



Advanced Google Searching CLIST Workshop 19th April 2012 14.00-16.00

John Nyman
St Mary's Campus Library
Imperial College London

Kate Clark
Royal College of Nursing

Learning outcomes – you will

- have reviewed how Google works
- understand the changing nature of Google
- be aware of some of the recent changes made to Google
- have tried out some techniques for undertaking more complex searches using Google
- be aware of some recommended alternatives to Google

Agenda

- Presentation (with some demos & discussion)
 - How Google works
 - Problems with using it
 - Tips for getting the best from it
 - Alternatives to Google
- Hands-on
- Feedback and discussion
- Action planning and evaluation

How does Google Work?

- Google's web crawling robot finds and retrieves webpages
- It passes them to the Google indexer
- The query processor compares your keywords to the index and retrieves relevant webpages



How does Google Work?



Your query retrieves what is in the Google database, not the real time web

Clicking on a result links to the website in real time

The >> displays the cache – an older stored version

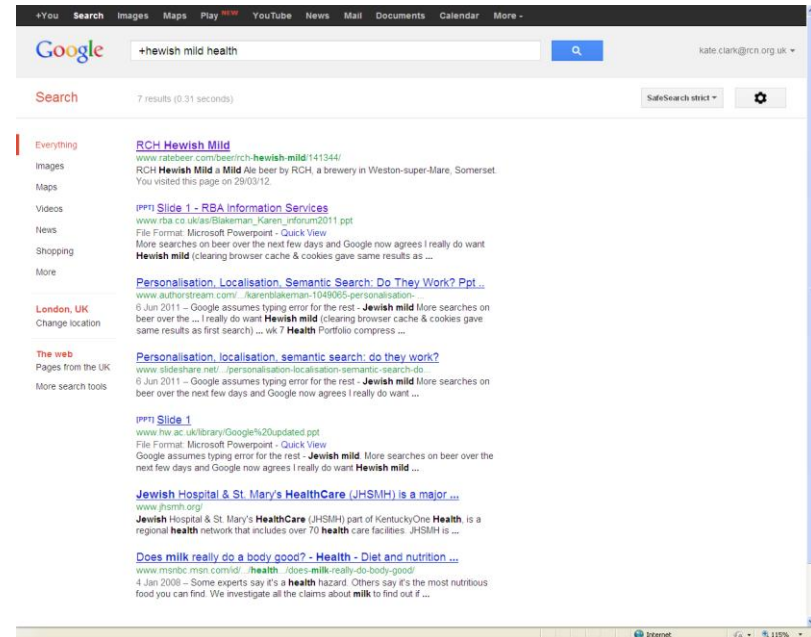
Why is the cached version useful?

- If there is:-
 - internet congestion
 - The site is down
 - The website has been removed
- The cached version will still be there
- If you get 'page unavailable' the website is either down or has been removed.
- The cache will still be there

“Each year Google changes its search algorithm up to 500-600 times.”

<http://www.seomoz.org/google-algorithm-change> Accessed 13/04/12

- + can no longer be used to force an exact match.
- Increasing attempts to personalise the service



“The searches you make on Google are usually improved to help minimize your time spent searching and get you the information you really want.

Here are some of the improvements Google makes:

- suggest **spelling corrections** and alternative spellings
- **personalize** your search by using information such as sites you've visited before
- include **synonyms** of your search terms to find related results
- find results that match **similar terms** to those in your query
- search for words with the same **stem**, like "running" when you search for [run]”

Taken from <http://support.google.com>, accessed 18/04/12

More on how Google Searches

- Google doesn't always search all your terms
- It will sometimes omit a word when the search retrieves no results to get something
- This soft ANDing is a way to get something, but you are not told this
- Unexpected results should alert you to think about changing your search strategy

Searching levels

- Basic
- Side Bar
- Advanced Search
- Command Line
- From experience you get to know which option is likely to retrieve the best results

Basic searching tips

- Imagine what you would like to appear in your ideal document and include those terms in your strategy
- Repeat key search terms in your strategy
- Change the order of your search terms as the order you put them in is given a higher weighting on what is retrieved
- “Speech marks” are used for phrases

Basic Search Examples

- *Find an Australian recipe that involves bananas peaches and pears for children*
- bananas peaches pears australia recipe (child OR kid)
- *A sculpture of Augustus John*
- “Augustus john” sculpture retrieves more accurate results than – Augustus John sculpture

Newish feature

- **Google side bar**
 - Options change, depending on tool selected

London, UK
Auto-detected

Everything
Images
Maps
Videos
News
Shopping
Books

London, UK
Change location

The web
Pages from the UK

Everything
Images
Maps
Videos
News
Shopping
Books
Places
Blogs
Discussions
Applications
Patents
Fewer




Any time
Past hour
Past 24 hours
Past week
Past month
Past year
Custom range...

All results
Sites with images
Related searches
Visited pages
Not yet visited
Dictionary
Reading level
Social
Translated foreign pages
Verbatim

Google side bar – most useful features... ...probably

- Verbatim
- Reading level
- Sites with images
- Images
- Videos

Advanced Google Search

- This option does not appear on the main screen. After a search the gear symbol appears which is the link . A link also appears at the bottom of the screen.
- This gives you various options to choose from, but does not allow you to use multiple field options, e.g. you can only choose one filetype from the drop down menu.
- The command line allows you to search various filetypes.

Advanced Search

Find pages with...

all these words:

this exact word or phrase:

any of these words:

none of these words:

numbers ranging from: to

To do this in the search box.

Type the important words: `tri-colour rat terrier`

Put exact words in quotes: `"rat terrier"`

Type OR between all the words you want: `miniature OR standard`

Put a minus sign just before words that you don't want: `-rodent, -"Jack Russell"`

Put two full stops between the numbers and add a unit of measurement: `10..35 kg, £300..£500, 2010..2011`

Then narrow your results by...

language:

region:

last update:

site or domain:

terms appearing:

SafeSearch:

reading level:

file type:

usage rights:

Find pages in the language that you select.

Find pages published in a particular region.

Find pages updated within the time that you specify.

Search one site (like `wikipedia.org`) or limit your results to a domain like `.edu`, `.org` or `.gov`

Search for terms in the whole page, page title or web address, or links to the page you're looking for.

Tell [SafeSearch](#) how much explicit sexual content to filter.

Find pages at one reading level or just view the level info.

Find pages in the format that you prefer.

Find pages that you are free to use yourself.

Advanced Search

Advanced search – most useful features - probably

- site or domain
- file type
- usage rights

Features only available from Advanced Search Screen

- language
- region
- safe searching
- usage rights

Command line tips

- Boolean Operators
 - AND – OR – NOT
- AND is the default – you don't have to enter it. All other operators you do
- OR or | (pipe)
 - Hiv (rnaï OR vaccine) Hiv (rnaï|vaccine|genetics)
- NOT use – (dash)
 - Diabetes insulin -obesity
 - Obesity high blood pressure -diabetes

More command line options

- Phrase Searching “ “ (Use Verbatim for multiple phrase searching to force this accuracy which may be lost in the command line search)
 - Obesity (“high blood pressure” OR hypertension) –diabetes
- Exact word = “ “ – replaces the + sign
 - “the” times or “the” onion
- Synonyms = ~
 - ~tutorial will find guides, documentation, introduction etc
- Any word = *
 - This is the night * crossing the border (is it mail or train)
- Changing the word sequence
 - This can improve relevancy

Focus the search

- Filetype - specific document type
 - filetype:ppt OR filetype:pdf OR filetype:doc
- Site - search within a specific site or domain
 - site:gov OR site:ac.uk
 - malaria site:www.who.int/en/
- Inurl - the term must appear in the URL
 - Inurl:media
- Year range - ..
 - 1990..1950
- In **Google News** - Location – country where the website originates
 - Elvis Presley location:de (Germany)

Think naturally

- what are the characteristics of...
- how much does...
- the author died on...
- who founded...
- a guide to...
- a checklist of...
- was created by...

Further thinking

- Be specific not general – this is different from medical database searching which is investigative and retrieves a relevant set of results
- An internet search is usually to find a specific piece of information
- You retrieve plenty of likely sites, so some browsing is required

Queries that Google finds difficult

- The population of Japan in 1490
- Why is the 'Old Man and the Sea' a Classic?
- Concepts that are related through judgment are not always found as keywords, making websites that reveal this relationship not easily searchable

Google Scholar

- Mainly journal articles
- You can set up Google Scholar to access your institution's password controlled e-journals for the full text
- Quicker route to journal articles than from Google
- Not all full-text, but does include the ever increasing free journals

Other resources 1

- Bing www.bing.com
 - reported to be often more up-to-date than Google by Karen Blakeman (personal communication 06/02/12).
 - Has option to restrict to pages from the UK
- Blekko <http://blekko.com>
 - spam free

Both recommended by Karen Blakeman, 06/02/12

Other resources 2

- DuckDuckGo <https://duckduckgo.com>
 - “less spam and clutter” (quote taken from their site)
- Zuula <http://www.zuula.com>
 - See results from 9 sites, each on a separate tab: Google, Bing, Yahoo, Gigablast, Exalead, Alexa, Entireweb, Mahala and Moheek

Both recommended by Karen Blakeman, 06/02/12

Other resources 3 - Internet Archive

- The WayBack Machine - <http://archive.org/web/web.php>
- You need to know the URL, e.g.
 - www.imperial.ac.uk/library
 - www.rcn.org.uk
 - www.theindependent.com

Authentic or Hoax?

- “*Every man should have a built-in automatic crap detector operating inside him.*” Ernest Hemingway
- 1 Search: **how does google work so fast**
- Evaluate: Google Technology
- 2 Search: **biological hazard detector**
- Evaluate: SKC Civil Defense Tools
- 3 Search: **biological hazard detector**
- Evaluate: McWortle Enterprises

Authentic or Hoax?

Google Our Search: Google Technology

[Home](#)
[About Google](#)
[Help Central](#)
[Google Features](#)

Our Technology
▶ [PigeonRank](#)

Find on this site:

The technology behind Google's great results

As a Google user, you're familiar with the speed and accuracy of a Google search. How exactly does Google manage to find the right results for every query as quickly as it does? The heart of Google's search technology is PigeonRank™, a system for ranking web pages developed by Google founders [Larry Page](#) and [Sergey Brin](#) at Stanford University.

PigeonRank System

Building upon the breakthrough work of [B. F. Skinner](#), Page and Brin reasoned that low cost pigeon clusters (PCs) could be used to compute the relative value of web pages faster than human editors or machine-based algorithms. And while Google has dozens of engineers working to improve every aspect of our service on a daily basis, PigeonRank continues to provide the basis for all of our web search tools.

Why Google's patented PigeonRank™ works so well

PigeonRank's success relies primarily on the superior trainability of the domestic pigeon (*Columba livia*) and its unique capacity to recognize objects regardless of [spatial orientation](#). The common gray pigeon can easily distinguish among items displaying only the minutest differences, an ability that enables it to select relevant web sites from among thousands of similar pages.

By collecting flocks of pigeons in dense clusters, Google is able to process search queries at speeds superior to traditional search engines, which typically rely on birds of prey, brooding hens or slow-moving waterfowl to do their relevance rankings.

When a search query is submitted to Google, it is routed to a data coop where monitors flash result pages at blazing speeds. When a relevant result is observed by one of the pigeons in the cluster, it strikes a rubber-coated steel bar with its beak, which assigns the page a PigeonRank value of one. For each peck, the PigeonRank increases. Those pages receiving the most pecks, are returned at the top of the user's results page with the other results displayed in pecking order.

Integrity

Google's pigeon-driven methods make tampering with our results extremely difficult. While some unscrupulous websites have tried to boost their ranking by including images on their pages of bread crumbs, bird seed and parrots posing seductively in resplendent plumage, Google's PigeonRank technology cannot be deceived by these techniques. A Google search is an easy, honest and objective way to find high-quality websites with information relevant to your search.

Authentic or Hoax?

World Leader in Sampling Technologies
Corporate Headquarters in the USA call 800-752-8472
863 Valley View Road, Eighty Four PA 15330 USA

Training Online Store Rentals Sales & Service Contact SKC

Headlines: Sample Bag Stability Data Green Building Testing Nanoparticle Measurement New! SKC Webinars/Podcast

Search the SKC Website

Homeland Security

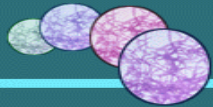
- [Important Links](#)
- [Emergency Assistance](#)
- [Products for Sampling](#)
- [SARS/Anthrax Sampling](#)
- [Lead Sampling](#)

Emergency Assistance

Business hours Monday through Friday:

SKC Inc. 800-752-8472	8:30 AM — 5:00 PM Eastern
SKC South 800-752-7684	8:30 AM — 5:00 PM Eastern
SKC Gulf Coast 800-225-1309	8:30 AM — 5:00 PM Central

Authentic or Hoax?



mcwhortle

About Us

Testimonials

Invest Now

Press

McWhortle Enterprises, Inc.

McWhortle Enterprises is an established and well-known manufacturer of biological defense mechanisms. Fortune 500 companies routinely use McWhortle Defense systems to protect their far-flung executives living in dangerous areas. These discreet, confidential safeguards have for years given employees and their families peace of mind.

Now, for the first time, McWhortle Enterprises is offering a product to the general public: the new **Bio-Hazard Alert Detector**. Running quietly on two double-A batteries, the Bio-Hazard Alert Detector emits an audible beep and flashes when in the presence of all known bio-hazards. The Bio-Hazard Alert Detector, measuring only 3 by 7 inches, is small enough to slip into a man's jacket pocket, a woman's purse or a child's backpack.

The Bio-Hazard Alert Detector works by detecting microscopic levels of hazardous bio-organisms and deadly virus organisms. It can detect even the finest-milled, weapons-grade biohazards from 50 feet, long before the risk of inhalation or cutaneous infection, by testing for the distinctive surface leptins. When equipped with a patented McWhortle computer micro-chip, the Alert Detector has a proven effectiveness of just .02 microns per cubic meter of air. The Bio-Hazard Detector gained instant acceptance in all test markets, which were spread throughout the Midwest. Within hours the product sold out of each store, with no advertising and only word-of-mouth endorsement within each community. Stock prices have soared as the markets have recognized the huge potential of the Detector for combating virus-borne illness world-wide.

Want to know more? [Click here.](#)



Website checklist

- Authority and Accuracy
- Purpose and Content
- Currency
- Design, Organization and Ease of Use

Summary

- Google is constantly changing.
- Check Google is doing what you told it to do
- Use natural phrases
- Be specific, not general
- You don't need to be over-complicated
- Use the different features for more accurate results – e.g. Maps – News
- Use Advanced Search and/or Side Bar and/or Command Line options for appropriate limits

Sites worth a look

- *Search now and search in the future* – David Russell 2012
Youtube - <http://www.youtube.com/watch?v=QDBhP7XTfTI>
- *Google Guide* - <http://www.googleguide.com/>
- Various commoncraft.com videos
- *Karen Blakeman's Blog* <http://www.rba.co.uk/wordpress/>
- *Phil Bradley's website: Making the net easier*
<http://www.philb.com/>