



FAST SMART CONNECTOR FOR DOCUMENTUM

version 4.0.1

MODULE GUIDE

Copyright

Copyright © 1997-2004 by Fast Search & Transfer, Inc. and its associated companies and licensors. All rights reserved. Fast Search & Transfer may hereinafter be referred to as FAST.

Information in this document is subject to change without notice. The software described in this document is furnished under a license agreement. The software may be used only in accordance with the terms of the agreements. No part of this document may be reproduced, stored in a retrieval system, or transmitted in any form or any means, electronic or mechanical, including photocopying and recording, for any purpose other than the purchaser's use, without the written permission of FAST.

Trademarks

FAST is a registered trademark of Fast Search & Transfer. All rights reserved.

FAST Search, and FAST Data Search are trademarks of Fast Search & Transfer. All rights reserved.

Sun, Sun Microsystems, all SPARC trademarks, Java and Solaris are trademarks or registered trademarks of Sun Microsystems, Inc. in the United States and other countries. All rights reserved.

Netscape is a registered trademark of Netscape Communications Corporation in the United States and other countries.

Windows, Visual Basic, and Internet Explorer are registered trademarks of Microsoft Corporation.

Red Hat is a registered trademark of Red Hat, Inc. All rights reserved.

Linux is a registered trademark of Linus Torvalds. All rights reserved.

UNIX is a registered trademark of The Open Group. All rights reserved.

AIX is a registered trademark of International Business Machines Corporation. All rights reserved.

HP and the names of HP products referenced herein are either trademarks and/or service marks or registered trademarks and/or service marks of HP and/or its subsidiaries.

Oracle is a registered trademark, and Oracle8 is a trademark of Oracle Corporation.

DB2, DB2 UDB, UDB, and MVS are all registered trademarks of the IBM Corporation.

Microsoft® is a registered trademark of Microsoft Corporation.

SQL Server 2000 is a trademark of Microsoft Corporation.

Documentum and all other Documentum product and service names and logos are either service marks, trademarks, or registered trademarks of Documentum, a division of EMC Corporation.

All other trademarks and copyrights referred to are the property of their respective owners.

Restricted Rights Legend

Software and accompanying documentation are provided to the U.S. government in a transaction subject to the Federal Acquisition Regulations with Restricted Rights. Use, duplication, or disclosure of the software by the government is subject to restrictions as set forth in FAR 52.227-19 Commercial Computer Software-Restricted Rights (June 1987).

Contents

FAST Support.....	v
About this Guide	vii
Chapter 1 Introduction	1
About the Documentum ECM Platform	1
About the Documentum Connector	2
About the Content Connector.....	2
Initial Extraction.....	3
Incremental Extraction.....	3
Handling Attachments	4
Operation.....	4
About the User Monitor Connector	5
Chapter 2 Installing the Documentum Connector	7
Before You Install	7
Setting Up the Documentum Runtime Environment.....	8
Installing the Documentum Connector	8
After You Install	9
Chapter 3 Configuring the Documentum Connector	11
Documentum.xml.....	11
Chapter 4 Configuring Documentum	15
Creating a Documentum Super-User	15
Registering the Save Event: dm_save	15
Registering the Delete Event: dm_destroy	17

Registering the Update Event: dm_checkin..... 17

Chapter 5 Configuring FAST Data Search 19

About Configuring FAST Data Search 19
 Configuring a Cluster 20
 Customizing an Index Profile 20
 Uploading an Index Profile..... 20
 Creating a Collection for Extracted Data..... 22

Chapter 6 Operation 27

Using the Content Connector..... 27
 Running the Content Connector 27
 Command Line Options 28
 Stopping a Running Connector..... 28
 Logging 28
 Error Conditions..... 29
 Documentum Is Not Responding..... 29
 Error Reading/Converting Documents..... 29
 Verifying that the Documentum Data is Loaded 30
 Using the User Monitor Connector..... 30
 Running the User Monitor Connector 30
 Logging 30

Appendix A Sample Index Profile 31

datasearch-4.0-documentum.xml 31

Appendix B Sample Configuration File 37

documentum.xml..... 37

Appendix C Documentum Meta Fields 39

FAST Support

Website

Please visit us at:

<http://www.fastsearch.com/>

Contacting FAST

Fast Search & Transfer Inc.

Cutler Lake Corporate Center

117 Kendrick Street, Suite 100

Needham, MA 02492 USA

Tel: +1 (781) 304-2400 (8:30am - 5:30pm EST)

Fax: +1 (781) 304-2410

Technical Support and Licensing Procedures

E-mail: fds-support@fastsearch.com

Product Training

E-mail: fastuniversity@fastsearch.com

Sales

E-mail: sales@fastsearch.com

About this Guide

Purpose of this Guide

This guide describes the FAST Smart Connector for Documentum and explains how to install and integrate it with your Fast Data Search installation.

Audience

This guide provides information for several types of users:

- **System Managers**, who need to understand how the Documentum Connector functions.
- **System Integrators**, who need to know how to integrate the Documentum Connector into existing installations of Fast Data Search and Documentum.

How this Guide Integrates with the Standard Fast Data Search Documentation

This guide is a supplement to the standard documentation delivered with Fast Data Search.

Conventions

This guide uses the following textual conventions:

- Terminal output, contents of plaintext ASCII files are represented using the following format:

```
Answer yes to place the node in the known_hosts file.
```

- Terminal input from operators is in the same but bold format:

```
chmod 755 $HOME
```

- Input of some logic meaning is enclosed in <> brackets:

```
setup_<OS>.tar.gz
```

where <OS> represents a specific operating system that must be entered.

- URLs, folder paths, commands, and the names of files, tags, and fields in paragraphs appear in the following format:

The default home folder is the *C:\DataSearch* folder.

- User Interface page/window texts, buttons, and lists appear in the following format:
Click **Next** and the **License Agreement** screen is displayed.
- *\$FASTSEARCH* (UNIX) or *%FASTSEARCH%* (Windows) refer to an environment variable set to the folder where Fast Data Search is installed.

Chapter 1

Introduction

About this Chapter

This chapter introduces the FAST Smart Connector for Documentum. It includes:

- About the Documentum ECM Platform
- About the Documentum Connector
- About the Content Connector
- About the User Monitor Connector

About the Documentum ECM Platform

Documentum's Enterprise Content Management solution provides a scalable, secure, and extensible platform that enables enterprises to standardize content applications on a common infrastructure. Documentum's architecture is based on a set of shared technologies: the Documentum content repository, content services, and content applications:

- Content Repository – Documentum provides a universal repository capable of storing and managing all content types including documents, scanned images, Web pages, XML, rich media, records, engineering drawings, reports, and many others.
- Content Services – The Documentum Content Server comprises the core of all content services provided by Documentum and provides a variety of content services common to all content types. Content services deliver native storage of diverse types of content, access control, versioning, search, workflow, and many other services.
- Content Applications – To access content in the Documentum repository and Documentum content services, content applications leverage the Documentum API.

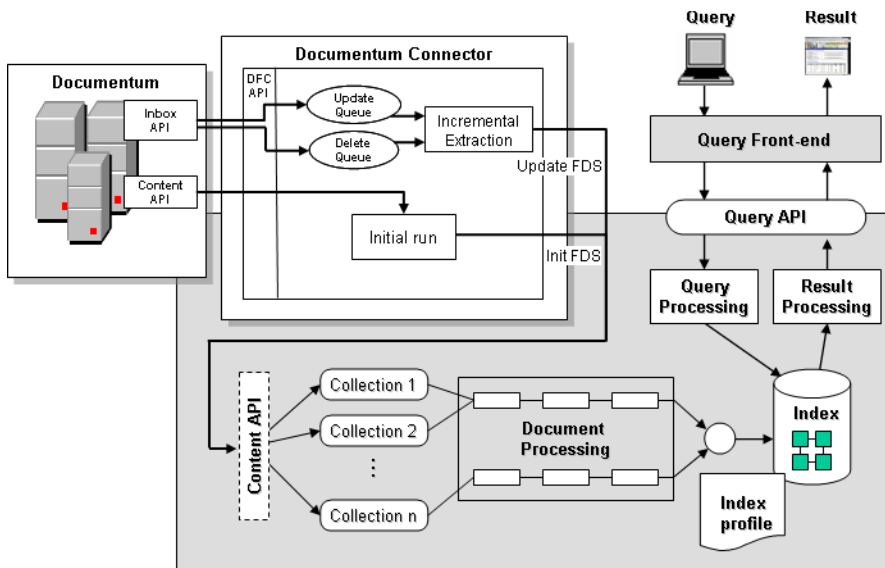
About the Documentum Connector

The Documentum Connector consists of two standalone programs that run from the command line:

- The **content connector** extracts data from Documentum and feeds the data directly to FAST Data Search. The extraction is done using a Java program that parses an XML configuration file that specifies the metadata fields to extract from Documentum.
- The **user monitor** connector supports extraction of authorization (access control) information for use with the FAST Data Search Security Access Module. This allows you to build secure search front-end applications.

About the Content Connector

The figure below shows how the content connector works.



The content connector extracts data from Documentum using the Documentum Foundation Classes (DFC) interface and sends the data to FAST Data Search using the Content API. The content connector has two distinct modes of operation:

- Initial Extraction
- Incremental Extraction

This section also discusses:

- Handling Attachments
- Operation

Initial Extraction

When you run the content connector for the first time, it extracts all documents from Documentum and indexes them. The next time you run the connector, it does “incremental extraction” as explained in *Incremental Extraction* on page 3.

Should you need to repeat the initial extraction, first delete all the files in the output directory defined in the configuration file (see Appendix B *Sample Configuration File*). The next time you run the content connector, it extracts all documents from Documentum and indexes them.

Incremental Extraction

The Documentum Connector supports incremental extraction based on events. Events are specific actions on specific documents, folders, cabinets, or other objects. When a document is updated, deleted or created, the connector immediately does the required actions in FAST Data Search to reflect the change.

- If a new document is created or updated without being given a new version number, it is fed into the FAST Data Search index.
- If document is updated and given a new version number, the new version of the document is fed into the FAST Data Search index. If the operation succeeds, the previous version of the document is removed from the index.
- If all versions of a document are deleted, the document is removed from the FAST Data Search index.
- If only the latest version or versions of a document are deleted, the Connector tries to find the current version and feeds it to the FAST Data Search index.

A Documentum user can register to receive events, after which event notifications are sent to that user’s *inbox*. (See Chapter 4 *Configuring Documentum* for instructions on registering events.)

The Documentum Connector is represented by a specific Documentum super-user and periodically polls that user’s inbox to check for event notifications. Each time it finds one, it puts a new item on one of two queues maintained by the connector, one for deletions,

and one for updates and new documents. The connector reads the queues periodically and performs the required actions.

When extracting content from a Documentum source, the Documentum Connector represents each database record as a FAST Data Search document and columns or fields as document attributes. The document attributes eventually end up as document elements and searchable fields in FAST Data Search. The elements to us as searchable information must be specified in:

- the Documentum Connector configuration file, as part of the DocumMetaFields and DocumMetaFolders (as described in Chapter 3 *Configuring the Documentum Connector*).
- the FAST Data Search index-profile (as described in Chapter 5 *Configuring FAST Data Search*).

When the connector has extracted all specified information from Documentum, based on the configuration file, it processes the Documentum metadata and retrieves the documents referenced. The metadata and file content from Documentum is sent to FAST Data Search.

Once the content is submitted to FAST Data Search, it is treated the same as any other content. In other words, it is sent through document processing pipelines (see FAST Data Search System Reference Guide, Chapter 4, Processing Documents) and then indexed to be searchable and filterable.

Handling Attachments

The extracted meta information from Documentum refers to a document stored in Documentum (Office document, PDF, etc.). The content of the file is indexed like an ordinary document. This allows the end-user to search for information contained partly in the Documentum Metadata and partly in the referenced file. This makes it possible to combine ACL information extracted from Documentum with the content of documents.

Operation

The content connector is designed to run in incremental extraction mode continuously and this is the recommended mode of operation. You can, however, run it only as needed.

About the User Monitor Connector

The user monitor connector:

- extracts all listed users in Documentum.
- extracts the group memberships for these users, including nested group memberships.
- passes this information to an instance of the generic user monitor in the Security Access Module.

Each time a query is submitted, the FAST Data Search query engine accesses the Security Access Module, which passes user identification to the user monitor in order to find its group memberships. That information is used to produce a filter that is applied to the search. The result is that users can only see the documents that they are authorized to see.

Refer to the FAST Data Search *Security Module Users Guide* for detailed security information.

Installing the Documentum Connector

About this Chapter

This chapter explains how to install and configure the FAST Smart Connector for Documentum as an add-on to an existing Fast Data Search installation.

This chapter includes:

- Before You Install
- Installing the Documentum Connector

Before You Install

- The machine on which you install the connector needs to be fully patched with Windows Update.
- Fast Data Search must be installed and running, not necessarily on the machine where the Documentum Connector is installed.
- The Documentum Connector requires the Documentum Foundation Classes (DFC) Runtime Environment 5.1 or later. (See [Setting Up the Documentum Runtime Environment](#) below).
- The Documentum server machine must be running version 5.2.

Setting Up the Documentum Runtime Environment

The Documentum Connector requires the Documentum Foundation Classes (DFC) Runtime Environment 5.1 or later. To verify that this is present:

- 1 Go to **Control Panel** -> **Add/Remove Programs**
- 2 If you do not have the DFC Runtime 5.1 it can be downloaded from your Documentum Download Center.
- 3 To determine the Java version on your machine, enter the following command from a command prompt:

```
C:\>java -version
```

- 4 If you do not have Java v1.4.0 it can be downloaded from <http://java.sun.com/>.

Installing the Documentum Connector

Note! Note! The FDS_Install_Directory is the name of the folder where Fast Data Search is installed; the configuration file in this example refers to this folder as C:\datasearch.

Use the following procedure to install the Documentum Connector:

- 1 Unpack the software kit into a folder created for the purpose of containing the Documentum Connector. For example:

```
C:\Documentum
```

- 2 Copy the file:

```
C:\Documentum\etc\datasearch-4.0-documentum.xml
```

to the folder:

```
<FDS_Install_Directory>\index-profiles\
```

- 3 Copy the Documentum DFC file:

```
C:\Program Files\Documentum\Shared\dfc.jar
```

to the root folder:

```
C:\Documentum\lib\
```

- 4 The Documentum Connector configuration file (documentum.xml) can be loaded from a local file or from the Fast Data Search configuration host. To use the configuration host, install the configuration file in the folder:

```
$FASTSEARCH/etc/config_data/DocumentumConnector/
```


After You Install

When finished the Documentum folder contains the following subfolders and files:

Folder	Files	Description
bin	DFCUsers.bat DFCUsers.sh	Script file used to run the extraction of User/Group information from Documentum and loading the Security Access Module with security information.
	DocumentumConnector.bat DocumentumConnector.sh	Script file used to run the connector, extraction meta information and file content from Documentum.
doc	DocumentumConnector401.pdf	This document.
etc	documentum.xml	XML configuration for DocumentumConnector and DFCUsers
	no.fast.connector toolkit.properties	Change detection configuration file.
	simplelog.properties	Log configuration.
	indexprofile-datasearch-4.0	Sample index profile.
		Note! See Appendix A <i>Sample Index Profile</i> .
lib	commons-logging.jar	Log methods.
	log4j.properties	Search log configuration file.
	DmtCon.jar	The Fast Data Search connector.
	FastMD5.jar	MD5 method.
	hsqldb.jar	Change detection DB.
	xmlrpc-1.1.jar xmlrpc-1.1-applet.jar	Content API communication method.
	java-getopt.jar	Library to extract command line options.

log4j.jar

Log methods.

mysql-connector-java

Methods used in internal change
detection.

Configuring the Documentum Connector

About this Chapter

This chapter describes the steps required to configure the Documentum Connector using the `Documentum.xml` file.

Documentum.xml

To configure the Documentum Connector, copy and edit the example XML configuration file provided in the distribution kit:

```
etc/documentum.xml
```

The name of the configuration file can be anything. You specify the name and path of the configuration file at run time, as described in *Command Line Options* on page 28.

A listing of the example configuration file is provided in *Appendix B Sample Configuration File*. The configuration parameters are described in the table below.

Element	Example value	Description
Accessor	Documentum	Id for the Documentum Connector.
OutputDataDir	FASTSEARCH/data/ documentum	Folder where the Documentum accessor writes output XML files to be loaded into FAST Data Search using the file traverser.
FDSConfigHost	pingle04.oslo.fast.no	Host where the FAST Data Search Configuration server runs. Used to load data directly using DSAPI.

Element	Example value	Description
FDSConfigPort	16005	Port where FAST Data Search Configuration server runs.
FDSNameservicePort	16099	FAST Data Search name service port, used in context communication between connector and FAST Data Search.
FDSCollection	documentum	Collection used to store and index data from the connector in Documentum Connector. The collection is only used with values in Host/Port.
SourceDrive	"D: "	Drive letter used on the local computer where Documentum is installed. This drive letter is retrieved from Documentum for the PATH of the documents in Documentum.
FileWebPrefix	\\myfileserver	<p>Value that replaces the content in: FileAttachRoot.</p> <p>The prefix is used by FAST Data Search to complete the path to access the file from a search result.</p> <p>If FileWebPrefix contains the word "HTTP", then the URI uses "/" as separator. If not, then the URI uses "\".</p>
webtopHostPort	webtophost.com:8080	If specified, the URI to the Documents extracted from Documentum is based on Documentum WebTop.
DocumMetaFields	object_name, title, subject	<p>Comma separated list of metadata fields to extract from Documentum.</p> <hr/> <p>Note! See Appendix C <i>Documentum Meta Fields.</i></p> <hr/>
DocumMetaFolders	Testdata	Comma separated list of data folders in Documentum. All documents in the specified folders and subfolders are selected.
DocumMetaDocBase	DocBase1	Documentum Docbase from which the connector retrieves data.

Element	Example value	Description
DocumUserName	Testuser	Documentum user used to connect to DocBase (see <i>Creating a Documentum Super-User</i> on page 15).
DocumUserPwd	Testpwd	Documentum password used to connect to DocBase.
DocumUserDomain	Windows Domain	Domain for the Documentum user.
DeleteThreshold	10	Secure parameter used with the connector's incremental support. Specifies the maximum percentage of rows that can be deleted. If the percentage is exceeded, a warning is displayed and no rows are deleted.
FDSSecurityHost	localhost	Host where the Security Access Module runs.
FDSSecurityPort	112	Port for Security Access Module.
FDSSecurityCertificate	c:\\datasearch\\etc \\ssl\\certs\\FDSSM _something.key	SSL certificate.
FDSSecurityCertPass	FDSDFS	SSL certificate password.

Chapter 4

Configuring Documentum

About this Chapter

This chapter explains how to configure Documentum to work with the FAST Smart Connector for Documentum. It includes:

- Creating a Documentum Super-User
- Registering the Save Event: `dm_save`
- Registering the Delete Event: `dm_destroy`
- Registering the Update Event: `dm_checkin`

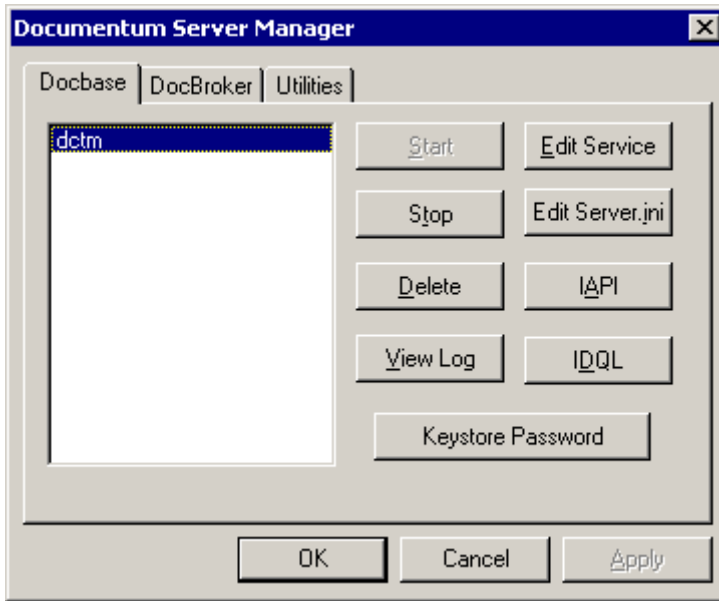
Creating a Documentum Super-User

Refer to the Documentum documentation for instructions on how to create a super-user identity for document processing. In this chapter, the super-user is *angeliqa*. That name also must be specified in the configuration file (see Chapter 3 *Configuring the Documentum Connector*).

Registering the Save Event: `dm_save`

The Save Event sends a notification to the Documentum inbox every time a new document is saved (new ones or updated ones).

- 1 Open the Documentum Server Manager window on the server where Documentum is installed.



- 2 Select the docbase for which to add the save event.
- 3 Press the IAPI button.
- 4 Log in with the Documentum super-user name and password.

```
Please enter a user (angeliqa):
Please enter password for angeliqa: *****

Documentum iapi - Interactive API interface
(c) Copyright Documentum, Inc., 1992 - 2004
All rights reserved.
Client Library Release 5.3.0.018 Win32

Connecting to Server using docbase dctm
IDM_SESSION_I_SESSION_START!info: "Session 0100270f8000c14a started for user an
geliqa."

Connected to Documentum Server running Release 5.3.0.016 Win32.SQLServer
Session id is s0
```

- 5 To retrieve the object ID for dm_document, type the command:

```
retrieve,c,dm_type where name='dm_document'
```

```
I API> retrieve,c,dm_type where name='dm_document'
---
0300270f80000129
```

- 6 Copy the retrieved object ID to the clipboard so that you can paste it for the registration step.

7 Type the command:

```
register,c,0300270f80000129,dm_save
```

replacing 0300270f80000129 with the actual object ID you copied.

8 At the SET> prompt, type the name of the event (in this case **New Document**).

```
API> register,c,0300270f80000129,dm_save
SET> New document
...
OK
```

9 To verify that the event is registered, type the command:

```
retrieve,c,dmi_registry where message='New Document'
```

```
API> retrieve,c,dmi_registry where message='New document'
...
2600270f80001103
```

10 You get an object id as a result.

Registering the Delete Event: dm_destroy

The Delete Event sends a notification to the Documentum inbox every time a document is deleted. Perform the same procedure as *Registering the Save Event: dm_save* on page 15 with the following exceptions:

- Change **dm_save** to **dm_destroy**
- The message should be something like **Document deleted** instead of **New Document**.

Registering the Update Event: dm_checkin

The Update Event sends a notification to the Documentum inbox every time a document is updated (a new version of the document is checked in). Perform the same procedure as *Registering the Save Event: dm_save* on page 15 with the following exceptions:

- Change **dm_save** to **dm_checkin**
- The message should be something like **Document updated** instead of **New Document**.

Configuring FAST Data Search

About This Chapter

In order to use the Documentum Connector, you must integrate the connector into your FAST Data Search installation. This chapter describes the required configuration procedures. It includes:

- About Configuring FAST Data Search
- Configuring a Cluster
- Creating a Collection for Extracted Data

About Configuring FAST Data Search

Note! This chapter provides a procedural description of how to configure FAST Data Search for the Documentum Connector. For detailed information, refer to the *Indexing Database Content and XML Guide*.

Configuring FAST Data Search consists of two procedures:

- Configuring a Cluster — you must configure a FAST Data Search cluster to function as a dedicated Documentum indexer by attaching it to an index profile.
- Creating a Collection — you must create at least one collection in FAST Data Search in order to receive extracted data.

Configuring a Cluster

The default FAST Data Search cluster (*webcluster*) is normally used for crawled web content.

- In a single node system, modify *webcluster* to index Documentum content instead of crawled web content.
- In a multiple node system, create a new cluster to index Documentum content (leaving *webcluster* to index crawled content).

Customizing an Index Profile

The Documentum Connector provides a sample index profile in the *etc* folder that you can copy and customize (see Appendix A *Sample Index Profile*). It is built on the *datasearch-default* index profile and has been extended with the meta fields that are extracted by a query to Documentum, and also with security related fields.

When you customize the default index profile, be sure that the Documentum Connector index profile is compatible with the FAST Data Search index profile. The following rules apply when you customize:

- To make the names unique, start the field prefix name with: "dmt"
- Use only alpha characters
- Transform characters to lowercase field mapping

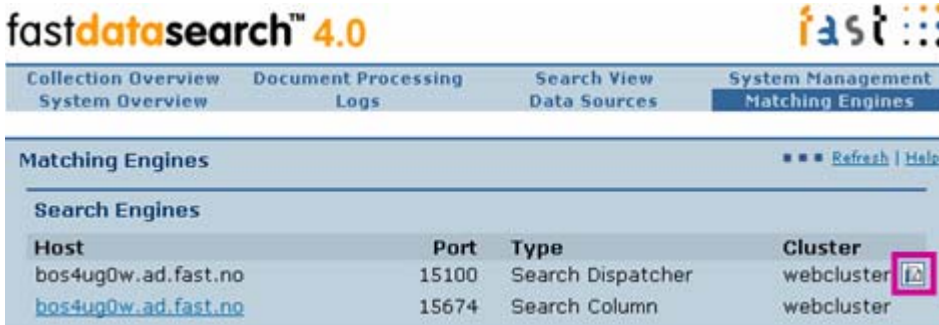
Uploading an Index Profile

To upload an index profile to FAST Data Search:

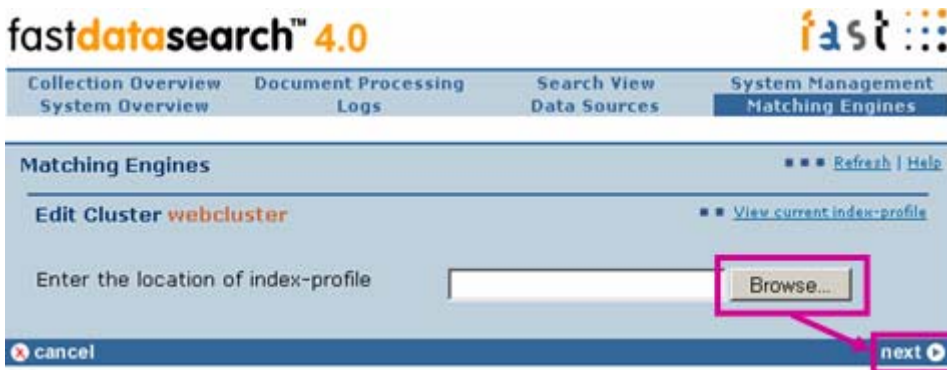
- 1 Open the FAST Data Search **Admin interface**.
- 2 Click **Matching Engines**.



- Click the **icon** next to the **Search Dispatcher** host entry.



- In the **Edit Cluster** screen, click the **Browse** button to identify the location of the index profile.

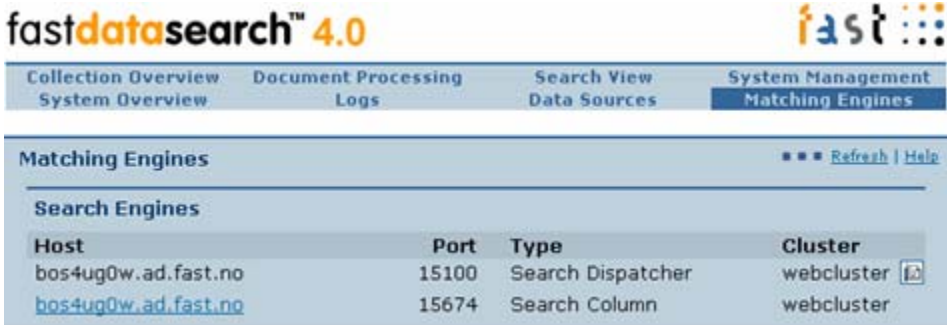


Note! The index profile must be accessible to the web browser. The file must be located either on the machine that is running the active web browser or that has a remote connection to the file system on which FAST Data Search is running.

- Verify that the index profile update has completed successfully and click **Ok**.



- 6 The admin interface returns to the Matching Engines dialog.



Creating a Collection for Extracted Data

- 1 Open the FAST Data Search **Admin interface**.
- 2 If necessary, click **Collection Overview** on the navigation bar.
- 3 Click **Create Collection**.



- 4 In the **Name** field enter Documentum.

fastdatasearch™ 4.0

Collection Overview | Document Processing | Search View | System Management
System Overview | Logs | Data Sources | Matching Engines

New Collection - Description Cluster Pipeline Data Source Refresh Help

Description

Name

Description

cancel next

- 5 Optionally fill in the **Description** field.
- 6 Click **next**.
- 7 In the **Cluster Configuration** screen, click the **Available Clusters** dropdown list button.

fastdatasearch™ 4.0

Collection Overview | Document Processing | Search View | System Management
System Overview | Logs | Data Sources | Matching Engines

New Collection Documentum - Description Cluster Pipeline Data Source Refresh Help

Cluster configuration

Available Clusters

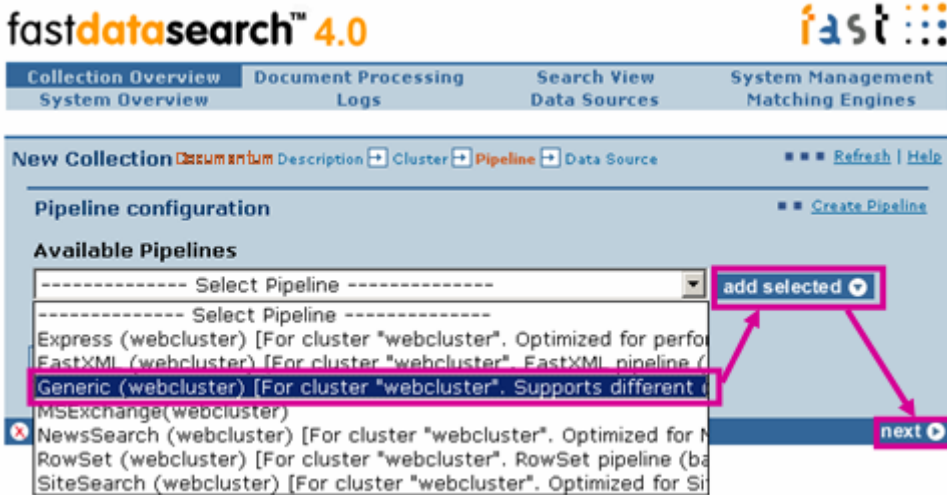
add selected

Selected Cluster

cancel next

- 8 Select a cluster from the list. In a single-node system, **select webcluster**. In a multiple-node system, select the cluster you created in *Configuring a Cluster* on page 20.
- 9 Click **add selected**.
- 10 Click **next**.

- 11 In the **Pipeline Configuration** screen, click the **Available Pipelines** dropdown list button.



- 12 Select **Generic (webcluster)**. This provides a thorough processing of the documents, including linguistics, document structure, teaser generation, language detection etc.
- 13 Click **add selected**.
- 14 Click **next**.
- 15 In the **Data Source Configuration** screen, click **ok**.



16 In the next screen, click **ok**.

The screenshot shows the 'Edit Collection' interface for 'Documentum'. The top navigation bar includes 'Collection Overview', 'Document Processing', 'Search View', and 'System Management'. The main content area is divided into several sections:

- Description:** 'No description set'.
- Control Panel:** A grid of buttons for 'Edit Description', 'Edit Cluster', 'Add Document', 'Search', 'Edit Data Sources', 'Edit Pipeline', 'Delete Document', and 'Delete Collection'.
- Status:** A flow diagram showing the status of components: 'Data Sources' (N/A, None configured), 'Content Distributor' (OK, Docs: 0), 'Pipeline' (OK, FastXML (webcluster)), and 'Matching Engines' (OK, Docs: 0).
- Modules:** A table listing modules and their status.
- Cluster:** 'webcluster'.

An 'ok' button is highlighted at the bottom center of the interface.

Module	Host	Port	Status
ContentDistributor (fds/contentdistributor)	bos4ug0w	N/A	Responding
ProcessorServer	bos4ug0w.ad.fast.no	16200	Responding
Search Engine (RTS Indexer)	bos4ug0w.ad.fast.no	15674	Responding
StatusService (fds/statusservice_0)	bos4ug0w	N/A	Responding

17 The admin interface returns to the Collection Overview dialog.

The screenshot shows the 'Collection Overview' interface. The top navigation bar is the same as in the previous screenshot. The main content area shows:

- Collection Overview:** 'Create Collection | Refresh | Help'.
- Overview:** 'There is just one Collection'.
- Table:** A table with columns 'Name', 'Description', 'Last input', and 'Docs'.

Name	Description	Last input	Docs
Documentum		N/A	0

Chapter 6

Operation

About this Chapter

This chapter explains how to operate the FAST Smart Connector for Documentum. It includes:

- Using the Content Connector
- Using the User Monitor Connector

Using the Content Connector

Running the Content Connector

- 1 Go to the <installation directory>.
- 2 From the command line, run the script as follows:

- using a local configuration file:

```
DocumentumConnector.bat -l ..\etc -f documentum.xml
```

- using a configuration file on a Fast Data Search configuration host:

```
DocumentumConnector.bat -c <FDS configuration host> -p 16005 -f documentum.xml
```

By default, the Connector does not run as part of the Fast Data Search node controller. It is therefore not displayed in the standard Fast Data Search administrator interface or controlled by the nctrl start/stop-commands.

Command Line Options

Option	Description
-h	Help - displays this information
-a [start stop]	Starts (default) or stops the connector. This option is used to remotely start and stop the connector
-c configserverhost	Fast Data Search configuration server hostname
-p configserverport	Fast Data Search configuration server host port number
-f configfilename	Connector configuration file name. If installed on the Fast Data Search configuration host the file is located in the folder: <code>etc/config_data/DocumentumConnector</code>
-l localfilepath	Specifies the path (absolute or relative) to a local configuration file. <hr/> Note! Only the path is provided with this parameter; the filename is provided with the -f option. <hr/>
-m maxDocumentSize	Configure memory usage (default is 70MB)

Stopping a Running Connector

To stop running the connector, you can either:

- Close the command window running the connector or
- Enter Ctrl-C in the command window

If the Connector is running in background, the connector can be stopped by issuing the command: `DocumentumConnector -a stop`

Logging

The Documentum Connector sends log messages to the standard output so you can monitor the behavior of the connector. Alternatively, you can send log messages to the Fast Data Search LogServer and watch the logs in the administration GUI of Fast Data Search.

- Go to the collections overview page in the administration GUI to verify that updates to the source are communicated to Fast Data Search
- Go to the search view to perform sample searches that verify that content has been extracted and indexed as expected.

The Documentum Connector logging can be configured by editing the file \$FAST-SEARCH/etc/log4j.properties file.

To change the log level to debug, change:

```
log4j.rootCategory=warn, A1, A2
```

to

```
log4j.rootCategory=debug, A1, A2
```

Error Conditions

This section describes some typical errors encountered in running the connector.

Documentum Is Not Responding

If Documentum fails to respond, it does not write any data to Fast Data Search.

Error message:

```
[ERROR] DocumentumConnector - -DfException:: THREAD: main; MSG:  
[DM_SESSION_E_RPC_ERROR]error:  
"RPC error 116 occurred: Unknown error code 116 (_nl_error_ = 0). Extended  
network error: 0"
```

Solution:

Log on to Documentum host and make sure that the Docbase/Docbroker runs, if not start

Error Reading/Converting Documents

Documentum and the Documentum Connector run as independent jobs. Documentum may therefore write zero, one or many files to the input folder, before the files are processed by the Connector.

Error message:

```
[ERROR] DocumentumConnector - -Error Converting document  
\documentum.node.com...\f1.ppt to  
base 64:\\ documentum.node.com \..\f1.ppt (Logon failure: unknown user  
name orbad password)  
[ERROR] DocumentumConnector - -Document \documentum.node.com...\f1.ppt  
wont be sent to FDS
```

Solution:

Map the drive "documentum.node.com" and log on with username and password.

Verifying that the Documentum Data is Loaded

To verify that your Documentum data has been loaded, go to the Fast Data Search Administrator Interface and follow the instructions below:

Standard Search

To verify that data is available and searchable in the index, go to the Search Views for the collection you added in Fast Data Search administrator interface.

Using the User Monitor Connector

Running the User Monitor Connector

To run the user monitor connector:

- 1 Go to the <installation directory>.
- 2 From the command line, run the DocumentumConnector.bat script as follows:

```
DocumentumConnector.bat -f <name of config file> -l <path to config file>
```

Example:

```
DocumentumConnector.bat -f config.xml -l etc
```

Logging

The Documentum Connector sends log messages to the standard output so you can monitor the behavior of the connector. Alternatively, you can send log messages to the DS LogServer and watch the logs in the administration GUI of Data Search.

You can also inspect the logs of the Security Access Module to verify that users and group information is extracted properly.

Appendix A

Sample Index Profile

datasearch-4.0-documentum.xml

```
<?xml version="1.0"?>
<!DOCTYPE index-profile SYSTEM "index-profile-3.0.dtd">
<index-profile name="datasearch">
<field-list>

  <field name="title" sort="yes" fullsort="yes" tokenize="auto"
lemmatize="yes" >
    <vectorize default="10:0"/>
  </field>
  <field name="body" tokenize="auto" max-result-size="1024"
    fallback-ref="teaser" result="dynamic" index="no"
lemmatize="yes">
    <vectorize default="5:5" alternative="{ja,ko,zh,szh,tzh}:5:0" />
  </field>
  <field name="teaser" index="no" />
  <field name="headings" tokenize="auto" lemmatize="yes" />
  <field name="description" result="no" />
  <field name="anchortext" result="no" tokenize="auto" />
  <field name="keywords" result="no" />
  <field name="contenttype" element-name="mime" />
  <field name="format" index="no"/>
  <field name="language" />
  <field name="languages" separator=";" />
  <field name="charset" />
  <field name="urls" />
  <field name="url" index="no" />
  <field name="domain" element-name="url.domain" result="no" />
  <field name="tld" element-name="url.tld" result="no" />
  <field name="path" element-name="url.path" result="no" />
```

```
<!-- Non-text fields -->
<field name="crawltime" type="datetime" fullsort="yes" />
<field name="processingtime" type="datetime" fullsort="yes" />
<field name="docdatetime" type="datetime" fullsort="yes" />

<field name="size" type="int32" fullsort="yes" />

<field name="generic1" />
<field name="generic2" />
<field name="generic3" />
<field name="generic4" result="no" />
<field name="igeneric1" type="int32" fullsort="yes" />
<field name="igeneric2" type="int32" fullsort="yes" />
<field name="dtgeneric1" type="datetime" fullsort="yes" />
<field name="dtgeneric2" type="datetime" fullsort="yes" />

<!-- News Entity Extraction Fields -->
<field name="companies" separator=";" />
<field name="locations" separator=";" />
<field name="personnames" separator=";" />
<field name="topics" separator=";" />
<field name="emails" separator=";" />
<field name="taxonomy" />
<field name="host" separator=";" />

<!-- Security field: Document Access Control List -->
<field name="docacl" result="no" />
<field name="docaclsystemid"/>

<!-- Documentum Connector fields, Documentum prefix = dmt -->

<field name="dmtrobjectid" />
<field name="dmtrobjectname" />
<field name="dmtrtitle" />
<field name="dmtrsubject" />
<field name="dmtrlanguagecode" />
<field name="dmtrownername" />
<field name="dmtrgroupname" />
<field name="dmtrlogentry" />
<field name="dmtracldomain" />
<field name="dmtraclname" />
<field name="dmtrcreationdate" />
<field name="dmtrmodifydate" />
<field name="dmtracontenttype" />

</field-list>
```



```
<composite-field name="content" rank="yes" default="yes" lemmas="yes"
  query-tokenize="auto">
  <field-ref name="body" level="1"/>
  <field-ref name="headings" level="2"/>
  <field-ref name="path" level="2"/>
  <field-ref name="description" level="2"/>
  <field-ref name="domain" level="3"/>
  <field-ref name="keywords" level="3"/>
  <field-ref name="title" level="4"/>
  <field-ref name="anchortext" type="external" level="5"/>

  <rank-profile name="default" rank-model="default">
    <authority weight="50" field-ref="anchortext" />
    <freshness weight="50" field-ref="docdatetime" auto="yes" />
    <proximity weight="50" />
    <context weight="50">
  <field-weight field-ref="body" value="5"/>
  <field-weight field-ref="headings" value="20"/>
  <field-weight field-ref="path" value="20"/>
  <field-weight field-ref="description" value="30"/>
  <field-weight field-ref="domain" value="50"/>
  <field-weight field-ref="keywords" value="50"/>
  <field-weight field-ref="title" value="60"/>
    </context>
  </rank-profile>

  <rank-profile name="news" rank-model="news">
    <authority weight="50" field-ref="anchortext" />
    <freshness weight="200" field-ref="docdatetime" auto="yes" />
    <proximity weight="50" />
    <context weight="50">
  <field-weight field-ref="body" value="5"/>
  <field-weight field-ref="headings" value="20"/>
  <field-weight field-ref="path" value="20"/>
  <field-weight field-ref="description" value="30"/>
  <field-weight field-ref="domain" value="50"/>
  <field-weight field-ref="keywords" value="50"/>
  <field-weight field-ref="title" value="60"/>
    </context>
  </rank-profile>

  <rank-profile name="site" rank-model="site">
    <authority weight="70" field-ref="anchortext" />
    <freshness weight="50" field-ref="docdatetime" auto="yes" />
    <proximity weight="60" />
    <context weight="70">
  <field-weight field-ref="body" value="5"/>
  <field-weight field-ref="headings" value="20"/>
  <field-weight field-ref="path" value="20"/>
  <field-weight field-ref="description" value="30"/>
```

```
<field-weight field-ref="domain" value="50"/>
<field-weight field-ref="keywords" value="50"/>
<field-weight field-ref="title" value="60"/>
  </context>
</rank-profile>

</composite-field>

<result-specification>

  <categorization name="default" sort-by="label">
    <field-ref name="taxonomy"/>
  </categorization>

  <clustering name="default" sort-by="none" size="10" threshold="0.30"/>

  <!-- Result proximity boosting: Set to "yes" to enable boosting per
  default -->
  <!-- set to "no" to generate necessary config to allow boosting on a per
  -->
  <!-- query basis but have boosting off per default -->
  <result-proximity boost="no">
    <field-ref name="body"/>
    <field-ref name="title"/>
    <field-ref name="anchortext"/>
  </result-proximity>

  <numeric-navigator name="sizenavigator"
    display="Size"
    unit="kB"
    divisor="1024"
    intervals="4"
    resolution="1024">
    <field-ref name="size"/>

    <range-label type="first" format="Less than %.2g" offset="0"/>
    <range-label type="middle" format="Between %.2g and %.2g" />
    <range-label type="last" format="More than %.2g" />

    <ignore-value value="0"/>
  </numeric-navigator>

  <numeric-navigator name="docdatetimenavigator"
    display="Document Time"
    unit="Date"
    intervals="4"
    resolution="1">
    <field-ref name="docdatetime"/>

    <range-label type="first" format="Before %s" />
```

```
<range-label type="middle" format="Between %s and %s" />
<range-label type="last" format="%s or after" />
</numeric-navigator>

<string-navigator name="contenttypenavigator" display="MIME">
  <field-ref name="contenttype"/>
</string-navigator>

<string-navigator name="charsetnavigator" display="Encoding">
  <field-ref name="charset"/>
</string-navigator>

<string-navigator name="languagesnavigator"
  display="Languages">
  <field-ref name="languages"/>
</string-navigator>

<!-- News Entity Navigators -->
<string-navigator name="companiesnavigator" display="Companies">
  <field-ref name="companies"/>
</string-navigator>

<string-navigator name="locationsnavigator" display="Locations">
  <field-ref name="locations"/>
</string-navigator>

<string-navigator name="personnamesnavigator" display="People">
  <field-ref name="personnames"/>
</string-navigator>

<string-navigator name="topicsnavigator" display="Topics">
  <field-ref name="topics"/>
</string-navigator>

<string-navigator name="emailsnavigator" display="Emails">
  <field-ref name="emails"/>
</string-navigator>

<string-navigator name="hostnavigator" display="Hosts">
  <field-ref name="host"/>
</string-navigator>

<result-view name="urls">
  <field-ref name="url"/>
  <field-ref name="urls"/>
</result-view>

</result-specification>

</index-profile>
```


Appendix B

Sample Configuration File

documentum.xml

```
<?xml version="1.0" encoding="UTF-8"?>
<AccessorConfig>
  <Accessor id="documentum">
    <parameter name="OutputDataDir">
      c:\\temp
    </parameter>
    <parameter name="FDSConfigHost">
      fds.host.com
    </parameter>
    <parameter name="FDSConfigPort">
      16005
    </parameter>
    <parameter name="FDSCollection">
      documentum
    </parameter>
    <parameter name="SourceDrive">
      D:
    </parameter>
    <parameter name="FileWebPrefix">
      \\documentum.host.com\d$
    </parameter>
    <parameter name="DocumMetaFields">
      object_name, title, subject
    </parameter>
    <parameter name="DocumMetaFolders">
      Testdata
    </parameter>
    <parameter name="DocumMetaDocBase">
      DocBase1
    </parameter>
    <parameter name="DocumUserName">
```

```
    dmtuses
  </parameter>
  <parameter name="DocumUserPwd">
    dmtpwd
  </parameter>
  <parameter name="DocumUserDomain">
    DMTDOMAIN
  </parameter>
  <parameter name="DeleteThreshold">
    10
  </parameter>
  <parameter name="FDSNameservicePort">
    16099
  </parameter>
  <parameter name="webtopHostPort"> <!-- <server_name>:8080 -->
    webtophost.com:8080
  </parameter>
</Accessor>

<Accessor id="documentum_users">
  <parameter name="OutputDataDir">
    c:\\temp
  </parameter>
  <parameter name="DocumMetaDocBase">
    DocBase1
  </parameter>
  <parameter name="DocumUserName">
    dmtuser
  </parameter>
  <parameter name="DocumUserPwd">
    dmtpwd
  </parameter>
  <parameter name="DocumUserDomain">
    DMTDOMAIN
  </parameter>
  <parameter name="FDSSecurityHost">
    localhost
  </parameter>
  <parameter name="FDSSecurityPort">
    112
  </parameter>
  <parameter name="FDSSecurityCertificate">
    c:\\datasearch\\etc\\ssl\\certs\\FDSSM_something.key
  </parameter>
  <parameter name="FDSSecurityCertPass">
    FDSFDS
  </parameter>
</Accessor>

</AccessorConfig>
```

Appendix C

Documentum Meta Fields

The following table is a complete list of the Documentum metadata fields:

Field	Type
object_name	CHAR(255)
r_object_type	CHAR(32)
title	CHAR(400)
subject	CHAR(192)
authors	CHAR(48)
keywords	CHAR(48)
a_application_type	CHAR(32)
a_status	CHAR(16)
r_creation_date	TIME
r_modify_date	TIME
r_modifier	CHAR(32)
r_access_date	TIME
a_is_hidden	BOOLEAN
i_is_deleted	BOOLEAN
a_retention_date	TIME
a_archive	BOOLEAN
a_compound_architecture	CHAR(16)
a_link_resolved	BOOLEAN

FAST Data Search

i_reference_cnt	INTEGER
i_has_folder	BOOLEAN
i_folder_id	ID
r_composite_id	ID
r_composite_label	CHAR(32)
r_component_label	CHAR(32)
r_order_no	INTEGER
r_link_cnt	INTEGER
r_link_high_cnt	INTEGER
r_assembled_from_id	ID
r_frzn_assembly_cnt	INTEGER
r_has_frzn_assembly	BOOLEAN
resolution_label	CHAR(32)
r_is_virtual_doc	INTEGER
i_contents_id	ID
a_content_type	CHAR(32)
r_page_cnt	INTEGER
r_content_size	INTEGER
a_full_text	BOOLEAN
a_storage_type	CHAR(32)
i_cabinet_id	ID
owner_name	CHAR(32)
owner_permit	INTEGER
group_name	CHAR(32)
group_permit	INTEGER
world_permit	INTEGER
i_antecedent_id	ID
i_chronicle_id	ID
i_latest_flag	BOOLEAN

r_lock_owner	CHAR(32)
r_lock_date	TIME
r_lock_machine	CHAR(32)
log_entry	CHAR(120)
r_version_label	CHAR(32)
i_branch_cnt	INTEGER
i_direct_dsc	BOOLEAN
r_immutable_flag	BOOLEAN
r_frozen_flag	BOOLEAN
r_has_events	BOOLEAN
acl_domain	CHAR(32)
acl_name	CHAR(32)
a_special_app	CHAR(32)
i_is_reference	BOOLEAN
r_creator_name	CHAR(32)
r_is_public	BOOLEAN
r_policy_id	ID
r_resume_state	INTEGER
r_current_state	INTEGER
r_alias_set_id	ID
a_effective_date	TIME
a_expiration_date	TIME
a_publish_formats	CHAR(32)
a_effective_label	CHAR(32)
a_effective_flag	CHAR(8)
a_category	CHAR(64)
language_code	CHAR(5)
a_is_template	BOOLEAN
a_controlling_app	CHAR(32)

<code>r_full_content_size</code>	DOUBLE
<code>a_extended_properties</code>	CHAR(32)
<code>i_is_replica</code>	BOOLEAN
<code>i_vstamp</code>	INTEGER

By default the connector extracts the fields in the list below. It is not necessary to include these fields in the XML configuration file (see Appendix B *Sample Configuration File*) even if you want it to be a part of the index. They must however be included in the index-profile (see Appendix A *Sample Index Profile*) if they should be a part of the FAST Data Search Index.

- `r_object_id`
- `i_contents_id`
- `i_is_deleted`
- `acl_domain`
- `acl_name`
- `r_modify_date`
- `owner_name`
- `owner_permit`
- `group_name`
- `group_permit`
- `language_code`
- `log_entry`
- `r_creation_date`
- `a_content_type`