

HAT Installation Guide

It is advised that the Cent O/S VM no longer be used, please use the Ubuntu image

Download HAT VM from test server <http://137.149.157.11/files/Ubu-CS482-Mar102011.ova>

- Install into VirtualBox 4.0 or greater
 - username: islandora
 - password: islandora
- use sudo su to change to root user
- change ProxyRequests value in /etc/apache2/sites-available/islandora on line 59 to On
 - sudo vim /etc/apache2/sites-available/islandora
- Disable the old proxy module, enable the new with commands:
 - a2dismod
 - proxy
 - a2enmod
 - proxy_http
- Restart apache
 - /etc/init.d/apache2 restart

Checkout latest HAT repository from <http://github.com/posif/HAT>

- Copy HAT/iiv into /usr/local/fedora/tomcat/webapps/iiv
- Change file permissions to ownership by fedora
 - sudo chown -R fedora:fedora /usr/local/fedora/tomcat/webapps/iiv
- Install the iiv_annotation module
 - copy HAT/iiv_annotation to /var/www/html/drupal/sites/all/modules/iiv_annotation on VM
 - sudo chown -R www-data:www-data /var/www/html/drupal/sites/all/modules/iiv_annotation
 - log into the drupal site at <http://192.168.56.101/user>
 - username: admin
 - password: islandora
 - using the administrative menu, select "Site Building-->Modules"
 - Under the "Fedora Repository" fieldset there will be a IIV_Annotation option, enable it, scroll to the bottom and submit
 - Be certain to set the Drupal permissions for the module
- Repeat module installation steps to install the guardian_solr_config module
 - Update the Solr Search Block admin page to output results from the guardian_solr_config module by visiting http://192.168.56.101/admin/settings/islandora_solr_search
 - Set Configuration Package to Guardian Output

Checkout the latest from from <http://github.com/posif/islandora>

- Copy posif islandora files into /var/www/html/drupal/sites/all/modules/islandora

Grab the latest IIV copy from <http://github.com/posif/iiv>

- follow readme and install war to /usr/local/fedora/tomcat/webapps

Update newspapers:viewerSdep-issueCModel in fedora

Go to <http://192.168.56.101:8080/fedora/admin> (default credentials are fedoraAdmin fedoraAdmin)

Delete the old dataobject newspapers:viewerSdep-issueCModel from Fedora

Inject the new object found in the root of the HAT repository

Update newspapers:viewerSdep-pageCModel in fedora

Go to <http://192.168.56.101:8080/fedora/admin> (default credentials are fedoraAdmin fedoraAdmin)

Delete the old dataobject newspapers:viewerSdep-pageCModel from Fedora

Inject the new object found in the root of the HAT repository

Solar Multiword Index Creation

Solr indexing allows efficient processing of strings in the wordsxy file.

NOTE: An equivalent index could be made a little more efficiently by using the basic OCR

NOTE: Careful if copy/pasting; double quotes may be messed up.

1. Stop fedora. (/etc/init.d/fedora stop)
2. Setup to index WORDSXY as dsm.WORDSXY
 - a. Add to (demo)FoxmlToSolr.xslt (\$FEDORA_HOME/tomcat/webapps/fedoragsearch/WEB-INF/classes/config/index/gsearch_solr/demoFoxmlToSolr.xslt) :

```
<xsl:for-each select="foxml:datastream[@ID='WORDSXY']">
  <field>
    <xsl:attribute name="name">
      <xsl:value-of select="concat('dsm.', @ID)"/>
    </xsl:attribute>
    <xsl:value-of select="exts:getDatastreamText($PID, $REPOSITORYNAME,
@ID, $FEDORASOAP, $FEDORAUSER, $FEDORAPASS, $TRUSTSTOREPATH,
$TRUSTSTOREPASS)"/>
  </field>
</xsl:for-each>
```

- b. Update schema.xml (\$FEDORA_HOME/gsearch_solr/solr/conf/schema.xml):

- i. Add type to schema.xml:

```
<fieldType name="wordsxy_text" class="solr.TextField">
  <analyzer>
    <charFilter class="solr.HTMLStripCharFilterFactory"/>
    <tokenizer class="solr.StandardTokenizerFactory"/>
    <filter class="solr.StandardFilterFactory"/>
    <filter class="solr.LowerCaseFilterFactory"/>
  </analyzer>
</fieldType>
```

```

        <!--Eliminate preceding number, quotes, and right angle
        bracket; may not be necessary, or in fact useless given the change
        to Standard Tokenizer/Filter...-->
        <filter class="solr.PatternReplaceFilterFactory"
        pattern="([0-9]*(\u0022|(\u0026amp;\u003B))(\u003E|(\u0026gt;\u003B))
        )" replacement="" replace="all" />
        <filter class="solr.TrimFilterFactory"
        updateOffsets="true"/>
    </analyzer>
</fieldType>

```

ii. Add field to schema.xml:

```

<field name="dsm.WORDSXY" type="wordsxy_text" indexed="true"
stored="true" multiValued="true" termVectors="true"
termPositions="true" termOffsets="
true"/>

```

3. Update solrconfig.xml (\$FEDORA_HOME/gsearch_solr/solr/conf/solrconfig.xml):

a. Add search component:

```

<searchComponent name="TVC"
class="org.apache.solr.handler.component.TermVectorComponent"/>

```

b. Add request handler:

```

<requestHandler name="guardian_words" class="solr.SearchHandler">
    <lst name="defaults">
        <str name="fl">PID, dc.title, type, score</str>
        <str name="fq">PID:guardian*</str>
        <str name="qf">PID^0.5 dc.title dsm.WORDSXY^2.0</str>
        <str name="pf">dsm.WORDSXY^4.0</str>
        <str name="mm">100%</str>
        <str name="ps">20</str>
    </lst>
    <arr name="last-components">
        <str>TVC</str>
    </arr>
</requestHandler>

```

4. Start fedora (/etc/init.d/fedora start)

5. Actually index the files

- Go to <islandora-url>:8080/fedoragsearch/rest
- Hit "update index"
- Enter the "Foxml files path:" field (/usr/local/fedora/data/objectStore)
- Hit "updateIndex from Foxml files"
 - This may take a while; should return to the same screen.

6. Modify configuration of "Islandora Solr Client" in Drupal

- Change to use the "proper" Solr instance: localhost:8080/solr
- Change to use the new request handler (requires submission of previous step)
- Change "facet fields" to: dc.title,dsm.WORDSXY

7. Test Islandora Solr Client:

- Search for "certain gentlemen took up the movement" (guardian:18901216-00s-003-001),

or some other phrase...