FAST SMART CONNECTOR FOR INTERWOVEN TEAMSITE

version 4.0.1

MODULE GUIDE

RELEASE CANDIDATE #1

MAY 24, 2005

Document Number: 927, Document Revision: B, May 24, 2005

Copyright

Trademarks

Restricted Rights Legend

Contents

Last update: Tuesday, May 24, 2005 2:11 pm

Chapter 1

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 Interwoven TeamSite and explains how to use it.

Audience

This guide provides information for all users of the FAST Smart Connector for Interwoven TeamSite.

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.

Modified on: Tuesday, May 24, 2005 2:11 pm

Chapter 1

Introducing the Interwoven Connector

About this Chapter

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

- Overview
- How the Interwoven Connector Works

Overview

Interwoven TeamSite is content management software for the enterprise. The FAST Smart Connector for Interwoven TeamSite is a program that extracts information from an Interwoven TeamSite server and feeds it to FAST Data Search for indexing. Making your Interwoven TeamSite content searchable consists of the following major steps:

1. Installing the Interwoven Connector.
2. Configuring the Interwoven Connector.
3. Configuring FAST Data Search to receive Interwoven TeamSite content. This involves:
   - defining an index profile specifying how to index TeamSite content
   - creating a cluster
   - creating a custom pipeline
   - creating a collection
4. Extracting and submitting the Interwoven TeamSite content to FAST Data Search

Supported Platforms

Operating Systems

- Solaris 2.8 and all required patches

Interwoven

- TeamSite 5.5.2 or 6.0
- OpenDeploy 5.6

FAST Data Search

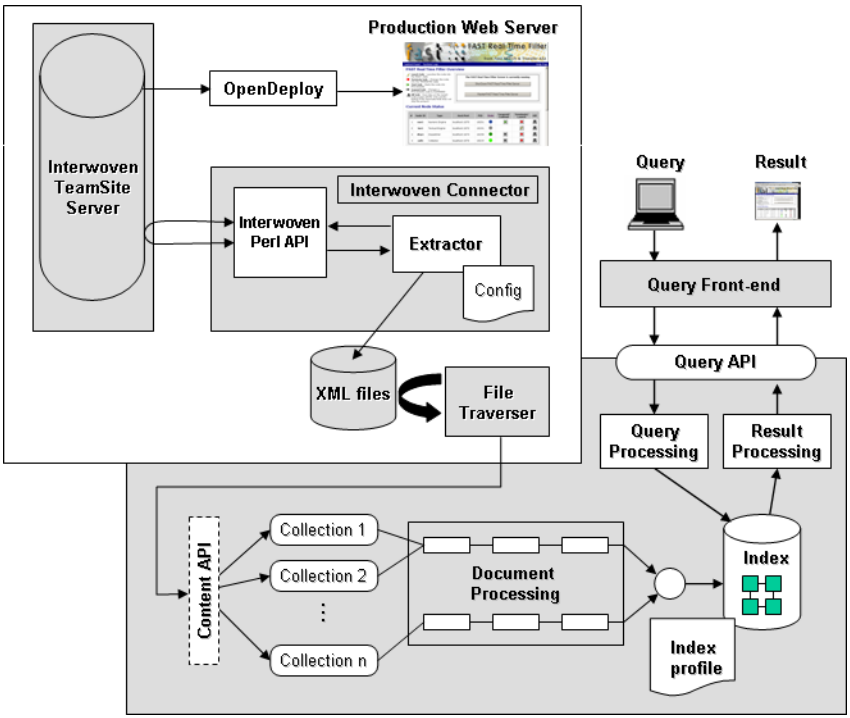
- 4.0 or later

How the Interwoven Connector Works

Documents maintained in TeamSite servers are deployed to production web servers using the OpenDeploy Distribution Server. The Interwoven Connector is placed in the default deployment workflow in TeamSite and is configured to run after the OpenDeploy stage has completed successfully.

The connector:

- uses Interwoven-Perl API calls to fetch information about which files and directories have changed and should be submitted to FDS
- creates FastXML files containing the content of those files
- sends the FastXML files to FDS using file traversal


Production Web Server
OpenDeploy
Interwoven TeamSite Server
Interwoven Connector
Interwoven Perl API
Extractor
Config
XML files
File Traverser
Query
Result
Query Front-end
Query API
Query Processing
Result Processing
Content API
Collection 1
Collection 2
Collection n
Document Processing
Index
Index profile

Modified on: Tuesday, May 24, 2005 2:11 pm

Chapter 2

Installing the Interwoven Connector

About this Chapter

This chapter describes how to install the FAST Smart Connector for Interwoven TeamSite. It includes:

- Before You Install
- Installing the Interwoven Connector on UNIX Platforms
- After You Install

Before You Install

Setting Environment Variables

- The library path variable must include ???.

Platform	Library Path Variable
Solaris	`LD_LIBRARY_PATH`

Message to Reviewer! Is the library path variable really necessary?

Depending on the platform and TeamSite, these variables may or may not be set at client installation time.

- `LM_LICENSE_FILE` must point to a FlexLM license file (see Installing the License Key).

Installing the License Key

- Include the license key for the Interwoven Connector in the FlexLM license file for FAST Data Search. Otherwise the Interwoven Connector asks you for it at run time.

  If the connector is not installed on a FAST Data Search node, you can use a license file that points to a remote license manager:

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

  The keyword ANY means that the license allows the software to run on any machine.

- The environment variable `LM_LICENSE_FILE` must point to the FlexLM license file containing valid FAST Data Search and Interwoven Connector licenses.

Verifying Sufficient Disk Space

The machine on which the Interwoven Connector runs must have enough temporary disk space to store documents to be sent to FAST Data Search. Enough disk space must be avaliable to store all the files being submitted temporarily. The files are encoded with base64, so temporary files can be up to 35% larger than the original content (+ metadata).

Installing the Interwoven Connector on UNIX Platforms

To unpack and install the Interwoven Connector on Unix platforms, perform the following procedure:

1. Create a FAST directory under the TeamSite installation directory and change to it.

   ```
   mkdir <TeamSite_installdir>\FAST
   cd <TeamSite_installdir>\FAST
   ```

2. Copy the file interwovenconnector-<*platform*>-<*version*>*.tgz* (or *.tar.gz*) from the installation disk to the current directory on the hard drive.

   <*platform*> identifies the type of platform on which you want to install the Connector.

   <*version*> identifies the version of the Interwoven Connector to be installed.

3. Uncompress the installation file:

   ```
   gzip -d interwovenconnector-<platform>-<version>.tgz
   ```

4. Extract the uncompressed *tar* file. The extraction creates the directory interwovenconnector.

   ```
   tar -xvf interwovenconnector-<platform>-<version>.tar
   ```

After You Install

When the installation is complete, the Interwoven Connector directory contains the following subdirectories.

Directory	Files and Descriptions
filetraverser	`filetraverser-<platform>-<version>.tgz` FDS 4.0 standalone filetraverser.
filetraverser/bin	`filetraverser` `filetraverser.sh`
filetraverser/doc	Documentation for the Interwoven Connector.

filetraverser/etc	Example configuration files.
	`fastsearch.lic`
	`omniorb.cfg`

filetraverser/lib	Libraries required by the File Traverser.
	`libcrypto.so.0.9.6`
	`libcrypto.so.0.9.7`
	`libexpat.so.0`
	`libgigabase_r.so.2`
	`libomniORB4.so.0`
	`libomnithread.so.3`
	`libssl.so.0.9.6`
	`libssl.so.0.9.7`

filetraverser/lib/python2.2	`_codecs.so`
	`_omnipymodule.so`
	`_socketmodule.so`
	`_weakref.so`
	`_xmlrpc.so`
	`binascii.so`
	`math.so`
	`md5.so`
	`pyexpat.so`
	`pygbhash.so`
	`select.so`
	`struct.so`
	`time.so`

workflow	`wft_fastdeploy.ipl`
	InterWoven-perl script that extracts data and generates data suitable for feeding to FDS. Is run as part of the submit workflow
	`configurable_default_submit_fast.cfg`
	Example configuration file for the submit workflow
	`configurable_default_submit_fast.wft`
	Example workflow file for the submit workflow

Modified on: Tuesday, May 24, 2005 2:11 pm

Chapter 3

Configuring the Interwoven Connector

About This Chapter

The Interwoven Connector is configured manually using configuration files as described in this chapter.

This chapter includes:

- About Configuring the Interwoven Connector
- Configuring the File Traverser
- Configuring the Workflow Templates
- Configuring the Workflow Definition

Message to Reviewer!	I have not performed any of the procedures in this chapter. It seems that some of these steps configure the connector and some of them configure TeamSite itself. I need access to a TeamSite server to verify that these are complete and correct.

About Configuring the Interwoven Connector

Some of the steps needed to configure the Interwoven Connector require modifying files provided with the connector itself. Other steps require modifying your TeamSite configuration.

Configuring the File Traverser

Modify the File Traverser files as described below. These files are in:

<*TeamSite_installdir*>\FAST\filetraverser\

1 Edit **etc\fastsearch.lic** and change **test603.oslo.fast.no** to your FAST Data Search installation:

```
SERVER test603.oslo.fast.no ANY
VENDOR FASTSRCH
USE_SERVER
```

Message to Reviewer! This doesn't seem to comply with the standard usage of FlexLM, where you can combine license keys for many different products in one license file.

2 Edit **etc/omniorb.cfg** and change the following line to point to the name service for your FAST Data Search installation.

```
InitRef = NameService=corbaname::ting.grenland.fast.no:16099
```

Message to Reviewer! Is the port always 16099?

3 Edit **bin\filetraverser.sh** to send to the correct content distributor node and port.

Message to Reviewer! Where do you specify the port?

```
#!/bin/sh
FASTSEARCH=/usr/fast/iw-home/connect/filetraverser
LD_LIBRARY_PATH=$FASTSEARCH/lib
PATH=$FASTSEARCH/bin:$PATH
export FASTSEARCH LD_LIBRARY_PATH PATH
cd /usr/fast/iw-home/connect/filetraverser/etc
filetraverser -c test -r . -x :document:filedir
```

4 Replace **test** with your collection name. See the File Traverser chapter of the *Configuration Guide* for more info on filetraverser arguments.

Configuring the Workflow Templates

To make TeamSite use the configurable workflow instead of the default:

1 Change <*TeamSite_installdir*>\local\config\wft\avaliable_templates.cfg to read:

```
<template_file name='Submit Workflow' path='solutions/
configurable_default_submit.wft'>
    <command_list>
      <command value='submit' />
      <command value='all' include='no' />
    </command_list>
    <role_list>
      <role value='author' include='no' allusers='yes'/>
      <role value='all' include='yes' allusers='yes'/>
    </role_list>
  </template_file>
```

Configuring the Workflow Definition

For these changes, refer to the files in <*TeamSite_installdir*>\FAST>.

1 Copy wft_fastdeploy.ipl to <*TeamSite_installdir*>\local\config\wft\solutions

2 Make the following changes (if they are not already there) to <*TeamSite_installdir*>\local\config\wft\solutions\configurable_default_submit.cfg

   a set deploy_task=yes

   b set ask_deploy_task=no

3 Make the following changes to <*TeamSite_installdir*>\local\config\wft\solutions\configurable_default_submit.wft

   a Just after:

   ```
   my $deploy_ipl = "$this_wft_dir/wft_opendeploy.ipl";
   ```

   add:

   ```
   # The path to the FAST deployment script.
   my $fastdeploy_ipl = "$this_wft_dir/wft_fastdeploy.ipl";
   ```

   b In the link tasks section, change

```
link_tasks('deploy', 'end', "Success");
```

to

```
link_tasks('deploy', 'fastdeploy', "Success");
```

and add

```
link_tasks('fastdeploy', 'end', "Success");
```

This makes sure that the fastdeploy task is run after the deploy task, but only if it succeds.

4 Add the following to the task definition list just after the definition of `deploy`:

```
fastdeploy => {
    name        => $localizer->format('fastdeploy_task.name'),
   description => $localizer->format('fastdeploy_task.description'),
    type        => "externaltask",
    owner       => $iw_areaowner,
    areavpath   => $iw_workarea,
    command     => "$iwperl $fastdeploy_ipl",
},
```

Modified on: Tuesday, May 24, 2005 2:11 pm

Chapter 4

Configuring FAST Data Search

About This Chapter

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

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

About Configuring FAST Data Search

In order to use the Interwoven Connector, you must configure a FAST Data Search cluster to function as a dedicated 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 TeamSite content instead of crawled web content.
- In a multiple node system, create a new cluster to index TeamSite 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 Interwoven 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*.

Note! The Interwoven Connector provides a sample index profile, as shown in Appendix 7 *Sample Index Profile*. The file is provided in the etc folder so that you can copy and customize it.

Creating an Index Profile

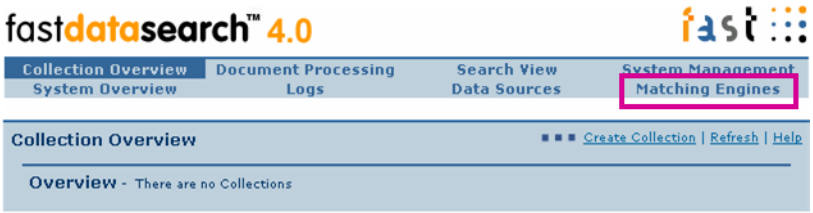
An index profile for the Interwoven Connector *must have the same fields as the* TeamSite *data being imported*. When you customize the default index profile, be sure to stay compatible with the FAST Data Search index profile rules:

- To make the names unique, start the field prefix name with: **db**
- Use only alphanumeric characters
- Transform characters to lowercase field mapping
- The parameter max-index-size specifies the maximum number of kilobytes that can be indexed for a document. The default value is “1024" (KB), which means that documents larger than that size will be truncated. To allow larger documents to be processed without truncation, specify a positive integer up to two gigabytes. See the FAST Data Search *Configuration Guide* for details.

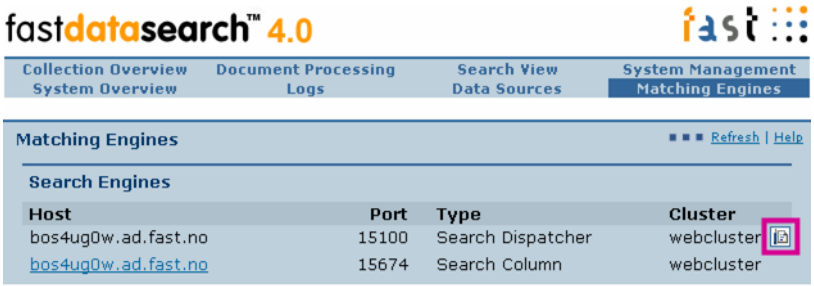
Uploading an Index Profile

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

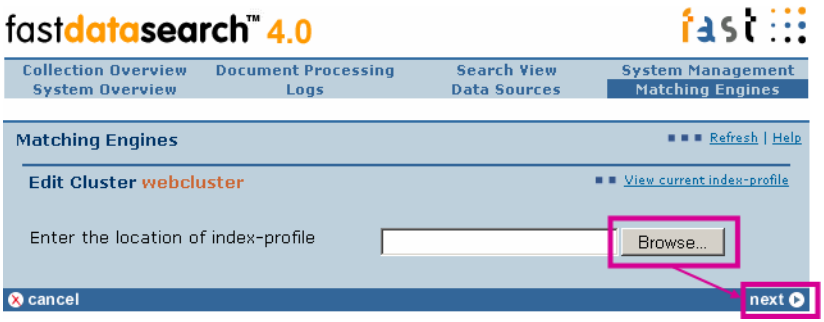
1. Open the FAST Data Search **Admin interface**.
2. Click **Matching Engines**.



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

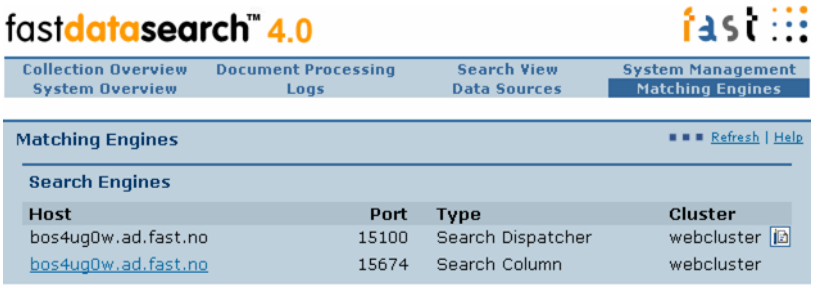


4 In the **Edit Cluster** screen, click the **Browse** button.



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

6 The admin interface returns to the Matching Engines dialog.



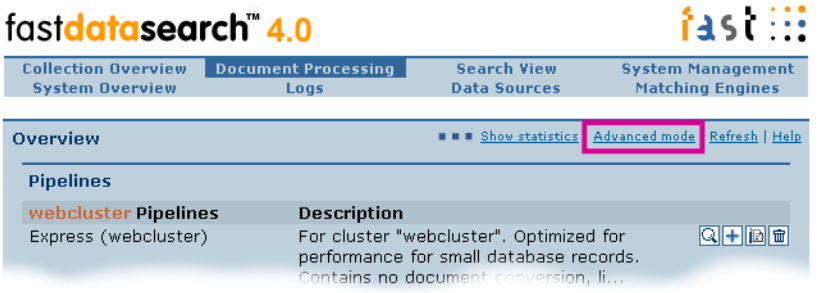
Creating a Custom Pipeline

Create a new pipeline using the generic pipeline as a base. There are two major steps involved in this task:

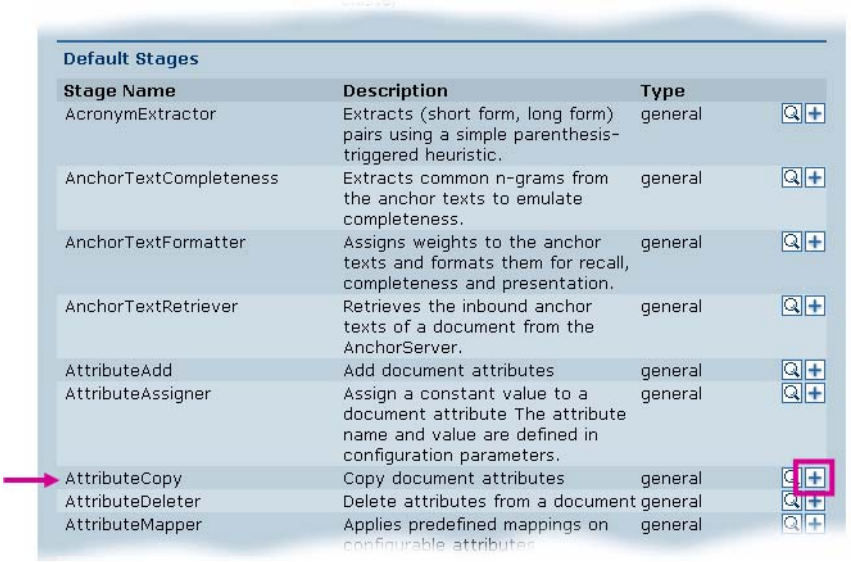
- Creating Custom Stages
- Adding Custom Stages to the Pipeline

Creating Custom Stages

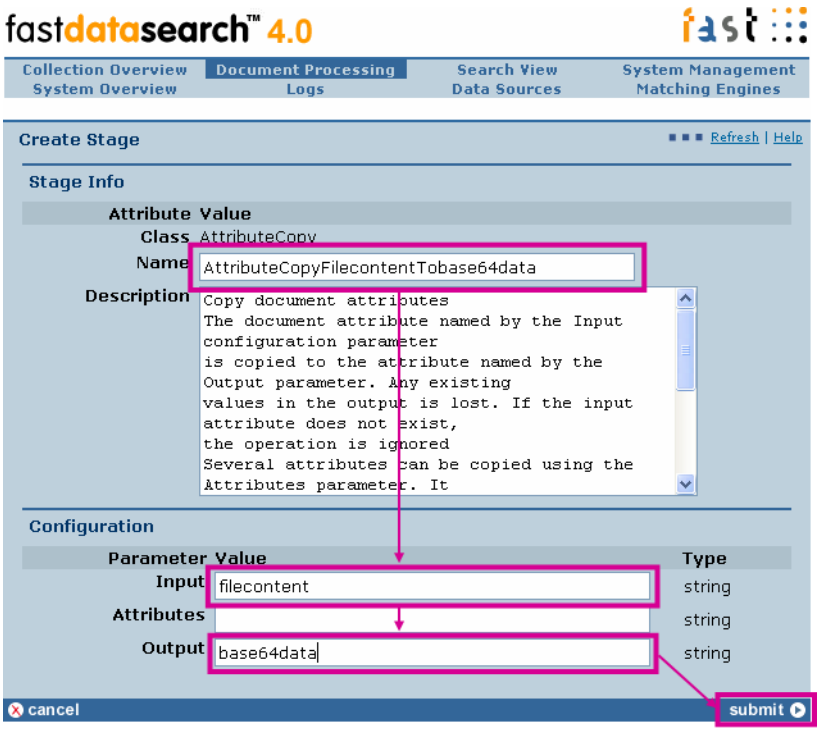
1. Open the FAST Data Search **Admin interface**.
2. Go into **Document Processing** and click **Advanced Mode**



3 Under **Default Stages**, find **AttributeCopy** and **click the plus sign (+)**.



4 In the **Create Stage** dialog, set the **Name** field to **AttributeCopyFilecontentTobase64data**.



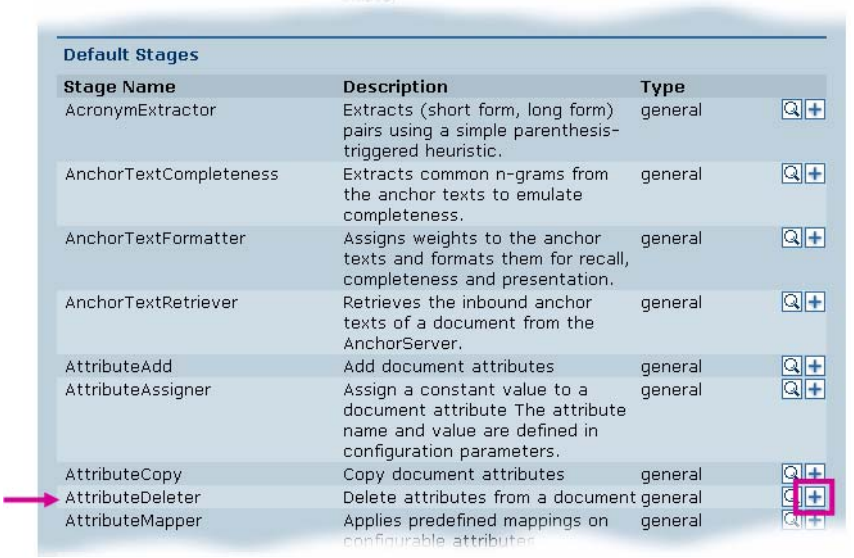
5 Set the **Input** field to **filecontent**.

6 Leave the **Attributes** field empty.

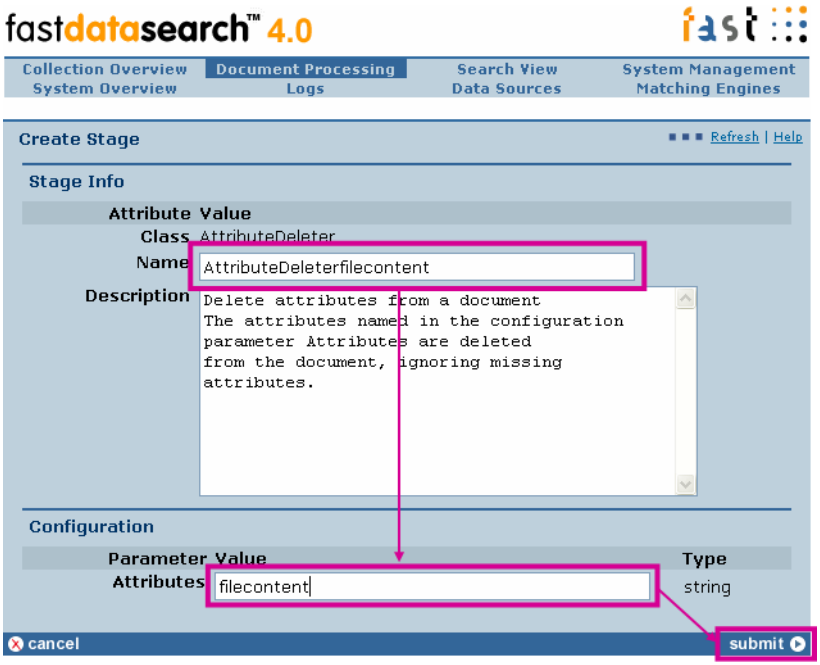
7 Set the **Output** field to **base64data**.

8 Click **submit**.

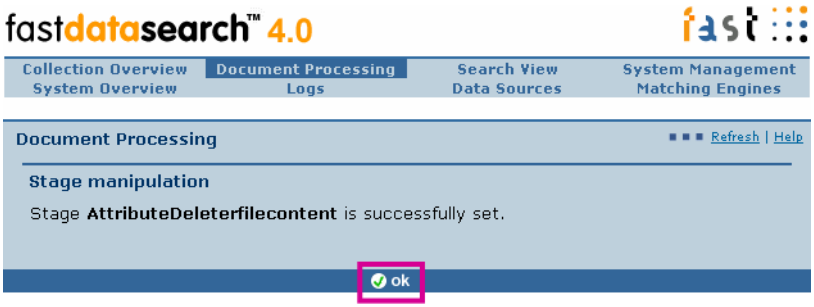
9 Under **Default Stages**, find **AttributeDeleter** and **click the plus sign (+)**.



10 In the **Create Stage** dialog, set the **Name** field to **AttributeDeleterfilecontent**.

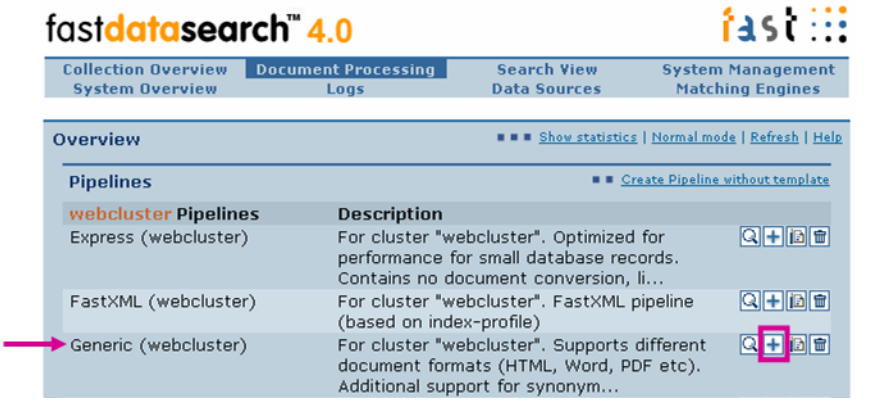


11 Click **submit**.

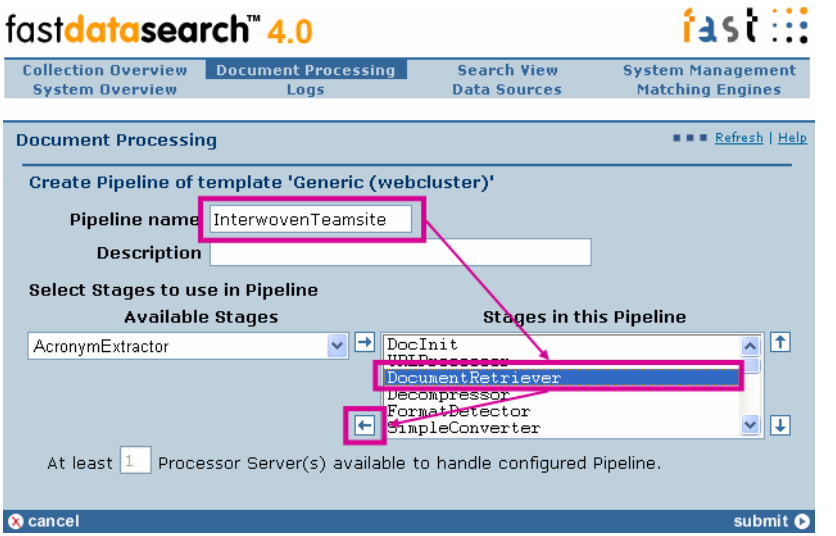


Adding Custom Stages to the Pipeline

1 In the **Document Processing** dialog, find the **Generic(webcluster)** pipeline and **click the plus sign (+)**.



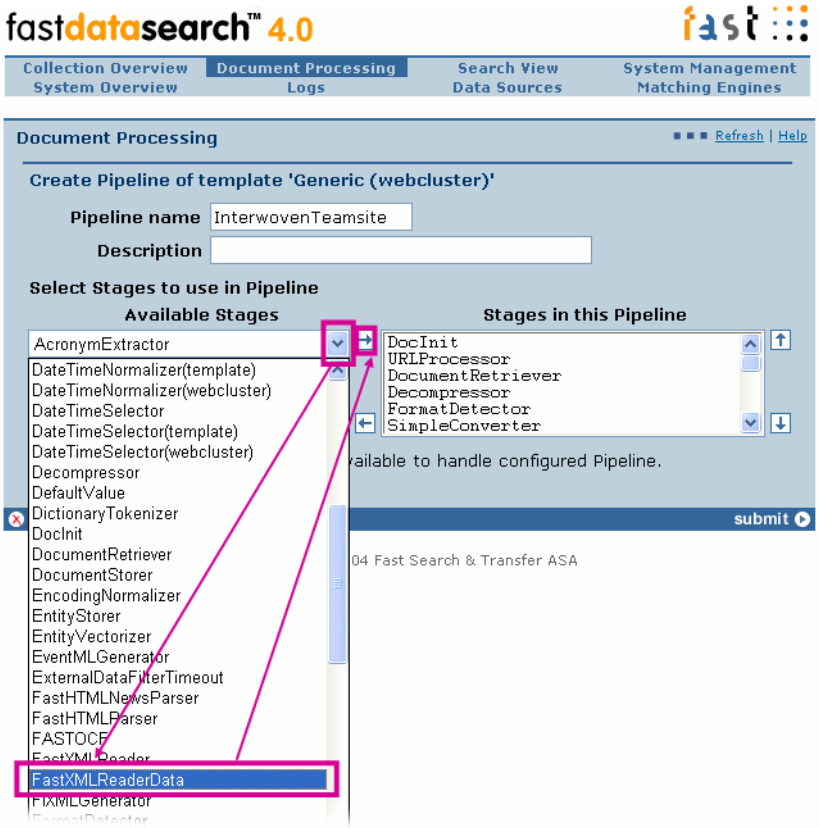
2 In the **Create Pipeline** dialog, set the **Name** field to **InterwovenTeamsite**.



3 In the **Stages in this Pipeline** list, click the **left-arrow** to remove:

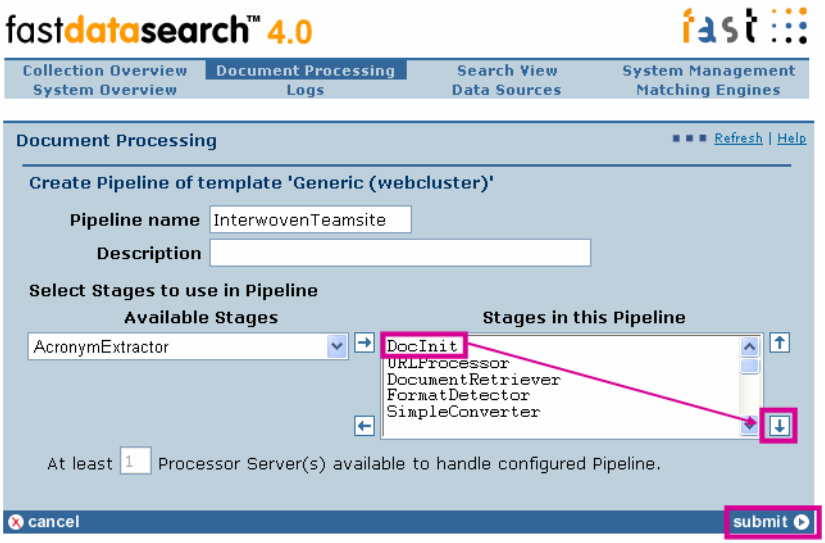
```
DocumentRetriever
Decompresser
FastHTMLParser
DateTimeNormalizer(webcluster)
DateTimeSelector(webcluster)
MapperTransformer
RankTuner
```

4 In the **Available Stages** dropdown list, click the **right-arrow** to add:




```
FASTXMLReaderData
AttributeCopyFilecontentTobase64data
AttributeDeleterfilecontent
Base64Decoder
HTMLParser
```

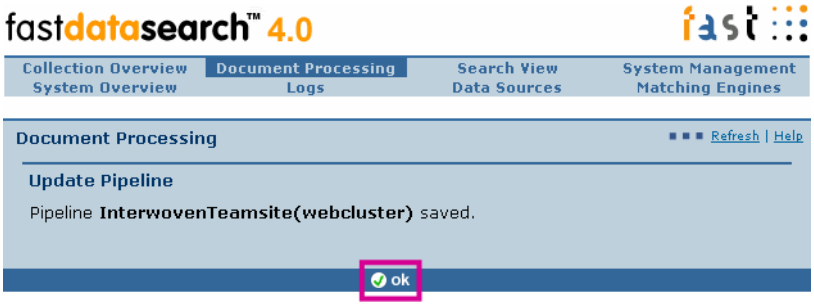
5 In the **Stages In this Pipeline list**, click the **up-arrow** and **down-arrow** to arrange the stages into the following order:




```
FastXMLReaderData
AttributeCopyFilecontentTobase64data
AttributeDeleterfilecontent
URLProcessor
DocInit
Base64Decoder
FormatDetector
SimpleConverter
SearchMLConverter
LanguageAndEncodingDetector
EncodingNormalizer
HTMLParser
TeaserGenerator
Tokenizer(webcluster)
Lemmatizer(webcluster)
Vectorizer(webcluster)
Boost
ATWImport
FIXMLGenerator
RTSOutput
```

6 Click **submit**.

7 Click **ok**.



Creating a Collection for Extracted Data

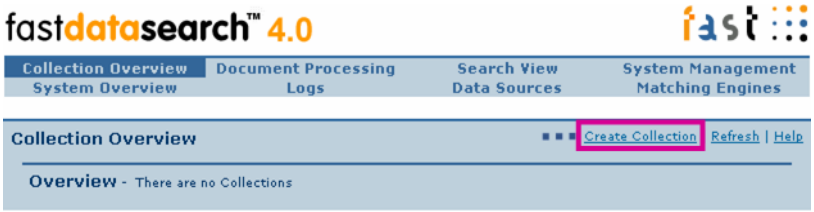
A collection is a logical group of documents. Collections are set up in order to group documents based on selected criteria such as semantics (for example, similar types of documents) and/or document processing (for example, through pipeline configuration).

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.

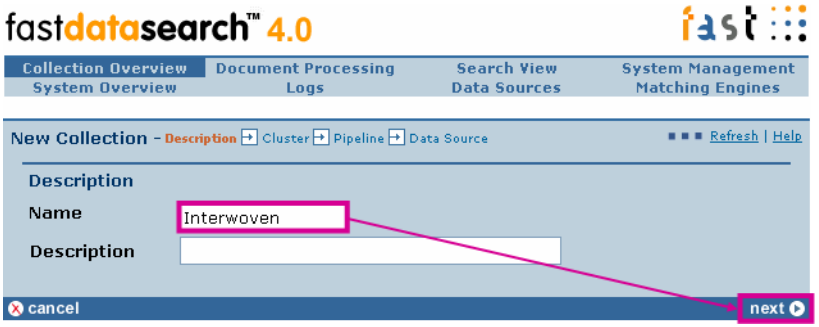
1 Open the FAST Data Search **Admin interface**.

2 If necessary, click **Collection Overview** on the navigation bar.

3 Click **Create Collection**.



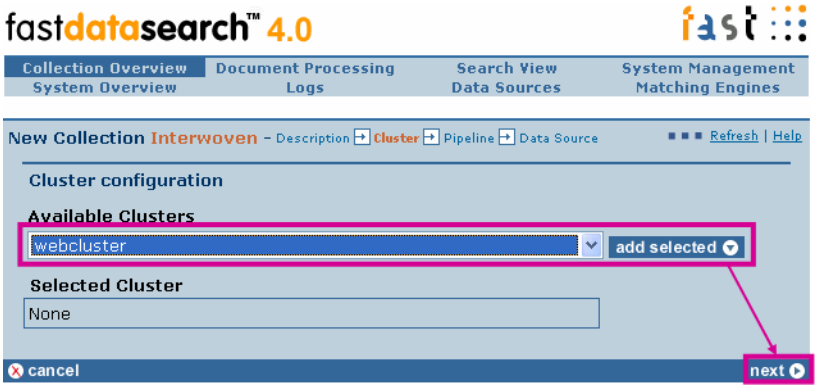
4 In the **Name** field enter **Interwoven**.



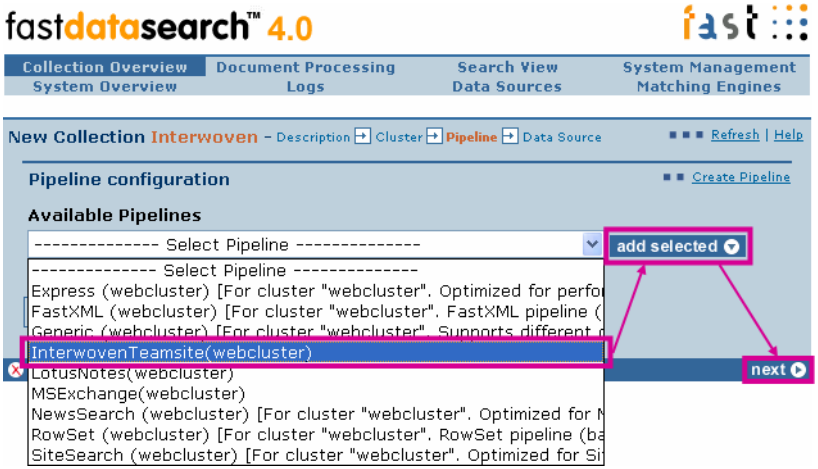
5 Optionally fill in the **Description** field.

6 Click **next**.

7 In the **Cluster Configuration** screen, click the **Available Clusters** dropdown list button and select **webcluster**.



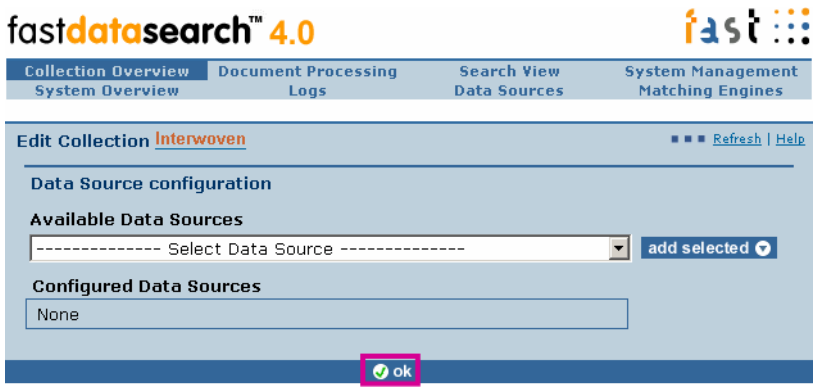
8 In the **Pipeline Configuration** screen, click the **Available Pipelines** dropdown list button and select **InterwovenTeamsite**.



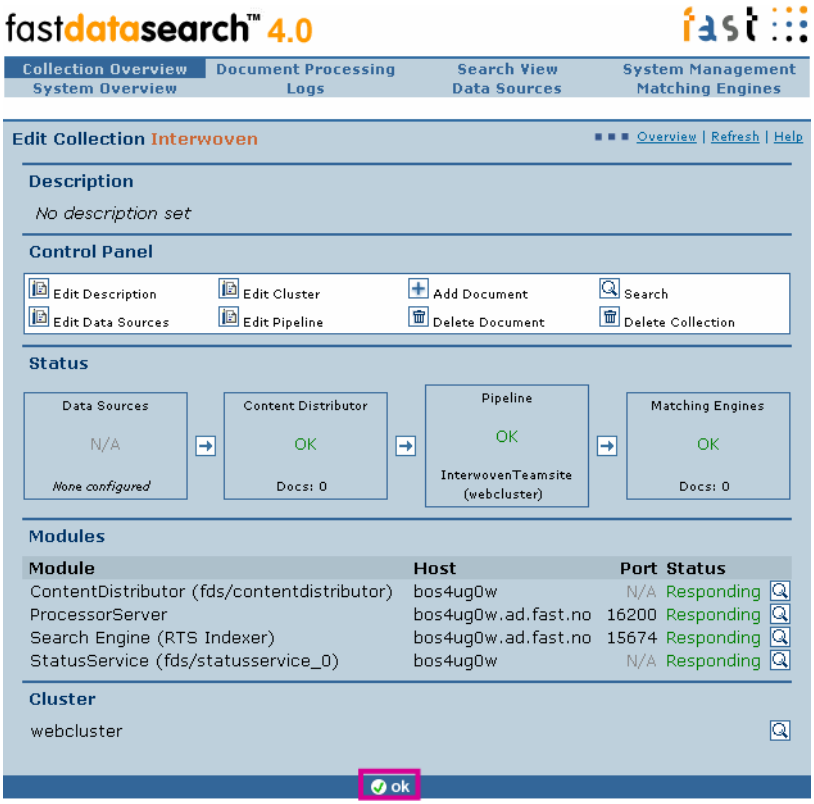
9 Click **add selected**.

10 Click **next**.

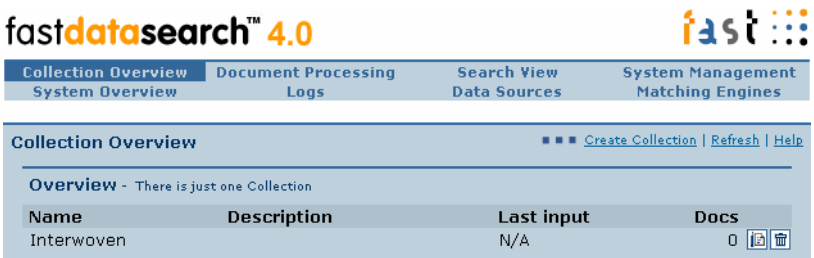
11 In the **Data Source Configuration** screen, click **ok**.



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



13 The admin interface returns to the Collection Overview dialog.



Congratulations. You have successfully configured FAST Data Search for the Interwoven Connector.

Modified on: Tuesday, May 24, 2005 2:11 pm

Chapter 5

Using the Interwoven Connector

About this Chapter

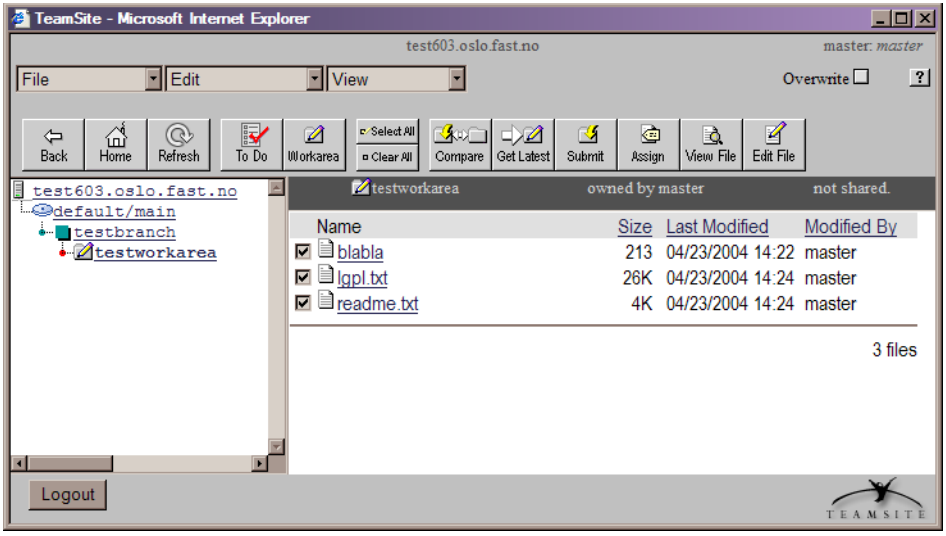
This chapter describes how to use the FAST Smart Connector for Interwoven TeamSite. It includes:

- Running the Interwoven Connector

Running the Interwoven Connector

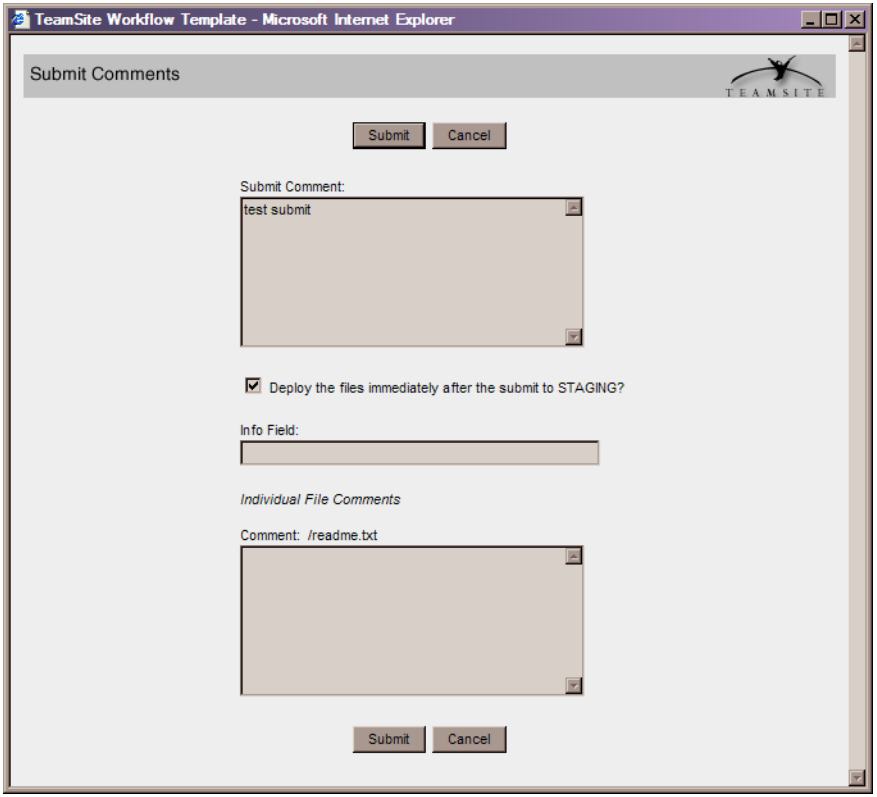
To run the Interwoven Connector:

1 Select the files to be indexed in your TeamSite workarea.



2 Click **Submit**.

3 A new window appears where you can enter comments on the deployment and on each file:



4 Click **Submit**.

5 Make sure that the documents reach FAST Data Search and are properly indexed. A search result should look like this:

rank	2117
docid	f89cee10506a269ab13e0ec43295d31f-test1
collection	test1
title	FAST SEARCH & TRANSFER INC
body	. © Fast Search & Transfer ASA. Commercial in confidence Page FAST Data Search 3.1 InterWoven Connector Introduction This...InterWoven/TeamSite connector that has been developed for FAST Data Search. The connector is based on an export facility...
teaser	. © Fast Search & Transfer ASA. Commercial in confidence Page FAST Data Search 3.1 InterWoven Connector Introduction This document describes the InterWoven/TeamSite connector that has been developed for FAST Data Search.

headings	foobar
contenttype	htmlexport/1146
language	en
charset	utf-8
urls	/default/main/Test1/WORKAREA/TestWA1/bla/InterWovenConnector.doc
url	/default/main/Test1/WORKAREA/TestWA1/bla/InterWovenConnector.doc
filedir	/default/main/Test1/WORKAREA/TestWA1/bla/InterWovenConnector.doc
name	InterWovenConnector.doc
areacid	0x00002100
nfsmode	0x000001b4
isxmlfile	FALSE
group	other
cid	0x00001008
lastuser	root
crdate	Tue Jun 29 08:01:19 2004
mutated	FALSE
compareflags	0x00000000
area	0x0000210000000000000021c
isdir	FALSE
revisionbranch	0x00002250000000000000200
archive	0x00002020000000000000064
prevversion	0x00000000000000000000000
objid	0x0000100840e15a0f40001007
istagged	FALSE
owner	root
flags	0x00000000
branch	0x00002250000000000000200
creator	root
currdirectory	0x0000100840e15a0640001006
submitevent	0x000022580000000000001bf2
fsmoddate	Tue Jun 29 08:01:26 2004
revisionnumber	1
attrmoddate	Tue Jun 29 08:01:26 2004
permissions	rw-rw-r--

moddate	Tue Jun 29 08:01:26 2004
filecontent	foobar
modified	FALSE
size	806400
generic1	foobar
generic2	foobar
generic3	foobar
igeneric1	foobar
docvector	[xsl, 1][self, 0.408451][meta data, 0.375587][data search, 0.328638][meta-data name, 0.258216][document, 0.239437][data value, 0.234742][metadata, 0.225352][template match, 0.211268][add, 0.197183][nasdaqs worst, 0.187793][value-of select, 0.187793][worst week, 0.187793][fast data, 0.187793][body content, 0.164319]

Modified on: Tuesday, May 24, 2005 2:11 pm

Chapter 6

Troubleshooting the Interwoven Connector

About This Chapter

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

- Logging

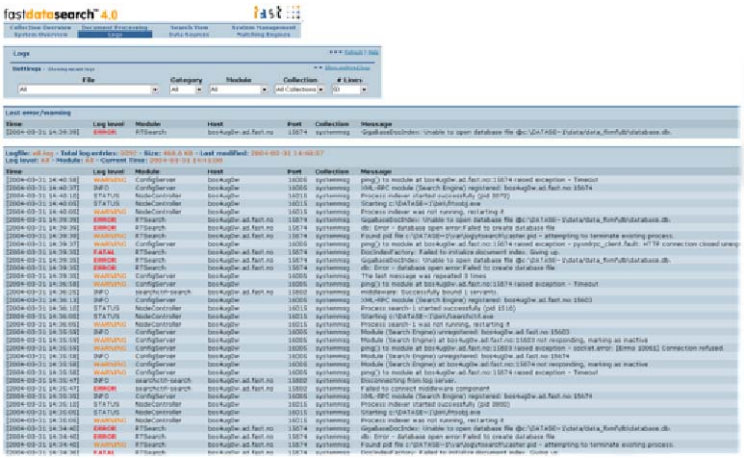
Logging

Log messages from the deploy script appear in **/var/adm/iwtrace.log**. For example:

```
[START 2004-04-26 11:09:14 /ld/work/morten/tmp/iw-home/local/config/
wft/solutions/wft_fastdeploy.ipl]
send_stuff_to_fds Task Information
  Task ID: 850
  Name: Deploy
  Description: Deploy Files
  Command: /ld/work/morten/tmp/iw-home/iw-perl/bin/iwperl
    /ld/work/morten/tmp/iw-home/local/config/wft/solutions/
wft_fastdeploy.i
    pl
  Area: /default/main/testbranch/WORKAREA/testworkarea
deleting /ld/work/morten/tmp/iw-home/FAST/xml/deleted.txt
deleting /ld/work/morten/tmp/iw-home/FAST/xml/iout_00000000.xml
  xml_filename = iout_00000000.xml
iwmnt_dir     = /iwmnt
file_location = /default/main/testbranch/WORKAREA/testworkarea/
readme.txt
fullpath      = /iwmnt/default/main/testbranch/WORKAREA/testworkarea/
readme.txt
xml_location  = /ld/work/morten/tmp/iw-home/FAST/xml
appending to /ld/work/morten/tmp/iw-home/FAST/xml/iout_00000000.xml
```

```
got 4838 bytes from /iwmnt/default/main/testbranch/WORKAREA/
testworkarea/readme.txt
starting run_filetraverser
filetraverser said '[2004-04-26 11:09:15] VERBOSE   :
Filetraverser@test603.oslo.fast.no: test: Starting filetraverser
[2004-04-26 11:09:15] VERBOSE   : Filetraverser@test603.oslo.fast.no:
test: Sending batch of 1 documents (original data size = 9.00 kB)
'
[END 2004-04-26 11:09:15 /ld/work/morten/tmp/iw-home/local/config/wft/
solutions/wft_fastdeploy.ipl]
```

Also examine the Fast Data Search logs for information.



Modified on: Tuesday, May 24, 2005 2:11 pm

Appendix 7

Sample Index Profile

teamsite40.xml

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

  <field name="title" 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,szh,tzh}:5:0"/>
  </field>
  <field name="teaser" index="no"/>
  <field name="headings" tokenize="auto" />
  <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" />
  <field name="filedir"/>
  <field name="modified"/>
  <field name="name"/>
  <field name="areacid"/>
```

```
    <field name="nfsmode"/>
    <field name="isxmlfile"/>
    <field name="group"/>
    <field name="cid"/>
    <field name="lastuser"/>
    <field name="crdate"/>
    <field name="mutated"/>
    <field name="compareflags"/>
    <field name="area"/>
    <field name="isdir"/>
    <field name="revisionbranch"/>
    <field name="archive"/>
    <field name="prevversion"/>
    <field name="objid"/>
    <field name="istagged"/>
    <field name="owner"/>
    <field name="flags"/>
    <field name="branch"/>
    <field name="creator"/>
    <field name="currdirectory"/>
    <field name="submitevent"/>
    <field name="fsmoddate"/>
    <field name="revisionnumber"/>
    <field name="attrmoddate"/>
    <field name="permissions"/>
    <field name="moddate"/>




  <!--  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=";" />



</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 %.0f" />
      <range-label type="middle" format="Between %.0f and %.0f" />
      <range-label type="last" format="More than %.0f" />

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

```
    <numeric-navigator name="docdatetimenavigator"
                       display="Document Time"
                       unit="Date"
                       intervals="4"
                       resolution="86400">
      <field-ref name="docdatetime"/>

      <range-label type="first" format="Before %.10s" />
      <range-label type="middle" format="Between %.10s and %.10s" />
      <range-label type="last" format="%.10s 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>
```