Reusing search statistics to enhance the use of library services

Morag Watson
Digital Library Development Manager
University of Edinburgh
Structure

- Background
  - Challenge
- Process
- Technology
- Tools
- Deliverables
Challenge

- How can we use this technology to encourage use of library services?
- How can we use this technology to present library services attractively?
Initial Thoughts

- Webcams
  - Views of the campus
  - Views of the city
- Images
  - Unique Items from Library collections
- Video
  - Graduation
  - Art Exhibitions
- and word clouds
Word Clouds (1)

- Library Web Site
  - Generate from top level pages
  - Regularly Updated

- Technology
  - Wordle
    - RSS / Atom Feed
    - Flat text file
    - Del.icio.us tags
Word Clouds (1)

- Library Web Site
  - Generate from top level pages
  - Regularly Updated

- Technology
  - Wordle
    - Use restrictions
    - Software Status
Word Clouds (2)

- Other Tools
  - Makecloud
  - TagCloud
  - TagCrowd
  - Etc. etc.
Your Tag Cloud is Ready!

Copy the code in the box below, then paste it into the HTML source of your blog or website where you want the tag cloud to appear.

```html
<!doctype html>
<title>Your Tag Cloud</title>
<style>
.tag-cloud {
  font-size: 12px;
  width: 200px;
  margin: 15px auto;
  padding: 5px;
}
</style>
      type="text/javascript">
</script>

Preview:
library edinburgh university special
irelande social donald turlough sites
resources sustaining masterpieces
museums stores fees divinity
catalogues collections able tag cloud.
Time for Review

- RSS / Atom feed
  - Only gave you word cloud based on ‘news’
- ‘Scraping’ a web page
  - Gives you that web page’s content
- Loading File
  - Snapshot of data at that point in time
Time for Review

- **Widgets**
  - Embedded on web pages
  - Performance
  - Reliability / Sustainability

- **Presentation**
  - Nothing beats wordle
New Vision

- Install open source word cloud tool
- Build our own word cloud tool

- Possible Technologies
  - PHP
  - JavaScript
  - HTML
  - JPG
New Vision

- Use search statistics from library services to drive word clouds
  - Library Catalogue
    - Voyager
  - Institutional Repository
    - DSpace
    - Edinburgh Research Archive
Voyager Data

- Opac search log tables
  - Turn opac logging on
  - Extract of data using sql
<table>
<thead>
<tr>
<th>DATE</th>
<th>ACTION</th>
<th>ID</th>
<th>CLIENT_IP</th>
<th>DBKEY</th>
</tr>
</thead>
<tbody>
<tr>
<td>16-APR-09</td>
<td>Webopac</td>
<td>20090416115517</td>
<td>Scotland in the twentieth century</td>
<td></td>
</tr>
<tr>
<td>16-APR-09</td>
<td>Webopac</td>
<td>20090416115517</td>
<td>Keyword Relevance Search</td>
<td></td>
</tr>
<tr>
<td>16-APR-09</td>
<td>Webopac</td>
<td>20090416115517</td>
<td>Left Anchored Title</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DATE</th>
<th>ACTION</th>
<th>ID</th>
<th>CLIENT_IP</th>
<th>DBKEY</th>
</tr>
</thead>
<tbody>
<tr>
<td>16-APR-09</td>
<td>Webopac</td>
<td>20090416115517</td>
<td>Scotland in the twentieth century</td>
<td></td>
</tr>
<tr>
<td>16-APR-09</td>
<td>Webopac</td>
<td>20090416115517</td>
<td>Keyword Relevance Search</td>
<td></td>
</tr>
<tr>
<td>16-APR-09</td>
<td>Webopac</td>
<td>20090416115517</td>
<td>Left Anchored Title</td>
<td></td>
</tr>
</tbody>
</table>
Voyager

- **Opac search log tables**
  - Lots of data
  - Lots of cleanup
    - Dates
    - Access method
      - Webopac
      - Z39.50
        - zzzzzzz
    - Search indexes
      - TALL,NALL,TKEY,NKEY,JALL,CALL, ISBN,ISSN
    - Boolean Search
      - AND, OR
    - Other Characters
      - , (,),"
'cleaned' voyager data
Voyager Word Cloud

- Automation
  - Cron job to extract data from opac search logs
  - Cron job to do data clean up
  - Transfer to server running PHP word cloud generator
  - Generation of word cloud
  - And refresh / restart
TagCrowd Word Cloud
Voyager Word Cloud

- **Dynamic Word Clouds**
  - extracts data directly from Oracle tables
    - Last 2000 items
  - clean the data
  - build the word clouds
  - handle presentation as HTML
  - Word cloud refreshed every time page is reloaded
  - Word cloud always pointing at real time data
return $catArray;

function &getCommonwords()
{
    $commonwordsString = 'the,be,to,of,and,a,in,that,have,i,it,for,not,on,with,he,as,you,do,at,this,but,hi,be,say,her,he,or,an,will,my,one,all,would,there,their,what,so,up,out,if,about,who,get,which,go,me,when,make,can'
    //...
Library Catalogue Search Terms

SCOTTISH EDUCATION  FINANCING  braun  romanze
nononsense  style  short  TWENTIETH  midday  L.A.
AUSTRALIAN  CAMBRIDGE  SERVE  EQUALITY
REGULATING
POLITICAL  PARTY  CAMPAIGN  DEALING  yehezekel
perspective  terrorism  phd  very  AUSTRALIA
HALSBURYS  STYLE  HISTORICAL  ELECTORAL
mesembrianthema  CENTURY  CALL  global
FUNDING  VOTES  DEMOCRACY  PHD  BRIbery
INTERNATIONAL  LAW  REALISING
EDUCATION
DSpace / Edinburgh Research Archive

- Modified database to collect search terms
- Data less dynamic
- Different types of searching
- Data clean up
  - Search syntax
  - Robots
  - OAI harvesting
(abstract:Statistical AND abstract:Theory AND abstract:Extreme AND abstract:Values)
(abstract:aspirin)
(abstract:cortisol)
(title:"Statistical Theory Of Extreme Values")
(abstract:test)
(abstract:alternative AND abstract:fuels AND abstract:electricity)
(abstract:Quantum AND abstract:Dots AND abstract:Theory)
Hurtado
(abstract:current AND abstract:opinions AND abstract:biotechnology)
(abstract:test)
(abstract:test)
(abstract:liquid AND abstract:desiccant)
(abstract:TX AND abstract:"ENVIRONMENTAL law" AND abstract:TX AND abstract:"ACTIONS & defenses" AND abstract:TX AND abstract:
(abstract:recession)
(abstract:hepatitis AND abstract:produce)
(author:T Furuya)
(abstract:conservation AND abstract:research)
(abstract:recession)
(abstract:forrester)
(abstract:Laboratory AND abstract:Techniques AND abstract:Biochemistry AND abstract:Molecular AND abstract:Biology)
(abstract:access AND abstract:surgery)
(abstract:Modeling AND abstract:Impact AND abstract:Initiation)
/* A map of the queries grouped by category. */
private Map<String, Integer> catMap;

/* A List version of the map in random order. */
private List<Map.Entry<String, Integer>> catList;

/* Just the unique values (i.e., categories) of the queries. */
private Set<Integer> valuesSet;

private int queryCount = 50;
private int catTotal = 5;

private Context context;

private static Logger log = Logger.getLogger(SearchQueryAnalyzer.class);

public SearchQueryAnalyzer(Context context, int queryCount, int catTotal)
{
    this.context = context;
    this.queryCount = queryCount;
    this.catTotal = catTotal;
Welcome to the Edinburgh Research Archive (ERA)

ERA is a digital repository of research produced at The University of Edinburgh. Here we present a selection of our best research including full-text digital Theses and Dissertations, book chapters, working papers, technical reports, journal pre-prints and peer-reviewed journal reprints.

If you are a member of the University of Edinburgh and would like to deposit your items, please send an email to prg-help@ed.ac.uk

If you believe that any material held in ERA infringes copyright, please contact prg-help@ed.ac.uk providing details and we will remove the work from the repository and investigate your claim.
Edinburgh Research Archive Search Terms

nature water control Synthesis vector management energy impact tower fasano human health space cancer mosquito borne solar gender computer agent GIS Heart cancer computers internet computation intelligent Med Age test gas wind aspirin eye database clinical care alessio journal disease web music celiac Clark Health plasma turbine nuclear Fasano springer Ageing
Holopro plus Word Cloud
What next?

- Integrated with Library services
  - Library website
  - Library Catalogue
  - Edinburgh Research Archive

- More Word Clouds
  - Aquabrowser
  - Webfeat
What next?

- Enhancement
  - Clickable search terms
  - More admin options
  - Browser support

- Service Comparison
  - Aquabrowser
  - Library Catalogue
Reusing search statistics to enhance the use of library services

Morag Watson
Digital Library Development Manager
University of Edinburgh
Morag.watson@ed.ac.uk