



FAST SMART CONNECTOR FOR FILENET IMAGE SERVICES

version 4.0.1

MODULE GUIDE

DRAFT #1
MONDAY, MAY 10, 2004

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.

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).

About this Guide

Purpose of this Guide

This guide describes the FAST Smart Connector for FileNet Image Services and explains how to use it.

Audience

This guide provides information for all users of the FAST Smart Connector for FileNet Image Services.

Conventions

This guide uses the following textual conventions:

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

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

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

```
chmod 755 $HOME
```

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

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

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

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

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

- 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 directory where FAST Data Search is installed.

Contents

	About this Guide	iii
	FAST Support.....	vii
Chapter 1	Introduction	1
	About FileNet Image Services	1
	About the FileNet Connector	2
	FileNet Connector Features	3
Chapter 2	Installation	5
	Before You Install	5
	Installing the FileNet Connector	6
Chapter 3	Configuration	7
	Configuring the Content Connector	7
	log4j.properties	9
	no.fast.connectortoolkit.properties	10
	Configuring the User Monitor Connector	10
Chapter 4	Configuring FAST Data Search	11
	About Configuring FAST Data Search.....	11
	Configuring an Index Profile	12
	Creating an Index Profile	12
	Uploading an Index Profile	13
	Creating a Collection for Extracted Data	15
	Mixing Content From Multiple Sources	18

Chapter 5	Operation	21
	Using the Content Connector	21
	Running the Content Connector	21
	Stopping a Running Connector	21
	Logging.....	22
	Using the User Monitor Connector	22
	Running the User Monitor Connector.....	22
	Logging.....	22
Appendix A	Sample Configuration File	25
	filename.XML.....	25

Last update: Monday, May 10, 2004 3:07 pm

FAST Support

Website

Please visit us at:

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

Contacting FAST

Fast Search & Transfer Inc.

Wellesley Gateway

93 Worcester Street

Wellesley, MA 02481-9181 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

Chapter 1

Introduction

About this Chapter

This chapter introduces the FAST Smart Connector for FileNet Image Services. It includes:

- About FileNet Image Services
- About the FileNet Connector
- FileNet Connector Features

About FileNet Image Services

FileNet Image Services provides comprehensive content management functionality, specifically tailored to supporting management of scanned documents/images. It is also possible to add electronic documents such as Word, PDFs etc., to the system. System administrators manage document classes, user groups with access rights to the content, and set up storage for the documents. Content editors/users can then add content to the system according to the rights they have been granted.

The document model is briefly described as follows:

- Documents are associated with a document class, which is defined by a set of indices (attributes) and in addition all documents have a fixed set of metadata. These include entry date, content format, etc.
- For image content (scanned documents, e.g.), the users can add annotations on a per page basis.
- Documents are added to libraries and folders (tree structured). A document may be associated with more than one folder.

- Users are members of groups, which again are subgroups of larger groups in a group hierarchy. Security is defined by referring to group memberships.
- A group can be granted access rights to documents of a certain document class.

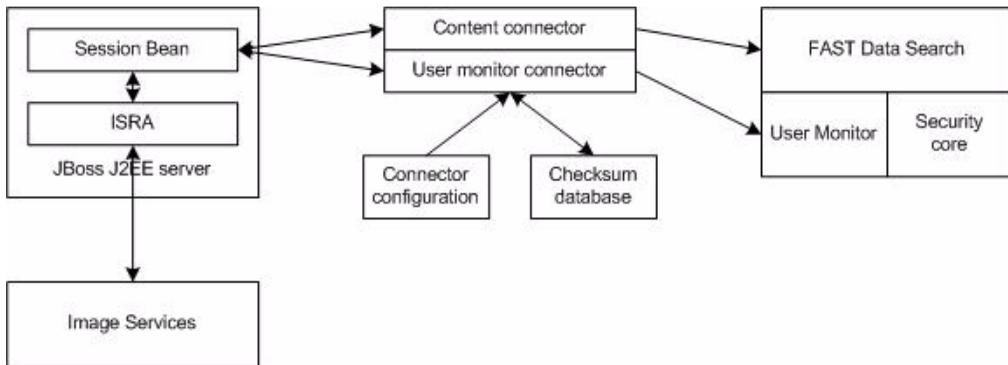
About the FileNet Connector

There are two parts to the FileNet Connector:

- The **content connector** supports data extraction and indexing from a FileNet Image Services system.
- The **user monitor** connector supports extraction of authorization (access control) information for use with the Fast Security Access Module (SAM).

The information that is extracted can be tailored to your needs.

The figure below shows how the content and user monitor connectors work.



The connectors are configured through the a single configuration file. The basic configuration includes the connection information to the Image Services System, as well as to Fast Data Search, and parameters that control which folders to extract data from, which metadata to extract for each document, whether to include the raw content or annotations, etc.

The connector and user monitor run from the command line and open a communication channel with the ISRA (Image Services Resource Adapter). The extracted data is formatted to meet the requirements of Fast Data Search (content) and a generic user monitor of the FAST Data Search Security Access Module (user and group information). The security module (SAM) is accessed by the query engine of Fast Data Search for each query. SAM then passes the user identification to the user monitor in order to find its group member-

ships, and uses this information 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.

FileNet Connector Features

The content connectors has the following main features:

- Extracts data according to a configuration parameter that contains an ISRA SELECT clause. Hence, both fixed metadata, and metadata defined as part of document classes can be extracted.
- Extracts data from a list of folders.
- Can optionally extract raw content for electronic documents, and annotations for images (scanned documents).
- For electronic documents, only content that belongs to a configurable list of formats is extracted. For other formats, only the metadata is extracted.
- Extracts the ACL information for each document, that is, which group has read access to the instances of the class.
- Can be configured with folder access rights of groups, so that the access is restricted to a combination of the document class and the folder where the document is located.
- Is able to do incremental updates based on checksum calculations of extracted data. A built in HSQL (HyperSonic) database, or an external MySQL database can be used to store the checksums.
- Performs extensive, configurable logging to stdout, file, and also to the LogServer of Data Search.

The user monitor connector has the following main features:

- It extracts all listed users in Image Services
- It extracts the group memberships for these users, and also the group/group memberships.
- It passes this information to an instance of the generic user monitor in SAM.

Chapter 2

Installation

About this Chapter

This chapter describes how to install the FAST Smart Connector for FileNet Image Services. It includes:

- Before You Install
- Installing the FileNet 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 FileNet Connector is installed.
- Include the FlexLM license key for the FileNet Connector in the license file for FAST Data Search. Otherwise you will be asked for it when running the FileNet Connector.

If necessary, you can use a license file that points to a remote license manager:

```
SERVER <host of license server> ANY  
VENDOR FASTSRCH  
USE_SERVER
```

- The FileNet Connector requires Java 1.4.1 or newer. Set JAVA_HOME to the installation directory.
- The FileNet Connector requires a FileNet Image Services Resource Adapter (ISRA).
- The FileNet Connector user account must have rights to retrieve all content desired, as well as user and group information. (See *FolderGroupAccess* on page 8.) Because of limitations of the ISRA API, the user account must be able to have a number of

concurrent sessions open. This also allows multiple instances of the connector to run at the same time.

- Add a mapping from the library that is accessed to the server IP address in the hosts file on the connector machine. On Windows this is:

```
C:\Windows\System32\Drivers\etc\hosts
```

For a domain 'Imaging' and organization 'FileNet', this could be

```
192.168.49.112 imaging-filenet-nch-server
```

Installing the FileNet Connector

The FileNet Connector is packaged in a zip file:

```
FileNetISConnector.zip
```

Unzip this package to your preferred location.

Note! Fast recommends that you limit access to the FileNet Connector directory to those who need it. Some configuration files may contain unencrypted passwords as described in Chapter 3 *Configuration*.

The directory has the following structure (if FileNetISConnector is the install directory):

Location	Content
FileNetISConnector	Run scripts
FileNetISConnector/doc	This document
FileNetISConnector/no	Subfolders include java classes for the connectors
FileNetISConnector/etc	Various configuration and properties files, certificate for user monitor.
FileNetISConnector/lib	Various libraries needed by the connectors
FileNetISConnector/docproc	A document processor for performing mappings of field names.

Chapter 3

Configuration

About this Chapter

This chapter describes how to configure the FAST Smart Connector for FileNet Image Services. It includes:

- Configuring the Content Connector
- Configuring the User Monitor Connector

Configuring the Content Connector

The connector is configured manually through an XML file:

```
<?xml version="1.0" encoding="UTF-8"?>
<AccessorConfig>
  <Accessor id="filenet_is">
    <parameter name="User">
      <![CDATA[user3]]>
    </parameter>
  </Accessor>
</AccessorConfig>
```

.... Other parameters

```
</Accessor>
</AccessorConfig>
```

The available parameters are listed below. For sample values, see the example configuration file in Appendix A *Sample Configuration File*.

Parameter	Description
User	User name in FileNet Image Services

Fast Data Search

Password	Password for the user
Domain	Domain or library name
Organization	Organization name
Query	SELECT that defines which data to extract.
IncludeContents	If true, include raw content for electronic documents
AcceptedFormats	Comma separated list of accepted formats. A value of ALL_FORMATS means that all document formats are to be accepted. <hr/> Note! When using ALL_FORMATS, consider that images are not processed meaningfully in a standard Data Search document processing pipeline. <hr/>
IncludeAnnotations	If true, include annotations for a scanned document
IncludeACLs	If true, include ACL information for extracted docs
SecurityDomainID	Unique three letter id for the security domain represented by FileNet Image Services
IncludeFolderNames	If true, include the names of the folders that the documents are associated with.
IndexFolders	Comma separated list of folder names that the query is applied to.
FolderGroupAccess	Grants groups persistent read access to folders. For example: <pre>/Diverse#Testgroup:Imaging:FileNet , /SecondMain#Testgroup:Imaging:FileNet</pre> grants Testgroup read access to folders /Diverse and /SecondMain and subfolders of these (unless there are conflicting access rights on those folders).
MaxMemory	Maximum memory in the internal queue.
MaxRowsFromIS	Maximum number of rows to extract from FileNet Image Services
CharacterEncoding	Character encoding used by FileNet Image Services.
PrimaryKey	The meta data that is used as a unique identifier for content. By default, this is F_DOCNUMBER.
FDSHostName	Host name for the DS node that has the config server.

FDSConfigPort	Port for the config server
FDSNameServicePort	Port number for the CORBA name service
OutputDirectory	The folder in which to store checksums for extracted data, as well as the user database.
CollectionName	The collection that will keep the data.
AbortPercent	A percent value, say 10, that limits the number of documents that can be deleted in a single run of the connector. If less than 10 percent of the docs are deleted, the operation is performed and committed, otherwise the documents are not deleted from the checksum database.
BatchSize	Batch size for sending new documents to Data Search.
BatchSizeFromIS	Batch size for extracting documents from ISRA.
UseBytesForString	If true, use bytes for string values. This will resolve some problems with some character formats as UTF8.
DataField	The field of the DS documents that will get the raw content when this is retrieved.
DataSourceName	Name of data source (optional)
ExcludeChecksumOn	Comma separated list of elements not to use in change detection. The default is <code>data,isannotations</code> which means that the raw content and the annotations are not checked for updates.

In addition to the configuration file, the following properties files can be modified:

log4j.properties

Properties that control logging to stdout and to Data Search. In the following sample, only warnings are logged (other options include info, debug), both to stdout, and to the log server at the host gre-tang.

```
log4j.rootCategory=warn, A1, A2
log4j.appender.A1=org.apache.log4j.ConsoleAppender
log4j.appender.A1.layout=org.apache.log4j.PatternLayout
log4j.appender.A1.layout.ConversionPattern=%-5p %d{ISO8601} %c{1}\n
%m%n%n
log4j.appender.A2=no.fast.connector toolkit.common.DataSearchAppender
log4j.appender.A2.logHost=gre-tang
log4j.appender.A2.logPort=26010
```

no.fast.connectortoolkit.properties

Properties that determines which checksum database to use (HSQL or MySQL), the connection information for the chosen database, whether to dump all data to DS in each run, and whether to enable change detection.

Note! Other properties files in the etc directory should be left untouched!

Configuring the User Monitor Connector

The user monitor connector is configured in the same configuration file. The additional parameters are as follows:

Parameter	Description
FDSSecurityHost	Comma separated list of hosts where user monitor is installed
FDSSecurityPort	Comma separated list of ports of hosts where user monitor is installed
FDSSecurityCertificate	Comma separated list of paths to security certificates to use.
FDSSecurityCertPass	Comma separated list of pass phrases for the corresponding certificates.

For generating certificates, see the security module documentation. You can also use the sample certificate that is contained within the connector zip, with the pass phrase found in the sample configuration file.

Chapter 4

Configuring FAST Data Search

About This Chapter

Before you feed any content to FAST Data Search, you must integrate the FileNet Connector into your FAST Data Search installation. This chapter includes:

- About Configuring FAST Data Search
- Configuring an Index Profile
- Creating a Collection for Extracted Data

About Configuring FAST Data Search

In order to use the FileNet Connector, you must configure a FAST Data Search cluster to function as a dedicated FileNet indexer by attaching it to an index profile. The default FAST Data Search cluster (*webcluster*) is used for crawled web content.

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

The indexer organizes documents according to how you configure clusters to make data searchable. When you select a cluster, you select in which set of search nodes the document will reside. Each cluster can have a number of collections.

For information about creating and editing clusters, refer to the *Indexing Database Content and XML Guide* chapter on *Document Processing and the Index Profile*.

Configuring an Index Profile

An index profile defines how content is to be indexed. This section provides a simplified description of how to set up an index profile for the FileNet Connector. For detailed information about creating and updating index profiles, refer to the:

- *Configuration Guide* chapter on *Index Profile*.
- *Indexing Database Content and XML Guide* chapter on *Document Processing and the Index Profile*.

Creating an Index Profile

The FileNet Connector provides a sample index profile in the *etc* folder that you can copy and modify. When you customize the default index profile, be sure that the FileNet Connector index profile is compatible with the FAST Data Search index profile.

If you want to index raw content, say, PDFs or Word documents, we suggest that you use one of the default index profiles. In the distribution of this connector (etc directory), you will find a sample index profile that is built on the datasearch-default index profile. This has been extended with the meta fields that is extracted by a query to FileNet Image Services, and also with security related fields. These are listed in the following.

NOTE: In order to secure search to be performed, the security related fields need to have the same name and semantics as specified below.

```
<!-- FileNet fields -->
  <field name="farchivedate" type="datetime" element-
name="F_ARCHIVEDATE"/>
  <field name="fclosed" element-name="F_CLOSED"/>
  <field name="fdeletedate" type="datetime" element-name="F_DELETEDATE"/>
  <field name="fdocclassname" element-name="F_DOCCLASSNAME"/>
  <field name="fdocclassnumber" type="int32" element-
name="F_DOCCLASSNUMBER"/>
  <field name="fdocformat" element-name="F_DOCFORMAT"/>
  <field name="fdoclocation" element-name="F_DOCLOCATION"/>
  <field name="fdocnumber" element-name="F_DOCNUMBER"/>
  <field name="fdoctype" element-name="F_DOCTYPE"/>
  <field name="fentrydate" type="datetime" element-name="F_ENTRYDATE"/>
  <field name="fpages" type="int32" element-name="F_PAGES"/>
  <field name="fretentbase" type="int32" element-name="F_RETENTBASE"/>
  <field name="fretentdisp" type="int32" element-name="F_RETENTDISP"/>
  <field name="fretentoffset" type="int32" element-name="F_RETENTOFFSET"/>
>
  <field name="ftestnumeric" type="int32" element-name="testnumeric"/>
  <field name="fteststring" element-name="teststring"/>
  <field name="ftestdate" type="datetime" element-name="testdate"/>
  <field name="ffoldernames" element-name="isfoldernames"/>
```

```

<field name="fannotations" element-name="isannotations"/>
<!-- Security fields -->
<field name="docacl"/>
<field name="docfolderacl"/>
<field name="docaclsystemid"/>

```

Uploading an Index Profile

To upload the FileNet Connector sample index profile to FAST Data Search:

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

The screenshot shows the FAST Data Search 4.0 Admin interface. The top navigation bar includes 'Collection Overview', 'Document Processing', 'Search View', and 'System Management'. Under 'System Management', the 'Matching Engines' link is highlighted with a pink box. Below the navigation bar, the 'Collection Overview' section is visible, showing 'Overview - There are no Collections'.

Copyright © 2004 Fast Search & Transfer ASA

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

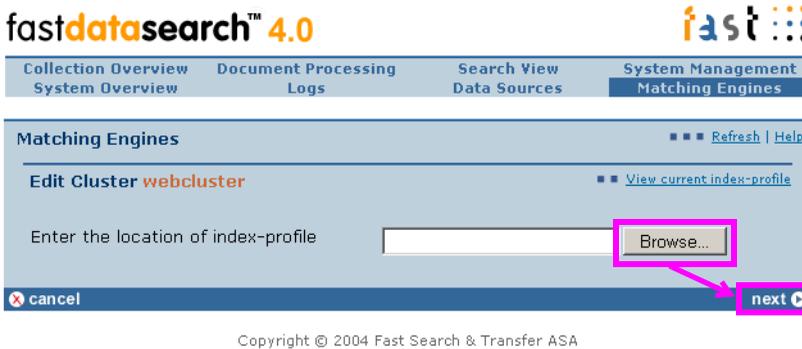
The screenshot shows the FAST Data Search 4.0 Admin interface with the 'Matching Engines' page selected. A table titled 'Search Engines' is displayed with the following data:

Host	Port	Type	Cluster
bos4ug0w.ad.fast.no	15100	Search Dispatcher	webcluster
bos4ug0w.ad.fast.no	15674	Search Column	webcluster

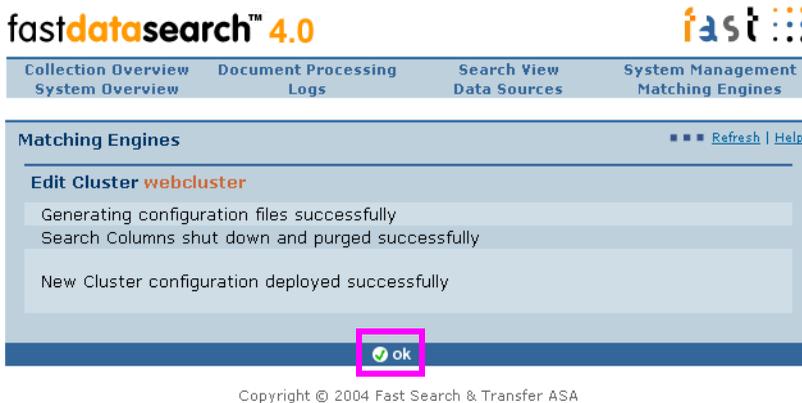
The icon next to the 'Search Dispatcher' host entry is highlighted with a pink box.

Copyright © 2004 Fast Search & Transfer ASA

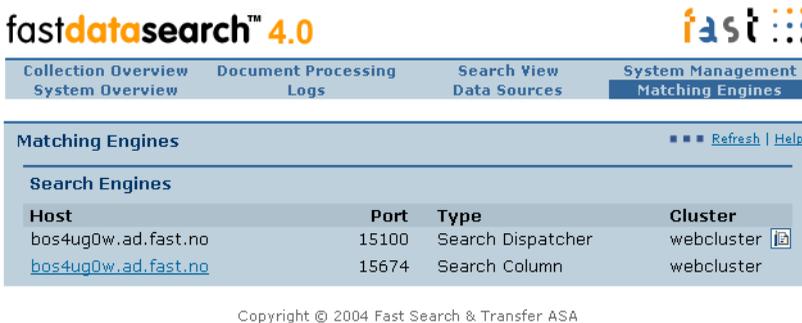
- 4 In the **Edit Cluster** screen, click the **Browse** button.



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



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

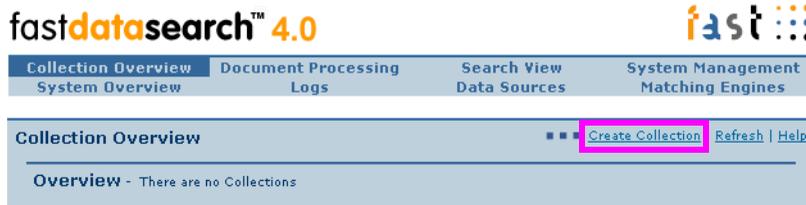


Creating a Collection for Extracted Data

A collection is a logical group of documents. Collections are set up in order to group documents based on selected criteria such as semantics (for example, similar types of documents) and/or document processing (for example, through pipeline configuration).

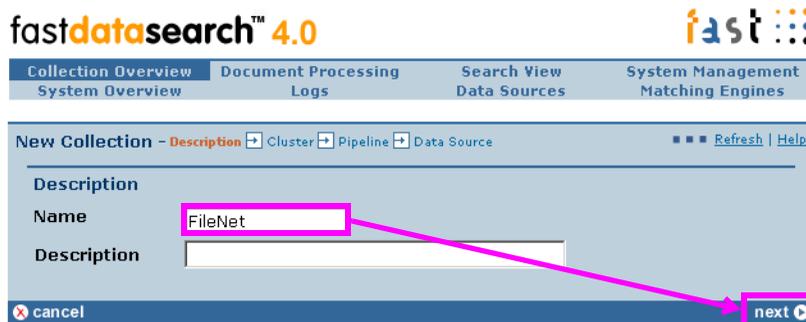
You need to create a minimum of one collection in FAST Data Search in order to receive extracted data. Complete this subsection to add a collection using the newly created pipeline. Refer to the *Indexing Database Content and XML Guide* chapter on *Document Processing and the Index Profile* for details.

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



Copyright © 2004 Fast Search & Transfer ASA

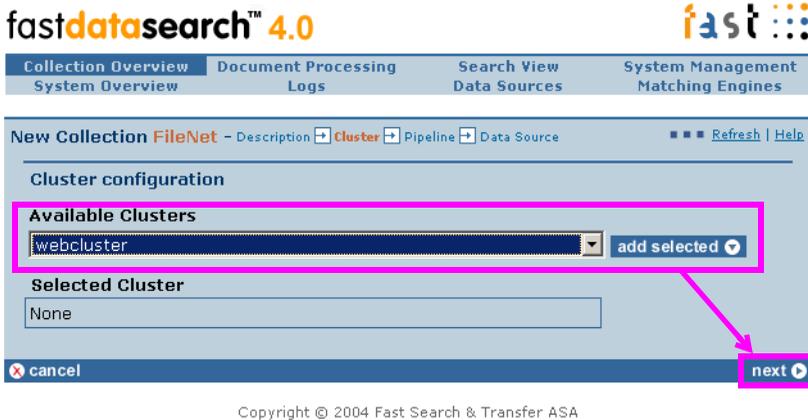
- 4 In the **Name** field enter **FileNet**.
- 5 Optionally fill in the **Description** field.
- 6 Click **next**.



Copyright © 2004 Fast Search & Transfer ASA

- 7 In the **Cluster Configuration** screen, click the **Available Clusters** dropdown list button.

- 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 an Index Profile* on page 12.
- 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.

Note! If you mix content from multiple sources, you may want to map fields specific to the sources to common names in the index profile. The FileNet Connector comes with a document processor stage that performs simple mappings from one name to another. See *Mixing Content From Multiple Sources* on page 18 for more information.

- 13 Click **add selected**.

14 Click **next**.

The screenshot shows the 'fastdatasearch™ 4.0' interface. The top navigation bar includes 'Collection Overview', 'Document Processing', 'Search View', and 'System Management'. The current view is 'New Collection FileNet - Description' with sub-tabs for 'Cluster', 'Pipeline', and 'Data Source'. The 'Pipeline configuration' section is active, showing a list of 'Available Pipelines'. The list includes: 'Express (webcluster)', 'FastWMI (webcluster)', 'Generic (webcluster)', 'NewsSearch (webcluster)', 'RowSet (webcluster)', and 'SiteSearch (webcluster)'. The 'Generic (webcluster)' pipeline is highlighted with a pink box. To the right of the list is an 'add selected' button. Below the list is a 'next' button, also highlighted with a pink box. Arrows point from the 'add selected' button to the 'next' button.

15 In the **Data Source Configuration** screen, click **ok**.

The screenshot shows the 'fastdatasearch™ 4.0' interface. The top navigation bar is the same as in the previous screenshot. The current view is 'New Collection FileNet - Description' with sub-tabs for 'Cluster', 'Pipeline', and 'Data Source'. The 'Data Source configuration' section is active, showing a list of 'Available Data Sources'. The list is currently empty, with a 'Select Data Source' dropdown menu and an 'add selected' button. Below the list is a 'Configured Data Sources' section, which is currently empty and labeled 'None'. At the bottom of the screen, there is an 'ok' button highlighted with a pink box.

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

fastdatasearch™ 4.0

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

Edit Collection FileNet Overview | Refresh | Help

Description
No description set

Control Panel

Edit Description | Edit Cluster | Add Document | Search
Edit Data Sources | Edit Pipeline | Delete Document | Delete Collection

Status

Data Sources: N/A (None configured) → Content Distributor: OK (Docs: 0) → Pipeline: OK (Generic (webcluster)) → Matching Engines: OK (Check logs)

Modules

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

Cluster
webcluster

ok

Copyright © 2004 Fast Search & Transfer ASA

17 The admin interface returns to the Collection Overview dialog.

fastdatasearch™ 4.0

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

Collection Overview Create Collection | Refresh | Help

Overview - There is just one Collection

Name	Description	Last input	Docs
FileNet		N/A	0

Copyright © 2004 Fast Search & Transfer ASA

Mixing Content From Multiple Sources

If you mix content from multiple sources, you may want to map fields specific to the sources to common names in the index profile. As an example, you may have an 'author' field in one source and a 'creator' field in another source. When searching across all

content sources, you may want to be able to search for authors by having a common author field. To this end, the FileNet Image Services connector comes with a document processor stage that performs simple mappings from one name to another. It is configured by the path to a configuration mapping file. An example of such a file is found in the etc directory of the distribution.

In order to install the mapping, do the following:

- 1** Copy the file `MultiAttributeMapper.py` to the directory `lib/python2.2/processors` of each document processing node in your DS installation.
- 2** Copy the file `MultiAttributeMapper.xml` to the `etc/processors` directory of each document processing node in your DS installation.
- 3** Restart all procservers on all document processor nodes.

Chapter 5

Operation

About this Chapter

This chapter describes how to use the FAST Smart Connector for FileNet Image Services. 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 runContentConnector script as follows:

```
runContentConnector start -f <name of config file> -l <path to config file>
```

Example:

```
runContentConnector start -f config.xml -l etc
```

Stopping a Running Connector

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

```
runContentConnector stop -f <name of config file> -l <path to config file>
```

Example:

```
runContentConnector stop -f config.xml -l etc
```

Message to Reviewer! <Document stop feature as well?>

Logging

The FileNet 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.

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

Message to Reviewer! <Explain how to do search with security pipeline, see CMS/FDSSM doc>

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 runumconnector script as follows:

```
runumconnector -f <name of config file> -l <path to config file>
```

Example:

```
runumconnector -f config.xml -l etc
```

Logging

The FileNet 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 (SAM) to verify that users and group information is extracted properly.

Appendix A

Sample Configuration File

filename.XML

The sample is also found in the etc directory of this distribution.

```
<?xml version="1.0" encoding="UTF-8"?>
<AccessorConfig>
  <Accessor id="filenet_is">
    <parameter name="User">
      <![CDATA[SysAdmin]]>
    </parameter>
    <parameter name="Password">
      <![CDATA[SysAdmin]]>
    </parameter>
    <parameter name="Domain">
      <![CDATA[Imaging]]>
    </parameter>
    <parameter name="Organisation">
      <![CDATA[FileNet]]>
    </parameter>
    <parameter name="IncludeContents">
      <![CDATA[true]]>
    </parameter>
    <parameter name="AcceptedFormats">
      <![CDATA[text/plain,application/octet-stream]]>
    </parameter>
    <parameter name="IncludeAnnotations">
      <![CDATA[false]]>
    </parameter>
    <parameter name="IncludeACLs">
      <![CDATA[true]]>
    </parameter>
    <parameter name="SecurityDomainID">
```

```
    <![CDATA[fis]]>
</parameter>
<parameter name="IncludeFolderNames">
    <![CDATA[true]]>
</parameter>
<parameter name="IndexFolders">
    <![CDATA[/]]>
</parameter>
<parameter name="FolderGroupAccess">
    <![CDATA[/Diverse#Testgroup:Imaging:FileNet, /
SecondMain#Testgroup:Imaging:FileNet]]>
</parameter>
<parameter name="MaxMemory">
    <![CDATA[200000000]]>
</parameter>
<parameter name="MaxRowsFromIS">
    <![CDATA[1000000]]>
</parameter>
<parameter name="Query">
    <![CDATA[SELECT
F_ARCHIVEDATE, F_CLOSED, F_DELETEDATE, F_DOCCLASSNAME, F_DOCCLASSNUMBER, F_DOC
FORMAT, F_DOCLOCATION, F_DOCNUMBER, F_DOCTYPE, F_ENTRYDATE, F_PAGES, teststring
FROM FnDocument]]>
</parameter>
<parameter name="CharacterEncoding">
    <![CDATA[iso-8859-1]]>
</parameter>
<parameter name="PrimaryKey">
    <![CDATA[F_DOCNUMBER]]>
</parameter>
<parameter name="FDSHostName">
    <![CDATA[gre-tang]]>
</parameter>
<parameter name="FDSConfigPort">
    <![CDATA[26005]]>
</parameter>
<parameter name="FDSNameServicePort">
    <![CDATA[26099]]>
</parameter>
<parameter name="OutputDirectory">
    <![CDATA[C:\output\checksums]]>
</parameter>
<parameter name="CollectionName">
    <![CDATA[filenet]]>
</parameter>
<parameter name="AbortPercent">
    <![CDATA[90]]>
</parameter>
<parameter name="BatchSize">
    <![CDATA[10]]>
```

```
</parameter>
<parameter name="UseBytesForString">
  <![CDATA[true]]>
</parameter>
<parameter name="DataField">
  <![CDATA[data]]>
</parameter>
  <parameter name="DataSourceName">
    <![CDATA[FileNetSource1]]>
  </parameter>
<parameter name="ExcludeChecksumOn">
  <![CDATA[data,isannotations]]>
</parameter>
<parameter name="FDSSecurityHost">
  <![CDATA[gre-geirw-1t]]>
</parameter>
<parameter name="FDSSecurityPort">
  <![CDATA[112]]>
</parameter>
<parameter name="FDSSecurityCertificate">
  <![CDATA[../etc/fdssm_client.key]]>
</parameter>
<parameter name="FDSSecurityCertPass">
  <![CDATA[FDSFDS]]>
</parameter>
</Accessor>
</AccessorConfig>
```

