Installation Guide Document
Product Name: Freestyle Search Engine
Version 3.2

Submitted to: NCI Center for Bioinformatics

Prepared by:

Larry Hebel
101 West Renner Road, Suite 130
Richardson, TX 75082
December 21, 2006

Contact Information:
E-mail   sentinel@scenpro.com
Telephone   972/437-5001
FAX   972/437-3611

Revision History

Author(s)/Responsible Party	Rev No.	Date	Page(s)
Larry Hebel	Original	07/06/06	All

Table of Contents

1 Introduction

This document details the Installation Steps and Site Configuration Options for the caDSR Freestyle Search Engine Version 3.2.

The Freestyle Search project containing source and related documentation can be accessed at http://gforge.nci.nih.gov/projects/freestylesearch. All caCORE companion projects can be referenced from http://gforge.nci.nih.gov.

Dependencies

1. These steps rely on the installation of the standard caCORE 3.2 Technology Stack for runtime and builds. **If the environment has not been previously used for one of the other caDSR Tools (e.g. CDE Curation, Browser, etc.), please refer to Section 5, Required Technology, for special notes and technology URL references.**

2. The caDSR database is installed and verified using Oracle SQL Plus or other appropriate database query tool.

3. The Oracle JAR should be copied from the Oracle Client installation to the JBoss 4.0.2 installation. The Freestyle Search accesses the caDSR database from the web server using a 'thin' connection which does not require the installation of Oracle on the JBoss environment, only the Oracle JAR. The file, ojdbc14.jar, for Oracle 9 is included in the open source package.

4. The Oracle Application Server can **not** be executed on the same machine as JBoss.

Document Definitions and Syntax

1. Download Package
    The zip file available from the NCICB GForge project site noted above.

2. Folder Path Separator ('/')
    The folder path separator on Windows is the character '\' and on Unix is the character '/'. For simplicity all folder (directory) paths in this document use the '/' character. Change as needed for your environment.

3. Folder name 'freestylesearch'
    All examples represent the root for the extracted package as 'freestylesearch'.

4. Folder name 'JBoss'
    All examples represent the root for the installed JBoss 4.0.2 web server as 'JBoss'.

5. Use of quotes (" ")
    The appearance of quotes is restricted to examples in which the quote should be used as a typed key.

6. Use of apostrophes (' ')
   Apostrophes are restricted to clarification within this document and do not appear in examples or any data used for the installation of the software.

7. References to freestylesearch/conf/prod and build-prod
   For simplicity in the documentation all instructions assume a deployment to the Production environment, however the freestylesearch/conf folder contains several folders specific to different environments, i.e. DEV, QA, Stage and Production. If a deployment is being prepared to any environment other than Production, please substitute "freestylesearch/conf/prod" with the appropriate folder matching the deployment, e.g. "freestylesearch/conf/stage" when deploying to Stage.

8. Case sensitivity
   Unless otherwise stated, all parameters, attributes, values, XML elements and examples are case sensitive.

2 Installation Steps

Should any errors occur during the completion of these steps, refer to Section 6 Troubleshooting, for assistance.

1. Extract the complete content of the **download package zip** file, preserving the folder names.

2. Modify 'freestylesearch/conf/prod/**seed.xml**'.

   This file is required by the Freestyle Index Table Seed process. To provide acceptable performance the Freestyle Search Engine requires several tables collectively called the Index Tables. This provides information about the data within the caDSR. These tables are kept current by running the Seed process. The Seed process script is described in more detail below. The Search Engine supports a configuration to host the Index Tables within the caDSR or within another database. The elements within the seed.xml are:

   'index.DSurl', this is the database alias, e.g. CBDEV, which hosts the Index Tables. If the Oracle thick client is required use the referenced defined in the TNSNAMES.ORA file. If the Oracle thin client is used enumerate the server, port and id in the standard form using colons, e.g. <server>:<port>:<SID>.

   'index.DSusername', this is the account to access the Index Table host database. It must have write privileges.

   'index.DSpassword', this is the password required by the index database account, above.

   'index.DSschema', this is the schema name within the index database which owns the Index Tables. The default is "sbrext" and should not be changed without notifying all users of the Search Engine API (freestylesearch.jar).

   'data.DSurl', this is the database alias for the caDSR. If the Index Tables are hosted by the caDSR use the same format as specified for index.DSurl above.

   'data.DSusername', this is the account to access the caDSR database. It must have read privileges to all table views.

   'data.DSpassword', this is the password required by the caDSR database account, above.

3. Modify 'freestylesearch/conf/web/**web.xml**'.

   The Search Engine is primarily an API, however a small Browser example is included and may be used to test the installation of the Index Tables and confirm the results obtained via the API. Currently the Browser interface requires the Index Tables be hosted by the caDSR.

   Only the 'init-param' values need to be reviewed and changed.

'jbossDataSource', this is the '<jndi-name>' specified in the oracle-ds.xml for JBoss. This is the preferred and recommended specification for connectivity to the caDSR database. An example of the '<datasources>' element to be placed in the oracle-ds.xml can be found in the 'freestylesearch/conf/common/**oracle-ds-example.xml**' file. The location of the oracle-ds.xml is defined by JBoss and is normally in the 'jboss/server/default/conf' folder.

'DSusername', this is the account for access to the caDSR, it must have read privileges to all table views.

'DSpassword', this is the password required by the database account, above.

4. Execute '**ant build-prod**' to build the WAR and JAR files.

    Upon completion a WAR and a JAR file are placed in the 'freestylesearch/distrib/war' folder.

5. Execute the freestylesearch/conf/prod/**load_options.sql** in your SQL IDE.

    This SQL script should not be edited. It sets a runtime value used by the Seed process to track the last successful completion of the process. Initially the Index Tables are empty and the date noted in the script causes a full parsing of the caDSR database and load of the Index Tables, approximately 2.5 hours. Refer to the Seed process script below for more details.

    This script can be run from Eclipse, SQL Plus, Toad or any SQL IDE connected to the desired caDSR database instance.

6. Execute '**ant build-doc**' to create the JavaDoc.

    The resulting temporary files are placed in the 'freestylesearch/tmp/freestylesearch/doc' folder. A ZIP version is also placed in the 'freestylesearch/distrib/doc' folder. Complete documentation for the API is contained in the JavaDoc.

7. Move the **freestyle.war** from 'freestylesearch/distrib/war' to the 'JBoss/server/default/deploy' folder.

8. Modify the **oracle-ds.xml** file

    As noted above this file contains the preferred caDSR connection configuration. An example of the '<datasources>' element can be found in 'freestylesearch/conf/common/oracle-ds-example.xml'. This is not a fully specified oracle-ds.xml file only the element needed to define the caDSR as a datasource. Refer to JBoss documentation for the full specification and location of the file if one does not already exist.

9. Modify the **log4j.xml** file for JBoss

Example content can be found in 'freestylesearch/conf/common/**jboss_log4j.xml**'. Review all parameter values for correctness.

10. Restart **JBoss** if necessary.

   Depending on the installation JBoss may be configured to automatically deploy changes and new applications. Check the JBoss server.log to verify the deployment of the freestyle.war file. Also check the Freestyle log as specified in the log4j.xml file to verify correct specification and use of the oracle-ds.xml and web.xml settings.

11. Open a browser window and enter **http://<server name>/freestylesearch/do/search** in the address.

   Replace <server name> with the appropriate reference to the server and port on which JBoss is listening. This should open the example Browser interface to Freestyle and confirms JBoss has deployed the application. At this point the Index Tables are still empty. Entering any search terms will result in nothing found. The Freestyle interface may be used even while the Seed process is loading or updating the Index Tables. However, until the process is complete, the results may not accurately reflect the content within the caDSR.

12. Create and execute a command script to run the **Seed** process.

   Copy the following lines into the file seed.sh and mark the script executable (chmod +x seed.sh).

```
#!/bin/bash

DATE=`date +%Y%m%d`
JAVA_HOME=/usr/jdk1.5.0_06
BASE_DIR=/local/content/freestyle/bin

export JAVA_HOME BASE_DIR

ORACLE_HOME=/app/oracle/product/dbhome/9.2.0
PATH=$ORACLE_HOME/bin:$PATH
LD_LIBRARY_PATH=$ORACLE_HOME/lib:$LD_LIBRARY_PATH
TNS_ADMIN=$ORACLE_HOME/network/admin
JAVA_PARMS='-Xms512m -Xmx512m -XX:PermSize=64m'

export JAVA_PARMS ORACLE_HOME TNS_ADMIN PATH LD_LIBRARY_PATH

echo "Executing Auto Run for Freestyle Search Engine"
echo "Executing job as `id`"
echo "Executing on `date`"

$JAVA_HOME/bin/java -client $JAVA_PARMS -classpath $BASE_DIR/cacore32-
client.jar:$BASE_DIR/hibernate3.jar:$BASE_DIR/spring.jar:$BASE_DIR/log4j-
1.2.13.jar:$BASE_DIR/ojdbc14.jar:$BASE_DIR/freestylesearch.jar gov.nih.nci.cadsr.freestylesearch.util.Seed
$BASE_DIR/log4j_seed.xml $BASE_DIR/seed.xml
```

   The path references should change to match the deploy environment. Note the JAR files are provided in the 'freestylesearch/lib' folder. The freestylesearch.jar is created by step 7, above, and is located in 'freestylesearch/distrib/war'.

   The first program argument, '$BASE_DIR/log4j.xml', contains the logging configuration and is **not** the same log4j.xml used by JBoss. A full and complete log4j.xml for the Seed

process can be found in 'freestylesearch/conf/common/**log4j.xml**'. Review all elements prior to executing the script.

The second program argument, '$BASE_DIR/seed.xml', specifies the caDSR database connection values as described in step 2, above.

The log file as specified in the log4j.xml contains status information. The first execution of the script must be done manually. Once the run completes and is successful, set up a system task.

13. Create a **scheduled task** to execute the command file created in step 12.

    This script should run at least once every 24 hours. More often if needed as dictated by the activity of changes in the caDSR. For a server located in the Eastern Standard Time Zone the recommended schedule time is between 3:00 and 5:00 AM or immediately after completion of any off hour backups.

14. Optionally, copy the **freestylesearch.jar** to an application project.

    The Freestyle Search API JAR file is created by step 7, above, and is located in 'freestylesearch/distrib/war'. Please note the distinction in the names of the JAR files in this folder. The content of the JAR is documented by executing step 7, above. Please refer to the Javadoc for package '**gov.nih.nci.cadsr.freestylesearch.util**' for features and use of the API.

15. Optionally, delete the '**freestylesearch**' folder and zip.

    After deployment the entire 'freestylesearch' folder and zip file may be deleted. The contents of 'freestylesearch/build', 'freestylesearch/tmp', 'freestylesearch/ospackage' and 'freestylesearch/test_results' are dynamic.

3 Site Configuration

As noted in section 2, the Freestyle Search package is built using the ANT script file build.xml. Possible build task values are 'build-proto', 'build-dev', 'build-qa', 'build-stage', 'build-prod', 'build-ospackage' and 'build-doc'. Please try the build.xml without modifications first and then only change it if necessary.

Operating Properties

The caDSR Freestyle Search has several configuration values. These are:

1. SEED.LASTUPDATE
    This Time Stamp is set when the Seed process successfully completes. During the process only changes after this date/time are processed. To rebuild the Freestyle tables and on the first installation use the date "1999-01-01 00:00:00.0", this is also the default value in the load_options.sql script. This value is maintained by the Seed process consequently this value should only be changed via the load_options.sql script under the direction of the development team.

2. URL
    The URL to access the Freestyle Search page, e.g. NCICB production is
    http://freestyle.nci.nih.gov.

4 Required Technology

The product versions which follow were used during the development and system testing of the caDSR Freestyle Search. Although new versions of the products have been released they have not yet been certified caCORE 3.2.

> **Note: deviating in any way from the products and versions listed below may cause errors. When attempting to deploy the Freestyle Search to a different technology stack, please first deploy as documented in this Installation Guide and then begin the migration.**

Product	Version	Web Site & Notes
JBoss	4.0.2	http://www.jboss.org/downloads/index
JDK	1.5.0_06	http://java.sun.com/j2se/1.5.0/ Recommend downloading the J2EE 1.5 SDK.
Struts	1.2.2	http://struts.apache.org/
Log4j	1.2.8	http://logging.apache.org/log4j
ANT	1.6.2	http://ant.apache.org/
Oracle	9i	http://www.oracle.com Purchase from Oracle.
Eclipse	3.0.1	http://www.eclipse.org/platform/ This is a development environment only and provides no runtime or build components to the Freestyle Search.
IE	6.0+	http://www.microsoft.com/windows/ie/downloads This is the latest Internet Explorer from Microsoft.

5 Troubleshooting

Following is a list of known errors and resolutions that may occur during the build and deployment process. If you have a question or situation not covered in this document please contact the NCICB Help Desk via email at ncicb@pop.nci.nih.gov.

1. **java.lang.NoClassDefFoundError: oracle/jdbc/driver/OracleDriver in JBoss server.log.**
   The message indicates the oracle ojdbc14.jar is missing or not in the class path. This file is included in the open source package. If the problem persists, find the JAR in the Oracle installation folders (%oracle_home%) and copy it to the appropriate JBoss lib folder or modify the class path to include the file. If Oracle is not installed on the machine contact your database or system administrator to acquire the file.

2. **Using Firefox, Mozilla and other web browsers**
   The Freestyle Search has only been certified with Microsoft Internet Explorer version 6.0 and higher.

If you have a question or situation not covered in this document please contact the NCICB Help Desk via email at ncicb@pop.nci.nih.gov.