Chapter 9 - Enabling Indexing & Searching with Solr

Islandora utilizes the Solr open source search platform to enable flexible and configurable indexing and searching. Solr uses the Lucene Java search library at its core for full-text indexing and search and offers hit highlighting, faceted search, dynamic clustering, database integration, and rich document (e.g., Word, PDF) handling as additional features.

On this page:

- How Islandora uses Solr/Lucene and Gsearch
- Installing Solr
- Installing GSearch
- Installing & Configuring the Solr Search Module

How Islandora uses Solr/Lucene and Gsearch

Islandora makes it possible to use the power of Solr/Lucene for discovery. Gsearch is used as a method for keeping indexes current. When an item is ingested, the FOXML is transformed by an XSLT file stored in Gsearch into a format that can be read by Solr's schema and returned based on the request handlers in our custom solrconfig.xml.

The Islandora Solr Search module is packaged with files that will support Islandora solution packs, but these can be modified if you are familiar with Solr. Specifically the Solr schema and corresponding Gsearch XSLT are a good starting point even if you do not use the Solr Search module. Additional information about Solr is presented in Chapter 4: Search and Discovery in Islandora (The Solr Module).

While Solr is not required to run Islandora, it is recommended.

Installing Solr

1. Download Solr to your local environment and unpack the downloaded file.

2. Create the following directories:

```
mkdir -p /opt/solr
```

3. Drag the .war file that is located under the dist directory of your unpacked download to the newly created /opt/solr directory.

4. Create a new file called solr.xml under $CATALINA_HOME/conf/Catalina/localhost/ and insert the following into the newly created solr.xml file:

```
<Context docBase="/opt/solr/apache-solr-1.4.1.war" debug="0" crossContext="true">

<Environment name="solr/home" type="java.lang.String" value="/usr/local/fedora/gsearch_solr/solr" override="
true" />

</Context>
```

5. Navigate to Fedora Home and create a gsearch_solr directory:

```
cd $FEDORA_HOME

mkdir gsearch_solr
```

6. Navigate into your newly created gsearch_solr directory:

```
cd gsearch_solr
```

7. Copy the entire Solr directory (located under exmple/solr from the unpacked zipped file) to the gsearch_solr directory you just created.

8. Navigate into the Solr directory and run a print working directory (pwd) command. The system response should yield usr/local/fedora/gsearch_solr/solr.

9. Restart your web server.

10. Solr should now be up and running. Verify this by going to http://server:8080/solr/admin.

Installing GSearch

The Fedora Generic Search Service, or GSearch, is a search service installed with Fedora that allows for automatic updating of the Lucene/Solr index. GSearch relies on JMS to receive messages that are sent when Fedora objects are ingested, modified or purged. This keeps the Lucene index in sync with the Fedora repository.

Pre-installation software checklist:
It is recommended that Solr be setup, configured and running prior to installing GSearch.

Installation Steps:
1. Download fedoraGSearch from SourceForge.net and extract the contents of the compressed file.

2. Copy the fedoragsearch.war file located in the genericsearch-2.2 directory of the downloaded file to your Fedora webapps directory:

```
cd genericsearch-2.2

cp fedoragsearch.war $CATALINA_HOME/webapps
```

3. Stop and restart your Fedora instance. When you restart, note that a fedoragsearch directory has been created in your Fedora webapps directory.

4. Navigate into the following directory:

```
cd $CATALINA_HOME/webapps/fedoragsearch/WEB-INF/classes
```

5. Move the contents of the config/DemoOnSolr directory into the config directory:

```
mv configDemoOnSolr/ config

cd config
```

6. To configure the GSearch service for automatic updating of the Solr index, a few GSearch configuration files must be modified. To do this, make the following edits to the fedoragsearch.properties file:

a. Set the uncommented fedoragsearch.soapBase = http://localhost:8080/fedoragsearch/services
b. Update the fedoragsearch.soapUser = YOURFEDORAUSERNAME
c. Update the fedoragsearch.soapPass = YOURFEDORAPASSWORD Note: If you have forgotten your Fedora password, it can be found in $FEDORA_HOME/server/config/fedora-users.xml
d. Update fedoragsearch.repositoryNames = gsearch_solr
e. Update fedoragsearch.indexNames = gsearch_solr

7. Now we'll make our GSearch directory using a move command:

```
cd repository

mv DemoAtDtu gsearch_solr

cd gsearch_solr
```

8. The following output details the changes made to the repository.properties file.

Note: Ensure that the TrustStore path matches the path on your system:

```
vi repository.properties

\# $Id: repository.properties 5732 2006-11-27 15:26:04Z gertsp $

fgsrepository.repositoryName = gsearch_solr

fgsrepository.fedoraSoap = http://localhost:8080/fedora/services

fgsrepository.fedoraUser = fedoraAdmin

fgsrepository.fedoraPass = fedoraAdmin

fgsrepository.fedoraObjectDir = /usr/local/fedora/data/objectStore

fgsrepository.fedoraVersion = 3.4

fgsrepository.defaultGetRepositoryInfoResultXslt = copyXml

fgsrepository.trustStorePath = TRUSTSTOREPATH/truststore                hint: usually /usr/local/fedora/server
/truststore

fgsrepository.trustStorePass = TRUSTSTOREPASS                                hint: usually tomcat
```

9. Next, move the contents of the DemoOnSolr directory into the gsearch_solr directory:

```
cd ../..

cd index

mv DemoOnSolr/ gsearch_solr

cd gsearch_solr
```

10. Modify the index.properties file, making the following changes:
  a. Update fgsindex.indexName = gsearch_solr
  b. Update fgsindex.indexBase = http://localhost:8080/solr
  c. Update fgsindex.indexDir = /usr/local/fedora/gsearch_solr/solr/data/index

11. Next, make the following changes (During this process you will be putting files that are packaged with the Islandora Solr module to ensure support for Islandora Solution Packs. These are: demoFoxmlToSolr.xslt, schema.xml):

```
cd /usr/local/fedora/tomcat/webapps/fedoragsearch/WEB-INF/classes/config

mkdir updater

cd updater

cp -R ../../configBasic/updater/* .

cd /usr/local/fedora/gsearch_solr/solr/conf

mv schema.xml schema.xml.bak

mv solrconfig.xml solrconfig.xml.bak

cp ../../../tomcat/webapps/fedoragsearch/WEB-INF/classes/config/index/gsearch_solr/conf/schema.xml .

cp ../../../tomcat/webapps/fedoragsearch/WEB-INF/classes/config/index/gsearch_solr/conf/solrconfig.xml .
```

12. Modify the solrconfig.xml file as follows:

Replace

```
${solr.data.dir:./solr/data}
```

within the <dataDir> tags with

```
/usr/local/fedora/gsearch_solr/solr/data
```

13. Next, we'll need to update all xslts in /usr/local/fedora/tomcat/webapps/fedoragsearch/WEB-INF/classes/config/rest
  a. Navigate to /usr/local/fedora/tomcat/webapps/fedoragsearch/WEB-INF/classes/config/rest
  b. Replace CONFIGPATH in each xslt with /usr/local/fedora/tomcat/webapps/fedoragsearch/WEB-INF/classes/config
  c. Performing a grep CONFIGPATH * will tell you where the references are for each file.

```
sed -i 's#CONFIGPATH#/usr/local/fedora/tomcat/webapps/fedoragsearch/WEB-INF/classes/config#g' .
```

14. Once done, copy the lucene jar files from the solr webapp lib directory to the fedoragsearch webapp lib directory using the following commands:

```
cd $CATALINA_HOME/webapps/fedoragsearch/WEB-INF/lib

cp ../../../solr/WEB-INF/lib/lucene-*.jar .
```

15. Remove the old lucene jar files from the fedoragsearch webapps lib directory. As well, if there are any other duplicates of lucene-* files, remove the older versions:

```
rm lucene-core-2.4.0.jar
```

16. Finally, make the following changes to the demoFoxmlToSolr.xslt:

```
cd $CATALINA_HOME/webapps/fedoragsearch/WEB-INF/classes/config/index/gsearch_solr/

vi demoFoxmlToSolr.xslt
```

Modify the following statement:

```
<xsl:if test="starts-with($PID,'demo')">

<xsl:apply-templates mode="activeDemoFedoraObject"/>

</xsl:if>
```

Remove the condition, so it looks like this:

```
<xsl:apply-templates mode="activeDemoFedoraObject"/>
```

17. Restart Fedora

18. GSearch and Solr should now be running properly and GSearch should be automatically updating the Solr index. You can see GSearch in action by visiting http://localhost:8080/fedoragsearch/rest

Installing & Configuring the Solr Search Module

The Islandora Solr search module allows you to search the Solr index. The Islandora Solr Sample Configuration module provides default display profiles to the module. The module makes four new blocks available; two for search and two for display. The other block is called the *Advanced Search Block* and does fielded searches against the Solr index. Both blocks would use whatever request handler is configured in the module settings. For information on how to configure the Solr module, see Chapter 4 - Search and Discovery in Islandora (The Solr Module).

Installation Steps:

1. Verify that fedoraGSearch and Solr are both installed and running.
2. Download the islandora_solr_search module and install as a Drupal module.
3. Then, download and uncompress the Apache Solr php client and copy the Solr directory under Apache from the archive to the islandora_solr_search module's folder.
4. Log in to your Drupal site to enable the islandora_solr_search module.
    a. Administer > Modules and enable Islandora Solr Search