UNDESA FEDERATED SEARCH APPLICATION
UNDESA CROSSWALK

INTRODUCTION

The UNDESA Crosswalk application, hereinafter called *tochi*, is a federated search application designed to ease the task of information discovery and extraction from multiple information repositories. One might ask the need for such an application when there are numerous search engines like Google; the answer is simple, due to the specificity of the information, tailor-made solutions like the crosswalk offer higher information discovery efficiency than generic solutions.

INSTALLATION INSTRUCTIONS

Presently, there is no release of tochi available, but one will be available with time. The following are the installation instructions; they assume basic knowledge of the Linux Operating System (in particular Ubuntu) but the instructions can be used on other systems as well.

REQUIRED LIBRARIES

Tochi requires certain libraries be available for successful compilation and deployment. These are

1. Spring.jar (2.5+)
2. spring-webmvc.jar (2.5+)
3. spring-aspects.jar
4. commons-collection.jar
5. commons-logging.jar
6. commons-fileupload.jar
7. commons-lang.jar
8. httpclient.jar
9. dspace-api.jar
10. dspace-api-lang.jar
11. dspace-api-jspui.jar
12. abdera-0.4-incubating.jar
13. abdera-client-0.4-incubating.jar
14. abdera-core-0.4-incubating.jar
15. abdera-extensions-main-0.4-incubating.jar
16. abdera-extensions-opensearch-0.4-incubating.jar
17. abdera-parser-0.4-incubating.jar
18. abdera-server-0.4-incubating.jar
19. activation.jar
20. handle.jar
21. xpp3_min.jar
22. xercesImpl.jar
23. xstream.jar (3+)
24. flexjson.jar
25. jaxen.jar
26. dom4j.jar
27. log4j.jar
28. mets.jar
29. mail.jar
30. marc4j.jar
31. lucene-core-2.3.jar
32. lucene-analyzers-2.3.jar

33. poi.jar
34. pdfbox.jar
35. postgresql-8.1.jar
36. tm-extractors.jar
37. axiom-api.jar
38. axiom-impl.jar
39. jdom.jar
40. htmlparser.jar
41. jempbox.jar
42. servlet-api-2.5.jar
43. xalan-2.7.jar
44. cglib-nodep.jar
45. aspectjrt.jar
46. aspectjweaver.jar
47. normalizer.jar
48. jstl.jar
49. slf4j-api.jar
50. slf4j-log4j.jar
51. standard.jar
52. ezmorph.jar
53. xwork.jar
54. xmlsec.jar

Most of these libraries are required by DSPACE as a DSPACE instance is executed within tochi.

Instructions:

1. Install Java JDK 1.5+ and set the *JDK_HOME* to point to the Java installation directory. This can be done with the apt-get command from a terminal

   $sudo apt-get install openjdk-6-jdk

   Then, set the *JDK_HOME* environment variable by:
   1. Open the file .bashrc in your home directory with your preferred editor
   2. Append the line:
      export JDK_HOME=/usr/lib/jvm/java-6-openjdk
   3. Save the file
2. Obtain the Java development tool ANT from the Apache website i.e. http://ant.apache.org version 1.6 and above.
   1. Extract the archive to a folder
   2. Edit the .bashrc file as in above

   3. append the lines:
      1. export ANT_HOME={extraction folder}
      2. export PATH=$PATH:$ANT_HOME/bin
      3. Save the file
2. Get apache Tomcat 6+ (or any servlet 2.5 complient Java Application server like Mortbay Jetty 6) from the apache website i.e. http://tomcat.apache.org. Extract the archive into a folder, which we will designate as *TOMCAT_HOME*. The *TOMCAT_HOME* environment variable MUST be

set.

3. Obtain and install DSPACE 1.5+ digital repository. Be sure to note the DSPACE_HOME environment variable i.e. the location of the Dspace installation, this environment variable DSPACE_HOME MUST be set. The Dspace installation directory MUST be in the same server that is to house tochi.
4. (Optional) Obtain and install Koha 3+ library application. Be sure to specify ZEBRA as the indexing option and enable the PazPar2 option during Koha configuration. This allows for remote access of the index server
5. Checkout the tochi source from the google code svn repository.
   svn co http://code.google.com/p/bungeni-dspace/trunk.
6. From the tochi folder (designated as *TOCHI_HOME*), build tochi
   $ant
7. Modify the dspace configuration file *DSPACE_HOME*/config/dspace.cfg as follows:
   In the *Search indexing settings* section, replace the <search.index.*> parameters with the following:
   search.index.1 = author:dc.contributor.*
   search.index.2 = author:dc.creator.*
   search.index.3 = title:dc.title
   search.index.4 = keyword:dc.subject.*
   search.index.5 = abstract:dc.description.abstract
   search.index.6 = author:dc.description.statementofresponsibility
   search.index.7 = series:dc.relation.ispartofseries
   search.index.8 = abstract:dc.description.tableofcontents
   search.index.9 = mime:dc.format.mimetype
   search.index.10 = sponsor:dc.description.sponsorship
   search.index.11 = govdocument:dc.identifier.govdoc
   search.index.12 = language:dc.language.iso
   search.index.13 = description:dc.description
   search.index.14 = issue_date:dc.date.issued

   Also, search for the occurance of the parameter *search.operator* and replace the value to AND i.e. search.operator = OR to search.operator = AND

   Re-index the index. This is done by executing the following command in DSPACE_HOME/bin

   $*DSPACE_HOME*/bin/index-init

   During re-indexing, make sure no instance of DSPACE is running.
   Make sure this command is run as the user account that will deploy DSPACE!

CONFIGURATION

Managing Repositories

Repositories information is kept in an XML file called repositories.xml located in *TOCHI_HOME*/WEB-INF folder. Adding or removing repositories to be searched is just a matter of adding/removing lines in this file and restarting the application.

The following is the structure of the repositories.xml file:

```
<repositories>
        <repository url="http://127.0.0.1:8080/crosswalk" family="dspace" default="true"/>
        <repository url="http://127.0.0.1:11001/biblios" family="koha"/>
</repositories>
```

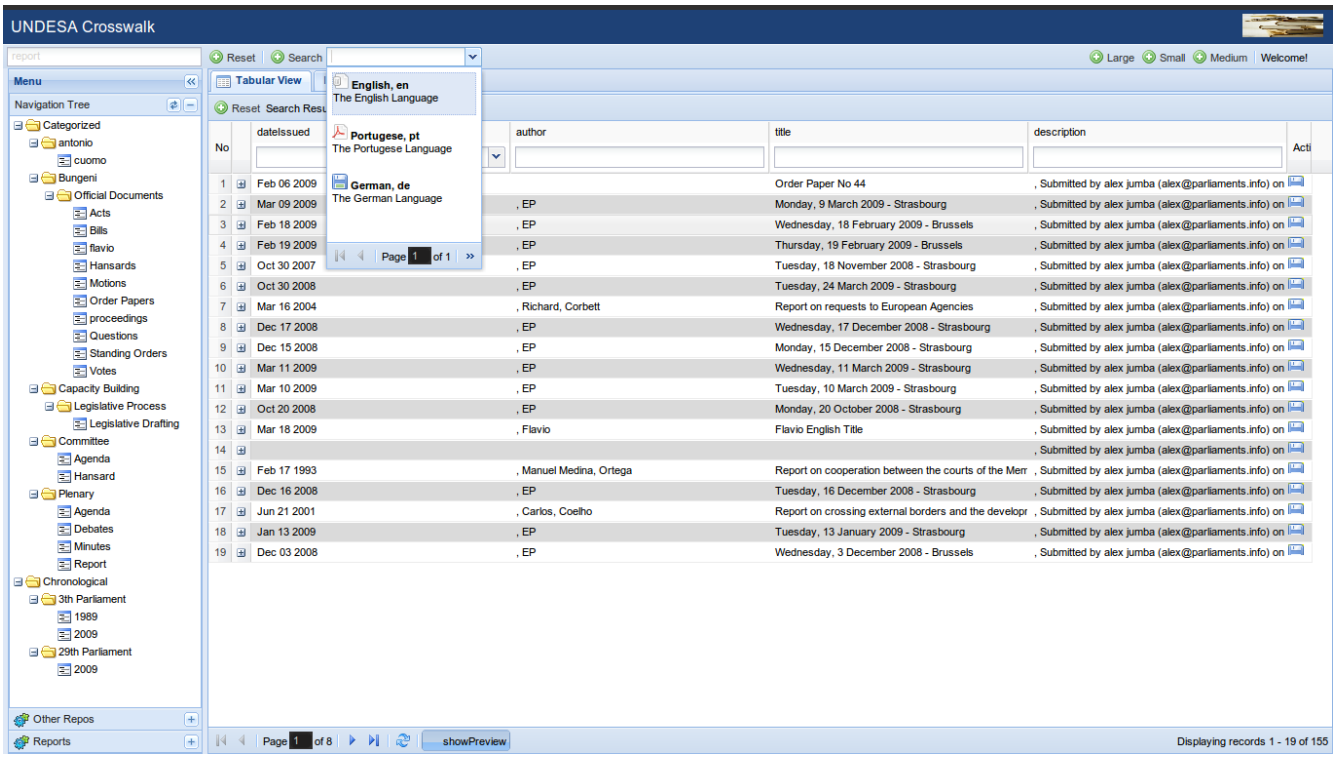
Managing Languages

Tochi is i18n enabled allowing it to be translated into several languages

Adding languages

Language records are kept in an XML file called languages.xml. Each entry in this file corresponds to a distinct language. In addition, there is a properties file with the structure messages_{language_code}.properties e.g. messages_en.properties for the English bundle. These properties files are located in *TOCHI_HOME*/WEB-INF/classes folder and contains the translations for the languages. The language code suffix of the properties files MUST correspond to the code attribute of the language tag of the appropriate language in languages.xml file. There is also a provision for the flags of the countries which speak the language. These flags are to be stored in the TOCHI_HOME/images/languages folder with the name being the code of the language and having the extension .png.

Language selection



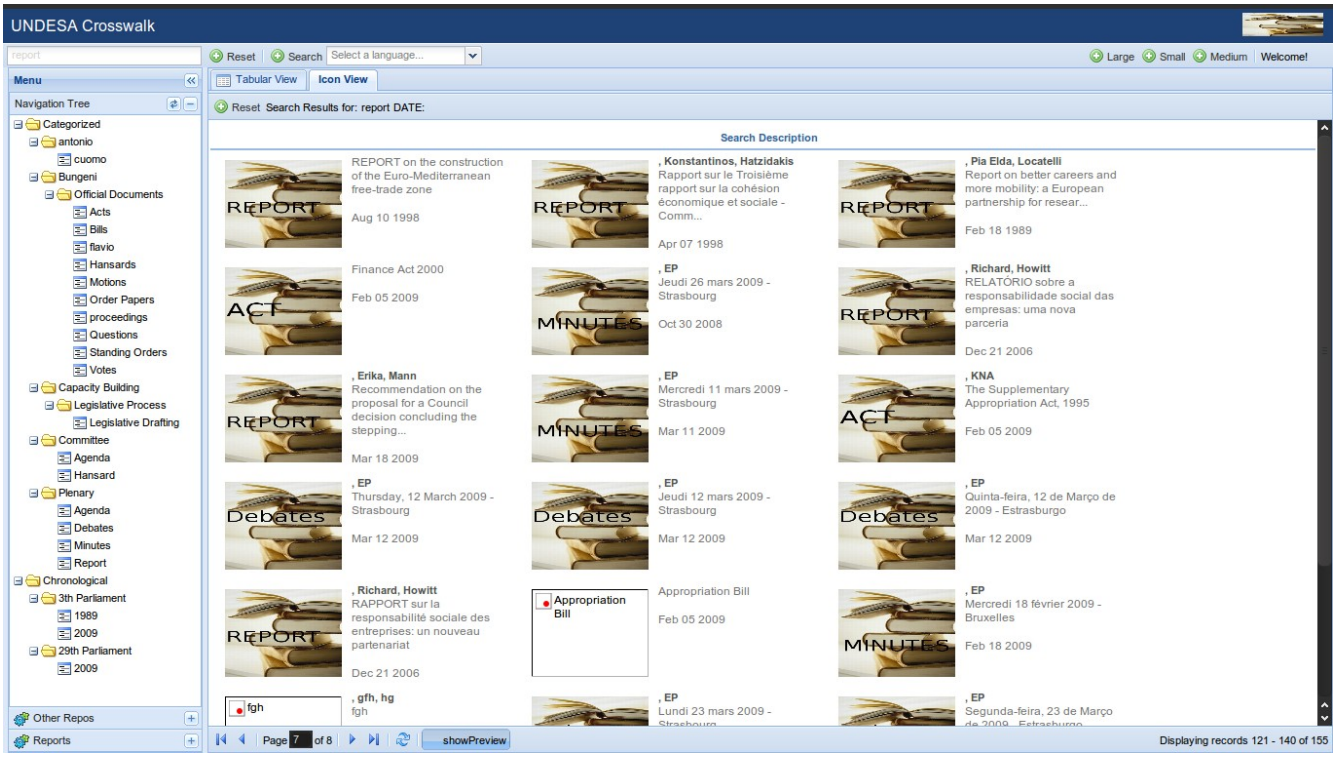
USAGE

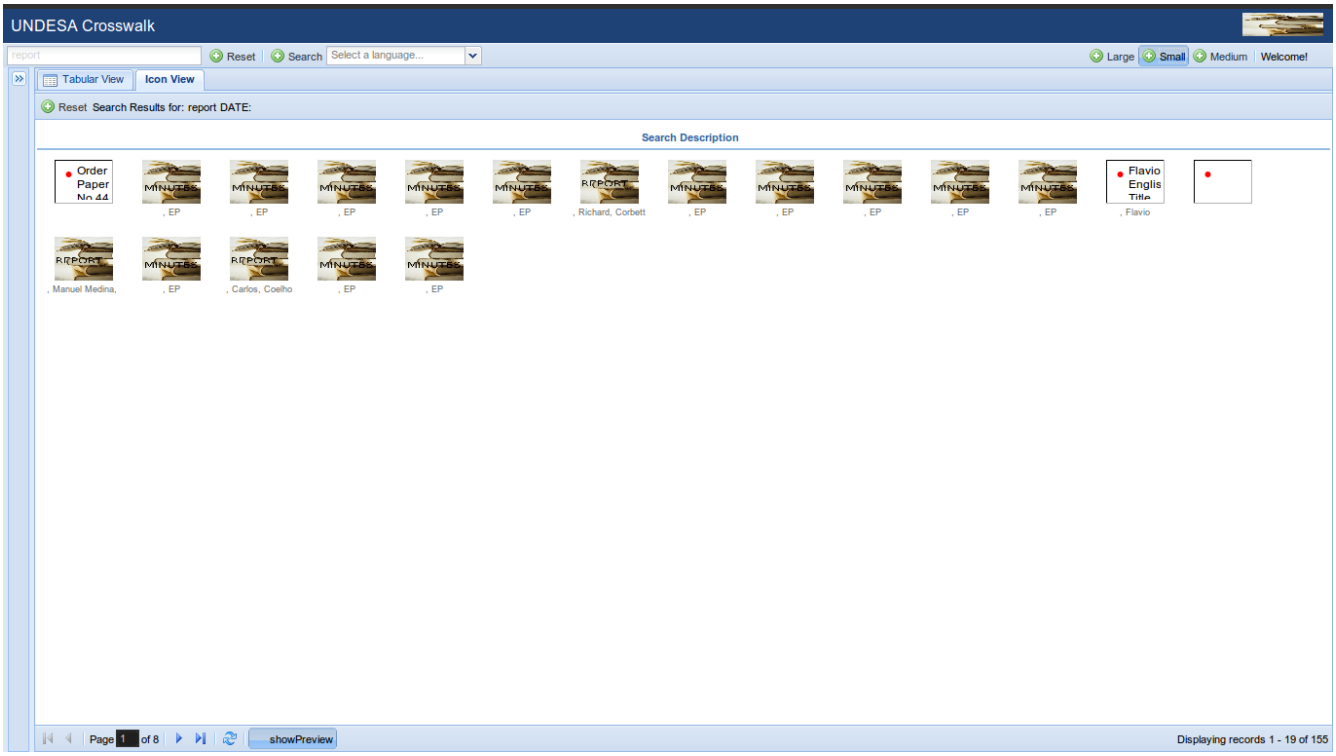
Tochi can be used both as a web–service as well as a web application.

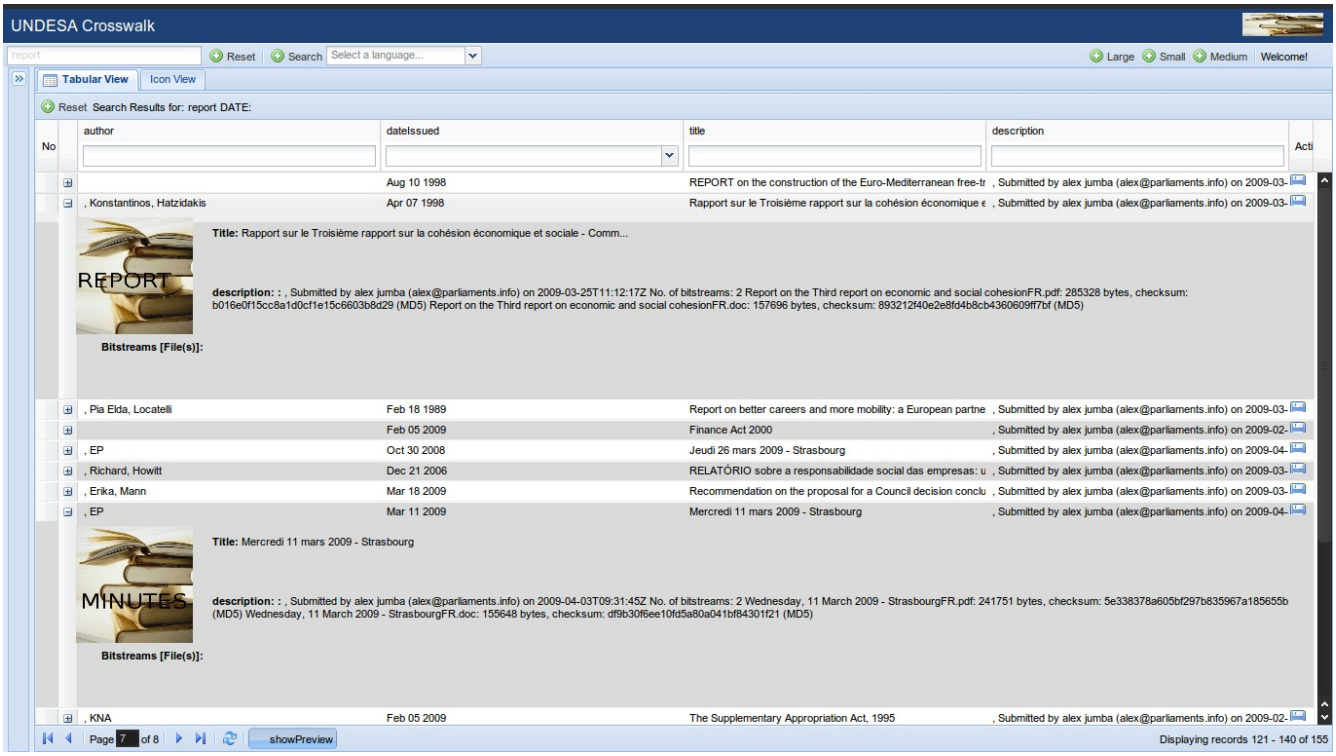
As a Web-Application

Tochi can be accessed in a browser by pointing the browser to the url: http://<hostname>:8080/crosswalk/. The top pane is the application banner as well as the master search controls and view mode controls.

The left pane houses an accordion which aids in the selection of the activities to be performed.




UNDESA Crosswalk
Reset
Search
Select a language...
Large
Small
Medium
Welcome!
Tabular View
Icon View
Reset Search Results for: report DATE:
Search Description
Order Paper No 44
Flavio Englis Title
, Flavio
, Richard, Corbett
, Manuel Medina,
, Carlos, Coelho
Page 1 of 8
showPreview
Displaying records 1 - 19 of 155


UNDESA Crosswalk
Reset
Search
Select a language...
Large
Small
Medium
Welcome!
Tabular View
Icon View
Reset Search Results for: report DATE:
No
author
dateIssued
title
description
Acti
Aug 10 1998
REPORT on the construction of the Euro-Mediterranean free-tr
, Submitted by alex jumba (alex@parliaments.info) on 2009-03-
, Konstantinos, Hatzidakis
Apr 07 1998
Rapport sur le Troisième rapport sur la cohésion économique e
, Submitted by alex jumba (alex@parliaments.info) on 2009-03-
Title: Rapport sur le Troisième rapport sur la cohésion économique et sociale - Comm...
description: : , Submitted by alex jumba (alex@parliaments.info) on 2009-03-25T11:12:17Z No. of bitstreams: 2 Report on the Third report on economic and social cohesionFR.pdf: 285328 bytes, checksum: b016e0f15cc8a1d0cf1e15c6603b8d29 (MD5) Report on the Third report on economic and social cohesionFR.doc: 157696 bytes, checksum: 893212f40e2e8fd4b8cb4360609ff7bf (MD5)
Bitstreams [File(s)]:
, Pia Elda, Locatelli
Feb 18 1989
Report on better careers and more mobility: a European partne
, Submitted by alex jumba (alex@parliaments.info) on 2009-03-
Feb 05 2009
Finance Act 2000
, Submitted by alex jumba (alex@parliaments.info) on 2009-02-
, EP
Oct 30 2008
Jeudi 26 mars 2009 - Strasbourg
, Submitted by alex jumba (alex@parliaments.info) on 2009-04-
, Richard, Howitt
Dec 21 2006
RELATÓRIO sobre a responsabilidade social das empresas: u
, Submitted by alex jumba (alex@parliaments.info) on 2009-03-
, Erika, Mann
Mar 18 2009
Recommendation on the proposal for a Council decision conclu
, Submitted by alex jumba (alex@parliaments.info) on 2009-03-
, EP
Mar 11 2009
Mercredi 11 mars 2009 - Strasbourg
, Submitted by alex jumba (alex@parliaments.info) on 2009-04-
Title: Mercredi 11 mars 2009 - Strasbourg
description: : , Submitted by alex jumba (alex@parliaments.info) on 2009-04-03T09:31:45Z No. of bitstreams: 2 Wednesday, 11 March 2009 - StrasbourgFR.pdf: 241751 bytes, checksum: 5e338378a605bf297b835967a185655b (MD5) Wednesday, 11 March 2009 - StrasbourgFR.doc: 155648 bytes, checksum: df9b30f6ee10fd5a80a041bf84301f21 (MD5)
Bitstreams [File(s)]:
, KNA
Feb 05 2009
The Supplementary Appropriation Act, 1995
, Submitted by alex jumba (alex@parliaments.info) on 2009-02-
Page 7 of 8
showPreview
Displaying records 121 - 140 of 155

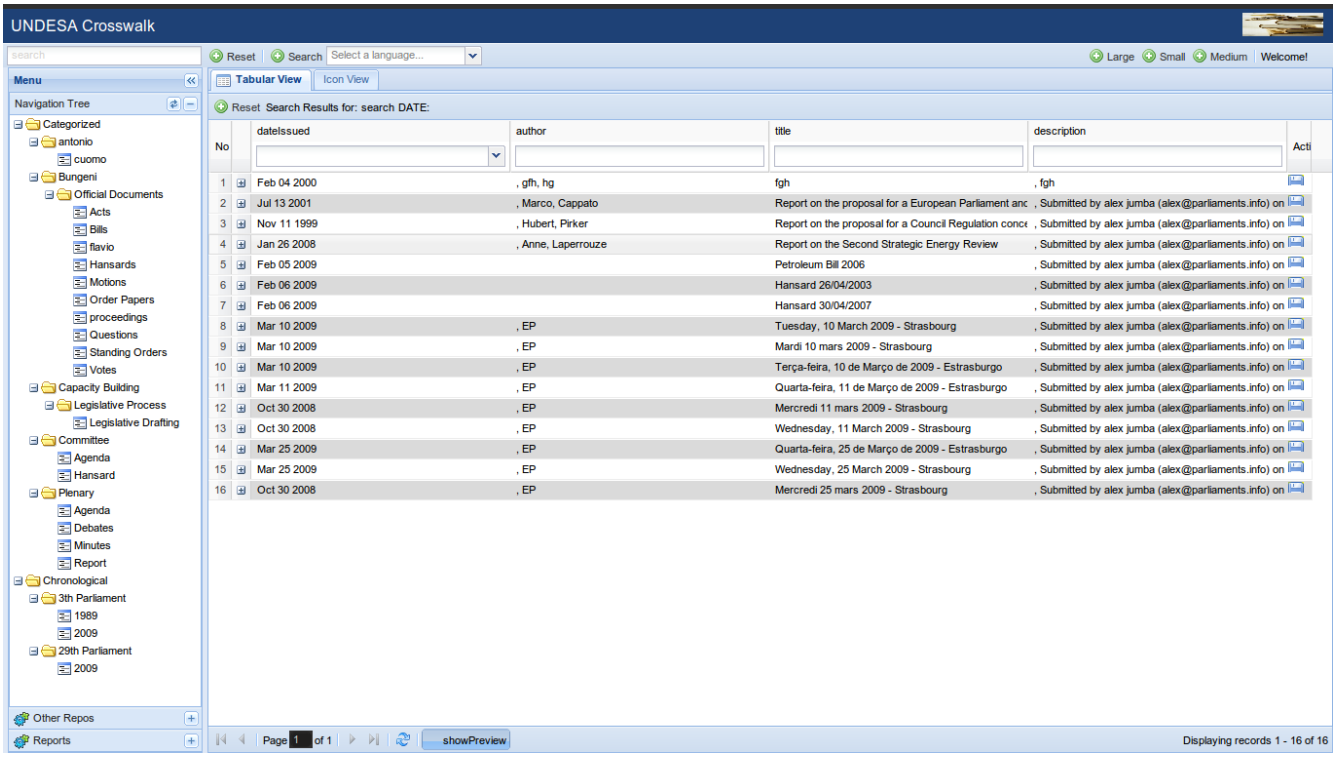
What parts of the url mean

The master base url does not change with the interactions of the user with the application. GET parameters are used to further customize the user experience. The following are the parameters:

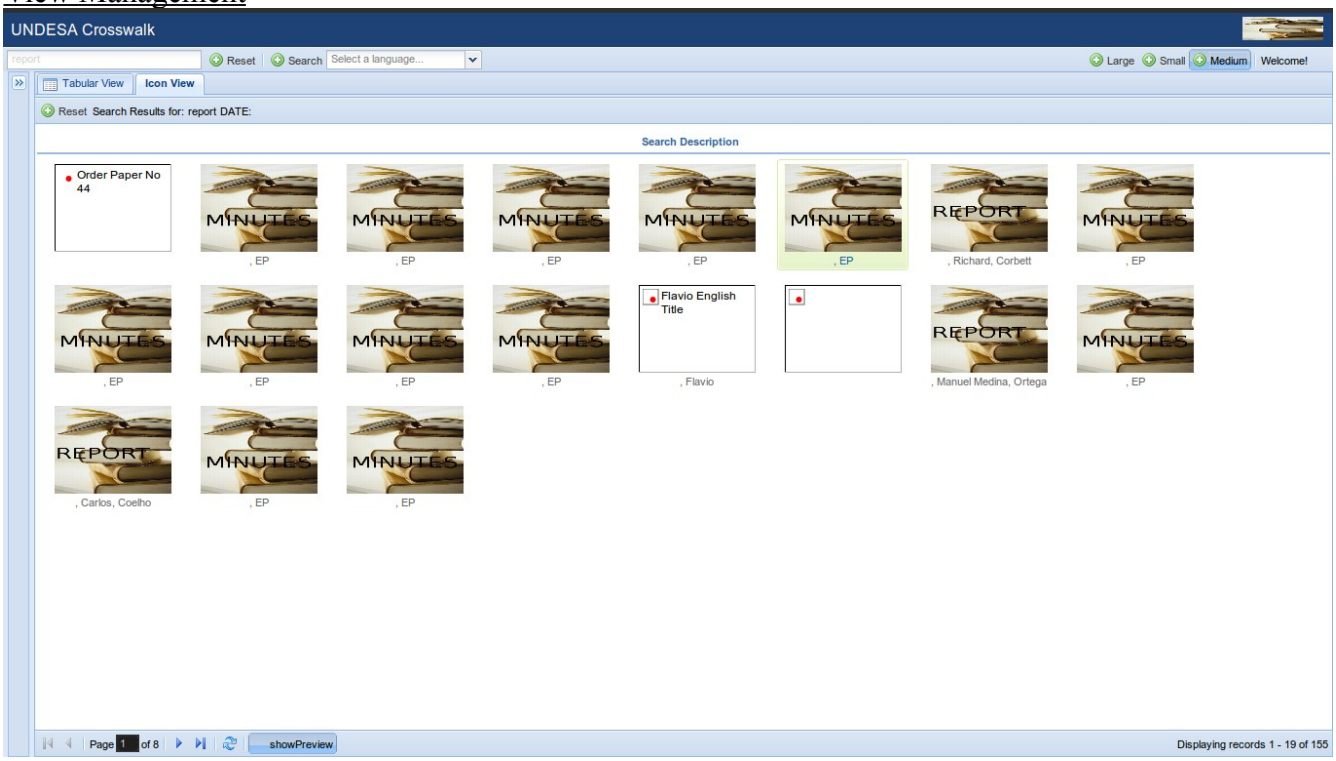
lang: this specifies the language code of the language you would like to switch to e.g. pt for Portuguese

search: this is the search term to be searched by the application. This is a convenience parameter as the resulting page will contain the results of this search.

Navigation



View Management



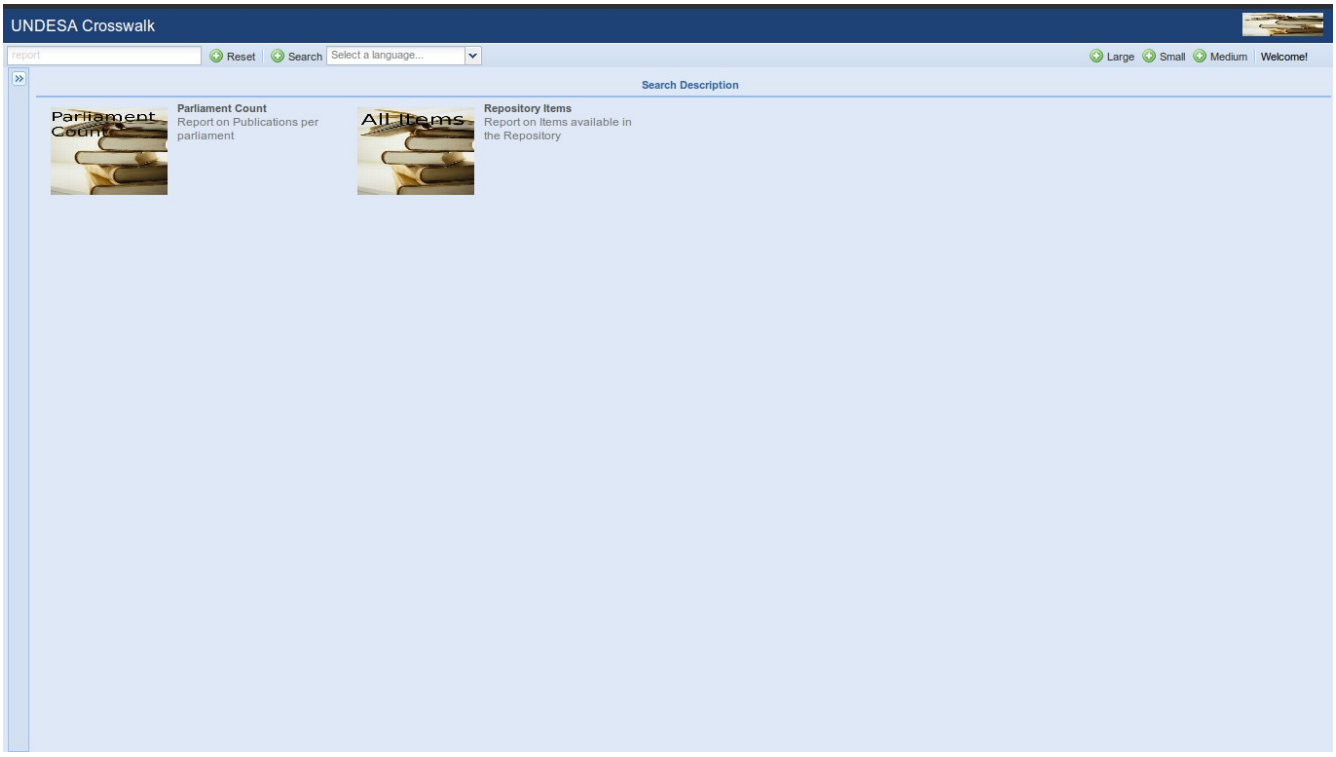
Reports

A reporting facility exists for the primary Dspace library managed by tochi. The reports are pluggable by the use of an XML file to specify the details of the reports and these reports will dynamically appear in the web user interface in the reports page. Tochi contains plugins to the dspace library that enable it to perform unique statistical analyses of the library items and relay this information to the reports exected by the reporting engine runtime.

The following is the structure of the reports.xml file located in the TOCHI_HOME/WEB-INF/ folder:

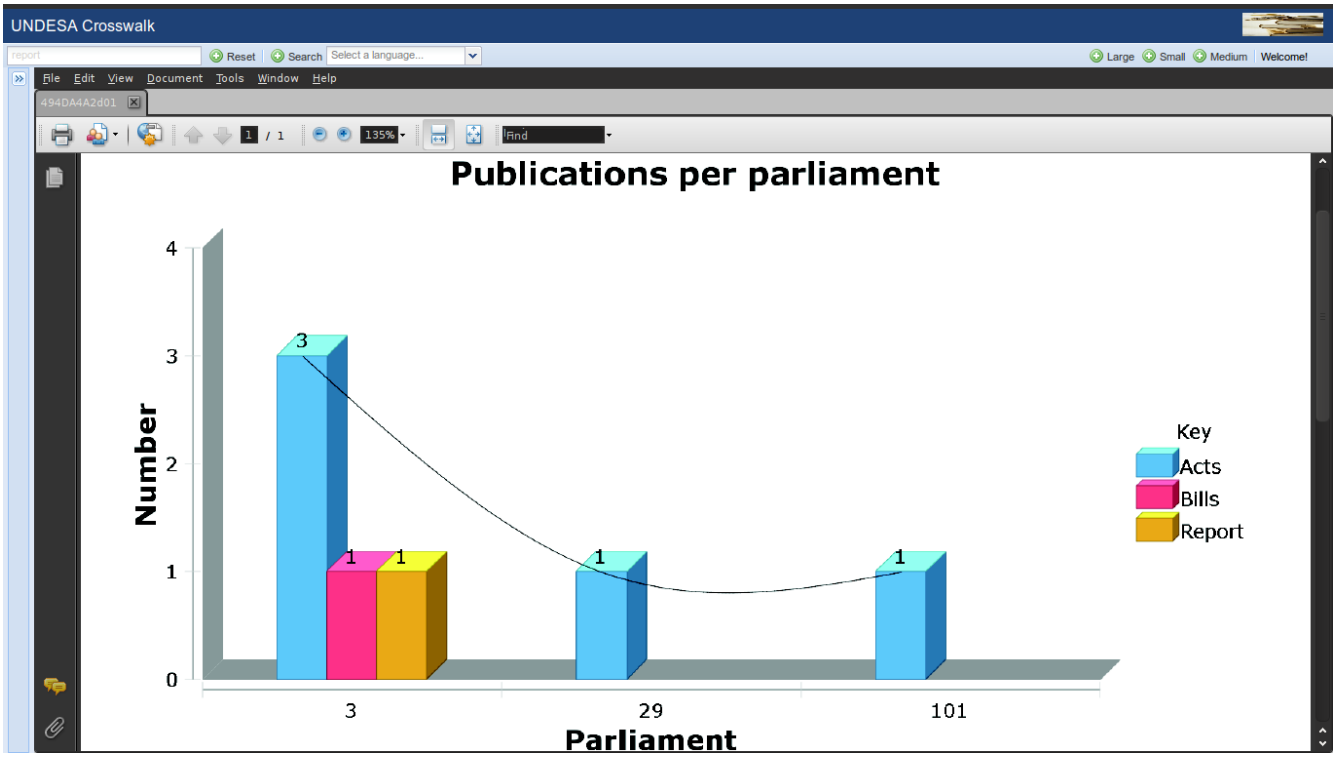
```
<reports>
        <report name="Parliament Count" description="Report on Publications per parliament" uri="http://127.0.0.1:8080/report/frameset?__report=reports/parliament_count.rptdesign&amp;amp;__format=pdf"/>
        <report name="Repository Items" description="Report on Items available in the Repository" uri="http://127.0.0.1:8080/report/frameset?__report=reports/dspaceitems.rptdesign&amp;amp;__format=pdf"/>
</reports>
```

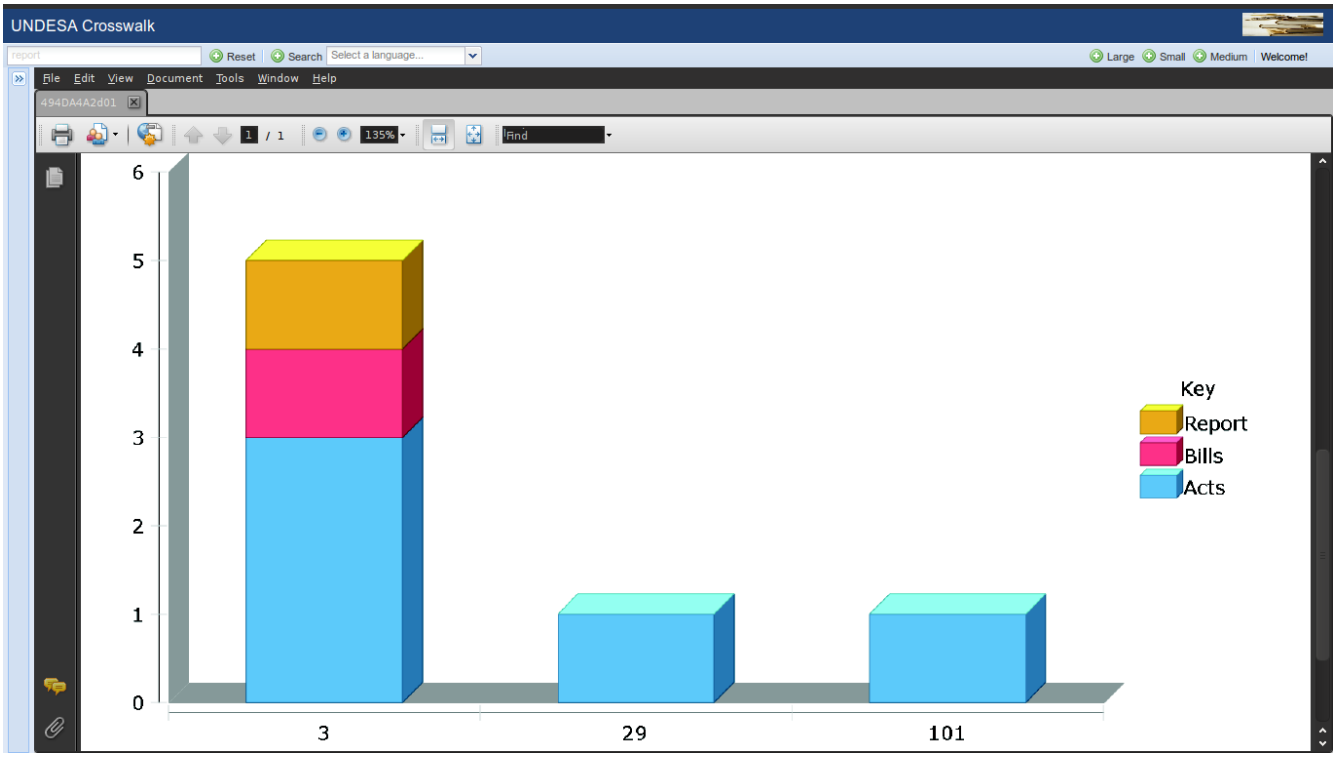
The reporting page is as follows:



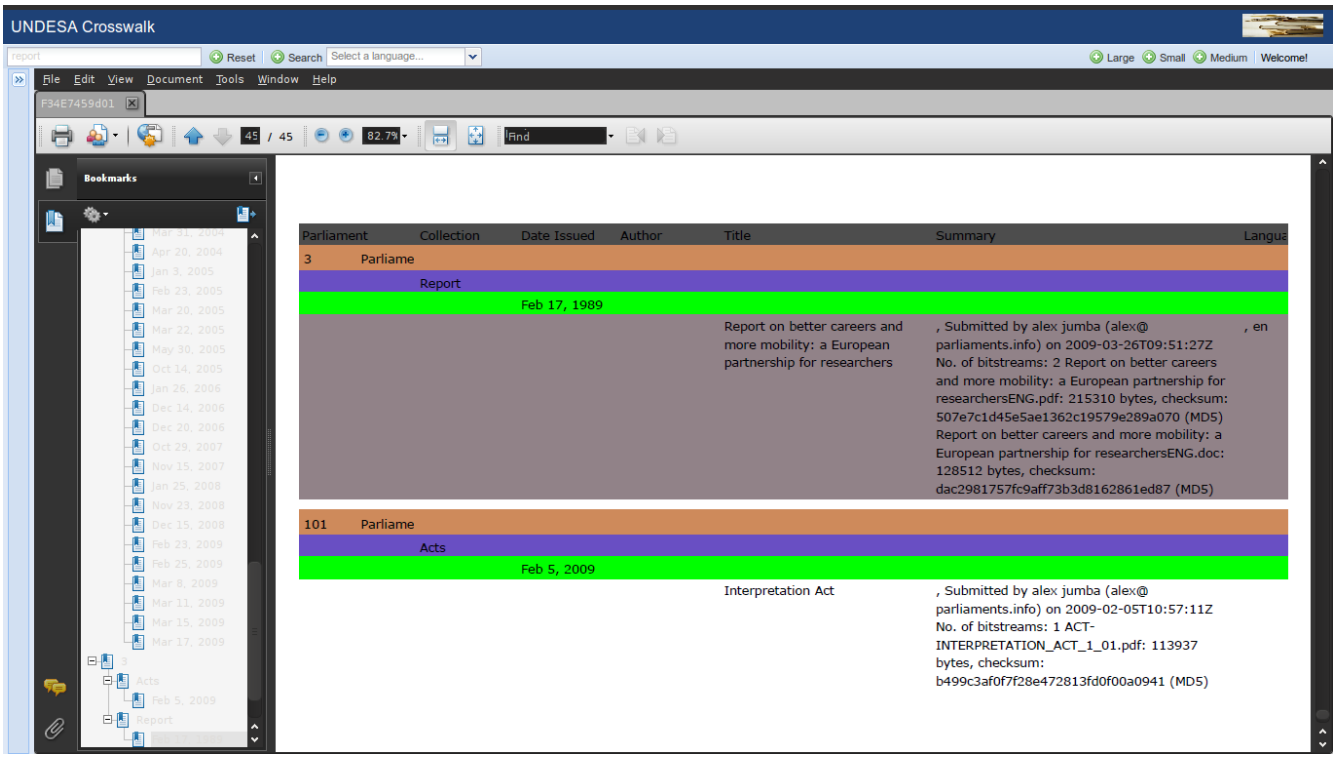
Tochi comes with several built-in reports. These are:

Report on Publications per parliament





Report on items available in the repository



As aWeb-Service

Tochi exposes several web-services that offer convinience to developers as a standard API for accessing tochi services without a user interface. The following are some of the web-services exposed:

- Tochi search
- Automatic Item submission into Dspace library
- Reporting functions

We'll go into details of each.

1. Tochi search
   This web-service leverages the opensearch API as well as Atom feeds, viewable by an Atom 1.0 compatible feed viewer as well as select browsers (Mozilla Firefox 2.0+). The following is the structure of the web-service url:

   Input
   /bungeni-opensearch?
   pagesize={page_size}&start={starting_offset}&field1=ANY&query1={search_term}

   Output:
   [os output]

<feed xmlns="http://www.w3.org/2005/Atom" xmlns:os="http://a9.com/-/spec/opensearch/1.1/">
<os:itemsPerPage>10</os:itemsPerPage>
<os:startIndex>0</os:startIndex>
<os:totalResults>136</os:totalResults>
<id>tag:alex.org,2007:/foo</id>
<title type="text">Test Feed</title>
<subtitle type="text">Feed subtitle</subtitle>
<updated>2009-07-13T14:23:28.758Z</updated>
<author><name>Alex Jumba</name></author>
<link href="http://alex.com" />
<link href="http://alex.com/foo" rel="self" />
<entry>
<published>2009-02-06T10:36:52.493Z</published>
<title type="text">Hansard 24/02/2004</title>
<summary type="text">, Submitted by alex jumba (alex@parliaments.info) on 2009-02-06T10:36:52Z
No. of bitstreams: 1
Hansard 24-FEB-2004.doc: 251392 bytes, checksum: be68e121adcd84b50aaebbc6c1aec4a5 (MD5)
</summary>
<id>123456789/48</id>
<content type="xhtml">

<dl class="xoxo">
<dt />
<h1 />
<dt />
<h2>Hansards</h2>
<dt />
<h3>
<ul class="bitstreams">
<li>
<a target="_blank" href="/crosswalk/bitstream/123456789/48/1/Hansard%2024-FEB-2004.doc">Hansard 24-FEB-2004.doc 245.5 kB Microsoft Word</a>
</li>
</ul>
</h3>
</dl>

</content>
</entry>
<entry xml:lang=", en ">
. . . . . .
</entry>
. . . . . .
</feed>

2. Automatic Item submission
   url: /crosswalk/bungeni-alternate
   This is a simple HTTP POST to the above url with the following parameters:
   collection: the collection in which the item will belong
   eperson: the eperson responsible for the item import
   dublin_core: the item metadata in the form of an xml file. Structure will be elaborated
   [binary files]: the bitstream files associated with the item. Can be multiple.

3. Reporting functions
   url: multiple as at now, will be coalesced in future.

Ubiquitous search

Tochi utilizes the opensearch protocols for universal searching across applications. The most visible advantage of leveraging the opensearch interface is the ability to perform a tochi search straight from the browser search box (Mozilla Firefox 2.0+, Microsoft Internet Explorer 7+).

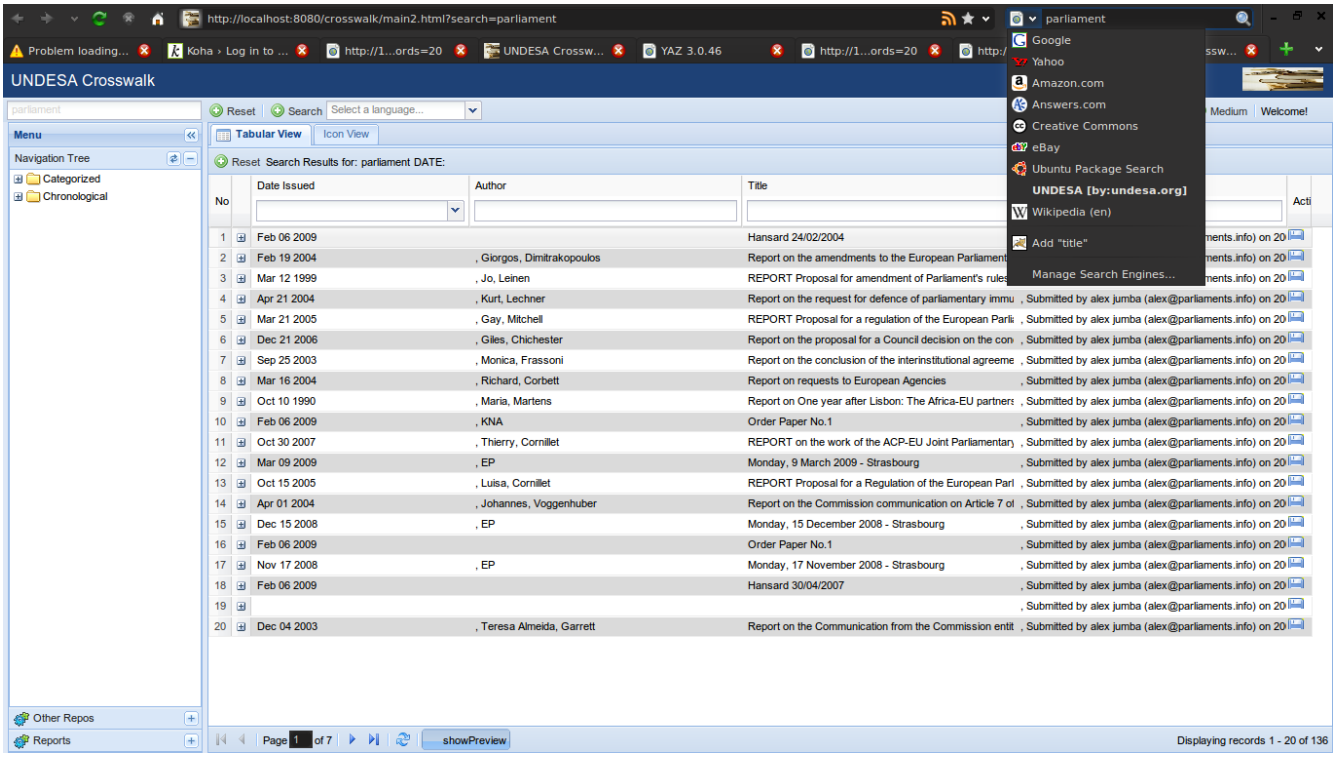
Configuring browser search

Mozilla Firefox:
<During Development>
Execute the following line in the terminal:

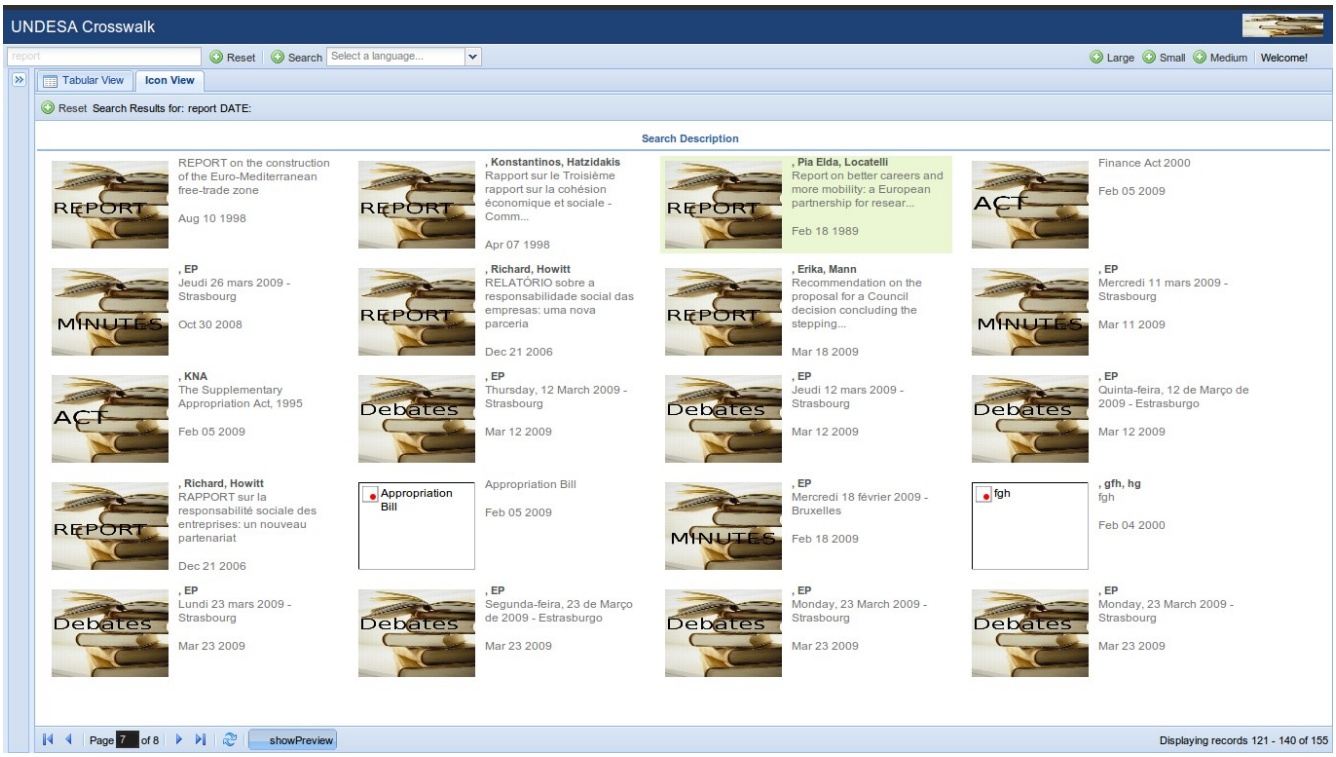
$sudo cp *TOCHI_HOME*/war/os.xml /usr/lib/firefox-3.5.1pre/searchplugins

The path may vary depending on the desired firefox version.

<Deployment>
To register this service with firefox, the web application must be attached to a web domain



Restart Mozilla Firefox.
The Firefox search box will now contain the crosswalk search as shown above.

Changing Collection icons

Collection icons are stored in the *TOCHI_HOME*/images/collection directory. Below is a screenshot with some icons missing viz. Bill