

IBM DB2 Information Integrator  
OmniFind Edition



# Release Notes

*Version 8.2.2*



IBM DB2 Information Integrator  
OmniFind Edition



# Release Notes

*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

<b>About the release notes . . . . .</b>	<b>1</b>	Changes to error messages . . . . .	13
1 Technical support information on the Web . . . . .	1	Updates to query syntax . . . . .	13
<b>New in this release . . . . .</b>	<b>3</b>	Collapsing URIs . . . . .	13
1 Enterprise search base annotator package. . . . .	4	Enabling direct document access for FileNet Content Services . . . . .	14
<b>Known limitations, problems, and workarounds . . . . .</b>	<b>7</b>	Generic crawler plug-in . . . . .	14
1 Document size limits . . . . .	7	Help for the My Profile page does not match the user interface. . . . .	15
1 Fix pack installation results in a full recrawl. . . . .	7	Help pages for the Web crawler include incorrect values . . . . .	15
1 Heap size limits for Java parser annotator . . . . .	7	indexBuildItem description . . . . .	16
1 Maximum size of processing engine archive files . . . . .	8	Mapping a common analysis structure to the index . . . . .	16
1 Notes crawler on AIX requires the I/O Completion Port module . . . . .	9	Migrating a model-based taxonomy from WebSphere Portal . . . . .	16
1 Querying secure Notes collections might result in inaccurate result document counts. . . . .	10	<b>Notices . . . . .</b>	<b>19</b>
1 User names can contain periods . . . . .	10	Trademarks . . . . .	21
1 Wild card characters in a fielded search . . . . .	11	<b>Contacting IBM . . . . .</b>	<b>23</b>
1 Index location on the server . . . . .	11	Obtaining product information . . . . .	23
1 Orphan URLFetcher process for Linux and AIX . . . . .	11	Providing comments on the documentation. . . . .	23
1 Shared memory segments in AIX . . . . .	12		
<b>Documentation updates . . . . .</b>	<b>13</b>		



---

## About the release notes

The release notes document describes what is new in this release, known limitations, problems, workarounds, and documentation updates for IBM® DB2® Information Integrator OmniFind™ Edition, Version 8.2.2 (version 8.2 plus Fix Pack 2).

DB2 II OmniFind Edition, Version 8.2.2 is an upgrade only. You must have version 8.2 or version 8.2.1 before you can install version 8.2.2.

Before you install on SuSE Linux® Enterprise Server 9, check the DB2 II OmniFind Edition, Version 8.2.2 Support Web site for workarounds and updates.

---

### 1 Technical support information on the Web

1 The DB2 II OmniFind Edition product was recently renamed WebSphere®  
1 Information Integrator OmniFind Edition. You might see references to WebSphere  
1 Information Integrator OmniFind Edition on product-related Web pages, but the  
1 product interfaces, documentation, and troubleshooting information (technotes) still  
1 reflect the DB2 brand.

1 You can find technical support information on the following Web sites:

1 **IBM DB2 Information Integrator OmniFind Edition**

1 [www.ibm.com/software/data/integration/db2ii/supportomni.html](http://www.ibm.com/software/data/integration/db2ii/supportomni.html)

1 **IBM WebSphere Information Integrator**

1 <http://www.ibm.com/software/data/integration/db2ii/support.html>

1 **IBM WebSphere Application Server**

1 <http://www.ibm.com/software/webservers/appserv/was/support/>

1 **IBM DB2 Universal Database™ for Linux, UNIX®, and Windows® Version 8  
1 products**

1 [www.ibm.com/software/data/db2/udb/winos2unix/support](http://www.ibm.com/software/data/db2/udb/winos2unix/support)



---

## New in this release

IBM DB2 Information Integrator OmniFind Edition, Version 8.2.2 provides the following features:

### Operating systems

Version 8.2.2 supports the following additional operating systems:

- AIX® 5.3 (requires Maintenance Level 1 and the August 2004 C++ Runtime for AIX PTF)
- SuSE Linux Enterprise Server 9 with Service Pack 1 (UnitedLinux SP1) (requires libstdc++-3.3.3-43.28)

Before you install on SuSE Linux Enterprise Server 9, check the DB2 II OmniFind Edition, Version 8.2.2 Support Web site for workarounds and updates.

### Large documents

Collections can include documents that are up to 32 MB in size

### Document-level security for Domino® documents

- Real time validation (during query processing) of a user's credentials with access controls that are managed by the native repository
- User validation through user profiles and an LDAP user registry. By storing credentials in an encrypted format, users do not need to specify credentials each time they query a Domino source.

### Crawlers

- Notes® crawler: the ability to crawl directories in addition to databases, views, and folders
- Web crawler:
  - User interface support for crawling password-protected Web sites (for sites that use HTTP basic authentication or form-based authentication)
  - The ability to crawl Web sites that are served by proxy servers
  - The ability to return custom messages for soft error pages

### Search quality

- Federated searching (the ability to search more than one collection at a time)
- Support for trailing wild card characters in queries
- Support for customizing the display length of summaries in the search results
- User interface support for collapsing search results from the same URI

### Linguistic analysis

- Support for unstructured information management architecture (UIMA), which provides the ability to do custom text analysis on the documents in a collection
- Support for native XML search
- Support for custom synonym dictionaries

### API and search application

- User interface support for configuring and monitoring custom client data listener applications

- A generic crawler plug-in, enabling the specification of custom security and business rules
- Many enhancements to the sample search application to demonstrate query functions, search result options, federated search, and document-level security

---

## Enterprise search base annotator package

You can use the enterprise search base annotator package to develop new annotators based on the output of the enterprise search annotators, and to test both the enterprise search annotators and your custom annotators within the UIMA Software Development Kit (SDK).

The enterprise search base annotator package is a ZIP file that contains the text analysis steps that are used in enterprise search. These annotators will always run prior to any custom analysis when parsing documents.

The UIMA Software Development Kit (SDK) is available on the IBM developerWorks Web site:  
<http://www.ibm.com/developerworks/db2/zones/db2ii/>.

To install the annotator package:

1. Find the annotator package `OF_base_annotators.zip` in your enterprise search (DB2 II OmniFind Edition) installation in the `ES_INSTALL_ROOT/packages/uima` directory.
2. Copy the ZIP file to the root directory of your UIMA SDK installation.
3. Extract the ZIP file to add the enterprise search base annotator files to the specified directory structure of your UIMA SDK installation.

The enterprise search base annotator package contains only the dictionaries that are required to process English documents. If you want to process other languages in your development environment, follow these steps:

1. Find the enterprise search dictionaries in your enterprise search installation in `ES_INSTALL_ROOT/configurations/parserservice/jediidata/frost4`.
2. Copy the contents of the directory to your local UIMA SDK installation in `UIMA_SDK_INSTALL/data/frost4`.

To verify that UIMA SDK was successfully installed:

1. Open the CAS Visual Debugger (CVD) in the following directory:  
`UIMA_SDK_INSTALL/bin/cvd[.bat/.sh]`.
2. Click **Run** → **load TAE**.
3. Find the TAE specifier file `of_tokenization.xml` in the `UIMA_SDK_INSTALL/docs/examples/descriptors/analysis_engine` directory.
4. Load a sample document and run the TAE. You will see annotations of type `uima.tt.TokenAnnotation` in the CVD.

To use the enterprise search annotators for your processing:

1. Include a reference to the file `of_typesystem.xml` in the `typeSystem` section of your custom annotator specifier if your custom annotator uses types that are defined by enterprise search annotators. The `of_typesystem.xml` file is in the `UIMA_SDK_INSTALL/docs/examples/descriptors/analysis_engine` directory. See the file `jtok.xml` in the `analysis_engine` directory for a sample of how to include references to files.

2. Create an aggregate TAE specifier that includes the file `of_tokenization.xml` and your custom annotator specifier.
3. Use the aggregate TAE specifier when testing within the UIMA SDK.

When you create a PEAR file to upload your annotator to enterprise search, ensure that this PEAR file includes only your custom annotator specifier. It might not include any of the following files: `of_tokenization.xml`, `of_typesystem.xml`, the contents of `of_base_annotators.zip`, or the aggregate specifier created in step 2.

For more information about the analysis capabilities in enterprise search, see the "Custom Text Analysis Integration" and "Text Analysis included in enterprise search" topics in the information center or in *Programming Guide and API Reference for Enterprise Search*.



---

## Known limitations, problems, and workarounds

This information describes known limitations, problems, and workarounds for DB2 Information Integrator OmniFind Edition, Version 8.2.2.

Any limitations and restrictions might or might not apply to other releases of the product.

---

### Document size limits

Documents up to 32 MB can be crawled and indexed. DB2 Universal Database tables with LOB data can be crawled and indexed only if the data adheres to the 32 MB overall document size limit.

Ignore statements in the crawler property help files that specify that the maximum page size can be larger than 32 MB.

---

### Fix pack installation results in a full recrawl

After you install DB2 II OmniFind Edition, Version 8.2.2, all documents except for those that were crawled by Web and NNTP crawlers will be recrawled.

The first time that you start the non-Web and non-NNTP crawlers after you install version 8.2.2, all documents in all data sources will be crawled and indexed again, not only documents that are new or that changed since the last time a data source was crawled.

After crawling all documents in each table, directory, or folder in the collection, the crawlers switch to the normal update processing and begin crawling only new or changed documents.

Usually, a full recrawl is necessary only when changes to the crawler's configuration require it. For example, if the display names for fields or columns in a data source change, then the crawler must recrawl all documents so that the new information can be indexed with the documents.

The full recrawl that occurs automatically after you install version 8.2.2 incurs time, but it does not impact the quality of the enterprise search index.

---

### Heap size limits for Java parser annotator

If OutOfMemory messages are displayed in the collection error log for the *collection\_ID.casprocessor* session running the Java parser annotator, increase the heap size to 200 MB.

Custom text analysis annotators that are uploaded to enterprise search are run in a Java session. For example, in enterprise search, these sessions are *collection\_ID.casprocessor* sessions, where *collection\_ID* is the collection ID that this annotator belongs to. These sessions have a heap size limit of 200 MB.

To increase the heap size:

1. Stop the enterprise search system. On all operating systems, enter the following command: `esadmin stop`
2. Stop CCL:
  - On Linux and AIX, enter the following command: `stopccl.sh`
  - On Windows, stop the DB2 Information Integrator OmniFind Edition Windows service: **Start** → **Control Panel** → **Services**.
3. Find the collection ID that your annotator belongs to. To find the collection ID, either run the command `esadmin check` or go to `ES_NODE_ROOT/master_config` directory where you will find a `collection_ID_config.ini` file for each collection that you created.
4. Increase the `max_heap` value for the `collection_ID.casprocessor` session in the `collection_ID_config.ini` file.
5. Sort the `collection_ID_config.ini` file and look for the CAS processor entries. For example on Linux, for `col7`:
 

```
% cd $ES_NODE_ROOT
% sort col7_config.ini
...

session3.collectionid=col7
session3.configDir=col7.casprocessor
session3.description=Fenced CAS Processor Session
session3.displayname=Fenced CAS Processor Server
session3.domain=.
session3.flags=0
session3.id=col7.casprocessor
session3.max_heap=200

...

```
6. Edit the line `session_number.max_heap=200` to include the new maximum heap size (200). For example, for `col7`: `session3.max_heap=200`
7. Restart CCL:
  - On Linux and AIX, enter the following command: `startccl.sh -bg`
  - On Windows, start the DB2 Information Integrator OmniFind Edition Windows service: **Start** → **Control Panel** → **Services**.
8. Restart enterprise search. On all operating systems, enter the following command: `esadmin start`
9. Start the parser for this collection from the enterprise search administration console. The annotator will be run on a JVM with the new maximum heap size.

---

## Maximum size of processing engine archive files

For custom text analysis, the maximum size of processing engine archive files (PEAR files) is 8 MB.

Only PEAR files that are smaller than 8 MB can be uploaded using the enterprise search administration console. PEAR files that are larger than 8 MB often contain resources, such as dictionaries. To upload such a PEAR file:

1. Change your annotator code such that it loads the resource either through the UIMA resource manager (using the `getResourceObject` method of the `AnnotatorContext`) or such that it uses the value of an annotator configuration parameter that points to the resource.
2. Copy the resource files into a dedicated directory to the Indexer node of your enterprise search system.

3. Change your annotator configuration (either the <externalResource> specification or the <namevaluePair> of the configuration parameter that you use) to include the absolute path to the resource by using the directory that you created in step 2.
4. Re-create your PEAR file with the updated configuration. Ensure that it does not contain the resources and upload it to enterprise search.

---

## Notes crawler on AIX requires the I/O Completion Port module

To use the Notes crawler and discovery processes on IBM AIX platforms, you must install the I/O Completion Port (IOCP) module (bos.iocp.rte) from the AIX product CD-ROM and set its status to **Available**.

After you install the bos.iocp.rte module, you must apply the appropriate fix for it. See the information at the following link for instructions:

<http://www-1.ibm.com/support/docview.wss?uid=swg21086556>

Without the IOCP module, the Notes discovery process will fail during the creation of a new crawler, and the following error message will be displayed:

```
FFQM0105E Recieved an error from the server -
Message: FFQG0024E An unexpected exception was caught: discover
```

In addition, the following message, which includes the ENOEXEC error, will be written to the *ES\_NODE\_ROOT/logs/system\_yyyymmdd.log* file. (Some of the message text is split onto multiple lines to improve readability.)

```
5/20/05 18:08:52.423 JST [Error] [ES_ERR_EXCEPTION_DEFAULT_MESSAGE] [] [discovery]
dominoserver.ibm.com:0:2108088751:control:ComponentDiscoveryW.java:
com.ibm.es.control.discovery.server.ComponentDiscoveryW.discover:86
FFQ00277E An exception was caught with the detail 'java.lang.UnsatisfiedLinkError:
/opt/lotus/notes/65010/ibmpow/liblsxbe_r.a:
load ENOEXEC on shared library(s) /opt/lotus/notes/latest/ibmpow/libnotes_r.a'
and a stack trace of 'java.lang.UnsatisfiedLinkError:
/opt/lotus/notes/65010/ibmpow/liblsxbe_r.a:
load ENOEXEC on shared library(s) /opt/lotus/notes/latest/ibmpow/libnotes_r.a
at java.lang.ClassLoader$NativeLibrary.load(Native Method)
at java.lang.ClassLoader.loadLibrary0(ClassLoader.java:2120)
at java.lang.ClassLoader.loadLibrary(ClassLoader.java:1998)
at java.lang.Runtime.loadLibrary0(Runtime.java:824)
at java.lang.System.loadLibrary(System.java:908)
at lotus.domino.NotesThread.load(NotesThread.java:306)
at lotus.domino.NotesThread.checkLoaded(NotesThread.java:327)
at lotus.domino.NotesThread.sinitThread(NotesThread.java:181)
at com.ibm.es.crawler.discovery.notes.NotesLibrary$NotesOperation.discover
(Unknown Source)
at com.ibm.es.crawler.discovery.api.DiscoveryAPI.discover(Unknown Source)
at com.ibm.es.control.discovery.server.ComponentDiscoveryW.discover
(ComponentDiscoveryW.java:72)
at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:85)
at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:58)
at sun.reflect.DelegatingMethodAccessorImpl.invoke
(DelegatingMethodAccessorImpl.java:60)
at java.lang.reflect.Method.invoke(Method.java:391)
at com.ibm.es.ccl.sessionwrapper.CallThread.run(CallThread.java:77)
```

Ensure that the IOCP module is correctly installed before you create a Notes crawler:

1. Enter the following `ls|pp` command to ensure that the IOCP module is installed on crawler server:

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

The output from the `lspp` command is similar to the following example:

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

2. Enter the following `lsdev` command to ensure that the IOCP port is **Available**:

```
$ lsdev -Cc iocp
```

The output from the `lsdev` command should match the following example:

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

3. If the IOCP port status is **Defined**, do the following steps to change the status to **Available**:
  - a. Log in to the crawler server as the root user and issue the following command:

```
# smit iocp
```
  - b. Select **Change / Show Characteristics of I/O Completion Ports** and change **STATE to be configured at system restart** from **Defined** to **Available**.
  - c. Restart the crawler server.
  - d. Enter the `lsdev` command again and confirm that the status of the IOCP port is changed to **Available**.

---

## Querying secure Notes collections might result in inaccurate result document counts

The number of result documents in a result set might not be accurate for queries that search secure collections in which the option to validate current credentials during query processing is selected.

When you search an enterprise search collection, the total number of result documents is shown in the search results. This number might be greater than the actual number of result documents if all of the following conditions are true:

- Security was enabled for the collection when it was created.
- The collection includes a Notes crawler.
- The option to validate current credentials during query processing was specified when document level security options were configured for the Notes crawler.

This problem affects the `SIAPISearchResultSet.getAvailableNumberOfResults` value and the `ESSearchApplication` sample search application.

---

## User names can contain periods

The online instructions for adding enterprise search administrator names are incorrect.

When you use the enterprise search administration console to add an enterprise search administrator ID to the system, instructions on the page specify that the user name cannot contain a period. That instruction is incorrect. User names with periods (such as `j_user@this.company.com`) are allowed.

---

## Wild card characters in a fielded search

Using *field:pref\** as a query term is generally allowed. DB2 II OmniFind Edition will retrieve documents that include words beginning with *pref* marked as members of the field named *field*. There is an exception for this rule: for a few special fields for which wild card queries are less relevant, a query term *field:pref\** is treated as if it were *field:pref*. Those special fields are:

- site
- link
- docid
- samegroupas
- url
- group (newsgroup, which is similar to *site*, but it is used for NNTP)

For example, the query *url:porta\** might not return any results, whereas *url:portal* will return results.

---

## Index location on the server

If you create a collection and you specify an index location other than the default location on the server data root directory, the disk space usage for this nondefault index location will not be monitored.

For a multiple server installation, when you create a collection, if you want to specify an index location other than the default location, the index location must first exist on all four servers.

---

## Orphan URLFetcher process for Linux and AIX

**Problem:** The Web crawler uses two processes. When one process exits, it signals a second process to exit. One process might sometimes exit unexpectedly and leave the other process running indefinitely.

You can determine that this happened if you see an orphan process like the following process from the output of the UNIX `ps` command:

```
F S UID      PID PPID  C PRI  NI ADDR  SZ WCHAN  STIME TTY      TIME CMD
0 S user1    26222  1  0  75  0   - 152641 schedu 10:34 pts/8    00:00:00
```

```
/opt/IBMJava2-141/jre/bin/java -classpath ../../URLFetcher.jar:...
```

Notice that:

- The PPID (parent process ID) of 1 indicates that the process was orphaned.
- The presence of URLFetcher.jar in the classpath indicates that this is a Web crawler process.

### Solution

Find the process ID (PID) of the orphaned process and kill it. In the previous example, the orphan process is 26222. On Linux and AIX, use the `kill` command with no argument. You might also see “child” entries in the process table. The child entries represent threads that are created by the orphan process. These child entries should be deleted when you kill their ancestor process.

When you see an orphan URLFetcher process, you should kill the process because if it is in the process table, the operating system reserves resources for it. You should regularly check whether these orphan processes exist, especially if you suspect that the Web crawler has terminated abnormally. If the processes exist, kill them.

**Important:** In DB2 II OmniFind Edition Version 8.2.1, orphaned processes rarely occur because the URL fetcher now self-terminates when it notices that the C++ component is gone. If the problem does occur, take care not to kill another user's httpauth.ini editing session by accident. The HTTPAuthEdit utility is a JVM that runs from the URLFetcher JAR file, and might look similar in the process table.

---

## Shared memory segments in AIX

If you see message SQL1224N on AIX, you might have exceeded the number of shared memory segments that are allowed by AIX.

To resolve this problem, see the topic SQL1224N Shared Memory Segments on the IBM Support Web site.

---

## Documentation updates

Documentation updates include corrections or minor additions to the version 8.2.2 documentation for DB2 Information Integrator OmniFind Edition.

---

## Changes to error messages

Error messages were updated after the documentation was sent out for translation.

| **FFQD1034E**

| ERROR: The environment variable CMBROOT / IBMCMROOT is not defined.  
| Verify that DB2 Information Integrator for Content was installed  
| correctly.

| **Explanation:**

| The environment variable CMBROOT / IBMCMROOT 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 cmbicmsrvs.ini file was not found. Verify that DB2  
| Information Integrator for Content was installed correctly.

| **Explanation:**

| The cmbicmsrvs.ini file was not found in the specified path. DB2  
| Information Integrator for Content might not be installed.

| **User Response:**

| Verify that DB2 Information Integrator for Content was installed  
| correctly.

---

## Updates to query syntax

A query consisting of a single – (minus) term is invalid. A minus term can only be used as a post-filter, not to generate results.

The topic “Query syntax,” which describes the types of syntax that you can apply to search queries, is in the *Programming Guide and API Reference (iiysp.pdf)* and the information center.

---

## Collapsing URIs

Topics about how to collapse URIs in search results contain misleading and incorrect information.

- The documentation implies that URIs for all data source types can be collapsed in the search results. However, the ability to collapse URIs is limited to documents that have URLs (that is, documents that were crawled by a Web or NNTP crawler).

- The documentation also shows incorrect usage for the `samegroupas` query syntax. To use this keyword in a query, you must specify a fully qualified URL; you cannot specify the name of a group that was created to collapse multiple URIs together.

For example, to query documents that belong to the same group (such as a group named News), specify the URI of one of the documents in the group, not the group name:

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

Incorrect: `samegroupas:news`

---

## Enabling direct document access for FileNet Content Services

When you use the VeniceBridge crawler to crawl FileNet Content Services repositories that are managed by the DB2 Information Integrator Content Edition (formerly VeniceBridge) server, the target FileNet Content Services library and FileNet Web Services work with the FileNet Content Services connector and must be configured to use direct document access. See the FileNet Web Services *Administrator's Guide* and FileNet Content Services administration documentation to set up direct document access support.

---

## Generic crawler plug-in

When you configure crawler properties, you can specify the name of a Java™ class that you want to use to enforce document-level access control and to associate metadata with documents in an enterprise search index. You cannot use a generic crawler plug-in with Web or NNTP crawlers.

The crawler plug-in is implemented in the product, but it is not described in the enterprise search documentation.

The generic crawler plug-in covers all the functions of the security token plug-in that was available in version 8.2 of DB2 II OmniFind Edition. Although the security token plug-in is still supported, it is recommended that you use the generic crawler plug-in instead.

To use the plug-in, you must create a new collection. You cannot use a new plug-in with collections that you previously created.

The plug-in consists of a Java class that is called for each document that the crawler crawls. The Java class is passed the document identifier (URI), assigned security tokens, and metadata. The class returns a new or modified set of security tokens and metadata, or the class can indicate that the document is to be ignored, or skipped.

After all documents in the crawl space have been crawled, the plug-in is called for only new or changed documents. If you want to change the security tokens and metadata for documents that have not been updated, recrawl all documents in the crawl space and reorganize the index.

### Using the plug-in to enforce security

Document-level security is accomplished by associating one or more security tokens (a comma-delimited string) with each document that a crawler crawls. Group identifiers are commonly used as the security tokens.

By default, each document is assigned a public token that makes the document available to everyone. The public token can be replaced with a value that is provided by the administrator or a value that is extracted from a field in the crawled document.

The plug-in allows you to apply your own business rules to determine the value of the security tokens for crawled documents. The security tokens that are associated with each document are stored in the index. They are used to filter documents that match the security tokens and ensure that only the documents that a user has permission to view are returned in the search results.

#### Using the plug-in to add metadata

Document metadata, such as last modified date data or the field data of Lotus Notes® documents, is created for all crawled documents. The crawler plug-in allows you to apply your own business rules to determine the value of the metadata that should be indexed for each document.

The metadata is defined as a name-value pair. You can search the metadata with a free-text query or with a query that specifies the field name.

See the DB2 II OmniFind Edition Support Web site for more information about this topic and for information about sample plug-in code that you can modify with your own logic.

---

## Help for the My Profile page does not match the user interface

The user interface for the sample search application was updated after the documentation was translated.

On the My Profile page in the ESSearchApplication search application, the **Data source credentials** and **Data source name** labels were changed to **Domain Credentials** and **Domain name**, respectively. The help file for this page still reflects the old label text.

---

## Help pages for the Web crawler include incorrect values

Incorrect default values are specified in the online help for configuring a Web crawler.

Refer to the following corrections when you configure a Web crawler.

#### Web Crawler Properties page

- In **Number of concurrent threads for retrieving documents**, the online help states that the default and maximum value is 200. The correct default value is 50. The maximum value is 200.
- In **Maximum page size**, the online help states that acceptable values range from 250 KB to 10000 KB. The maximum value is 32 MB.

#### Advanced Web Crawler Properties page

- In **Maximum number of URLs**, the online help states that the default value is 2 000 000 000 URLs. The correct default value is 3 000 000.
- In **Maximum number of active hosts**, the online help states that the default value is 200. The correct default value is 500.
- In **Maximum number of unique documents**, the online help states that the default value is 7 000 000. The correct default value is 1 000 000.

- In **Percentage of new documents to crawl**, the online help states that the default value is 99 percent. The correct default value is 75 percent.
- In **Maximum number of crawled documents in temporary storage**, the online help states that the default value is 5 000 000 documents. The correct default value is 500 000.

#### Web Crawl Space page

- In **URL path depth**, the online help states that default value is 25. The correct default value is 16.

---

## indexBuildItem description

In the *Programming Guide and API Reference*, there is a topic called "Writing the index build configuration file." The description for the subsection called "The <indexBuildItem> element" is incorrect.

The text should read as follows:

If type A is a subtype of type B (for example, ?Person? as a subtype to ?Entity?), and there are <indexBuildItem> elements Ia and Ib defined for both types, the processing is as follows:

- Each index rule that is defined in Ib will be applied to feature structures of type B and feature structures of type A.
- Each index rule that is defined in Ia will be applied to feature structures of type A.

With this processing, one rule can be defined that stores all "Entity" annotations under a span "entity," which applies also for "Person" annotations, and another rule for "Person" annotations that creates an additional span named "Person." Each rule is applied independently of other rules, so they can contain different <attributemapping> elements for the same features.

---

## Mapping a common analysis structure to the index

The XML file for mapping a common analysis structure to the index must use xmlns, not xmlns:namespace.

The XML file must start with the following statements:

```
<?xmlversion="1.0" encoding="UTF-8"?>
<indexBuildSpecification xmlns="http://www.ibm.com/of/822/consumer/index/xml">
```

Do not attach :namespace to the xmlns attribute. Otherwise, the error message FFQT0138E will occur when the parser is started.

Use <name> instead of <typeName> in the <indexBuildItem> element. The sample configuration contains an incorrect tag. It should read:

```
<indexBuildItem><name>com.ibm.uima.tt.PersonAnnotation</name>
```

instead of

```
<indexBuildItem><typeName>com.ibm.uima.tt.PersonAnnotation</typeName>
```

---

## Migrating a model-based taxonomy from WebSphere Portal

In *Administering Enterprise Search*, the topic called "Migrating a model-based taxonomy from WebSphere Portal" requires additional information.

When the migration utility runs for the first time, it needs the path to the WebSphere Portal installation. Then, it will update enterprise search configuration files. For these configuration updates to work, the enterprise search system must be stopped. Otherwise, the option to migrate model-based taxonomy files will be greyed out. For subsequent runs of the migration utility, the enterprise search system can be active.



---

## 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  
Hummingbird  
Informix  
Lotus  
Lotus Notes  
Notes  
OmniFind  
POWER4  
POWER5  
RISC System/6000  
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.



---

## 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).







Printed in USA



**Java**<sup>™</sup>  
**COMPATIBLE**