

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



Main Library Redevelopment Project



Latest progress

- News articles
- Monthly update
- Construction bulletin



The redevelopment

- Floor plans
- Image gallery
- Why redevelop
- Timeline



Project information

- Contact
- Health & Safety
- Project partners



Current News

None updates...



Vision

Read about the vision and project goals...



Environment

How will the redevelopment help the environment?





© 2008 Design: H&P (P) Robert W. Herbst

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

Main users
Freshers Ground
weekends extra Floor
Square Saturdays Medicine
October.Medicine Online.For next
Sundays.Conference Changes
open Edinburgh Veterinary
University August basis
hours self-service
five held 16-18 GeorgeWeek
Library

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:

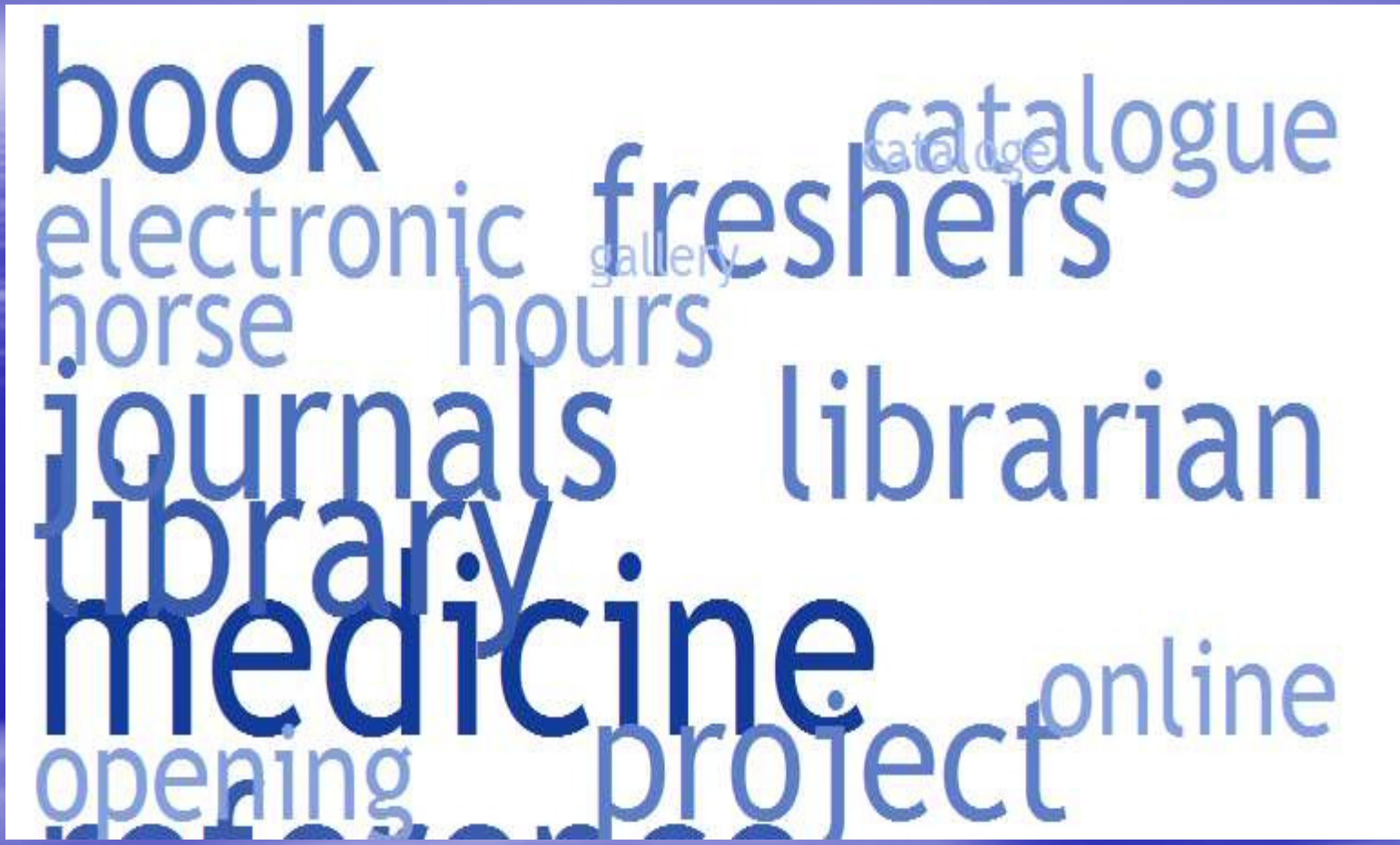
```
<!--tag cloud by makedoud.com--><script charset="ISO 8859-1" src="http://www.makedoud.com/doud2.php?url=http%3
```

Preview:

library edinburgh university special
irelande social donald turlough sites
resources sustaining masterpieces
museums stores fees divinity
catalogues collections david help tag cloud

A word cloud of library-related terms on a blue background. The words are in various shades of blue and sizes. The largest words are 'information' and 'library'. Other prominent words include 'online', 'research', 'resources', 'search', 'reference', 'opening project', 'medicine', 'journals', 'inter-library', 'librarian', 'loans', 'hours', 'homepage', and 'science'.

homepage hours
information inter-library
journals librarian
library loans online
medicine opening project
reference research
resources science
search

A word cloud on a white rectangular background, set against a blue gradient background. The words are in various shades of blue and sizes. The most prominent words are 'book', 'freshers', 'catalogue', 'journals', 'library', 'medicine', 'project', and 'librarian'. Other visible words include 'electronic', 'horse', 'hours', 'gallery', 'opening', and 'online'.

book
catalogue
freshers
electronic
horse
hours
gallery
journals
librarian
library
medicine
opening
project
online

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

Main users
Freshers Ground
weekends extra Floor
Square Saturdays Medicine
October.Medicine Online.For next
Sundays.Conference Changes
open Edinburgh Veterinary
University August basis
hours self-service
five held 16-18 GeorgeWeek
Library

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

Raw Voyager Search Data

I	R	H	HITS	S	C	CLIENT_IP	DBKEY
16-APR-09	webopac					20090416115517	Keyword Relevance Search
scotland in the twentieth century							
Y							

SEARCH_DA	STAT_STRING	SESSION_ID	SEARCH_TYPE
SEARCH_STRING			

L
-
LIMIT_STRING

I	R	H	HITS	S	C	CLIENT_IP	DBKEY
TYPE=as							
K	Y	N	1240	1	W	129.215.253.173	EDDB20020903165307

SEARCH_DA	STAT_STRING	SESSION_ID	SEARCH_TYPE
SEARCH_STRING			

L
-
LIMIT_STRING

I	R	H	HITS	S	C	CLIENT_IP	DBKEY
16-APR-09	webopac					20090416115517	Left Anchored Title
TALL SCOTLAND IN THE TWENTIETH CENTURY?							
N							

SEARCH_DA	STAT_STRING	SESSION_ID	SEARCH_TYPE
SEARCH_STRING			

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

french influence on scottish gardens

ESSAYS PRESENTED TO CHARLES WILLIAMS

ABOLITION OF MAN

MIND OF THE MAKER

williams charles

data acquisitions and processing in biology

data acquisition and processing in biology

comparative operating systems AND symposium

SELF-SERVICE LIBRARIES

SELF-SERVICE

self-service libraries

self-service libraries RFIDs

libraries AND self AND service

self service

international joint conference AND artificial intelligence AND proceedings

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

amour archaeology berheim century change comparative curr
developmental disorders education farming french gardens
german hills hiroshima history humphreys hunter-gatherers influence inquiry international
interpretation intepretation italian journal laing law libraries limit macrosocial malvern manual
perspectives planning pleadings primary re reproductive scotland
scottish self-service soir state style surrealist tradition twentieth update
vendredi

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

PHP

```
        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,his,by
    'we,say,her,she,or,an,will,my,one,all,would,there,their,what,so,up,out,if,about,who,get,which,go,me,when,make,can,
    'time,no,just,him,know,take,people,into,year,your,good,some,could,them,see,other,than,then,now,look,only,come,its,
    'think,also,back,after,use,two,how,our,work,first,well,way,even,new,want,because,any,these,give,day,most,us';

    $commonwords = explode(',',$commonwordsstring);

    // Todo : sort array into natural order to allow binary searching.

    return $commonwords;
}

/*
 * return an array whose elements are shuffled in random order.
 */
function &custom_shuffle($my_array = array()) {
    $copy = array();
    while (count($my_array)) {
        // takes a rand array elements by its key
        $element = array_rand($my_array);
        // assign the array and its value to an another array
        $copy[$element] = $my_array[$element];
        //delete the element from source array
        unset($my_array[$element]);
    }

    return $copy;
}
```


Voyager Word Cloud

Library Catalogue Search Terms

DEALING EDUCATION LAW COMPANION
terrorism CAMBRIDGE romanze global PARTY PhD Pat SERVE
HALSBURYS I.A. nonsense bruch INTERNATIONAL THEOLOGY
POSTMODERN very REALISING LAWS CAMPAIGN
yehezkel FINANCING SCOTTISH BRIBERY
DEMOCRACY guide AUSTRALIA short SYSTEMS
ELECTORAL VOTES mormonism CENTURY STYLE
TWENTIETH mesembrianthema CALL EQUALITY kol PERSPECTIVE
FUNDING lehmann nidre midday braun POLITICAL
AUSTRALIAN

Voyager Word Cloud (2)

Library Catalogue Search Terms

SCOTTISH EDUCATION FINANCING braun romanze
nononsense style short TWENTIETH midday LA EQUALITY
AUSTRALIAN CAMBRIDGE SERVE REGULATING
POLITICAL PARTY CAMPAIGN DEALING vehezekel
PERSPECTIVE JOURNAL terrorism phd bruch very AUSTRALIA
mesembrianthema STYLE HISTORICAL ELECTORAL
HALSBURYS CENTURY CALL global Pat DEMOCRACY PHD BRIBERY
guide FUNDING VOTES SYSTEMS mormonism LAW LAWS
REALISING INTERNATIONAL 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

DSpace Search Data

```
(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)
(title:Statistical AND title:Theory AND title:Extreme AND title:Values)
Hurtado
(abstract:Nguyen, AND abstract:Q. AND abstract:N.; AND abstract:Chavli, AND abstract:R. AND abstract:V.; AND abstract:
(abstract:current AND abstract:opinions AND abstract:biotechnology)
(abstract:test)
(abstract:test)
(abstract:sorghum, AND abstract:sweet AND abstract:sorghum, AND abstract:sorghum AND abstract:syrup, AND abstract:sorg
(abstract:liquid AND abstract:desiccant)
(abstract:TX AND abstract:"ENVIRONMENTAL law" AND abstract:TX AND abstract:"ACTIONS & defenses" AND abstract:TX AND ab
(abstract:recession)
(abstract:hepatitis AND abstract:produce)
(author:T Furuya)
(abstract:conservation AND abstract:research)
(abstract:recession)
(abstract:forrestal)
(abstract:Laboratory AND abstract:Techniques AND abstract:Biochemistry AND abstract:Molecular AND abstract:Biology)
(abstract:acess AND abstract:surgery)
(abstract:Modeling AND abstract:Impact AND abstract:Initiation)
```

Java

```
// A map of the queries grouped by category.
private Map<String, Integer> catMap;

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

// Just the unique values (ie 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;
}
```


ERA Home Page



THE UNIVERSITY of EDINBURGH
EDINBURGH RESEARCH ARCHIVE

Search ERA

[Advanced Search](#)

› [Home](#)

Browse

› [Communities
& Collections](#)

› [Titles](#)

› [Authors](#)

› [By Date](#)

Edinburgh Research Archive >

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.

Word Cloud of popular searches

We have generated a [word cloud](#) to visually depict what people are currently searching for in ERA. The size of the text in the word cloud

DSpace Word Cloud

Edinburgh Research Archive Search Terms

nature water control Synthesis vector management energy
impact tower fasano human health earth cancer mosquito borne space
solar gender computer agent gis Heart 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