



FAST SMART CONNECTOR FOR MS EXCHANGE

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.

Microsoft Exchange® 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
Overview.....	1
Features	1
Supported Platforms.....	2
How the Exchange Connector Works.....	2
About Security	3
About Incremental Indexing.....	3
About Real-Time Indexing.....	4
Chapter 2 Installing and Configuring the Exchange Connector	5
Before You Install	5
Installing the Connector.....	6
Changing the FAST Data Search Configuration.....	8
Uninstalling the Exchange Connector	8
Chapter 3 Configuring FAST Data Search to Index MS Exchange Data	9
Configuring the Index Profile	9
Create the Custom Stages	12
Create the Document Processing Pipeline.....	17
Create a Collection.....	18

Chapter 4 Operation 23

- Running the Exchange Connector for the First Time..... 24
- Command Line Options 25
- Folders Configuration File..... 28
 - Exclude Folders 29
- Reinitializing the Exchange Connector 30

Chapter 5 Troubleshooting the Exchange Connector 31

- About Troubleshooting the Exchange Connector 31
- Verifying the Exchange Version..... 31
- Inspecting the XML Query Responses and Documents 32
- Logging 32
 - Configuring the Logging Levels 32
 - Example Logging Configuration File..... 33
- Fast Data Search Logging 34
- Common Error Situations..... 34
 - Incorrect File Names for Attachments..... 34
 - Cannot Query Documents 34
 - Callback-Related Errors Occur During a Run..... 35
 - Invalid or Missing Certificate..... 35
 - Message Size Exceed Limit Error..... 36
 - Cannot Access Files When Crawling an OWA Site..... 36

Appendix A Index Profile 37

- datasearch-4.0-exchange.xml..... 37

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 MS Exchange 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 Exchange Connector functions.
- **System Integrators and Administrators**, who need to know how to integrate the Exchange Connector into existing installations of FAST Data Search and MS Exchange.

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 and contents of plaintext ASCII files are represented using the following format:

```
All hostnames should be fully qualified domain names.
```

```
What is the first node in the system?
```

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

```
What is the first node in the system? host1.mysite.com
```

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

Which port should the log server run on (default:16100)? **<enter>**

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

Run *Setup.exe* and follow the instructions.

- User Interface page/window texts, buttons, and lists appear in the following format:

Click on **Start Menu**.

Chapter 1

Introduction

About this Chapter

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

- Overview
- Features
- Supported Platforms
- How the Exchange Connector Works

Overview

The FAST Smart Connector for MS Exchange is a stand alone module that exports documents from MS Exchange Server to FAST Data Search 4.0. The Exchange Connector traverses and indexes the contents of an existing folder on an Microsoft Exchange Server. It can run on the Exchange Server machine or another machine.

Features

The Exchange Connector provides the following features:

- Extracts E-mail and attachments from MS Exchange Server user accounts
- Extracts documents and attachments from MS Exchange Server public folders
- Extracts document content as well as message metadata (To: From: Subject:)
- Supports incremental updates to the Fast Data Search index.
- Supports remote access to MS Exchange Server

- Supports deletion of documents from the index when deleted from the server
- Supports FAST Data Search Security Access Module

Supported Platforms

FAST Smart Connector for MS Exchange

- Any Windows platform on which Fast Data Search 4.0 is supported

Microsoft Exchange Server

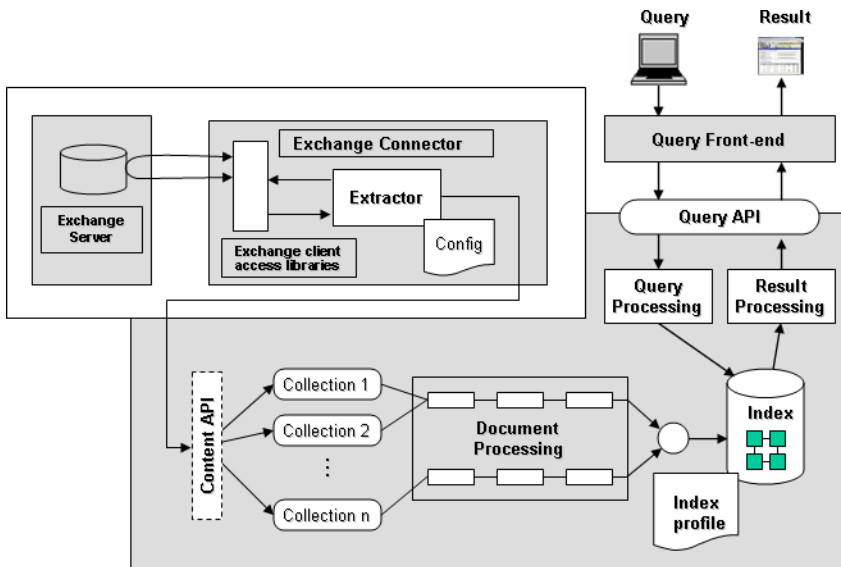
- Exchange Server 2000 or 2003 on any platform

FAST Data Search

- Version 4.0 or later on any platform

How the Exchange Connector Works

The Exchange Connector is a command-line utility that extracts content from Exchange and submits it directly to FAST Data Search using the Content API. The architecture of the connector is shown below.



The Exchange Connector is started from the command line. It uses the Exchange SDK (version 1.0) to connect to Exchange and uses WebDAV and Collaboration Data Objects (CDO) to communicate with Exchange. It traverses folders recursively and extracts all records matching the filters and the associated security. It extracts:

- mail messages and attachments in private mailboxes
- public folder postings with attachments
- documents posted directly to folders

Note! This version of the Exchange Connector does not index contacts, calendar entries, or tasks. Contact Fast Data Search if you need this functionality.

The Exchange Connector represents each record from the Microsoft Exchange Server as an individual document. Columns (or fields) such as sender, subject, recipient list are stored as document attributes and are mapped to searchable fields within FAST Data Search.

Once the content is submitted to FAST Data Search, it is treated as any other content. In other words, it is sent through document processing pipelines and then indexed to be searchable and filterable. See the *FAST Data Search System Reference Guide, Processing Documents* for more information.

The Exchange Connector supports all document formats that are supported by FAST Data Search. The connector downloads binary documents and feeds them raw to FAST Data Search which processes them in the standard way. For more details on what document formats are supported by FAST Data Search, consult the FAST Data Search standard documentation set.

About Security

The Exchange Connector leverages the security mechanisms within Windows and Active Directory by indexing the Access Control Lists (ACLs) for every document (mail message, attachment, etc.) in addition to the content. This allows for efficient searching where users only receive search results that they have permissions to read.

About Incremental Indexing

By default, the Exchange Connector supports incremental indexing by tracking which documents have already been indexed. No checksums are used. This method avoids the overhead of downloading headers for messages that are already indexed and results in a significant performance gain.

The default behavior is based on the assumption that mail messages are considered to be historical data that does not change. However, this may not be the case at your site. Exchange allows users to change posted documents (messages or other documents such as Word files) and the ACLs of those documents. To cover that situation, you can use the `-n` command line option to enable the use of checksums to detect changes. See *Command Line Options* on page 25 for more information about the `-f` option.

The behavior of the connector is identical for private and public folders. However, for private folders, you should always set the `owneruid` parameter as described in *Folders Configuration File* on page 28. Do not set this parameter for public folders.

About Real-Time Indexing

If you need instant indexing of all content that enters or leaves your Exchange Server, contact FAST for more information. For example, you can get real time alerts on your mobile phone if an email containing specific words enters your mailbox, or your management can be notified if an employee sends a message containing inappropriate content.

Chapter 2

Installing and Configuring the Exchange Connector

About this Chapter

This chapter explains how to install the Exchange Connector as an add-on to an existing FAST Data Search installation. It includes:

- Before You Install
- Installing the Connector
- Changing the FAST Data Search Configuration
- Uninstalling the Exchange Connector

Before You Install

- A license file is needed to run the Exchange Connector as described in *Running the Exchange Connector for the First Time* on page 24. Contact FAST at fds-support@fastsearch.com to obtain further information on how to receive a license file.
- Fast Data Search must be installed and running, not necessarily on the machine where the Exchange Connector is installed. If the Exchange Connector is installed on a machine that does not run Fast Data Search:
 - a** Copy the file `$FASTSEARCH/etc/omniorb.cfg` from your Fast Data Search installation to the machine where the connector runs.
 - b** Set the environment variable `OMNIORB_CONFIG` to point to this file.
Examples:

```
OMNIORB_CONFIG=C:\Program Files\FAST\omniorb.cfg
```

```
OMNIORB_CONFIG=C:\datasearch\etc\omniorb.cfg
```

```
OMNIORB_CONFIG=C:\ Program Files\FAST\ExchangeConnector  
\omniorb.cfg
```

- An Exchange 2000 Server system must be fully patched with all updates from Microsoft including SP3 and post SP3 hotfixes in order to be able to correctly render all attachment links. To obtain the latest Exchange Server Service Pack:

<http://support.microsoft.com/default.aspx?scid=kb;EN-US;301378>

- Copy the file `cdoex.dll` from your Exchange server to the machine on which the connector runs. The minimum version number is 6.0.6511.0, which comes with Exchange 2000, SP4. Copy the file to:

```
C:\Program Files\Common Files\Microsoft Shared\CDO
```

and register it using:

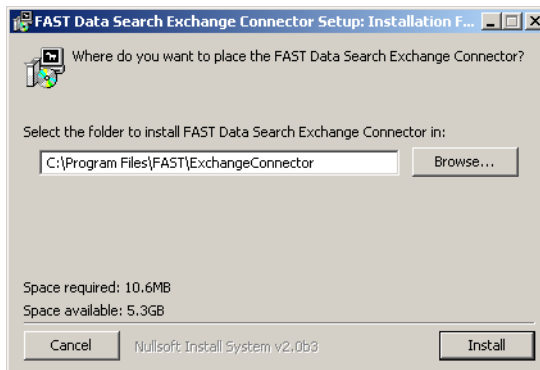
```
C:\Program Files\Common Files\Microsoft Shared\CDO> regsvr32 cdoex.dll
```

- Install one of the following on the machine running the Exchange Connector:
 - Outlook 2003
 - *Office 2003 Research Service Software Development Kit (SDK)*
<http://www.microsoft.com/downloads/details.aspx?FamilyID=d3fc8129-63f7-43b5-8d99-de4058ade0ec&displaylang=en>

Installing the Connector

Use the following procedure to install the Exchange Connector.

- 1 Launch the setup file *Setup_FDSExchangeConnector.exe*.
- 2 Select the directory in which to install the connector and **Click** the *Install* button.



- 3 The following command prompt appears only if you have no other Fast Data Search products installed on this machine. If the command prompt appears, type “y” and

press Enter unless you prefer to specify the configuration information each time you run the connector.

```
*****
* Exchange Connector setup *
*****

Do you want to save your FDS system configuration to registry? [y/n]
No means that you will have to provide all parameters at command line
```

4 In the command prompt window, enter the:

- Exchange Server system id (**win**). The system id is needed if you intend to use the Exchange Connector with the Security Access Module
- Exchange Server host name (fully qualified name)
- FAST Data Search name service host name (fully qualified name)
- FAST Data Search name service port
- FAST Data Search collection name (exchange if you followed the instructions in the previous chapter. “Create a Collection” on page 18.)

```
[14:13:09] [0x14b41] -S==T==A==R==T
*****
* Exchange Connector setup *
*****

Please provide a system id for the Exchange Connector:
█

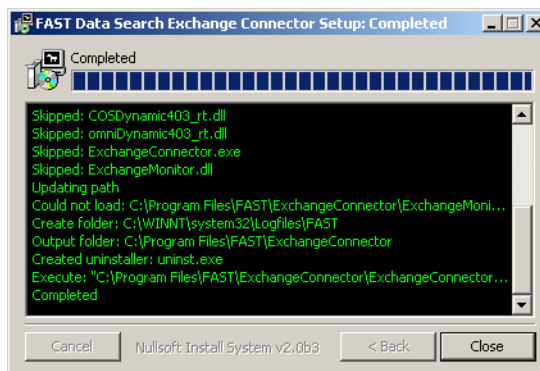
Please provide the fully qualified name of your Exchange Server:
█

Please provide the host name where the FDS name service is running:
█

Please provide the port for the FDS name service:
Port Idefault: 160991:

Please provide the name of your Exchange Collection:
█
```

5 The following screen appears. **Click** on the *Close* button to finish.



Changing the FAST Data Search Configuration

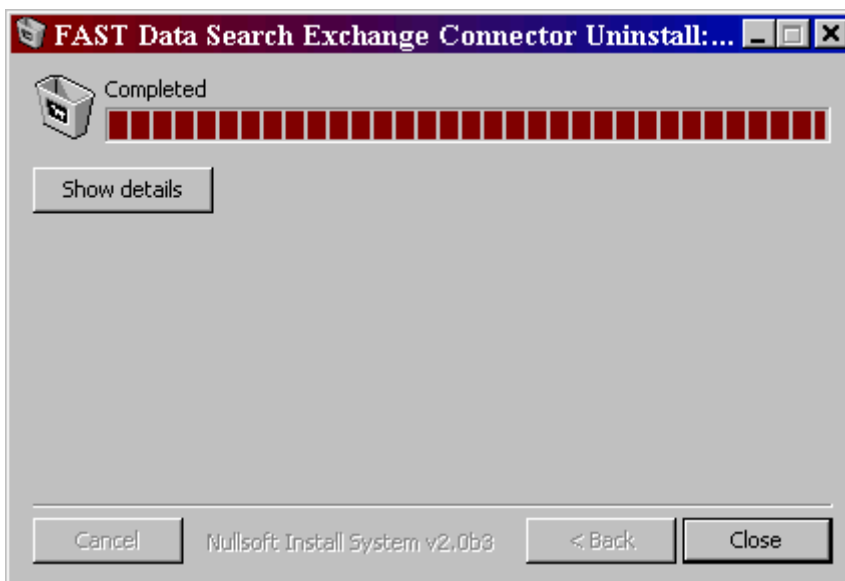
If you want to change the configuration parameters that were stored in the registry, you can run the command:

```
exchangeconnector -install
```

Uninstalling the Exchange Connector

Use the following procedure to uninstall the connector.

- 1 Go to Start -> Settings -> Control Panel -> Add/Remove Programs.
- 2 Highlight *FAST Exchange Connector* and click the **Change/Remove** button. The following screen appears.



- 3 Click on the **Close** button.
- 4 Delete the **changes.db** file.
- 5 Reverse all the **user privileges** that were set during the installation.

Message to Reviewer!

Chapter 3

Configuring FAST Data Search to Index MS Exchange Data

About This Chapter

FAST Data Search needs to be configured to index data from the FAST Smart Connector for MS Exchange. The following chapter describes how to setup FAST Data Search to index this data. It includes:

- Configuring the Index Profile
- Create the Custom Stages
- Create the Document Processing Pipeline
- Create a Collection

Configuring the Index Profile

The index profile in FAST Data Search must contain certain fields that are specific to security in Exchange:

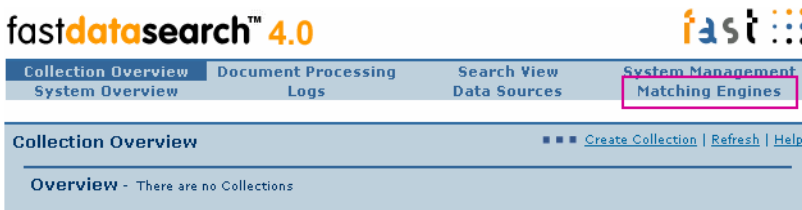
```
<!-- Exchange Connector fields -->
<field name="mailfrom" />           <!-- Sender -->
<field name="mailto" />            <!-- Receiver -->
<field name="mailsubject" />       <!-- Subject -->
<field name="mailbody" />          <!-- HTML version of message -->
<field name="mailattachmentcontainer" />
                                     <!-- Message owning this attachment -->
<field name="mailattachmentcontenttype" />
                                     <!-- MIME type for attachment -->
<!-- Security field: Document Access Control List -->
<field name="docacl" />
```

```
<!-- Security field: System id for domain to which this document belongs -->  
<!-- Required in result set for post processing -->  
<field name="docaclsystemid" />
```

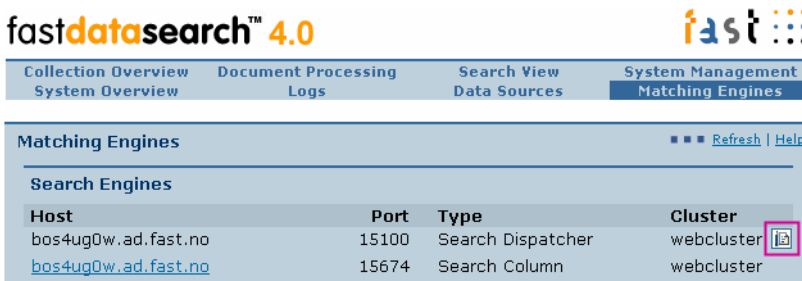
These fields are included in the index profile provided with the Exchange Connector (see Appendix A *Index Profile*).

Note! If you are currently using a custom index profile, it is recommended that you contact FAST Professional Services before modifying the existing index profile to support the exchange fields.

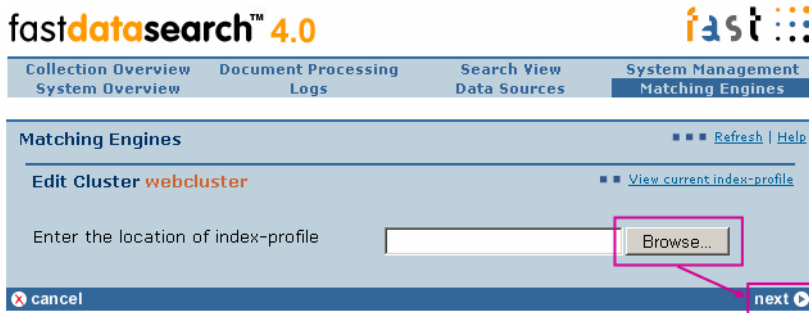
- 1 In the FAST Data Search administrator interface go to Matching Engines.



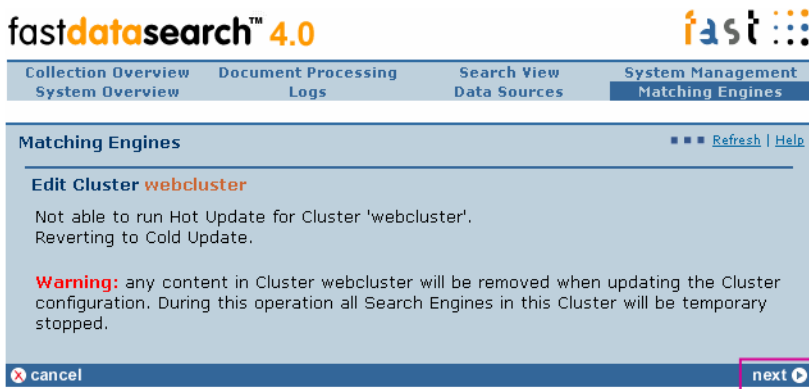
- 2 Click on the **edit** icon to update cluster configuration. (You may use an additional pre-defined cluster if necessary. Refer to the *FAST Data Search Configuration Guide* for further details).



- 3 Browse for the index profile that is supplied with the Exchange Connector distribution then click the **Next** button.



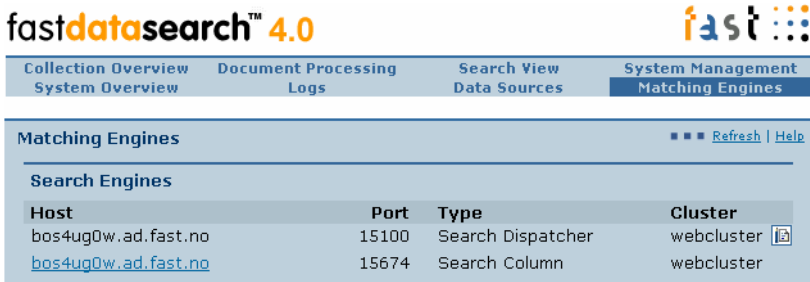
- 4 If you see this warning screen, click the **Next** button.



- 5 Verify that the index profile update has completed successfully.



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




fastdatasearch™ 4.0

fast

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

Matching Engines Refresh Help

Search Engines

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

Create the Custom Stages

Before creating a pipeline for MS Exchange, you need to create custom stages to be used in the pipeline. Use the following procedure to create these custom stages.

- 1 In the administrator interface, click on **Document Processing**.



fastdatasearch™ 4.0

fast

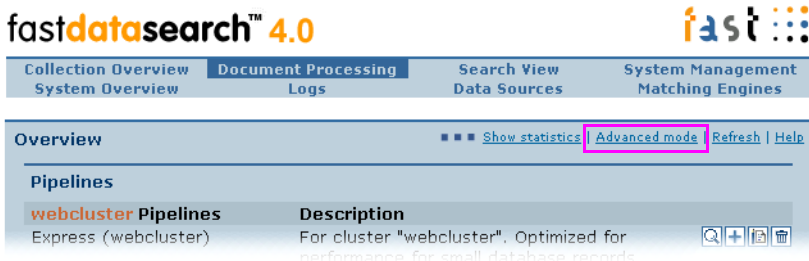
Collection Overview System Overview | Document Processing Logs | Search View Data Sources | System Management Matching Engines

Collection Overview Create Collection Refresh Help

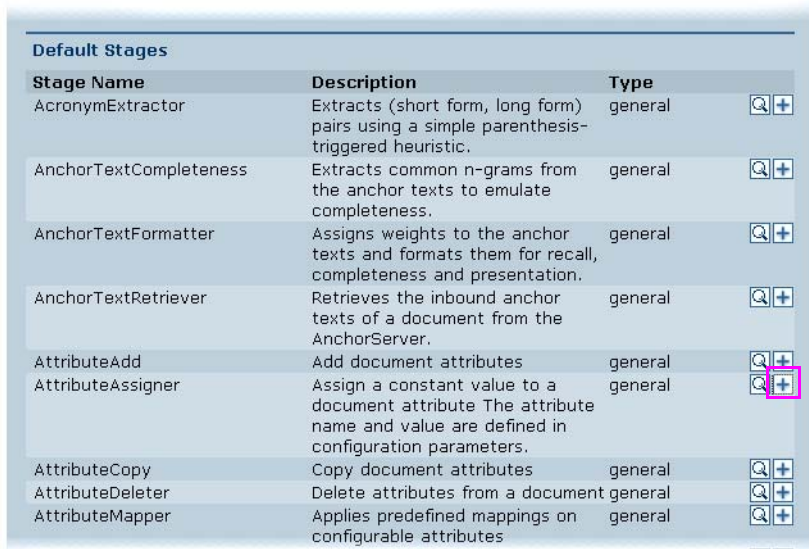
Overview - There are no Collections

Copyright © 2004 Fast Search & Transfer ASA

- 2 Click on the **Advanced mode** link.



- 3 Go to **Default Stages**, find **AttributeAssigner** and click the +.



- 4 Enter **AttributeAssignerExchange** in the *Name* field and enter **data** in the *Attribute* field. Click the **submit** button.

fastdatasearch™ 4.0 **fast** ::::

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

Create Stage Refresh | Help

Stage Info

Attribute Value

Class: AttributeAssigner

Name:

Description:

Configuration

Parameter	Value	Type
Attribute	<input type="text" value="data"/>	string
Value	<input type="text"/>	string

Copyright © 2004 Fast Search & Transfer ASA

- 5 Click the **ok** button.

fastdatasearch™ 4.0 **fast** ::::

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



















Document Processing Refresh | Help

Stage manipulation

Stage **AttributeAssignerExchange** is successfully set.

Copyright © 2004 Fast Search & Transfer ASA

- 6 Go to **Default Stages** under Stage Name and find *AttributeCopy* and click the +.

Default Stages			
Stage Name	Description	Type	
AcronymExtractor	Extracts (short form, long form) pairs using a simple parenthesis-triggered heuristic.	general	 
AnchorTextCompleteness	Extracts common n-grams from the anchor texts to emulate completeness.	general	 
AnchorTextFormatter	Assigns weights to the anchor texts and formats them for recall, completeness and presentation.	general	 
AnchorTextRetriever	Retrieves the inbound anchor texts of a document from the AnchorServer.	general	 
AttributeAdd	Add document attributes	general	 
AttributeAssigner	Assign a constant value to a document attribute The attribute name and value are defined in configuration parameters.	general	 
AttributeCopy	Copy document attributes	general	 
AttributeDeleter	Delete attributes from a document	general	 
AttributeMapper	Applies predefined mappings on configurable attributes	general	 

- 7 Enter **AttributeCopyExchange** in the *Name* field and enter the following string in the *Attributes* field.

```
mailbody:html mailattachmentcontenttype:mime
mailattachmentcontent:data mailattachmentcontainer:generic1
mailsubject:title
```

Note! Although the string is shown on multiple lines, enter it all on one line.

8 Click the **submit** button.

fastdatasearch™ 4.0 **fast** ::::

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

Create Stage Refresh | Help

Stage Info

Attribute	Value
Class	AttributeCopy
Name	AttributeCopyExchange
Description	Copy document attributes The document attribute named by the Input configuration parameter is copied to the attribute named by the Output parameter. Any existing values in the output is lost. If the input attribute does not exist, the operation is ignored. Several attributes can be copied using the Attributes parameter. It

Configuration

Parameter	Value	Type
Input		string
Attributes	ntcontent:data mailattachmentcontainer:generic1	string
Output		string

cancel submit

Copyright © 2004 Fast Search & Transfer ASA

9 Click the **ok** button.

fastdatasearch™ 4.0 **fast** ::::

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

Document Processing Refresh | Help










Stage manipulation

Stage **AttributeCopyExchange** is successfully set.

ok

Copyright © 2004 Fast Search & Transfer ASA

10 Verify that the custom stages have been created.

Custom Stages			
Stage Name	Description	Type	
AttributeAssignerExchange	Assign a constant value to a document attribute. The attribute name and value are defined in configuration parameters.	general	  
AttributeCopyExchange	Copy document attributes	general	  
BodySizeFilter	This filter discards documents	general	  

Create the Document Processing Pipeline













Data going into Fast Data Search is processed by a pipeline consisting of a number of Document Processors, each performing one or more operations/transformations on the data (e.g. the Lemmatizer processor performs lemmatization on the document). To learn more about Document processing, see the *Configuration Guide*.

Create a new pipeline as follows:

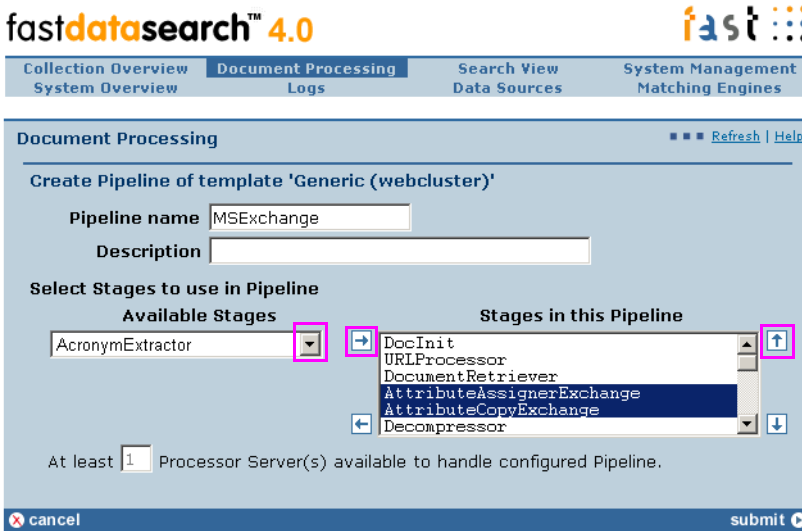
- 1 In **Document Processing Advanced Mode**, find the Generic (webcluster) pipeline and click the plus sign (+) icon. A pipeline defines a configured set of processing applied to a collection prior to indexing.

fastdatasearch™ 4.0

fast

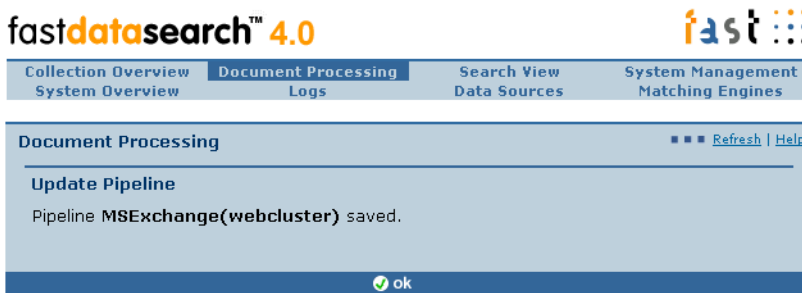
Collection Overview	Document Processing	Search View	System Management
System Overview	Logs	Data Sources	Matching Engines
Overview Show statistics Normal mode Refresh Help 			
Pipelines Create Pipeline without template 			
webcluster Pipelines		Description	
Express (webcluster)	For cluster "webcluster". Optimized for performance for small database records. Contains no document conversion, li...	  	
FastXML (webcluster)	For cluster "webcluster". FastXML pipeline (based on index-profile)	  	
Generic (webcluster)	For cluster "webcluster". Supports different document formats (HTML, Word, PDF etc). Additional support for synonym...	  	
NewsSearch (webcluster)	For cluster "webcluster". Optimized for News Search applications. Additional support for	  	

- 2 Enter **MSExchange** in the *Pipeline name* field.



Copyright © 2004 Fast Search & Transfer ASA

- 3 Select the **AttributeAssignerExchange** and **AttributeCopyExchange** stages from the dropdown list and add them to the Stages in this Pipeline list.
- 4 Use the up-arrow to place them right after the DocumentRetriever stage.
- 5 Click the **Submit** button.
- 6 Verify that the new document pipeline has been created and click the **Ok** button.



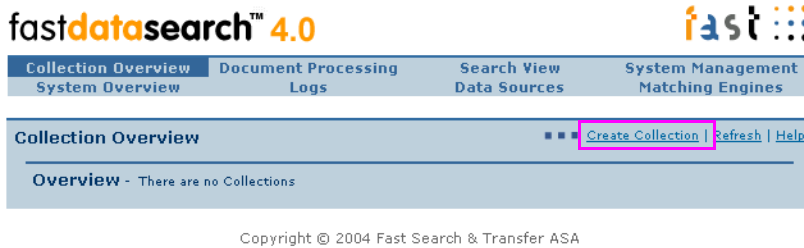
Copyright © 2004 Fast Search & Transfer ASA

Create a Collection

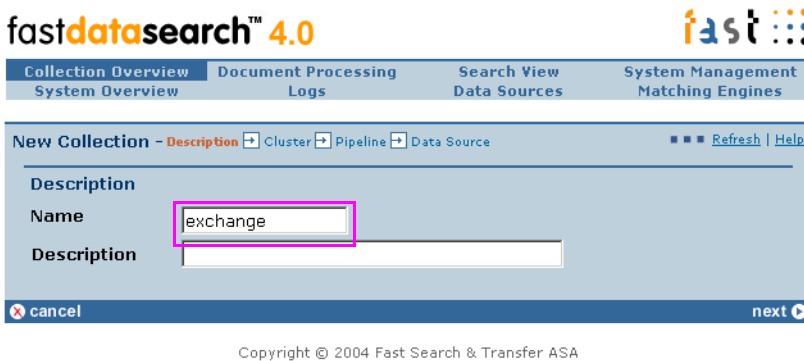
A collection is a logical group of documents. Collections are set up in order to group documents based on selected criteria such as semantics (for example, similar types of documents) and/or document processing (for example, through pipeline configuration).

You need to create a collection in FAST Data Search for feeding the MS Exchange content into. This collection should use the pipeline that you created in the previous section.

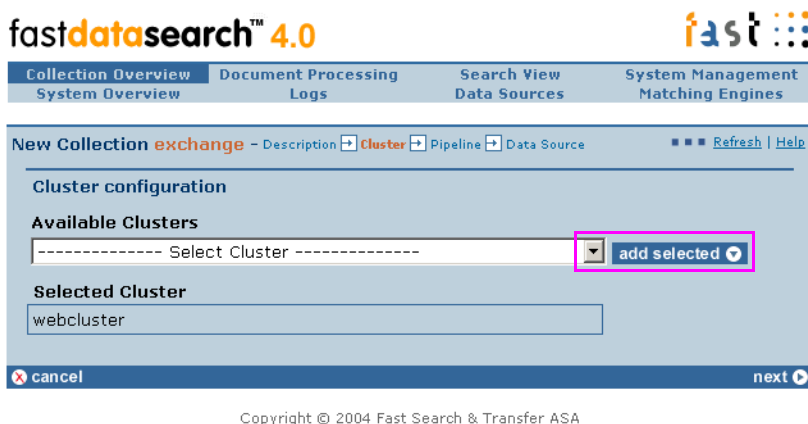
- 1 Select **Collection Overview** on the navigation bar and click **Create Collection**:



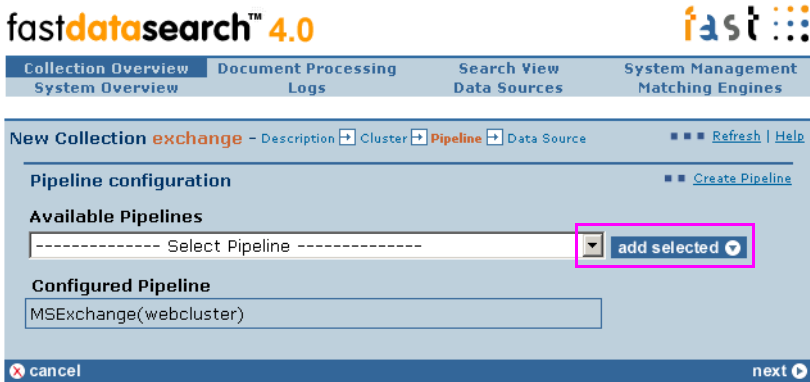
- 2 Enter **exchange** in the *Name* field (enter description if you like) then click the **Next** button.



- 3 In Available Clusters, select **webcluster**, click **add selected**, and click the **Next** button.

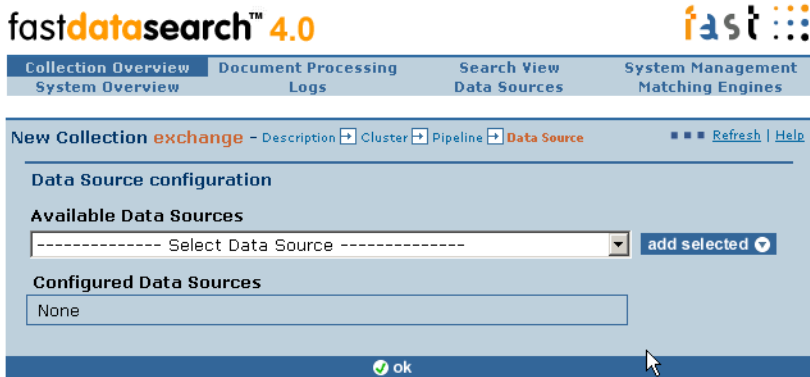


- 4 In Available Pipelines, select **MSEExchange (webcluster)**, click **add selected**, and click the **Next** button.



Copyright © 2004 Fast Search & Transfer ASA

- 5 In Data Source Configuration, click the **Next** button.



Copyright © 2004 Fast Search & Transfer ASA

- Verify that the new collection has been created and click the **Ok** button.

The screenshot shows the 'Edit Collection' interface for a collection named 'exchange'. The interface includes a navigation bar with tabs for '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, MSExchange (webcluster)), and 'Matching Engines' (OK, Docs: 0).
- Modules:** A table listing modules and their status.
- Cluster:** 'webcluster'.

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

At the bottom of the interface, there is a green checkmark and the text 'ok'.

Copyright © 2004 Fast Search & Transfer ASA

- The administrator interface returns to the Collection Overview dialog.

The screenshot shows the 'Collection Overview' interface. It features a navigation bar with tabs for 'Collection Overview', 'Document Processing', 'Search View', and 'System Management'. The main content area includes:

- Overview:** 'There is just one Collection'.
- Table:** A table showing the details of the 'exchange' collection.

Name	Description	Last input	Docs
exchange		N/A	0

Copyright © 2004 Fast Search & Transfer ASA

Your collection is now complete and ready to accept Exchange data from the connector.

Chapter 4

Operation

About this Chapter

This chapter explains how to operate the Exchange Connector. It includes:

- Running the Exchange Connector for the First Time
- Command Line Options
- Folders Configuration File
- Reinitializing the Exchange Connector

Running the Exchange Connector for the First Time

When you run the Exchange Connector for the first time, it prompts you for license information.

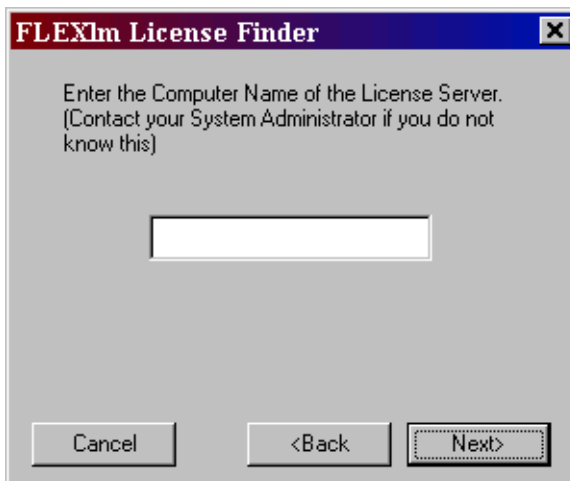
- 1 Run *ExchangeConnector.exe*. The following screen appears.



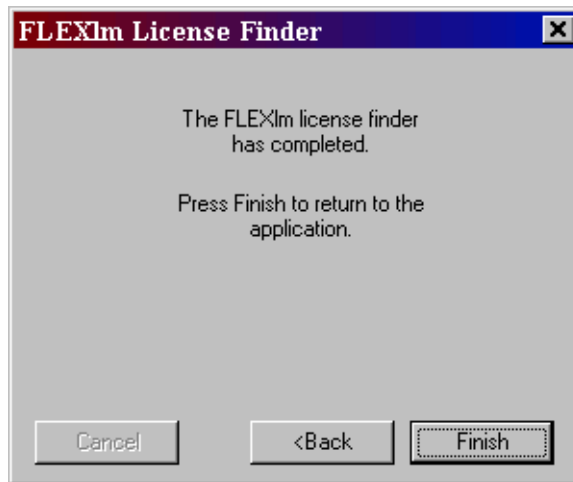
- 2 Choose the *Specify the License Server* radio button and click **Next**.

Note! The *Specify the License File* radio button is not applicable to the Exchange Connector.

The following screen appears.



- 3 Enter the computer name of the license server (refer to The *Licensing Manager* section in *Chapter 9* of the *FAST Data Search System Reference Guide*) and click the **Next** button. The following screen appears.



- 4 Setup is complete. Click on the **Finish** button.

Command Line Options

The Exchange Connector supports many command line options for flexibility. Typically, however, only a few options are needed.

Usage

```
ExchangeConnector [options...]
```

Note! Options stored during installation are not required on the command line

-F <FoldersConfig.xml>	Specifies an XML file containing a list of all folders/mailboxes to index. (See <i>Folders Configuration File</i> on page 28.)
-c <collection>	Specifies the name of the collection (see <i>Create a Collection</i> on page 18). Example: ExchangeConnector -c ExchangeColl99

`-a` Adds a specific user's mailbox to the FoldersConfig.xml file. This entry can then be manually modified to add excluded folders or duplicated for other folders, etc. Example:

```
ExchangeConnector -F FoldersConfig.xml  
-x https://mysite.com/exchange/jsmith/inbox  
-u mydomain\jsmith -p mypassword  
-a
```

`-x <URL>` Specifies the URL of a folder to index. Example:

```
ExchangeConnector -x https://mysite.com/public/
```

`-u <username>` Name of user (for use when the Exchange Connector user does not have access). The username must include a domain. Example:

```
ExchangeConnector -F d:\FoldersConfig.xml  
-u mydomain\jsmith -p mypassword
```

`-p <password>` Password (for use when the Exchange Connector user does not have access). The password is encrypted before it is stored. Example:

```
ExchangeConnector -F d:\FoldersConfig.xml  
-u mydomain\jsmith -p mypassword
```

Note! The folders configuration file allows you to specify username/password information for each folder. The `-p` option applies only to those folders that do not have a specified username/password.

-D	<p>Deletes all messages from the index that have been deleted on the Exchange server.</p> <hr/> <p>Note! If you run with -D, you must also use the -F option to specify a folders configuration file.</p> <hr/> <p>The Exchange Connector processes all folders in the same session that it does deletions. It cannot process individual folders. Example:</p> <pre>ExchangeConnector -F FoldersConfig.xml -D</pre> <p>You can add content from individual folders with the -x parameter but make sure those folders are also added to the XML file. Otherwise the content will be removed once you run with the -D option.</p> <hr/> <p>Note! Use the -D option with caution!</p> <hr/>
-d <service_portnumber>	<p>Naming service port.</p> <p>Example: -d 16099</p>
-h <hostname>	<p>Host name of the Fast Data Search naming service machine.</p> <p>Example: -h sfo-jsmith-lt</p>
-n	<p>Enables change detection on ACLs and on content.</p> <hr/> <p>Note! This option dramatically slows down incremental indexing, but should be used if any ACL changes have happened to any content that has been previously indexed.</p> <hr/>
-P	<p>Produces an encrypted password that you can insert into the pw fields in the the folders configuration file.</p>
-S <system_id>	<p>System id</p> <p>Example: -S win</p>
-T	<p>Enables verbose logging, including ACL content. Allows you to store all the XML query responses as they are extracted from the Exchange server to the connector folder, so they can be inspected.</p> <hr/> <p>Note! Do not use the -T option in a production environment.</p> <hr/>

-x	Stores all of the documents processed in FastXML format in the ExchangeConnector directory.
----	---

Folders Configuration File

The recommended way of running the Exchange Connector is to use the -F command line option. In other words, you maintain a FoldersConfig.xml file that specifies the folders and mailboxes to index and supplies the necessary credentials. This allows you to extract many folders in a single run.

Note! You can use the FoldersConfig.xml file while also specifying folders on the command line. However, you cannot specify folders on the command line when using the -D (delete messages) option.

The folders configuration file can have any name but must follow the format shown in the example below:

```
<?xml version="1.0"?>
<ExchangeFolders>
  <Folder url="https://mysite.com/public/Technology"/>
  <Folder url="https://mysite.com/exchange/jsmith/inbox"
    user="mydomain\jsmith" pw="xxxxx"/>
  <Folder url="https://mysite.com/public/Technology"
    user="mydomain\jsmith" pw="xxxxx"/>
</ExchangeFolders>
```

You can maintain the FoldersConfig.xml file manually. For private mailboxes, however, you can use the -u, -p, and -a command line options that allow the Exchange Connector to update the FoldersConfig.xml file for you. For example:

```
ExchangeConnector
-x https://bosowa.fast.no/exchange/ragnarb/inbox/AttachmentTest/
-u ad\ragnarb -p paSSw0rd -F FoldersConfig.xml -a
```

The result of this run is an entry added to the FoldersConfig.xml:

```
<Folder
  url="https://bosowa.fast.no/exchange/ragnarb/inbox/AttachmentTest/"
  user="ad\ragnarb" pw="aveljwlron"
  owneruid="aecqaaaaaaakfiaaaalckgbo1zogy3kxsnutswtbeaaa" />
```

The owneruid tag is the user id of the person owning this mailbox. Since only a single user should have access to search his or her own mailbox, and the uid is static, it is stored in the

configuration file to increase performance. This way there is no need to look up permissions for every mail message.

The new entry can then be manually modified to add excluded folders or duplicated for other folders, etc.

Exclude Folders

The folders configuration file allows you to specify folders to exclude from indexing. The specifications are recursive. In other words, excluding a given folder also excludes its subfolders. However, you can specify specific subfolders to index. For example, you can do something like:

```
include folder a
exclude folder a/b
include folder a/b/c
```

Here is a sample folders configuration file that demonstrates how to use exclude folders with Exchange. The connector is quite particular about how it does its check against the list of folders:

- names are case sensitive
- use escaped URLs (ie %20 for space)
- make sure you include the slash “/” character at the end

```
<?xml version="1.0"?>
<ExchangeFolders>

  <Folder url="https://fast.no/public/Technology"
    user="<domain>\<user>" pw="<encrypted pw>"

    <ExcludeFolder url="https://fast.no/public/Technology/Tech%20Email/
Tech%20Nuggets/" />

    <ExcludeFolder url="https://fast.no/public/Technology/Tech%20Email/
News%20Items/" />

    <Folder url="https://fast.no/public/Technology/Tech%20Email/
News%20Items/special/" />

  </Folder>

</ExchangeFolders>
```

Reinitializing the Exchange Connector

If you make changes to the command line options and want to start over again, delete the changes.db file before entering the modified command. The changes.db file contains a record of what was previously indexed. By deleting it you ensure that you will get a 100% reproducible behavior.

Troubleshooting the Exchange Connector

About This Chapter

This chapter describes how to troubleshoot problems encountered in using the Exchange Connector. It includes:

- About Troubleshooting the Exchange Connector
- Inspecting the XML Query Responses and Documents
- Logging
- Fast Data Search Logging
- Common Error Situations

About Troubleshooting the Exchange Connector

Errors are handled via return values and error messages to `STDERR`. The reported error messages are specific, and the most common errors result from invalid or incorrect input parameters to the extraction program. Erroneous configuration files are also likely to be a source of error.

Verifying the Exchange Version

To confirm that you are attempting to connect to a supported version of MS Exchange Server:

- 1 Log in to the Exchange Server console or with Terminal Services.
- 2 Click Programs > Microsoft Exchange > System Manager.

- 3 Expand the "Servers" node.
- 4 Right-click the server name and select "properties" to display the version number and service pack level.

Inspecting the XML Query Responses and Documents

- Use the -T option (super verbose) to store all XML query responses, as they are extracted from the Exchange server, to the connector folder so that they can be inspected.
- Use the -X option to store all of the documents processed in FastXML format in the ExchangeConnector directory.

Logging

The logging policy can be set by editing the following XML file:

```
\Program Files\FAST\ExchangeConnector\LoggerConfig.xml
```

The name of the file is used to distinguish the log source from other sources/modules.

The log files are located in:

```
<MSExchangeConnector_Install_Folder>\ExchangeConnector.log
```

Configuring the Logging Levels

The following logging severity levels are supported:

- VERBOSE
- INFO
- WARNING
- PROGRESS
- ERROR
- CRITICAL
- FATAL
- ALL

Example Logging Configuration File

The following is an example of how you can configure logging.

```
<?xml version="1.0"?>
<!DOCTYPE logConfigs SYSTEM "http://www.fast.no/logConfig1_2.dtd">
<logConfigs>
  <logConfig name="release">
    <stderr>
      <messageType name="ERROR"/>
      <messageType name="CRITICAL"/>
      <messageType name="FATAL"/>
    </stderr>
    <stdout>
      <messageType name="INFO"/>
      <messageType name="PROGRESS"/>
      <messageType name="DEBUG"/>
    </stdout>
    <file name="ExchangeConnector.log" maxBackups="5">
      <messageType name="INFO"/>
      <messageType name="PROGRESS"/>
      <messageType name="DEBUG"/>
      <messageType name="ERROR"/>
      <messageType name="CRITICAL"/>
      <messageType name="FATAL"/>
    </file>
    <!-- email smtpHost="mysmtphost.mydomain.tld" toAdr="me@mydomain.tld"
fromAdr="me@mydomain.tld">
      <messageType name="FATAL"/>
    </email -->
    <!--network host="localhost" port="14770">
      <messageType name="ALL"/>
    </network-->
  </logConfig>
</logConfigs>
```

Fast Data Search Logging

In addition to the Exchange Connector logs, examine the Fast Data Search logs for information.



Common Error Situations

Incorrect File Names for Attachments

From Microsoft (<http://support.microsoft.com/default.aspx?scid=kb;en-us;812256>):

When you right-click an attachment in a message and then click Save as), the suggested file name for the attachment appears similar to the following file name, where filename is the original name of the attached file: 1_multipart_xF8FF_2_filename

This issue disappears if you patch Exchange.

Cannot Query Documents

The document count increments in the **Collection Overview**, but it seems that I cannot query any documents in the index.

From the FAST Data Search administrator interface, select **Matching Engines**, and select the Search Engine that is configured to receive the data. Select **System Logs**. Check for connected feeds. If the last operation performed says 'update xxxxxxxx-Collection-NAME', then select the system logs for the indexer. It will display where the indexing process may be failing.

Callback-Related Errors Occur During a Run

Examine the Fast Data Search logs for information.

Invalid or Missing Certificate

If a user running the Exchange connector (username/password specified) does not have a valid SSL certificate on the Exchange server that the connector tries to connect to, something like the following can be seen in the log:

```
[2004-12-01 22:50:33] ERROR      : ExchangeConnector:
TraverseRemoteFolder(https://bores01.ad.fast.no/public) failed; last
checkpoint: Attempt XMLHTTP.
send [failure indicates server not available]
[2004-12-01 22:50:33] ERROR      : ExchangeConnector: Code = 800c0008
      Code meaning = Unknown error 0x800C0008
The download of the specified resource has failed.
```

The problem can also be seen when trying to open the URL in a web browser. If a certificate is lacking, the attached popup window appears:



Message Size Exceed Limit Error

The `MARSHAL_MessageSizeExceedLimitOnClient` error may indicate that you have individual documents larger than 200 MB. (The maximum message size in FDS's `omniorb.cfg` has been set to 200MB).

If the `omniorb.cfg` file and the `OMNIORB_CONFIG` environment variable have been set correctly (see *Before You Install* on page 5), edit the `omniorb.cfg` file and increase the maximum message size as needed.

Cannot Access Files When Crawling an OWA Site

The following `DEBUG` message does *not* mean there is a problem extracting the document with your existing connector's credentials, but rather indicates that the ACL type for that specific document (email, attachment, etc.) does not allow anonymous access:

```
[2004-12-15 10:30:40] DEBUG    : ExchangeConnector: Deny: NT AUTHORITY\ANONYMOUS LOGON [winaeqaaaaaaaakbyaaaaa]:
```

Appendix A

Index Profile

datasearch-4.0-exchange.xml

```
<?xml version="1.0"?>
<!DOCTYPE index-profile SYSTEM "index-profile-3.1.dtd">
<index-profile name="datasearch">
<field-list>

  <field name="title" sort="yes" fullsort="yes" tokenize="auto">
    <vectorize default="10:0"/>
  </field>
  <field name="body" tokenize="auto" max-result-size="1024"
    fallback-ref="teaser" result="dynamic" index="no">
    <vectorize default="5:5" alternative="{ja,ko,zh,shz,tzh}:5:0"/>
  </field>
  <field name="teaser" index="no"/>
  <field name="headings" tokenize="auto" />
  <field name="description" result="no" />
  <field name="anchortext" result="no" />
  <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" />

  <!-- Security field: Document Access Control List -->
  <field name="docacl" />
</field-list>
</index-profile>
</xml>
```

```
    <!-- Security field: System id for domain which this document belong
to -->
    <!-- Required in result set for post processing -->
    <field name="docaclsystemid" />

<!-- 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=";" />

<!-- Exchange Connector fields -->
<field name="mailfrom" />           <!-- Sender -->
<field name="mailto" />           <!-- Receiver -->
<field name="mailsubject" />     <!-- Subject -->
<field name="mailbody" />       <!-- HTML version of message -->
<field name="mailattachmentcontainer" /> <!-- Message owning this
attachment -->
<field name="mailattachmentcontenttype" /> <!-- MIME type for attachment
-->

</field-list>

<composite-field name="content" rank="yes" default="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>

<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>
```

