-in.

## FFQD3132W

This document will have empty security tokens because crawler plug-in returned empty string as new security tokens.

**Explanation:**

This document will have empty security tokens because crawler plug-in returned empty string as new security tokens.

**User Response:**

Crawler plug-in should return some string for new security tokens. Confirm the specification of crawler plug-in.

## FFQD3133E

Crawler plug-in error: *message*

**Explanation:**

There is an crawler plug-in error.

**User Response:**

Refer to the attached message and see the log files for more information about where the problem occurred.

## FFQD3134W

Crawler plug-in warning: *message*

**Explanation:**

There is an crawler plug-in warning.

**User Response:**

Refer to the attached message and see the log files for more information about where the problem occurred.

## FFQD3135E

The crawler has already started.

**Explanation:**

The specified crawler has already started.

**User Response:**

Refresh the information on Administrator page.

## FFQD3136E

The crawler has already stopped.

**Explanation:**

The specified crawler has already stopped.

**User Response:**

Refresh the information on Administrator page.

## FFQD3137E

The crawler status could not be shown on the monitor.

**Explanation:**

The crawler status was not reported correctly by the crawler.

**User Response:**

See the log files for more information about where the problem occurred.

**FFQD3138E**

A timeout error occurred sending a request to the crawler.

**Explanation:**

The crawler process might fail to be stopped successfully.

**User Response:**

See the log files for more information about where the problem occurred.

**FFQD3500E**

JDBC class *jdbc\_driver\_class* is not found.

**Explanation:**

JDBC driver class *jdbc\_driver\_class* was not found in the CLASSPATH.

**User Response:**

Verify that enterprise search was successfully installed and try again.

**FFQD3501W**

Exceptions occurred while altering *table\_name* table.

**Explanation:**

One or more exceptions occurred while trying to alter *table\_name* table.

**User Response:**

See the exceptions (typically SQLException from DB2) to determine the actual cause of this error.

**FFQD3502W**

Exceptions occurred while creating *table\_name* table.

**Explanation:**

One or more exceptions occurred while trying to create *table\_name* table.

**User Response:**

See the exceptions (typically SQLException from DB2) to determine the actual cause of this error.

**FFQD3503W**

Transaction rollback failed.

**Explanation:**

One or more exceptions occurred while trying to rollback the transaction.

**User Response:**

See the exceptions (typically SQLException from DB2) to determine the actual cause of this error.

## FFQD3504W

Transaction commit failed.

**Explanation:**

One or more exceptions occurred while trying to commit the transaction.

**User Response:**

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.

**Explanation:**

One or more exceptions occurred while 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.

**User Response:**

No action is required.

## FFQD3506W

Exceptions occurred while deleting all data from *table\_name* table.

**Explanation:**

One or more exceptions occurred while trying to delete all the data from *table\_name* table.

**User Response:**

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.

**Explanation:**

One or more exceptions occurred while trying to insert data into *table\_name* table.

**User Response:**

See the exceptions (typically SQLException from DB2) to determine the actual cause of this error.

## FFQD3508W

Exceptions occurred while dropping *table\_name* table.

**Explanation:**

One or more exceptions occurred while trying to drop *table\_name* table.

**User Response:**

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.

**Explanation:**

The crawled document was not stored into internal table because of too long URI *URI*.

**User Response:**

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.

**Explanation:**

The crawled document was not stored into internal table because of too large document.

**User Response:**

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.

**Explanation:**

The crawled document was not stored into an internal table because the metadata was too extensive.

**User Response:**

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*.

**Explanation:**

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.

**User Response:**

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*.

**Explanation:**

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.

**User Response:**

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*.

**Explanation:**

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.

**User Response:**

If this message appears again, verify that enterprise search was successfully installed and try again.

## FFQD3515W

Exceptions occurred while querying *table\_name* table.

**Explanation:**

One or more exceptions occurred while trying to execute select statement against *table\_name* table.

**User Response:**

See the exceptions (typically SQLException from DB2) to determine the actual cause of this error.

## FFQD3516W

Exceptions occurred while inserting into *table\_name* table.

**Explanation:**

One or more exceptions occurred while trying to executing insert statement against *table\_name* table.

**User Response:**

See the exceptions (typically SQLException from DB2) to determine the actual cause of this error.

## FFQD3517W

Exceptions occurred while updating *table\_name* table.

**Explanation:**

One or more exceptions occurred while trying to execute update statement against *table\_name* table.

**User Response:**

See the exceptions (typically SQLException from DB2) to determine the actual cause of this error.

## FFQD3518W

Exceptions occurred while deleting from *table\_name* table.



**Explanation:**

One or more exceptions occurred while trying to execute delete statement against *table\_name* table.

**User Response:**

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*.

**Explanation:**

One or more exceptions occurred while preparing the select statement for the *table\_name* table.

**User Response:**

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*.

**Explanation:**

One or more exceptions occurred while preparing insert statement against *table\_name* table.

**User Response:**

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*.

**Explanation:**

One or more exceptions occurred while preparing update statement against *table\_name* table.

**User Response:**

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*.

**Explanation:**

One or more exceptions occurred while preparing delete statement for *table\_name* table.

**User Response:**

See the exceptions (typically SQLException from DB2) to determine the actual cause of this error.

## FFQD3523W

The connection manager was already initialized.

**Explanation:**

One of the internal manager which managing database connection has already been initialized.

**User Response:**

No action is required.

## FFQD3524E

JNDI look-up failed for *jndi\_name*.

**Explanation:**

JNDI lookup was failed for *jndi\_name*.

**User Response:**

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*.

**Explanation:**

One or more exceptions occurred while changing the internal state form *from\_state* to *to\_state*.

**User Response:**

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.

**Explanation:**

Failed to get file lock for crawler that uses file lock facility.

**User Response:**

See the log files for more information about where the problem occurred.

## FFQD3541E

Failed to get file lock *file\_name* because of FileNotFoundException.

**Explanation:**

Failed to get file lock for crawler that uses file lock facility.

**User Response:**

See the log files for more information about where the problem occurred.

## FFQD3543E

Failed to release file lock *file\_name*.

**Explanation:**

Failed to release file lock for crawler that uses file lock facility.

**User Response:**

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.

**Explanation:**

Failed to release file lock for crawler because no lock was obtained.

**User Response:**

See the log files for more information about where the problem occurred.

**FFQD3545E**

Algorithm *algorithm\_name* for loading truststore file was not found.

**Explanation:**

The algorithm for loading truststore file was not contained in the JRE.

**User Response:**

Verify that DB2 Information Integrator OmniFind Edition (enterprise search) was successfully installed and try again.

**FFQD3546E**

No certificate was loaded.

**Explanation:**

The specified truststore file did not contain a suitable certificate.

**User Response:**

Verify that the truststore file contains a valid certificate and try again.

**FFQD3547E**

Unknown keystore type *keystore\_type* is detected.

**Explanation:**

An unexpected truststore file type was detected.

**User Response:**

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*.

**Explanation:**

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.

**User Response:**

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*.

**Explanation:**

One or more I/O errors occurred while closing the truststore file.

**User Response:**

Verify that the truststore file can be accessed and try again.

**FFQD3550E**

Failed to initialize javax.net.ssl.TrustManager.

**Explanation:**

One or more errors occurred while initializing the javax.net.ssl.TrustManager object.

**User Response:**

See the error messages in the log files for more information.

**FFQD3551E**

Failed to initialize javax.net.ssl.SSLContext.

**Explanation:**

One or more errors occurred while initializing the javax.net.ssl.SSLContext object.

**User Response:**

See the error messages in the log files for more information.

**FFQD3553E**

Exceptions occurred while creating connection to the database *database\_name*.

**Explanation:**

One or more exceptions occurred while creating connection to the database *database\_name*.

**User Response:**

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.

**Explanation:**

Necessary Notes Java class cannot be loaded. Setup procedure is required for Notes crawler.

**User Response:**

Perform setup procedure to work Notes crawler.

## FFQD4001E

Notes NRPC class (Notes.jar) is not found. Setup for Notes NRPC crawler is required.

**Explanation:**

Necessary Notes Java class for NRPC (Notes Remote Procedure Call) crawl cannot be loaded. Setup procedure is required for Notes crawler.

**User Response:**

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.

**Explanation:**

Necessary Notes library for NRPC( Notes Remote Procedure Call) crawl cannot be loaded. Setup procedure is required for Notes crawler.

**User Response:**

Perform setup procedure to work Notes crawler.

## FFQD4003E

Notes DIIOP class (NCS0.jar) is not found. Setup for Notes DIIOP crawler is required.

**Explanation:**

Necessary Notes Java class for DIIOP (Domino Internet Inter-ORB Protocol) crawl cannot be loaded. Setup procedure is required for Notes crawler.

**User Response:**

Perform setup procedure to work Notes crawler.

## FFQD4004E

Cannot determine Notes library name from detected platform *platform*.

**Explanation:**

Necessary Notes library name could not be determined for NRPC (Notes Remote Procedure Call) crawl.

**User Response:**

Ensure that the Notes crawler runs on the supported platform.

## FFQD4005E

Information for server, target, or both were not specified.

**Explanation:**

Domino server or target was not set properly.

**User Response:**

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.

**Explanation:**

Server string *server\_string* does not conform the format like:  
"server\_name!!database\_name."

**User Response:**

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.

**Explanation:**

Server string *server\_string* does not contain database name.

**User Response:**

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.

**Explanation:**

The connection to the server *server\_name* is already established and the crawler tried to connect to the server without disconnecting.

**User Response:**

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*.

**Explanation:**

the Notes crawler cannot connect to the server because of the invalid crawler setting or server setting.

**User Response:**

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*.

**Explanation:**

The Notes crawler cannot change the ID to *id\_file\_name*.

**User Response:**

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*.

**Explanation:**

The Domino server or the database could not disconnect.

**User Response:**

See the log files for more information about where the problem occurred.

## **FFQD4012E**

Cannot change the ID to *id\_file\_name*.

**Explanation:**

The Notes crawler cannot change the ID to *id\_file\_name*.

**User Response:**

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*.

**Explanation:**

The Notes crawler cannot verify whether the Notes document exists while the crawler fetches documents from the specified Notes database.

**User Response:**

See the log files for more information about where the problem occurred.

## **FFQD4014E**

Cannot change the ID to *id\_file\_name*.

**Explanation:**

Notes crawler cannot change the ID to *id\_file\_name*.

**User Response:**

See the log files for more information about where the problem occurred.

## **FFQD4015E**

Cannot create the URI.

**Explanation:**

An error occurred while creating a uniform resource identifier (URI).

**User Response:**

See the log files for more information about where the problem occurred.

## **FFQD4016E**

Cannot gather the unchanged attachment names since the last crawl.

**Explanation:**

An error occurred while gathering file names from unchanged documents in the Notes database.

**User Response:**

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*.

**Explanation:**

An error occurred while fetching the target content from database *database\_name* on the server *server\_name*.

**User Response:**

See the log files for more information about where the problem occurred.

**FFQD4018E**

Undefined record status *status\_code* was found in the internal repository.

**Explanation:**

An unexpected status code *status\_code* was returned from an enterprise search internal table.

**User Response:**

See the log files for more information about where the problem occurred.

**FFQD4019E**

Cannot obtain the attachment file name.

**Explanation:**

An error occurred while obtaining the attachment file name.

**User Response:**

See the log files for more information about where the problem occurred.

**FFQD4020E**

Cannot write to the temporary file *filename*.

**Explanation:**

An error occurred while writing content to the temporary file.

**User Response:**

Check that the appropriate privilege is given to temporary directory where the temporary file will be generated.

**FFQD4021E**

Cannot fetch the attachment file.

**Explanation:**

An error occurred while fetching attachment file from Notes document.



**User Response:**

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.

**Explanation:**

An I/O error occurred while reading or writing binary content.

**User Response:**

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*.

**Explanation:**

An unexpected status code *status\_code* was returned from enterprise search internal table.

**User Response:**

See the log files for more information about where the problem occurred.

## **FFQD4024E**

Cannot evaluate the string length because of an I/O error.

**Explanation:**

Notes crawler cannot evaluate string length because of an I/O error.

**User Response:**

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.

**Explanation:**

Fetches Date/Time field does not have a valid value. 0 is used as this field value.

**User Response:**

No action is required.

## **FFQD4026W**

Cannot obtain the creation date from the Lotus Notes document. 0 will be used instead.

**Explanation:**

Fetches Date/Time field does not have a valid value. 0 is used as this field value.

**User Response:**

No action is required.

## FFQD4027W

Cannot obtain the security field value.

**Explanation:**

Specified field as security value does not exist or does not have any value.

**User Response:**

No action is required.

## FFQD4028W

Cannot obtain file name for the data source metadata.

**Explanation:**

Specified field as security value does not exist or does not have any value.

**User Response:**

No action is required.

## FFQD4029W

Cannot obtain last modified date from the Lotus Notes document. 0 will be used instead.

**Explanation:**

Fetch Date/Time field does not have a valid value. 0 is used as this field value.

**User Response:**

No action is required.

## FFQD4030W

Cannot obtain creation date from the Lotus Notes document. 0 will be used instead.

**Explanation:**

Fetch Date/Time field does not have a valid value. 0 is used as this field value.

**User Response:**

No action is required.

## FFQD4031E

The connection is already active for server *server\_name* using user *user\_name*.

**Explanation:**

The connection is already active for the crawler.

**User Response:**

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*.

**Explanation:**

Server refuses the access from crawler user.

**User Response:**

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*.

**Explanation:**

Server refuses accessing crawler due to invalid user name or password.

**User Response:**

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*.

**Explanation:**

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.

**User Response:**

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\_name*. Check that you can access the database and that the crawl target is available to user *user\_name*.

**Explanation:**

The Notes crawler cannot access the Domino server and its database due to the invalid setting of the crawler or Domino server.

**User Response:**

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\_name*. Check that you can access the database and that the crawl target is available to user *user\_name*.

**Explanation:**

**User Response:**

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*.

**Explanation:**

The specified formula is not valid as a Notes formula. Or the Domino server or database was denied access to the crawler.

**User Response:**

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*.

**Explanation:**

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.

**User Response:**

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*.

**Explanation:**

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.

**User Response:**

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*.

**Explanation:**

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.

**User Response:**

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.

**Explanation:**

Notes thread is not started due to an error.

**User Response:**

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*.

**Explanation:**

The Notes crawler could not extract the Notes user ID from the server session.

**User Response:**

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*.

**Explanation:**

The connection is already active for the crawler.

**User Response:**

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*.

**Explanation:**

Notes crawler cannot have access to the Domino server and its database due to the invalid setting of the crawler or Domino server.

**User Response:**

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*.

### Explanation:

Notes crawler cannot have access to the Domino server and its database due to the invalid setting of the crawler or Domino server.

### User Response:

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*.

### Explanation:

Specified formula is not valid as Notes Formula. Or Domino server or database is prohibited to have access from crawler.

### User Response:

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*.

### Explanation:

Specified target (Notes View or Folder) is not valid or appropriate privilege is not given to the target to read document from crawler.

### User Response:

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*.

**Explanation:**

Specified target (Notes View or Folder) is not valid or appropriate privilege is not given to the target to read document from crawler.

**User Response:**

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*.

**Explanation:**

Specified target (Notes View or Folder) is not valid or appropriate privilege is not given to the target to read document from crawler.

**User Response:**

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.

**Explanation:**

An error occurred while fetching document from specified target. specified target is null.

**User Response:**

See the log files for more information about where the problem occurred.

## FFQD4051E

Cannot fetch Document from View.

**Explanation:**

An error occurred while fetching document from specified target.

**User Response:**

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.

**Explanation:**

An error occurred while fetching document from specified target.

**User Response:**

See the log files for more information about where the problem occurred.

### **FFQD4053E**

Cannot fetch Document from DocumentCollection.

**Explanation:**

An error occurred while fetching document from specified target.

**User Response:**

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.

**Explanation:**

An error occurred while fetching document from specified target.

**User Response:**

See the log files for more information about where the problem occurred.

### **FFQD4055E**

Current Document is null.

**Explanation:**

An error occurred while getting created date of the Notes document.

**User Response:**

See the log files for more information about where the problem occurred.

### **FFQD4056E**

Cannot obtain created date from current Document.

**Explanation:**

An error occurred while getting created date of the Notes document.

**User Response:**

See the log files for more information about where the problem occurred.

### **FFQD4057E**

Current document is null.

**Explanation:**

An error occurred while getting last modified date of the Notes document.

**User Response:**

See the log files for more information about where the problem occurred.

### **FFQD4058E**

Cannot obtain last modified date from current document.



**Explanation:**

An error occurred while getting last modified date of the Notes document.

**User Response:**

See the log files for more information about where the problem occurred.

## **FFQD4059E**

Cannot obtain current document size.

**Explanation:**

An error occurred while getting the size of the Notes document.

**User Response:**

See the log files for more information about where the problem occurred.

## **FFQD4060E**

Cannot fetch field from current document.

**Explanation:**

An error occurred while getting document field values of the Notes document.

**User Response:**

See the log files for more information about where the problem occurred.

## **FFQD4061E**

Cannot recycle Item.

**Explanation:**

An error occurred while releasing memory of the Notes object.

**User Response:**

See the log files for more information about where the problem occurred.

## **FFQD4062E**

Cannot fetch field from current document.

**Explanation:**

An error occurred while getting attachment attribute from the Notes document.

**User Response:**

See the log files for more information about where the problem occurred.

## **FFQD4063E**

Cannot recycle item.

**Explanation:**

An error occurred while releasing memory of the Notes object.

**User Response:**

See the log files for more information about where the problem occurred.

## FFQD4064E

Current EmbeddedObject is null.

**Explanation:**

An error occurred while getting input stream from Notes document attachment file.

**User Response:**

See the log files for more information about where the problem occurred.

## FFQD4065E

Cannot get InputStream from current EmbeddedObject.

**Explanation:**

An error occurred while getting input stream from Notes document attachment file.

**User Response:**

See the log files for more information about where the problem occurred.

## FFQD4066E

Current EmbeddedObject is null.

**Explanation:**

An error occurred while getting attachment file name from Notes document.

**User Response:**

See the log files for more information about where the problem occurred.

## FFQD4067E

Cannot get file name from current EmbeddedObject.

**Explanation:**

An error occurred while getting attachment file name from Notes document.

**User Response:**

See the log files for more information about where the problem occurred.

## FFQD4068E

Current EmbeddedObject is null.

**Explanation:**

An error occurred while getting attachment file size from Notes document.

**User Response:**

See the log files for more information about where the problem occurred.

## FFQD4069E

Cannot get the file size from current EmbeddedObject.

**Explanation:**

An error occurred while getting attachment file size from Notes document.

**User Response:**

See the log files for more information about where the problem occurred.

## **FFQD4070E**

Current document is null.

**Explanation:**

An error occurred while getting form name defined in the Notes document.

**User Response:**

See the log files for more information about where the problem occurred.

## **FFQD4071E**

Cannot get the form name from current document.

**Explanation:**

An error occurred while getting form name defined in the Notes document.

**User Response:**

See the log files for more information about where the problem occurred.

## **FFQD4072E**

Current document is null.

**Explanation:**

An error occurred while getting Universal Notes ID (UNID) defined in the Notes document.

**User Response:**

See the log files for more information about where the problem occurred.

## **FFQD4073E**

Cannot get document universal Notes ID.

**Explanation:**

An error occurred while getting Universal Notes ID (UNID) defined in the Notes document.

**User Response:**

See the log files for more information about where the problem occurred.

## **FFQD4074E**

Current view or DocumentCollection is null.

**Explanation:**

An error occurred while getting Universal Notes ID (UNID) defined in the Notes view or folder.

**User Response:**

See the log files for more information about where the problem occurred.

## FFQD4075E

Cannot get the universal ID for the View.

**Explanation:**

An error occurred while getting Universal Notes ID (UNID) defined in the Notes view or folder.

**User Response:**

See the log files for more information about where the problem occurred.

## FFQD4076E

Current database is null.

**Explanation:**

An error occurred while getting Notes replica ID defined in the Notes database.

**User Response:**

See the log files for more information about where the problem occurred.

## FFQD4077E

Cannot get database replica ID.

**Explanation:**

An error occurred while getting Notes replica ID defined in the Notes database.

**User Response:**

See the log files for more information about where the problem occurred.

## FFQD4078E

Cannot gather items to be crawled.

**Explanation:**

An error occurred while getting field value(s) stored in the Notes document.

**User Response:**

See the log files for more information about where the problem occurred.

## FFQD4079E

ID file *id\_filename* does not exist.

**Explanation:**

Specified ID file does not exist.

**User Response:**

Specify a valid Notes ID file from administration console.

## FFQD4080E

ID authorization failed for the ID file *id\_filename*.

**Explanation:**

Specified password is not valid for given Notes user ID file.

**User Response:**

Specify valid password for given Notes ID file from administration console.

## **FFQD4081E**

Cannot create Notes session.

**Explanation:**

An error occurred while creating Notes session.

**User Response:**

See the log files for more information about where the problem occurred.

## **FFQD4082E**

Cannot recycle registration.

**Explanation:**

An error occurred while releasing memory of the Notes object.

**User Response:**

See the log files for more information about where the problem occurred.

## **FFQD4083E**

Cannot change ID to *id\_filename*.

**Explanation:**

Notes crawler cannot switch to specified ID.

**User Response:**

See the log files for more information about where the problem occurred.

## **FFQD4084E**

This session is on DIIOP.

**Explanation:**

Notes crawler tried to change ID though the crawler is for DIIOP crawling.

**User Response:**

See the log files for more information about where the problem occurred.

## **FFQD4085E**

Cannot recycle session.

**Explanation:**

An error occurred while releasing memory of the Notes object.

**User Response:**

See the log files for more information about where the problem occurred.

## FFQD4086E

Cannot recycle current document.

**Explanation:**

An error occurred while releasing memory of the Notes object.

**User Response:**

See the log files for more information about where the problem occurred.

## FFQD4087E

Cannot recycle View/DocumentCollection.

**Explanation:**

An error occurred while releasing memory of the Notes object.

**User Response:**

See the log files for more information about where the problem occurred.

## FFQD4088E

Cannot recycle current document.

**Explanation:**

An error occurred while releasing memory of the Notes object.

**User Response:**

See the log files for more information about where the problem occurred.

## FFQD4089E

Cannot recycle current EmbeddedObject.

**Explanation:**

An error occurred while releasing memory of the Notes object.

**User Response:**

See the log files for more information about where the problem occurred.

## FFQD4090E

Cannot recycle current EmbeddedObject.

**Explanation:**

An error occurred while releasing memory of the Notes object.

**User Response:**

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*.

**Explanation:**

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.

**User Response:**

No action is required. To reduce the URI length, reduce database relative path from Domino server data directory and reduce attachment file name.

**FFQD4092E**

Communication to the server failed.

**Explanation:**

An error occurred while communicating to the server.

**User Response:**

Check the network connection to the server. See the log files for more information about where the problem occurred.

**FFQD4093E**

Cannot get Notes database ACL. Connection to the target database does not exist.

**Explanation:**

Cannot get Notes database ACL because connection to the target database does not exist.

**User Response:**

See the log files for more information about where the problem occurred.

**FFQD4094E**

An error occurred while getting Notes database ACL.

**Explanation:**

An error occurred while getting Notes database ACL.

**User Response:**

See the log files for more information about where the problem occurred.

**FFQD4095E**

Cannot get Notes database list from server. Connection to the target server does not exist.

**Explanation:**

Cannot get Notes Database list because connection to the target Domino server does not exist.

**User Response:**

See the log files for more information about where the problem occurred.

**FFQD4096E**

An error occurred while getting Notes database list.

**Explanation:**

An error occurred while getting Notes database list from Domino server.

**User Response:**

See the log files for more information about where the problem occurred.

## FFQD4097E

Cannot convert Notes hierarchical name. Notes session does not exist.

**Explanation:**

Cannot convert Notes hierarchical name because Notes session does not exist.

**User Response:**

See the log files for more information about where the problem occurred.

## FFQD4098E

An error occurred while converting Notes hierarchical name.

**Explanation:**

An error occurred while converting Notes hierarchical name.

**User Response:**

See the log files for more information about where the problem occurred.

## FFQD4099E

An error occurred while getting server ACL.

**Explanation:**

An error occurred while getting server ACL from server document on domino directory.

**User Response:**

See the log files for more information about where the problem occurred.

## FFQD4100E

An error occurred while comparing a document collection selected by a formula. Server: *server\_name*. ID:*id*. Database:*database\_name*. Formula: *formula*.

**Explanation:**

An error occurred while comparing DocumentCollection selected by a formula.

**User Response:**

See the log files for more information about where the problem occurred.

## FFQD4101W

Target database returned no results for specified formula. Crawler will not crawl any documents. Ensure that formula clause for the target is valid. Server: *server\_name*. ID:*id*. Database:*database\_name*. Formula: *formula*.

**Explanation:**

The target database returned no results for the specified formula. The crawler will not crawl any documents.

**User Response:**

See the log files for more information about where the problem occurred. Ensure that the formula clause for the target is valid.



## FFQD4102E

The server name must be registered to the Domino directory. Server name: *server\_name*

**Explanation:**

The specified server name *server\_name* was not registered to any server document in the Domino directory on the target Domino server.

**User Response:**

Specify the correct, fully qualified Internet host name. See the server document in the Domino directory (names.nsf file) on the target Domino server and ensure that server name *server\_name* is the same as the field value of fully qualified Internet host name.

## FFQD4200I

Connected to *database\_URL* using user *user\_name*.

**Explanation:**

The DB2 crawler connected to the specified database.

**User Response:**

This message is for your information only. No action is required.

## FFQD4201I

Disconnected from *database\_URL*.

**Explanation:**

The DB2 crawler disconnected from the specified database.

**User Response:**

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*.

**Explanation:**

The DB2 crawler truncated a metadata data because its size was larger than the size limit.

**User Response:**

No action is required.

## FFQD4203W

The content is ignored because it exceeds the size limit *data\_size\_limit*.

**Explanation:**

The DB2 crawler ignored content because total record size was larger than the size limit.

**User Response:**

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*.

**Explanation:**

The DB2 crawler ignored a row because null data is detected in the column specified as a unique key column.

**User Response:**

No action is required.

## FFQD4205W

A row was ignored because the URI is too long. URI: *URI*.

**Explanation:**

The DB2 crawler ignored a row because its URI is too long.

**User Response:**

No action is required.

## FFQD4207E

The fields specification is not valid.

**Explanation:**

The DB2 crawler configuration file contained incorrect column definition.

**User Response:**

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.

**Explanation:**

The DB2 crawler configuration file contained incorrect user id or password.

**User Response:**

Verify that the user name and the password is up-to-date.

## FFQD4209E

Database *database\_URL* was not found.

**Explanation:**

The DB2 crawler could not find the database.

**User Response:**

Verify that the database exists.

## FFQD4210E

Table or view *database\_URL* was not found.

**Explanation:**

The DB2 crawler could not find the table or view.

**User Response:**

Verify that the database exists.

**FFQD4211E**

Column was not found.

**Explanation:**

The DB2 crawler configuration file contained incorrect column definition.

**User Response:**

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*.

**Explanation:**

The SQL statement for selecting rows was incorrect. The DB2 crawler configuration file might contain an incorrect WHERE clause.

**User Response:**

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*.

**Explanation:**

The user did not have privileges to access the table.

**User Response:**

Grant CONNECT and SELECT privileges to the user.

**FFQD4214E**

Database connection to *database\_URL* is full.

**Explanation:**

The DB2 crawler could not establish the connection because the number of connection to the database had reached to the maximum.

**User Response:**

Increase the maximum number of connection or try again after a while.

**FFQD4215E**

Connection was terminated from the database *database\_URL*.

**Explanation:**

The database connection was terminated from the database unexpectedly.

**User Response:**

Verify the database is working correctly and try again.

## FFQD4216E

Connection to *database\_URL* failed.

**Explanation:**

One or more errors occurred while connecting to the database.

**User Response:**

See the log files for more information about where the problem occurred.

## FFQD4217E

Disconnection from *database\_URL* failed.

**Explanation:**

One or more errors occurred while disconnecting from the database.

**User Response:**

See the log files for more information about where the problem occurred.

## FFQD4218E

An I/O error occurred. Database: *database\_URL*. Table: *table\_name*.

**Explanation:**

One or more errors occurred while transferring data from the database.

**User Response:**

See the log files for more information about where the problem occurred.

## FFQD4219E

Database error occurred. Database: *database\_URL*. SQLCODE: *SQLCODE*.  
SQLSTATE: *SQLSTATE*.

**Explanation:**

One or more errors occurred while accessing to the database.

**User Response:**

See the log files for more information about where the problem occurred.

## FFQD4220E

There is no database connection.

**Explanation:**

The DB2 crawler had started successfully, but the database connection to crawl data was null unexpectedly.

**User Response:**

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*.

**Explanation:**

The DB2 crawler configuration file did not contain any column names.

**User Response:**

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*.

**Explanation:**

The security field was not a string column.

**User Response:**

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*.

**Explanation:**

The date field was not a string column.

**User Response:**

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*.

**Explanation:**

The DB2 crawler detected that the configuration file contained an unsupported data type column.

**User Response:**

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*.

**Explanation:**

The DB2 crawler detected that a not LOB column was specified as content

**User Response:**

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*.

**Explanation:**

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

**User Response:**

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*.

**Explanation:**

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

**User Response:**

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*.

**Explanation:**

The DB2 crawler detected that two or more content.

**User Response:**

Verify that the DB2 crawler configuration file corresponds with the current target table.

**FFQD4230E**

The value of unique identifiers must be unique through the table. Database: *database\_URL*. Table: *table\_name*. (*key\_value*)

**Explanation:**

DB2 crawler detected the value of unique identifiers are not unique. Unique identifiers are used for generating URI of the document and it must be unique.

**User Response:**

Verify unique identifiers are unique.

**FFQD4231E**

LONG VARCHAR and LONG VARGRAPHIC data type are not supported for unique identifiers of this table. Database: *database\_URL*. Table: *table\_name*.

**Explanation:**

LONG VARCHAR and LONG VARGRAPHIC data type cannot be used for unique identifiers if i) binary column is specified as content and ii) target data source doesn't support CURSOR WITH HOLD.

**User Response:**

Specify other columns for unique identifiers.

**FFQD4232E**

The column type of specified field for update detection is not supported. Database: *database\_URL*. Table: *table\_name*. Column: *column\_name*. Type: *date\_type*.

**Explanation:**

The column type of the field indicating update of row was not supported. The supported types are: TIMESTAMP, CHARACTER(8) FOR BIT DATA, VARCHAR(8) FOR BIT DATA

**User Response:**

Specify correct column.

**FFQD4300E**

An internal error occurred (another connection exists).

**Explanation:**

Unexpected error occurred.

**User Response:**

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.

**Explanation:**

The item type which was specified in configuration file did not exist in the Content Manager server or it was a part item type.

**User Response:**

Log on to the administration consol and edit the crawl space to remove the item type.

## FFQD4302E

Connection failed. Database: *database*. User: *user*.

**Explanation:**

A connection to the specified Content Manager server failed.

**User Response:**

Try the operation again. Make sure that the server is running, and the password is valid.

## FFQD4303E

Disconnection failed.

**Explanation:**

A disconnection from the Content Manager server failed.

**User Response:**

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*.

**Explanation:**

An error occurred while fetching content from the Content Manager server.

**User Response:**

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.

**Explanation:**

The error occurred when the crawler tried to store a large content as a temporary file.

**User Response:**

Refer to the crawler log for more information.

## FFQD4306E

Database connection does not exist.

**Explanation:**

A connection was not established for the Content Manager server. Connection might be destroyed.

**User Response:**

Try the operation again. Check the server is running.



## FFQD4307E

An error occurred while executing a query command *cmd*.

**Explanation:**

An error occurred while executing a query command of Content Manager.

**User Response:**

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*.

**Explanation:**

An error occurred while fetching a record from the Content Manager server.

**User Response:**

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*.

**Explanation:**

An error occurred while fetching a part from the Content Manager server.

**User Response:**

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.

**Explanation:**

The attribute which was specified in the configuration file did not exist in the Content Manager server.

**User Response:**

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.

**Explanation:**

The attribute which was specified in the configuration file did not exist or the attribute was a BLOB type.

**User Response:**

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.

**Explanation:**

The component which was specified in the configuration file did not exist in the Content Manager server.

**User Response:**

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.

**Explanation:**

An error occurred while fetching a value of attribute from the Content Manager server.

**User Response:**

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.

**Explanation:**

An error occurred while retrieving the last modified date from the Content Manager server.

**User Response:**

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.

**Explanation:**

An error occurred while retrieving a creation date from the Content Manager server.

**User Response:**

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*.

**Explanation:**

An error occurred while parsing a date. This is an unexpected error.

**User Response:**

Refer to the crawler log for more information.

**FFQD4317E**

An error occurred while retrieving an item type *item\_type*.

**Explanation:**

An error occurred while retrieving an item type from the Content Manager server.

**User Response:**

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.

**Explanation:**

An error occurred while retrieving an original file name from the Content Manager server.

**User Response:**

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.

**Explanation:**

The attribute which was specified as the field to control access in the configuration file did not exist in the Content Manager server.

**User Response:**

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.

**Explanation:**

The attribute which was specified as the document date in the configuration file was neither a `TIMESTAMP` type nor a `DATE` type.

**User Response:**

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.

**Explanation:**

The attribute which was specified as the document date in the configuration file did not exist in the Content Manager server.

**User Response:**

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*.

**Explanation:**

An error occurred while retrieving a document date from the Content Manager server.

**User Response:**

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*.

**Explanation:**

The generated URI was too long. This row will be ignored.

**User Response:**

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*.

**Explanation:**

The crawler server successfully connected to the Content Manager server.

**User Response:**

This message is for your information only. No action is required.

## FFQD4325I

Disconnected from server.

**Explanation:**

The crawler server successfully disconnected from the Content Manager server.

**User Response:**

This message is for your information only. No action is required.

## FFQD4326E

Resource Manager is not online or available.

**Explanation:**

The Resource Manager was not online or available.

**User Response:**

Start the Resource Manager and try the operation again.

## FFQD4327W

An error occurred while fetching content. Skip this document. PID:*pid*.

**Explanation:**

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.

**User Response:**

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.

## FFQD4328E

An error occurred while checking connection to Content Manager Server.

**Explanation:**

An error occurred while checking connection to Content Manager Server. The connection might be terminated.

**User Response:**

See IBM DB2 Content Manager for Multiplatforms/IBM DB2 Information Integrator for Content for Multiplatforms, Messages and Codes for more information.

## FFQD4402W

Property *property\_name* cannot be changed now. Restart the crawler.

**Explanation:**

This property *property\_name* cannot be changed while the crawler is running.

**User Response:**

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*.

**Explanation:**

NNTP crawler went into paused state automatically because the crawler reached to the number of articles specified as maximum number of articles.

**User Response:**

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.

**Explanation:**

NNTP crawler will sleep for a while because a network error was detected.

**User Response:**

If this is because of temporary network unavailability, wait for a while, and crawler will start crawling again. If not, make sure that the network is available on this crawler server.

## **FFQD4414E**

Updating the server status failed. The crawler will try again later.

**Explanation:**

NNTP crawler failed to update server status on the internal repository. Crawler will sleep for a while and try again.

**User Response:**

See other log messages to determine the cause of this error.

## **FFQD4415E**

Updating the newsgroup status failed. The crawler will try again later.

**Explanation:**

NNTP crawler failed to update newsgroup status on the internal repository. Crawler will sleep for a while and try again.

**User Response:**

See other log messages to determine the cause of this error.

## **FFQD4416E**

Inserting the message ID failed. The crawler will try again later.

**Explanation:**

NNTP crawler failed to insert message-id of an article to the internal repository. Crawler will sleep for a while and try again.

**User Response:**

See other log messages to determine the cause of this error.

## **FFQD4417E**

Inserting the news article failed. The crawler will try again later.

**Explanation:**

NNTP crawler failed to insert an article to the internal repository. Crawler will sleep for a while and try again.

**User Response:**

See other log messages to determine the cause of this error.

## FFQD4418E

Inserting deleted news article failed. The crawler will try again later.

**Explanation:**

NNTP crawler failed to insert information of a deleted article to the internal repository. Crawler will sleep for a while and try again.

**User Response:**

See other log messages to determine the cause of this error.

## FFQD4419W

The crawler will be stopped despite failure to update the server status.

**Explanation:**

NNTP crawler could not update server status. You might see obsolete or inconsistent server information in the enterprise search administration console. This information will be updated the next time that the crawler is started.

**User Response:**

No action is required.

## FFQD4420W

The crawler will be stopped despite failure to update newsgroup status. You might need to recrawl some articles.

**Explanation:**

The NNTP crawler could not update the newsgroup status. You might see obsolete or inconsistent server information in the enterprise search administration console. This information will be updated the next time that the crawler is started.

**User Response:**

No action is required.

## FFQD4421W

The crawler will be stopped despite failure to update message ID. You might need to recrawl some articles.

**Explanation:**

NNTP crawler could not record the message ID into an internal repository. The crawler will crawl the same article next time, but it will not affect to search result because the recrawled article will be eliminated if it a duplicate.

**User Response:**

No action is required.

## FFQD4422W

The crawler will be stopped despite failure to insert an article.

**Explanation:**

NNTP crawler could not store the article into an internal repository. However, it will be crawled when the crawler is started next time.

**User Response:**  
No action is required.

## **FFQD4423W**

The crawler will be stopped despite failure to insert a deleted article.

**Explanation:**  
NNTP crawler could not store the information of a deleted article into the internal repository. However, it will store the article the next time that the crawler is started.

**User Response:**  
No action is required.

## **FFQD4427E**

The newsgroup list could not be retrieved because of an unknown host *host\_name*.

**Explanation:**  
NNTP crawler cannot access the specified server *host\_name* when it tries to retrieve the list of newsgroups on the server because it is unknown host.

**User Response:**  
Verify that the host name is correct.

## **FFQD4428E**

The newsgroup list could not be retrieved because of a network error or server error on *host\_name*

**Explanation:**  
NNTP crawler could not retrieve the list of newsgroups on the server because of a network error occurred.

**User Response:**  
See the messages 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.

**Explanation:**  
NNTP crawler was stopped while writing information of deleted documents into an internal repository. This task will be resumed when the crawler is started next time.

**User Response:**  
No action is required.

## **FFQD4434W**

Message ID was not found in this article *message\_id*. This article will be ignored.



**Explanation:**

NNTP crawler detected an article that has an incorrect format. This article will be ignored.

**User Response:**

No action is required.

**FFQD4435W**

Message ID is inconsistent (*message\_id1* and *message\_id2* must be identical). This article will be ignored.

**Explanation:**

NNTP crawler detected an article that has an incorrect format. This article will be ignored.

**User Response:**

No action is required.

**FFQD4436W**

Unexpected date format *date*.

**Explanation:**

NNTP crawler detected unexpected date format in an article.

**User Response:**

No action is required.

**FFQD4437E**

An I/O error occurred while parsing article.

**Explanation:**

A network I/O error occurred while parsing an article retrieved from news server.

**User Response:**

See the messages in the log, verify the network availability, and confirm the news server is active.

**FFQD4449E**

Errors occurred in *api\_name*.

**Explanation:**

An error occurred while executing *api\_name* API.

**User Response:**

See the log files for more information about where the problem occurred.

**FFQD4450E**

File *file\_name* was not found.

**Explanation:**

One of the configuration file *file\_name* was not found. NNTP crawler could not start crawling.

**User Response:**

Verify that enterprise search was successfully installed and try again.

## **FFQD4451E**

Network I/O error occurred.

**Explanation:**

Network I/O error occurred while accessing to the server.

**User Response:**

See the messages 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*.

**Explanation:**

NNTP crawler received unexpected response *server\_response* from server for NNTP command *nntp\_command*.

**User Response:**

Check that the news server is serving accordingly.

## **FFQD4453W**

The status of an article pointer could not be retrieved. Return code was *return\_code*.

**Explanation:**

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 a server.

**User Response:**

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*.

**Explanation:**

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 is unavailable.

**User Response:**

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*.

**Explanation:**

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 is unavailable.

**User Response:**

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\_name*. Return code was *return\_code*.

**Explanation:**

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 is unavailable.

**User Response:**

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*.

**Explanation:**

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 is unavailable.

**User Response:**

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.

**Explanation:**

NNTP crawler could not resolve the server address for server *server\_name*.

**User Response:**

Remove the server from crawl space if the server does not exist. If it exists, check the network environment on the crawler server.

**FFQD4461E**

The crawler received service disconnected message from server.

**Explanation:**

NNTP crawler received service disconnected message from server. The connection to the server will be disconnected.

**User Response:**

Verify the availability of the server.

## FFQD4500E

The *directory\_name* directory was not found.

**Explanation:**

The specified directory was not found.

**User Response:**

Verify the directory on the crawler server and configure crawl space page on the administration console.

## FFQD4501E

The *directory\_name* directory was not found. Error message : *detail\_message*

**Explanation:**

The specified directory was not found.

**User Response:**

Verify the directory on the crawler server and configure the crawl space page on administration console.

## FFQD4502E

Failed to open *filename* file. The file might be used by another program.

**Explanation:**

The file might be used by other programs. The file was not crawled.

**User Response:**

Stop the other program that is using the crawled file and restart the crawler.

## FFQD4503E

An error occurred while reading *filename* file.

**Explanation:**

The file might be used by other programs. The file was not crawled.

**User Response:**

See the log files for more information about where the problem occurred.

## FFQD4504E

The subdirectory level *levels of subdirectories* is incorrect.

**Explanation:**

The specified levels of subdirectories was not valid.

**User Response:**

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*.

**Explanation:**

The specified directory name was too long to be stored.

**User Response:**

Move the directory to be crawled and configure the crawl space page on the administration console.

**FFQD4506W**

The length of the URI name is too long to be stored. URI: *directory\_name*.

**Explanation:**

The specified URI was too long to be stored. The length of URI name is calculated as UTF-8 byte string.

**User Response:**

Move the directory or rename the file name to be crawled and configure the crawl space page on administration console.

**FFQD4507E**

The network drive cannot be used. Error message : *detail\_message*

**Explanation:**

The network drive was not available.

**User Response:**

Verify the network and the directory on the crawler server.

**FFQD4508E**

The user ID *user\_name* or the password is incorrect. Network folder:*directory\_name*.

**Explanation:**

An incorrect user ID, an incorrect password or both were specified.

**User Response:**

Verify the user name and the password and configure crawl space page on Administration console.

**FFQD4509E**

The temporary file *temporary\_file* was not created.

**Explanation:**

The File System crawler could not create the temporary file.

**User Response:**

Verify the free disk spaces on the crawler server and restart the crawler.

**FFQD4510E**

The temporary file *temporary\_file* failed to be loaded.

**Explanation:**

The File System crawler could not read the temporary file.

**User Response:**

Verify the free disk spaces on the crawler server and restart the crawler.

## FFQD4511E

The temporary file *temporary\_file* failed to be stored.

**Explanation:**

The File System crawler could not write the temporary file.

**User Response:**

Verify the free disk spaces on the crawler server and restart the crawler.

## FFQD4600E

An I/O error occurred while receiving a response from the URL *public\_folder\_URL*.

**Explanation:**

One or more errors occurred while receiving a response from the public folder server.

**User Response:**

See the log files for more information about where the problem occurred.

## FFQD4601E

An XML parser configuration error occurred.

**Explanation:**

Could not create a XML parser for parsing responses from the public folder server.

**User Response:**

Verify that enterprise search was successfully installed and try again.

## FFQD4602E

WebDAV response from URL *public\_folder\_URL* is an incorrect XML document.

**Explanation:**

The public folder server returned a XML document which might contain incorrect elements and characters.

**User Response:**

Verify the public folder does not contain broken documents.

## FFQD4603I

Connected to *public\_folder\_URL* using user *user\_name*.

**Explanation:**

The Exchange crawler connected to the specified URL.

**User Response:**

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*.

**Explanation:**

The Exchange crawler connected to the specified URL.

**User Response:**

This message is for your information only. No action is required.

## **FFQD4605I**

Disconnected from *public\_folder\_URL*.

**Explanation:**

The Exchange crawler disconnected from the specified URL.

**User Response:**

This message is for your information only. No action is required.

## **FFQD4607E**

Connecting to *public\_folder\_URL* failed.

**Explanation:**

One or more errors occurred while connecting to the specified URL.

**User Response:**

See the log files for more information about where the problem occurred.

## **FFQD4608E**

Checking next record failed.

**Explanation:**

One or more errors occurred while getting a document from the specified URL.

**User Response:**

See the log files for more information about where the problem occurred.

## **FFQD4609E**

Error occurred while getting content from URL *uniform\_resource\_locator*.

**Explanation:**

One or more errors occurred while transferring data from the specified URL.

**User Response:**

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*.

**Explanation:**

One or more errors occurred while transferring data from the specified URL.

**User Response:**

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*.

**Explanation:**

One or more errors occurred while transferring data from the specified URL.

**User Response:**

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*.

**Explanation:**

The public folder server returned an unexpected HTTP response.

**User 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*.

**Explanation:**

The public folder server returned a response which contained an incorrect date string.

**User Response:**

Verify that the crawler which you want to run was configured correctly and try again.

## FFQD4614W

The specified URL *document\_URL* already exists.

**Explanation:**

The Exchange crawler detected duplicated URL.

**User Response:**

No action is required.

## FFQD4616E

Temporary file *temporary\_file* does not exist.

**Explanation:**

The Exchange crawler could not find the temporary file.

**User Response:**

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*.

**Explanation:**

The Exchange crawler could not write data to the temporary file.

**User Response:**

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*.

**Explanation:**

One or more I/O errors occurred while closing the temporary file.

**User Response:**

Verify that the crawler machine have free disk space and try again.

## FFQD4619E

An I/O error occurred while closing the file *temporary\_file*.

**Explanation:**

One or more I/O errors occurred while closing the temporary file.

**User Response:**

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*.

**Explanation:**

One or more I/O errors occurred while closing the temporary file.

**User Response:**

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*.

**Explanation:**

One or more I/O errors occurred while closing the temporary file.

**User Response:**

Verify that the crawler machine is working correctly and try again.

## FFQD4622E

An unknown host name *host\_name* was detected.

**Explanation:**

The host name in the specified public folder URL was not found.

**User Response:**

Verify that the host name is correct.

**FFQD4623E**

The URL *public\_folder\_URL* is incorrect.

**Explanation:**

The specified public folder URL was not found.

**User Response:**

Verify that the public folder URL is correct.

**FFQD4624E**

Could not refresh connection to *public\_folder\_URL*.

**Explanation:**

The Exchange crawler could not reset the connection to the public folder server.

**User Response:**

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*.

**Explanation:**

One or more errors occurred while sending a request to the public folder server.

**User Response:**

See the log files for more information about where the problem occurred.

**FFQD4626E**

Connecting to *public\_folder\_URL* failed.

**Explanation:**

One or more errors occurred while connecting to the specified URL.

**User Response:**

See the log files for more information about where the problem occurred.

**FFQD4627E**

A truststore file must be specified to use SSL.

**Explanation:**

To specify a truststore file is required for using SSL.

**User Response:**

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.

**Explanation:**

To specify a password for the truststore file is required for using SSL.

**User Response:**

Verify that the crawler which you want to run was configured correctly and try again.

## FFQD4629E

The truststore file *truststore\_file* does not exist.

**Explanation:**

The specified truststore file did not exist or you did not have permission to read the file.

**User Response:**

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*.

**Explanation:**

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.

**User Response:**

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*.

**Explanation:**

The Exchange crawler ignored a document because its URI is too long.

**User Response:**

No action is required.

## FFQD4633W

A folder was ignored because the URI is too long. URI: *URI*.

**Explanation:**

The Exchange crawler ignored a folder because its URI is too long.

**User Response:**

No action is required.

## FFQD4707E

Server name *server\_string* in configuration file does not represent VeniceBridge server and repository name.

**Explanation:**

The server string *server\_string* was not correct format.

**User Response:**

Configure VeniceBridge crawler setting from the administration console.

## FFQD4708E

VeniceBridge API error occurred.

**Explanation:**

The error occurred while accessing to the VeniceBridge server.

**User Response:**

See the log files for more information about where the problem occurred.

## FFQD4709E

VeniceBridge API is not licensed.

**Explanation:**

The VeniceBridge crawler could not access the VeniceBridge server because some of the VeniceBridge component was not licensed.

**User Response:**

See the log files for more information about where the problem occurred.

## FFQD4710E

VeniceBridge server instance cannot be created.

**Explanation:**

The VeniceBridge crawler could not access the VeniceBridge server because the specified JNDI properties were incorrect.

**User Response:**

See the log files for more information about where the problem occurred.

## FFQD4711E

VeniceBridge API configuration error occurred.

**Explanation:**

The error occurred while accessing to the VeniceBridge server.

**User Response:**

See the log files for more information about where the problem occurred.

## FFQD4712E

VeniceBridge server cannot be located.

**Explanation:**

The error occurred while accessing to the VeniceBridge server.

**User Response:**

See the log files for more information about where the problem occurred.

## **FFQD4713E**

Communication related error occurred.

**Explanation:**

The error occurred while accessing to the VeniceBridge server.

**User Response:**

See the log files for more information about where the problem occurred.

## **FFQD4714E**

The repository *repository\_name* is not found.

**Explanation:**

The *repository\_name* repository did not exist in the VeniceBridge server.

**User Response:**

Verify that the VeniceBridge crawler configuration corresponds with the current target repository.

## **FFQD4715E**

The VeniceBridge user instance was not initialized.

**Explanation:**

The error occurred while accessing to the VeniceBridge server.

**User Response:**

See the log files for more information about where the problem occurred.

## **FFQD4716E**

Logon to *repository\_name* repository failed.

**Explanation:**

The VeniceBridge crawler could not logon to the *repository\_name* repository.

**User Response:**

Check a correct user name and password are specified.

## **FFQD4718E**

The item class *item\_class* was not found for repository *repository\_name*.

**Explanation:**

The item class that was specified in configuration file did not exist in the VeniceBridge server.

**User Response:**

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*.

**Explanation:**

The property that was specified in configuration file did not exist in the VeniceBridge server.

**User Response:**

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*.

**Explanation:**

The property that was specified as the document data in configuration file did not exist in the VeniceBridge server.

**User Response:**

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*.

**Explanation:**

The property that was specified as the security field in configuration file did not exist in the VeniceBridge server.

**User Response:**

Verify that the VeniceBridge crawler configuration corresponds with the current target item class.

## FFQD4722E

The specified selection criteria *condition\_statement* has a problem.

**Explanation:**

The condition statement that was specified in configuration file was incorrect syntax.

**User Response:**

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*.

**Explanation:**

The property that was specified as the document data in configuration file was neither a DATE type nor a DATETIME type.

**User Response:**

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*.

**Explanation:**

The property that was specified in configuration file was an unsupported data type.

**User Response:**

Verify that the VeniceBridge crawler configuration corresponds with the current target item class.

**FFQD4725E**

The executed query has a problem.

**Explanation:**

The error occurred while executing a query command to the VeniceBridge server.

**User Response:**

See the log files for more information about where the problem occurred.

**FFQD4726E**

No order key is found. Key: *order\_key\_property*.

**Explanation:**

The property that was specified as order key in configuration file does not exist in the VeniceBridge server.

**User Response:**

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*.

**Explanation:**

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.

**User Response:**

No action is required.

**FFQD4729W**

The property *property* was not found.

**Explanation:**

The property that was specified in configuration file was not found.

**User Response:**

Verify that the VeniceBridge crawler configuration corresponds with the current target item class.

## **FFQD4730E**

An I/O error occurred while writing content.

**Explanation:**

The error occurred while transferring data from the VeniceBridge server.

**User Response:**

See the log files for more information about where the problem occurred.

## **FFQD4731E**

Ranking date field must be suitable as a selection property within a query. Repository: *repository\_name*. Item class: *item\_class*. Property: *property*.

**Explanation:**

The property that was specified as the document data in configuration file was not suitable as a selection property within a query.

**User Response:**

Verify that the VeniceBridge crawler configuration corresponds with the current target item class.

## **FFQD4732E**

Security field must be suitable as a selection property within a query. Repository: *repository\_name*. Item class: *item\_class*. Property: *property*.

**Explanation:**

The property that was specified as the security field in configuration file was not suitable as a selection property within a query.

**User Response:**

Verify that the VeniceBridge crawler configuration corresponds with the current target item class.

## **FFQD4733E**

Revision date property must be suitable as a selection property within a query. Repository: *repository\_name*. Item class: *item\_class*. Property: *property*.

**Explanation:**

The property that was specified as the revision date in configuration file was not suitable as a selection property within a query.

**User Response:**

Verify that the VeniceBridge crawler configuration corresponds with the current target item class.



## FFQD4734E

Creation date property must be suitable as a selection property within a query. Repository: *repository\_name*. Item class: *item\_class*. Property: *property*.

**Explanation:**

The property that was specified as the creation date in configuration file was not suitable as a selection property within a query.

**User Response:**

Verify that the VeniceBridge crawler configuration corresponds with the current target item class.

## FFQD4735E

Content size property must be suitable as a selection property within a query. Repository: *repository\_name*. Item class: *item\_class*. Property: *property*.

**Explanation:**

The property that was specified as the content size in configuration file was not suitable as a selection property within a query.

**User Response:**

Verify that the VeniceBridge crawler configuration corresponds with the current target item class.

## FFQD4736E

Sort key property must be suitable as a selection property within a query. Repository: *repository\_name*. Item class: *item\_class*. Property: *property*.

**Explanation:**

The property that was specified as the sort key in configuration file was not suitable as a selection property within a query.

**User Response:**

Verify that the VeniceBridge crawler configuration corresponds with the current target item class.

## FFQD4737W

The native content that was specified was not found. This record will be ignored. URI: *uri*.

**Explanation:**

The native content that was specified by URI was not found.

**User Response:**

Verify that the backend content management system is working correctly.

## FFQD4738W

The content that was specified was not found. This record will be ignored. URI: *uri*.

**Explanation:**

The content that was specified by URI was not found.

**User Response:**

Verify that the backend content management system is working correctly.

## **FFQD4739E**

The number of ignored documents exceeds the limit *warning\_limit*.

**Explanation:**

The number of ignored content and native content exceeds the limit.

**User Response:**

Verify that the backend content management system is working correctly.

## **FFQD4741W**

The error occurred while retrieving folders in which the specified document is filed. URI: *uri*.

**Explanation:**

The error occurred while retrieving folders in which the specified content is filed. The containing folder metadata of one or more documents are missing.

**User Response:**

Verify that the backend content management system is working correctly.

---

## **General system messages (FFQG)**

### **FFQG0026E**

The input and output files must be different files. Correct the appropriate file setting and resubmit the request. - Input File: *input\_file* Output File: *output\_file*

**Explanation:**

The same file was used for both input and output. The input and output files must be different.

**User Response:**

Resubmit the request with the correct input or output file.

---

## **Operating system subsystem (OSS) messages (FFQO)**

### **FFQO0034E**

The path *path\_name* does not exist.

**Explanation:**

The directory path specified by *path\_name* does not exist. The operation requires a directory path, but no path was found.

**User Response:**

Ensure that the path exists. If the path exists, ensure that the directory has the appropriate permissions. Often, this is an expected error and so this error can be safely ignored.

## FFQO0086E

The trace dump file *file\_name* failed to validate with OSSe Error Code: *err\_code* OSSe Error Name: *err\_name* OSSe Message: *err\_msg*.

### Explanation:

Validation of a specified trace dump file has failed with an error may be due to an incomplete trace dump file.

### User Response:

Ensure that the trace session has ended and has completed the trace dump file. To ensure that the session has ended, issue `stopccl` command or stop the running trace session.

## FFQO0141E

A socket error occurred with error code *error\_code* with address *host\_name* and port *port\_number*.

### Explanation:

A socket operation has resulted in the specified *error\_code*.

### User Response:

If *error\_code* is **22** (UNIX) or **10022** (Windows), in the case of run-time shutdown, you can safely ignore this error. If **10093** (Windows), you can also ignore this error.

## FFQO0164E

Unable to load properties file *file\_name*.

### Explanation:

The configuration file *file\_name* could not be loaded.

### User Response:

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*.

### Explanation:

The configuration file *file\_name* that was included by another configuration file could not be loaded.

### User Response:

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*.

### Explanation:

A configuration could not be saved to the file *file\_name*.

| **User Response:**

| Verify that the directory or file system into which the configuration file  
| was to be saved exists and has write permission. Check whether a  
| write-protected file with the same name already exists.

| **FFQO0193W**

| Temporary pool *memory\_pool\_name* was deinitialized while *number* of its  
| chunks are still in use.

| **Explanation:**

| A portion of the temporary memory pool named *memory\_pool\_name* was  
| still in use at the time that pool was deinitialized.

| **User Response:**

| Check the logs for additional messages. No further action is necessary.

| **FFQO0223E**

| One character passwords are not allowed.

| **Explanation:**

| Is it not allowed to specify a password with only one character in length.

| **User Response:**

| Specify a password longer than one character.

| **FFQO0258E**

| Session *session\_name* abnormally terminated.

| **Explanation:**

| The session *session\_name* was terminated abnormally.

| **User Response:**

| Check the log file related to *session\_name* to see why the session  
| abnormally terminated.

| **FFQO0259E**

| Requested operation timed out.

| **Explanation:**

| The requested operation was timed out.

| **User Response:**

| Try the operation again.

| **FFQO0273W**

| Internal warning *exception\_message*.

| **Explanation:**

| Encountered unexpected exceptions.

| **User Response:**

| If this message repeats, contact IBM Software Support.

---

## Parser messages (FFQT)

### FFQT0059E

The Stellent parser failed to open a document handle with code *stellent\_error\_code*. The URL of the document is *url*.

**Explanation:**

The document with URL *url* could not be parsed by the Stellent parser, and its content will not be indexed.

**User Response:**

Verify that the document at *url* is a valid document. Try to read it with an appropriate viewer for its file type. If it is not valid, replace it with a valid version, and recrawl the URL.

### FFQT0060E

The Stellent parser failed to open a content handle with code *stellent\_error\_code*. The URL of the document is *url*.

**Explanation:**

The document with URL *url* could not be parsed by the Stellent parser, and its content will not be indexed.

**User Response:**

Verify that the document at *url* is a valid document; try to read it with an appropriate viewer for its file type. If it is invalid, replace it with a valid version, and recrawl the URL.

### FFQT0061E

The Stellent parser failed to retrieve the first content segment with code *stellent\_error\_code*. The URL of the document is *url*.

**Explanation:**

The document with URL *url* could not be parsed by the Stellent parser, and its content will not be indexed.

**User Response:**

Verify that the document at *url* is a valid document. Try to read it with an appropriate viewer for its file type. If it is not valid, replace it with a valid version, and recrawl the URL.

### FFQT0062E

The Stellent parser failed to retrieve the next content segment with code *stellent\_error\_code*. The URL of the document is *url*.

**Explanation:**

The document with URL *url* could not be parsed completely by the Stellent parser, and only an initial part of its content will be indexed. This was probably caused by truncation of the document.

**User Response:**

Verify that the document at *url* is a valid document. Try to read it with an appropriate viewer for its file type. If it is not valid, replace it with a valid version, and recrawl the URL.

## FFQT0065I

The Stellent parser received a document of supported file type *stellent\_file\_type* (*doc\_type*). The URL of the document is *url*.

**Explanation:**

The document with URL *url* was recognized as of type *stellent\_file\_type*. It will be searchable with a document type of *doc\_type*.

**User Response:**

No action is required.

## FFQT0066I

The Stellent parser received a document of unsupported file type *stellent\_file\_type*. The URL of the document is *url*.

**Explanation:**

The document with URL *url* was recognized as of type *stellent\_file\_type*, which is not mapped to a searchable document type. Its content will not be indexed.

**User Response:**

No action is required.

## FFQT0067I

The Stellent parser received a document of file type *stellent\_file\_type* (*doc\_type*) that is supported by a native parser. The URL of the document is *url*.

**Explanation:**

The document with URL *url* was recognized as of type *stellent\_file\_type*. This type is supported by one of the built-in parsers, and Stellent will not be used to parse its content. It will be searchable with a document type of *doc\_type*.

**User Response:**

No action is required.

## FFQT0072I

The Stellent parser could not recognize the file type of the document with URL *url*.

**Explanation:**

The Stellent parser was not able to recognize the type of the document with URL *url*. Its content will not be indexed.

**User Response:**

No action is required.

## FFQT0078W

The Stellent parser failed to parse a document of size *document\_size*. The URL of the document is *url*.

**Explanation:**

The document with URL *url* could not be parsed successfully. All or parts of its content will not be indexed.

**User Response:**

No action is required.

**FFQT0084E**

Timeout waiting for a response from the Stellent parser server. The server is probably looping and needs to be restarted.

**Explanation:**

The Stellent parser is not responding. It is probably in an undefined state and needs to be restarted.

**User Response:**

Check the log whether the Stellent parser was automatically restarted. If not, stop and restart the parser.

**FFQT0085E**

Lost connection to the Stellent parser server while processing a document. The server might have stopped and needs to be restarted.

**Explanation:**

The connection to the Stellent parser was interrupted. It has probably stopped or was terminated abnormally, and needs to be restarted.

**User Response:**

Check the log whether the Stellent parser was automatically restarted. If not, stop and restart the parser.

**FFQT0086W**

The Stellent parser has processed *number\_of\_docs* documents. You should recycle the server to clean up its resources.

**Explanation:**

The Stellent parser might become unstable over time after it processes large numbers of documents. Therefore, it should be restarted after it processes the specified number of documents.

**User Response:**

Check the log whether the Stellent parser was automatically restarted. If not, stop and restart the parser.

**FFQT0138E**

An error occurred while parsing mapping file *file*.

**Explanation:**

The document with name *file* could not be parsed successfully.

**User Response:**

Make sure that the structure and elements of the XML file are correct. Ensure that the `indexBuildSpecification` tag has the proper namespace information: `xmlns="http://www.ibm.com/of/822/consumer/index/xml"`,

and not  
xmlns:namespace="http://www.ibm.com/of/822/consumer/index/xml".

---

## Utility messages (FFQU)

### FFQU0269E

The calculator encountered a syntax error for variable "*variable\_name*" containing expression "*variable\_expression*" at character position *position\_within\_the\_expression*.

**Explanation:**

The calculator cannot continue its evaluation of the variable *variable\_name* because it contains an expression *variable\_expression* with a syntax error at location *position\_within\_the\_expression*.

**User Response:**

If this message repeats, contact IBM Software Support.

### FFQU0270E

The calculator encountered an invalid operand "*invalid\_operand*" for variable "*variable\_name*" containing expression "*variable\_expression*" at character position *position\_within\_the\_expression*.

**Explanation:**

The calculator cannot continue its evaluation of the variable *variable\_name* because it contains an expression *variable\_expression* with an invalid operand at location *position\_within\_the\_expression*. A valid operand is either a variable name or a constant, but the calculator encountered *invalid\_operand*, which is not recognized as a variable or a constant.

**User Response:**

If this message repeats, contact IBM Software Support.

### FFQU0271E

The calculator encountered an invalid constant "*invalid\_constant*" for variable "*variable\_name*" containing expression "*variable\_expression*" at character position *position\_within\_the\_expression*.

**Explanation:**

The calculator cannot continue its evaluation of the variable *variable\_name* because it contains an expression *variable\_expression* with an invalid constant at location *position\_within\_the\_expression*. A valid constant is a literal floating point in decimal format, but the calculator encountered *invalid\_constant*, which is not a valid constant.

**User Response:**

If this message repeats, contact IBM Software Support.

### FFQU0272E

The calculator encountered an undefined variable "*variable\_name*".



| **Explanation:**

| The calculator cannot continue its evaluation of the variable *variable\_name*  
| because its expression is missing.

| **User Response:**

| If this message repeats, contact IBM Software Support.

| **FFQU0273E**

| The calculator encountered an infinite loop because of the circular  
| dependency of variable "*variable\_name*".

| **Explanation:**

| The calculator cannot continue its evaluation of the variable *variable\_name*  
| because it contains an expression that eventually references itself, which is  
| an infinite loop.

| **User Response:**

| If this message repeats, contact IBM Software Support.

---

| **Synonym dictionary messages (FFQV)**

| **FFQV0001E**

| Frost initialization failed.

| **Explanation:**

| An exception occurred during the initialization of Frost, perhaps Frost  
| libraries (dlt40.jar and dlt40.dll) are missing

| **User Response:**

| Verify that the dlt40.jar file and the dlt40.dll file are present and that your  
| system's PATH variable includes the dlt40.dll file.

| **FFQV0002E**

| Synonym XML file *file\_name* could not be found.

| **Explanation:**

| The specified synonym file could not be found.

| **User Response:**

| Specify an existing synonym file and check that its file permissions are set  
| correctly.

| **FFQV0003E**

| Synonym XML file *file\_name* could not be accessed.

| **Explanation:**

| The specified synonym file could not be accessed. The file might not have  
| the appropriate permissions set.

| **User Response:**

| Check file permissions of the specified synonym file and validate that the  
| file is not corrupted.

## FFQV0004E

Synonym XML file *file\_name* is not a valid XML file.

**Explanation:**

The specified synonym file is either not well-formed or does not comply with the XML schema synonyms.xsd.

**User Response:**

Use an XML editor to validate your XML file.

## FFQV0005E

An error occurred while storing the next synonym set to the dictionary file *file\_name*.

**Explanation:**

The specified synonym file is either corrupt or there is no disk space available to store the dictionary file.

**User Response:**

Use an XML editor to validate your XML file. Check that the file has the appropriate permissions and that there is enough disk space.

## FFQV0006E

An error occurred while trying to initialize the synonym dictionary file *file\_name*.

**Explanation:**

The specified dictionary file either already exists or could not be created due to incorrect permissions.

**User Response:**

Check whether the specified path is correct and has write permission.

## FFQV0007E

An error occurred while trying to close the dictionary file *file\_name*.

**Explanation:**

The specified dictionary could not be closed.

**User Response:**

Check whether there is enough disk space.

## FFQV0008E

The dictionary file *file\_name* already exists.

**Explanation:**

Existing dictionary files will not be overwritten.

**User Response:**

Specify a different dictionary file name, or delete or rename the existing one.

## FFQV0009E

The log file *file\_name* could not be created.

**Explanation:**

An exception occurred, but it could not be logged because the log file could not be accessed.

**User Response:**

Check that the log file has the appropriate permissions and that there is enough disk space.

## FFQV0010E

*file\_name* seems to be an invalid XML synonym file. The synonym files must have the extension "\*.xml".

**Explanation:**

Synonym files must have the extension "\*.xml".

**User Response:**

Specify a valid XML synonym file that has the extension "\*.xml" or rename the existing file.

## FFQV0011E

*file\_name* is not a valid extension for a dictionary file. Dictionary files must have the extension "\*.dic".

**Explanation:**

Dictionary files must have the extension "\*.dic".

**User Response:**

Use a file name that has the extension "\*.dic".

## FFQV0012E

An error occurred while trying to log to the system console.

**Explanation:**

An error occurred while trying to generate a message and write it to the system console.

**User Response:**

Check log file for error details.

## FFQV0021E

The xsd schema file (synonyms.xsd) could not be found in the current classpath.

**Explanation:**

To validate the synonyms-file, a file named synonyms.xsd is needed which could not be found in the current classpath.

**User Response:**

Add the synonyms.xsd file to the classpath or use the script `essyndictbuilder` to setup the classpath.

---

## Migration wizard messages (FFQW)

### FFQW0001E

No message was found for the *message\_key* key. Contact the enterprise search administrator.

**Explanation:**

The message file that is used by the wizard was not found.

**User Response:**

Verify that enterprise search components of DB2 Information Integrator OmniFind Edition were installed correctly.

### FFQW0002E

File *file\_name* cannot be read. Change your selection.

**Explanation:**

The selected file or directory cannot be read by the user that started the wizard.

**User Response:**

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*.

**Explanation:**

Each collection that you want to migrate must have a collection name that does not already exist in enterprise search.

**User Response:**

Specify a collection name that does not exist in enterprise search.

### FFQW0008E

The user ID or password is invalid.

**Explanation:**

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.

**User Response:**

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.

**Explanation:**

The document contains unexpected content. It might not be a valid collection configuration file, or the file was corrupted.

**User Response:**

Check whether the file is the one that you exported from WebSphere Portal. If it is the correct file, 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*.

**Explanation:**

An error occurred while creating the enterprise search collection.

**User Response:**

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*.

**Explanation:**

An error occurred while creating the Web crawler.

**User Response:**

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*.

**Explanation:**

The wizard could not connect to enterprise search.

**User Response:**

Ensure that the enterprise search system is running.

**FFQW0016E**

An error occurred while creating collection *collection\_name*: *error\_message*.

**Explanation:**

An error occurred while creating the enterprise search collection.

**User Response:**

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*

**Explanation:**

An error occurred while creating a Web crawler configuration file.

**User Response:**

See the wizard log file for more information. Correct the problem and try again.

**FFQW0120I**

The migration completed successfully. See the log file for more information.

**Explanation:****User Response:**

This message is for your information only. No action is required.

**FFQW0140I**

The collection was imported from the WebSphere Portal configuration file *file*.

**Explanation:****User Response:**

This message is for your information only. No action is required.

**FFQW0141I**

Collection *collection\_name* was created successfully.

**Explanation:****User Response:**

This message is for your information only. No action is required.

**FFQW0142I**

The Web crawler for collection *collection\_name* was created successfully.

**Explanation:****User Response:**

This message is for your information only. No action is required.

**FFQW0154I**

Creating the crawler data for collection *collection\_name*.

**Explanation:****User Response:**

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*.

**Explanation:**

**User Response:**

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*.

**Explanation:**

**User Response:**

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*.

**Explanation:**

**User Response:**

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*.

**Explanation:**

**User Response:**

This message is for your information only. No action is required.

**FFQW0159I**

The configuration file *file* was created. Return code: *return\_code*.

**Explanation:**

**User Response:**

This message is for your information only. No action is required.





---

## Glossary of terms for enterprise search

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. A phrase or a sentence that describes 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 3. 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 4. 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 5. 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 5. 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 6. 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 6. 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 7. 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 8. 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 8. 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 9. 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 10. 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 351.

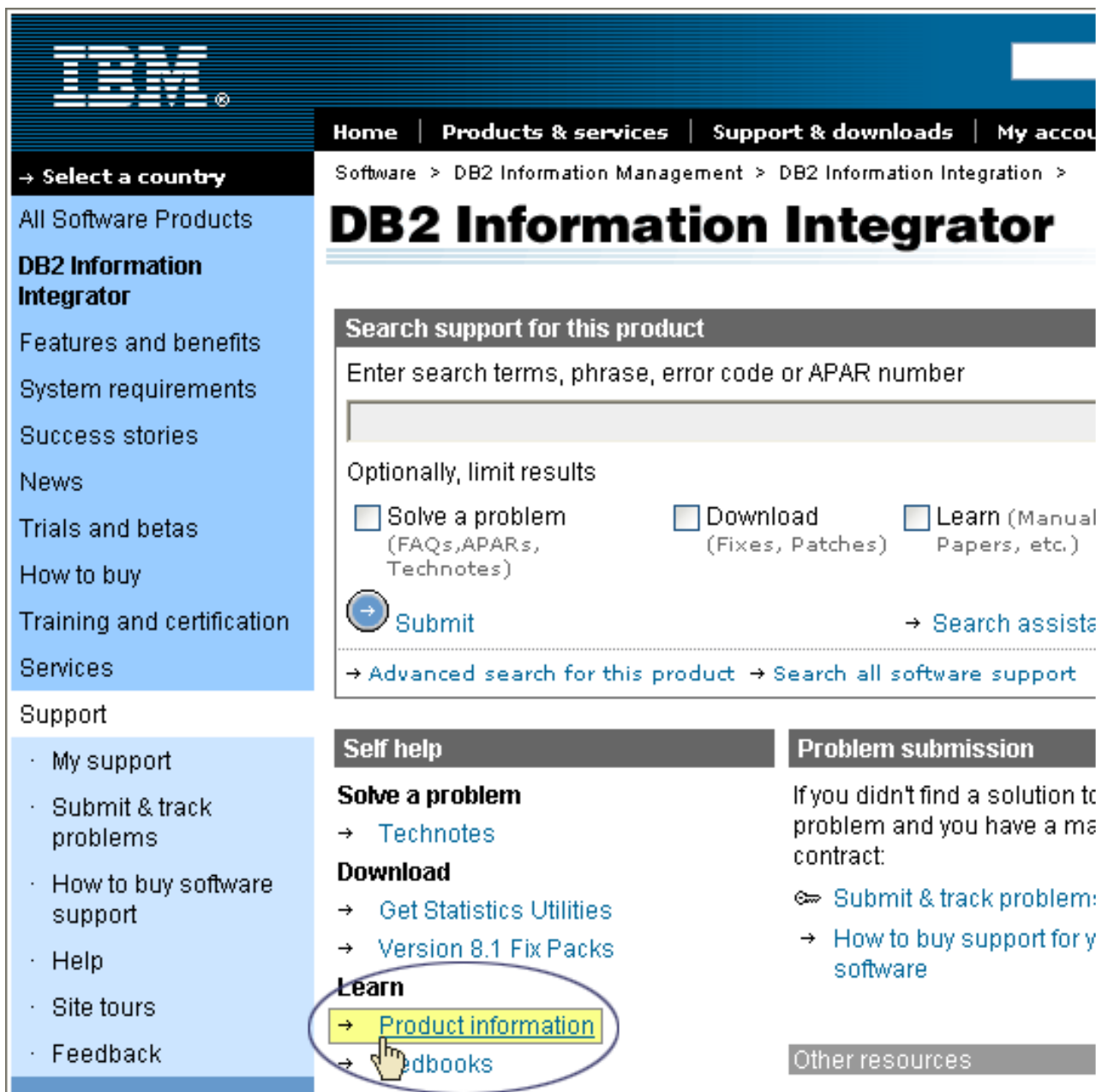


Figure 2. Product information link on the DB2 Information Integrator Support Web site

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 354.
- DB2 documentation is provided in an accessible format. For more information, see “Accessible documentation” on page 354.

---

## 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.

---

## 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 products is available by telephone or on the Web.

Information about DB2 Information Integrator products 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).



---

## 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 Stellant, 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  
Domino.doc  
Informix  
Lotus  
Lotus Notes  
Notes  
OmniFind  
POWER4  
POWER5  
Tivoli  
WebSphere  
Workplace  
xSeries  
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.





---

# Index

## A

- access controls
  - current validation of 148
  - description 142
  - disabling for a collection 153
  - document-level security 147, 157
  - Portal Search Engine support 157
  - user profiles 149
- Active Directory Server 150
- active Web sites, monitoring 174, 175
- address rules for Web crawlers 56
- administration console
  - description 8
  - interface 15
  - logging in 18
  - task summary 15
- administrative roles
  - collection administrator 140, 141
  - configuring 141
  - description 140
  - enterprise search administrator 140, 141
  - monitor 140, 141
  - operator 140, 141
- administrator password
  - changing on a single server 19
  - changing on multiple servers 20
- ads.add.ldif file 150
- AIX
  - configuring Content Manager crawlers 37
  - configuring Notes crawlers 45
  - configuring VeniceBridge crawlers 52
- alerts
  - collection-level 169, 188
  - description 187
  - documents crawled 188
  - documents indexed 188
  - e-mail options 188, 189
  - free space on servers 189
  - index limits 169
  - receiving e-mail for 192
  - search response times 188
  - SMTP server configuration 191
  - system-level 189
- analysis engines
  - associating with collections 90
  - associating with the system 89
  - description 88
  - mapping analysis results 93
  - mapping XML elements 91
- anchor text analysis
  - collection security 145
  - description 139
  - global analysis 145
  - indexing documents 145
- annotators 88
- APIs
  - Data Listener 1, 10
  - description 10
  - Search and Index 10, 125

- authentication
  - description 142
  - disabling 143
- B**
- backing up enterprise search 197
- backup scripts
  - description 197
  - running 197, 198
- C**
- categories
  - categorization type 82
  - category trees 81
  - creating 83
  - description 78
  - migrating from WebSphere Portal 159, 160
  - model-based 80, 82
  - nesting subcategories 81
  - rule-based 79, 82
  - searching 78
  - URI formats 71
- categorization type
  - model-based 80
  - rule-based 79
  - selecting 26, 82
- category rules
  - configuring 83
  - document content 79, 83
  - URI patterns 79, 83
- category trees
  - description 81
  - migrating from WebSphere Portal 159, 160
- checking system resources 170
- Chinese
  - n-gram segmentation 95
  - parsing new line characters 95
- collapsed search results
  - configuring 112
  - description 111
- collapsed URIs
  - configuring 112
  - description 111
- collection administrator
  - description 140
  - role configuration 141
- collection ID, syntax rules 26
- Collection wizard 25
- collection-level security
  - anchor text analysis 145
  - application IDs 146
  - description 139, 144
  - duplicate document analysis 144
  - enabling 26
- collections
  - anchor text security 145

- collections (*continued*)
  - application ID security 146
  - associating with search applications 126
  - bypassing document-level access controls 153
  - creating with Collection wizard 25
  - creating with Collections view 26
  - default migration settings 162
  - deleting 29
  - description 2
  - draft 25
  - duplicate document security 144
  - editing 28
  - estimating resources 169
  - estimating the size 26
  - federation 25
  - migrating from WebSphere Portal 159, 160
  - MigrationWizard.log file 160
  - monitoring 171
  - parsing 77
  - Search servers 115
  - searching 133
  - security 144
  - system status 171
  - ways to create 25
- Collections view
  - creating collections 26
  - description 15
- common analysis structures
  - description 88
  - mapping to the index 93
  - mapping XML elements to 91
- concurrent index builds 106
- config.properties file 128
- Content Manager crawlers
  - configuring 36
  - setting up on AIX 37
  - setting up on Linux 38
  - setting up on Windows 40
  - URI formats 71
- cookies for Web crawling
  - configuring 66
  - description 65
  - format 65
- cookies.ini file
  - configuring 66
  - description 65
  - format 65
- Crawl page, description 15
- crawl rate, monitoring 176
- crawl space
  - alerts about 188
  - description 3
  - editing 34
  - Web crawler configuration 56
- crawl.rules file 66
- crawler history reports
  - creating 177
  - description 174

- crawler history reports (*continued*)
  - HTTP return code report 177
  - Site report 177
- crawler properties
  - description 3
  - editing 34
- crawler types
  - combining in a collection 31
  - selecting base values for 31
- crawlers
  - base values for 31
  - combining crawler types 31
  - configuration overview 31
  - Content Manager 36
  - creating 33
  - Data Listener applications 70
  - DB2 41
  - default migration settings 162
  - deleting 35
  - description 3
  - document types 71
  - document-level security 147
  - editing crawl spaces 34
  - editing crawler properties 34
  - enabling document-level security 31
  - Exchange Server 42
  - initial values for 33
  - monitoring 173
  - NNTP 43
  - Notes 44
  - scheduling 31
  - starting 165, 173
  - stopping 167, 173
  - support for external 10, 70
  - system status 173
  - UNIX file system 49
  - URI formats 71
  - VeniceBridge 50
  - Web 54
  - Windows file system 69
- creating
  - collections 25, 26
  - crawlers 33
  - HTML search fields 87
  - quick links 122
  - rule-based categories 83
  - scopes 110
  - Web crawler reports 177
  - XML search fields 85
- custom text analysis
  - analysis engines 89, 90
  - description 88
  - mapping analysis results to the index 93
  - mapping common analysis structures 93
  - mapping XML elements 91

**D**

- data flow, enterprise search system 12
- Data Listener
  - configuring 70
  - monitoring 185
  - restarting 70, 185
- Data Listener API 1, 10

- data source types
  - DB2 Content Manager item types 36
  - DB2 databases 41
  - Exchange Server public folders 42
  - NNTP news groups 43
  - Notes databases 44
  - relational databases 41
  - support for external 1, 10
  - supported by enterprise search 1
  - UNIX file systems 49
  - VeniceBridge item classes 50
  - Web sites 54
  - Windows file systems 69
- DB2 crawlers
  - configuring 41
  - URI formats 71
- DB2 II OmniFind Edition
  - administration console 8
  - APIs 10
  - components 2
  - Crawler servers 3
  - data flow diagram 12
  - Index servers 5
  - integration with WebSphere Portal 155
  - messages 201
  - overview 1
  - parsers 4
  - portlet deployment 157
  - search applications 11
  - Search servers 7
- Default search application 128
- deleting
  - collections 29
  - crawlers 35
  - indexes from the queue 183
- DIC files 120
- DIIOP protocol, Notes crawler
  - configuration 49
- disabling index schedules 105
- discovery 3
- document importance
  - enabling for a collection 26
  - in migrated collections 160
- Document Search portlet 157
- document types
  - defaults for parser services 98
  - for parser services 96
  - for Stellent sessions 98, 101
  - parsing 96
- document-level security
  - crawler configuration 31
  - current credential validation 148
  - description 139, 147
  - identity management 150
  - indexed access controls 147
  - LDAP server configuration 151
  - Portal Search Engine support 157
  - real time validation 148
  - security tokens 147
  - user profiles 149
- domain rules for Web crawlers 56
- duplicate document analysis
  - description 139
  - enabling security 144
  - global analysis 144
- dynamic ranking 135

- dynamic summarization 136

## E

- e-mail notifications
  - for alerts 192
  - for messages 192
  - SMTP server configuration 191
- editing
  - collections 28
  - crawl spaces 34
  - crawler properties 34
  - Data Listener applications 70
  - search application properties 128
- enabling index schedules 105
- enterprise applications in WebSphere
  - Application Server 143
- enterprise search
  - administration console 8
  - administrative roles 140
  - APIs 10
  - backing up 197
  - collection-level security 144
  - components 2
  - configuring Portal Search Engine portlets for 157
  - crawler servers 3
  - data flow diagram 12
  - document-level security 147
  - Index servers 5
  - integration with WebSphere Portal 155
  - log files 187
  - messages 201
  - monitoring 169
  - overview 1
  - parsers 4
  - portlet deployment 157
  - portlets 155
  - restoring 197
  - restoring from a backup 198
  - restoring system files 198
  - search applications 11
  - Search servers 7
  - security 139
  - starting the components 165
  - stopping the components 165, 167
  - URI formats 71
- enterprise search administrator
  - changing the password on a single server 19
  - changing the password on multiple servers 20
  - description 140
  - role configuration 141
- error messages
  - enterprise search 201
  - receiving e-mail for 190, 192
  - SMTP server configuration 191
  - viewing log files 194
- ES\_INSTALL\_ROOT, description 19, 20
- ES\_NODE\_ROOT, description 19, 20
- es.cfg file 19, 20
- ESAdmin application
  - disabling security 143
  - logging in to 18
- esbackup.bat script 197, 198

- esbackup.sh script 197, 198
- eschangepw script 19, 20
- escrcm.sh script 37, 38
- escrcm.vbs script 40
- escrnote.sh script 45
- escrnte.vbs script 47
- escrvbr.sh script 52
- escrvbr.vbs script 53
- esrestore.bat script 198
- esrestore.sh script 198
- ESSearchApplication application
  - disabling security 143
  - enabling security 130
  - properties file 128
  - restarting 128, 130
- ESSearchServer application 143
- essyndictbuilder.bat script 120
- essyndictbuilder.sh script 120
- estimating system resources 169
- Exchange Server crawlers
  - configuring 42
  - URI formats 71
- external crawlers
  - configuring 70
  - Data Listener API 10
  - Data Listener applications 70

## F

- federated collections 25
- fielded queries 133
- file extensions
  - excluding from Web crawl spaces 56
  - supported by collection parsers 96, 98
  - supported by Stellent sessions 98, 101
- followindex.rules file
  - configuring 68
  - description 68
- form-based authentication 63, 64
- free space alerts 189
- free-text queries 133

## G

- global analysis
  - anchor text analysis 139, 145
  - description 5
  - duplicate document analysis 139, 144
- global Web crawl space 66
- global.rules file 66

## H

- HTML documents, searching 87
- HTML search fields
  - creating 87
  - description 87
  - mapping elements to 87
- HTTP basic authentication 63
- HTTP proxy servers 62
- HTTP return codes
  - received by Web crawlers 177
  - Web crawler report 177

## I

- identity management 150
- index builds
  - starting 182
  - stopping 182, 183
  - system status 183
- Index page, description 15
- index queue 183
- index refresh
  - description 5, 103
  - scheduling 104, 105
- index reorganization
  - description 5, 103
  - scheduling 104, 105
- indexes
  - alerts about 188
  - anchor text 145
  - changing the schedule 105
  - collapsed URIs 107, 111, 112
  - concurrent builds 106
  - deleting from the queue 183
  - description 5, 103
  - disabling the schedule 105, 182
  - enabling the schedule 105, 182
  - monitoring 182, 183
  - removing URIs 107, 113
  - scheduling 104
  - scopes 107, 110
  - starting 165
  - stopping 167
  - wildcard characters 107, 108, 109
- IP address rules for Web crawlers 56

## J

- Japanese
  - n-gram segmentation 95
  - parsing new line characters 95
- Java connector for Content Manager 37, 38, 40
- JavaScript support in Web crawlers 56

## K

- keywords in quick links 121, 122
- Korean, n-gram segmentation 95

## L

- LDAP servers
  - Active Directory Server 150
  - configuration 151
  - importing schemas 150
  - support for identity management 150
  - Tivoli Directory Server 150
  - user profiles 149, 151
- ldapadd command 150
- ldifde command 150
- limiting the Web crawl space 56
- linguistic support, synonym dictionaries 116
- Linux
  - configuring Content Manager crawlers 38
  - configuring Notes crawlers 45

Linux (*continued*)

- configuring VeniceBridge crawlers 52
- log files
  - default location 187
  - description 187
  - e-mail options 192
  - filtering 194
  - maximum size 190
  - migration wizard 163
  - monitoring 194
  - rotating 190
  - severity levels 190
  - SMTP server configuration 191
  - viewing 194
- Log page, description 15
- logging in to the administration console 18

## M

- mapping
  - common analysis structures 93
  - HTML search fields 87
  - XML elements to common analysis structures 91
  - XML search fields 85
- maximum crawl depth 56
- maximum recrawl interval 60
- messages for enterprise search 201
- migrating
  - collections 160
  - model-based taxonomy 159
  - rule-based taxonomy 160
- migration wizard
  - collections 160
  - default collection settings 162
  - default crawler settings 162
  - description 159
  - log file 163
  - model-based taxonomies 159
  - rule-based taxonomies 160
  - starting 159, 160
- MIME types, including in Web crawl spaces 56
- minimum recrawl interval 60
- model-based categories
  - description 80
  - selecting the categorization type 82
- model-based taxonomy, migrating from WebSphere Portal 159
- monitor
  - description 140
  - role configuration 141
- Monitor view, description 15
- monitoring
  - collections 171
  - crawlers 173
  - Data Listener 185
  - enterprise search 169
  - log files 194
  - parsers 181
  - popular queries 184
  - recent queries 184
  - response time history 184
  - Search servers 184
  - Web crawler active sites 175
  - Web crawler crawl rate 176

- monitoring (*continued*)
  - Web crawler thread details 175
  - Web crawlers 174

## N

- n-gram segmentation 95
- native XML search 94
- NNTP crawlers, configuring 43
- no-follow directives
  - configuring 68
  - description 68
- no-index directives
  - configuring 68
  - description 68
- Notes crawlers
  - configuring 44
  - DIIOP protocol configuration 49
  - NRPC protocol 45, 47
  - setting up on AIX 45
  - setting up on Linux 45
  - setting up on Windows 47
  - URI formats 71
- NRPC protocol, Notes crawler configuration 45, 47

## O

- operator
  - description 140
  - role configuration 141

## P

- parametric queries 133
- Parse page, description 15
- parsers
  - data analysis tasks 4
  - description 4, 77
  - document types for parser services 96, 98
  - document types for Stellent sessions 98, 101
  - monitoring 181
  - n-gram segmentation 95
  - native XML search 94
  - new line characters 95
  - parsing document types 96
  - starting 165, 181
  - stopping 167, 181
  - system status 181
  - white space 95
- parserTypes.cfg file 96
- password-protected Web sites
  - form-based authentication 64
  - HTTP basic authentication 63
- password, enterprise search administrator 19, 20
- popular queries, monitoring 184
- Portal Search Engine
  - configuring for enterprise search 157
  - description 155
  - document-level security 157
  - passing security tokens 157
- portlets
  - catalog Web site 157

- portlets (*continued*)
  - deploying 157
  - description 155
  - enterprise search 155
  - supported versions of WebSphere Portal 155
  - Taxonomy Management Portlet 159
- prefix rules for Web crawlers 56
- proxy servers 62

## Q

- queries
  - advanced techniques 133
  - fielded search 133
  - free-text search 133
  - parametric 133
  - spell checking 133
- query validation 148
- quick links
  - creating 122
  - description 121
  - searching 121
  - URI formats 71

## R

- ranking
  - dynamic 135
  - static 136
  - text-based scoring 135
- recent queries, monitoring 184
- recently crawled URLs, monitoring 174
- recrawl intervals for Web crawlers 60
- refreshing indexes 5, 103
- removing URIs from an index 113
- reorganizing indexes 5, 103
- response time history, monitoring 184
- restore scripts
  - description 197
  - running 198
- restoring
  - from a backup 198
  - system files to new servers 198
- revisiting URLs as soon as possible 60
- Robots Exclusion protocol 55
- robots.txt file 55
- rule-based categories
  - creating 83
  - description 79
  - selecting the categorization type 82
- rule-based taxonomy, migrating from WebSphere Portal 159, 160
- runtime-generic.properties file 137

## S

- sample search application
  - accessing 129
  - default deployment 128
  - description 11, 126
  - disabling security 143
  - enabling security 130
  - property file 128
  - search functions 125, 126
  - starting 129

- sample search application (*continued*)
  - WebSphere global security 130
- scheduling
  - crawlers 31
  - index builds 104, 105
- schemas, LDAP 150
- scopes
  - creating 110
  - description 110
  - searching 110
  - URI formats 71
  - URI patterns 110
- scripts
  - esbackup.bat 197, 198
  - esbackup.sh 197, 198
  - escrm.sh 37, 38
  - escrm.vbs 40
  - escrnote.sh 45
  - escrnte.vbs 47
  - escrvbr.sh 52
  - escrvbr.vbs 53
  - esrestore.bat 198
  - esrestore.sh 198
  - essyndictbuilder 120
  - startcl 198
- Search and Index API 10, 125
- search applications
  - application IDs 146
  - associating with collections 126
  - collection-level security 146
  - custom 125
  - description 11
  - sample 125, 126
  - synonym support 118
- search cache
  - configuring 116
  - description 116
- Search Center for WebSphere Portal
  - description 155
  - registration portlet 155
- Search page, description 15
- search response time
  - alerts about 188
  - monitoring 184
- search results
  - collapsing 111, 112
  - customizing summaries 137
  - description 133
  - dynamic ranking 135
  - dynamic summarization 136
  - grouping 111, 112
  - static ranking 136
  - summaries 137
  - text-based scoring 135
  - wildcard expansion 108, 109
- Search servers
  - associating synonym dictionaries 120, 121
  - creating synonym dictionaries 120
  - description 7, 115
  - monitoring 184
  - popular queries 184
  - recent queries 184
  - response time history 184
  - search cache 116
  - starting 165, 184
  - stopping 167, 184

- Search servers (*continued*)
  - synonym dictionaries 116
  - synonym XML files 119
  - system status 184
- searching
  - categories 78
  - collections 133
  - HTML documents 87
  - quick links 121
  - XML documents 84, 85, 91
- security
  - access controls 142
  - administrative roles 141
  - anchor text analysis 145
  - authentication 142, 143
  - bypassing document-level access controls 153
  - collection-level 144, 153
  - description 139
  - document-level 147, 148, 151, 153
  - duplicate document analysis 144
  - enabling for a collection 26, 139
  - enterprise applications in WebSphere
    - Application Server 143
    - LDAP schemas 150
    - LDAP servers 150, 151
    - sample search application 130
    - search application IDs 146
    - user profiles 149
    - WebSphere global security 143
  - security tokens
    - crawler configuration 147
    - disabling for a collection 153
    - document-level security 147, 157
    - Portal Search Engine processing 157
- Security view, description 15
- seed URLs for Web crawlers 56, 60
- semantic search 88, 91
- SIAPI (Search and Index API) 10, 125
- site history reports
  - creating 177
  - description 174
- SMTP server configuration 191
- soft error pages, Web crawlers 61
- spell checking in queries 133
- startccl script 198
- starting
  - crawlers 173
  - Data Listener 185
  - enterprise search components 165
  - index builds 182
  - migration wizard 159, 160
  - parsers 181
  - sample search application 129
  - Search servers 184
- static ranking
  - description 136
  - enabling for a collection 26
  - in migrated collections 160
- Stellent sessions
  - associating document types 98
  - default document types 101
  - parsing document types 96
- stellent.properties file 98
- stellentypes.cfg file 98
- stopping
  - crawlers 173

- stopping (*continued*)
  - enterprise search components 165, 167
  - index builds 182, 183
  - parsers 181
  - Search servers 184
- summaries
  - customizing 137
  - dynamic 136
- synonym dictionaries
  - associating with a collection 121
  - associating with the system 120
  - creating a DIC file 120
  - creating an XML file 119
  - description 116
  - search application support 118
- synonyms.xml file 159
- system backup 197
- system resources
  - checking 170
  - estimating 169
- system restore 197, 198
- system status
  - collections 171
  - crawlers 173
  - index builds 183
  - parsers 181
  - Search servers 184
  - Web crawlers 174
- System view, description 15

## T

- task summary, administration console 15
- taxonomies, migrating from WebSphere Portal 159, 160
- Taxonomy Management Portlet 159
- tds.add.ldif file 150
- text analysis
  - analysis engines 88, 89, 90
  - annotators 88
  - common analysis structures 88, 93
  - mapping XML elements 91
- text-based scoring 135
- thread details, monitoring 174, 175
- titles.xml file 159
- Tivoli Directory Server 150
- treenodes.xml file 159

## U

- UIMA
  - associating analysis engines 89, 90
  - common analysis structures 93
  - description 88
  - mapping analysis results to the index 93
  - mapping XML elements 91
- UNIX file system crawlers
  - configuring 49
  - URI formats 71
- URIs
  - category rules 79, 83
  - collapsed in search results 111, 112
  - formats in enterprise search 71

- URIs (*continued*)
  - quick links 121, 122
  - removing from an index 113
  - scopes 110
  - viewing details about 172
- user agents 55
- user profiles 149

## V

- validation of current credentials 148
- vbr\_access\_services.jar file 52, 53
- VeniceBridge crawlers
  - configuring 50
  - setting up on AIX 52
  - setting up on Linux 52
  - setting up on Windows 53
  - URI formats 71
- viewing
  - log files 194
  - URI details 172
- visiting URLs as soon as possible 60

## W

- Web crawlers
  - active sites 174, 175
  - configuring 54
  - cookie configuration 66
  - cookie format 65
  - cookies 65
  - crawl rate 176
  - crawler history 174
  - crawling rules 56
  - creating reports about 177
  - followindex.rules file 68
  - global crawl space 66
  - JavaScript support 56
  - limiting the crawl space 56
  - monitoring 174
  - no-follow directives 68
  - no-index directives 68
  - password-protected Web sites 63, 64
  - proxy servers 62
  - recently crawled URLs 174
  - recrawl intervals 60
  - seed URLs 56, 60
  - site history 174
  - soft error pages 61
  - system status 174
  - thread details 174, 175
  - URL status 174
  - user agents 55
  - visiting URLs as soon as possible 60
- WebSphere global security
  - disabling 143
  - search application properties 128
- WebSphere Portal
  - catalog Web site 157
  - category tree migration 159
  - collection migration 159
  - default migration settings 162
  - integration with enterprise search 155
  - model-based taxonomies 159
  - portlet deployment 157

- WebSphere Portal (*continued*)
  - Taxonomy Management Portlet 159
  - taxonomy migration 159
- WebSphere Portal Search Center 155
  - description 155
  - registration portlet 155
- wildcard characters
  - index expansion 108, 109
  - query expansion 108, 109
- Windows file system crawlers
  - configuring 69
  - URI formats 71
- Windows operating system
  - configuring Content Manager crawlers 40
  - configuring Notes crawlers 47
  - configuring VeniceBridge crawlers 53
- WpsMigratorLog.log file 163

## X

- XML documents, searching 85
- XML elements
  - mapping to common analysis structures 91
  - searching 84, 91
- XML fragments, native XML search 94
- XML query syntax, native 94
- XML search fields
  - creating 85
  - description 84, 91
  - mapping elements to 84, 85, 91
- XPath, native XML search 94







Printed in USA



**Java**<sup>™</sup>  
**COMPATIBLE**

SC18-9283-01





Spine information:



DB2 II OmniFind Edition

Administering Enterprise Search

Version 8.2.2