Using Google for academic research

Cathy Scutt – Bodleian Education Librarian
Ian Chilvers – Senior Library Assistant, SSL

Adapted from Karen Blakeman’s Google and Beyond
http://www.rba.co.uk/as/
https://creativecommons.org/licenses/by/3.0/
Search strategies

Start simple or straight for more complex strategies?

• Start with a simple search string
• Refine search with added commands and options

• Google does not use Boolean logic in the strict sense.
  • Commands allow you to mimic Boolean logic

• You will need to understand how different search engines work – especially Google
1. How Google ranks your search
2. Your search is personalized
3. “Google knows best”
4. Google has two indexes
5. Result replication is not guaranteed
1. How Google ranks your search

- Hundreds of criteria used, including:
  - Popularity, i.e. the number and type of links to a page
  - Where your search terms appear in a document and how often
  - The “authority” of a page

- You can change the ranking of results by:
  - Changing the order of the search terms
  - Repeating one or more of your search terms one or more times

- Personalization is an important part of how results are ranked and presented
Country versions and local information

- Country versions of Google give priority to local content

- Useful if you are researching a person, company, or sector in another country

- Go to the relevant country version of Google

- Google International Domains – a list of country and language codes
2. Your search is personalized

- Your search will be personalized based on:
  - Location
  - Past search history (from your Google account)
  - Past browsing history (local cookies)
  - Activity in other areas of Google (e.g. YouTube, Blogger, Images, &c.)
  - What other people have clicked on for similar searches
  - The device you are using
  - Content from contacts in your personal networks may be given priority (possibly)
What does Google know about you?

- Sign in to your Google account and go to: www.google.com/dashboard
- Check your ad preferences: www.google.com/ads/preferences
- If you do not have a Google account there will only be information about you in the right hand column
How to “un-personalize” your search

• Switch off web/search history
• Log out of your Google account
• Clear cookies

Or just use:
• **Use private/incognito browsing**
  • Firefox: Ctrl+Shift+P
  • Chrome: Ctrl+Shift+N
  • Opera: Ctrl+Shift+N
  • Internet Explorer: Ctrl+Shift+P

• Private browsing will not remove country personalization
3. “Google knows best”

Google doesn’t always respect your search terms

• Google will ignore terms if it thinks you do not have enough results.
  • Most of the time it will tell you
• “…” around terms, phrases, names, titles of documents does not always work
• Google will automatically look for:
  • Variations of words you have used
    • prevent, prevents, prevented, prevention
  • Synonyms of words you have used
    • prevent, stop, avert, block, halt, arrest, intercept
  • But you won’t know which variations and synonyms it is using
Google *doesn’t* always know best

- To force an exact match and inclusion of a term in search, prefix the term with **intext:**
  - *research methods* **intext:** *qualitative*

- Use **Verbatim** to do an exact phrase search

- Use **OR** to tell Google to use a particular synonym

- Google does not recognise **NOT**, but you can use the minus sign “-” to exclude terms
  - *-quantitative*
Missing terms

[Image: Screenshot of Google search results for "agricultural occupational asthma site:nhs.uk"]

- **Occupational Asthma Claim - Fentons.co.uk**
  - Claim Compensation For Asthma. Asthma At Work Claims Specialists.
  - [Case Studies](#)
  - [About Us](#)

- **Diagnosis, management and prevention of occupational asthma**
  - Concise guidance to good practice series. Diagnosis, management and prevention of occupational asthma. Setting higher standards. April 2012...

- **Occupational asthma - Royal Brompton & Harefield NHS Trust**
  - Occupational asthma is caused directly by work and is usually the result of an allergy to something inhaled in the workplace over a period of time. It is also...
  - Missing: agricultural

- **asthma - Health for Work Adviceline**
  - 3 Oct 2013 - Those working in construction, agriculture or the fishing industries, for example, are arguably more likely to be affected by occupational illness...
Using Verbatim
4. Google has two indexes

Google does not search everything in its database by default

- Google has two indexes
  - Main index
  - Supplementary index

- The supplementary index contains less popular, unusual and specialist material

- Supplemental index comes into play when Google thinks your search has returned too few results

- Verbatim and some advanced search commands seems to invoke a search in the supplementary index
Indexes: a comparison

“Normal search”
1,555,500 results

Search after Verbatim is applied
35,500,000 results
5. Non-replicability of results

Replication of results using the same search, at different times, by different people is not guaranteed

- Google changes its algorithms several hundred times a year
- Google tests new ideas for improving search on live searchers

We are all Google’s lab rats
USING GOOGLE SEARCH

1. Commands
2. Options
Commands: General

• ““ around phrases
  • “Metadata mega mess in Google Scholar”

• - to exclude a term
  • “research methods” qualitative – quantitative

• OR to specify variations and synonyms
  • oil OR petroleum
  • Sometimes it is better to run separate searches
Commands: General

• * to stand in for one or more words
  • solar * panels picks up solar PV panels, solar photovoltaic panels
  • * DOES NOT TRUNCATE

• intitle: to search the title of a webpage
  • intitle: “quantitative methods”
  • Similar to searching titles or abstracts in bibliographic databases
Commands: File types

Think about file formats

• PDF
  • For research papers, lengthy documents, government reports, industry papers

• PPT and PPTX
  • For presentations, tracking down an expert on a topic

• filetype:
  • zeolites environmental remediation filetype:pdf
  • "north sea" deep water drilling filetype:ppt
  • "north sea" deep water drilling filetype:pptx
Commands: Site search

- For searching large websites, or groups of sites by type for example: government, NHS, academic
  - agricultural occupational asthma UK site:nhs.uk
  - agricultural occupational asthma UK site:ac.uk
  - agricultural occupational asthma UK site:gov.uk

- You can also exclude sites using a minus sign
  - agricultural occupational asthma UK site:gov.uk
    "site:hse.gov.uk"

- N.B. site: does not seem to work with web archives such as
  http://www.nationalarchives.gov.uk/webarchive/
Commands: Numeric range

• Use for anything to do with numbers and quantities
  • years, temperatures, weights, distances, prices, &c.

• Type in your numbers separated by two dots

  • TV advertising spend forecasts 2014..2020
  • world oil demand forecasts 80..100 mb/d 2015..2030
  • toblerone 1..5 kg
Commands: Date range

• Restrict your results to a date range of your own specification
• Uses the Julian date format (ignoring fractions)

• For pages between June 20th and June 26th 2012 talking about the Statoil/Rosneft cooperation we would type:
  • \texttt{daterange:2456098-2456104 Statoil Rosneft}

• Julian Date Converter
  \url{http://aa.usno.navy.mil/data/docs/JulianDate.php/}
Date range – the easy way

There are third party tools for using daterange
Type your search in http://gmacker.com/web/content/gDateRange/gdr.htm then apply Verbatim
Options: Date

- Restrict your results to information that has been published within the last hour, day, week, month, year or your own date range
- Does not work with Verbatim
  - Use `daterange:` instead
Options: Reading level

- Changes the type of material that is returned
- Nothing to do with publishers assigned reading age
Reading level: a comparison

Reading level: Basic

Reading level: Advanced
1. Things to consider
2. Commands

GOOGLE SCHOLAR
Google Scholar

http://scholar.google.com/

“Google Scholar provides a simple way to broadly search for scholarly literature. From one place, you can search across many disciplines and sources: articles, theses, books, abstracts and court opinions, from academic publishers, professional societies, online repositories, universities and other web sites. Google Scholar helps you find relevant work across the world of scholarly research”.
Google Scholar

- Does not cover all key journals in all subjects – no source list
- Top publications for subjects and languages under Metrics link on home page or http://scholar.google.co.uk/citations?view_op=top_venues&hl=en
- Scholar indexes the full text but you may have to pay to view the whole article
- Groups different versions of an article together
Google Scholar

- Includes open access material, pre-prints, institutional repositories (but not necessarily author self archived papers on personal websites)

- Includes material that is NOT peer reviewed but is structured and looks like an academic article (title in large font, authors, affiliations, abstract, keywords, citations)

- Pre-prints and IR copies may differ from final published version – charts and images may be redacted because of copyright restrictions
Find it @ Oxford

Library Links

- Settings > Library Links
- Adds a [Find it @ Oxford](#) link for results you can access through library subscriptions
Google Scholar

- Does NOT use the publishers’ metadata
- Sometimes gets the author wrong
- Beware the advanced search screen and commands
  - Date and author search looks in the area of the document where those elements are usually found
  - Page numbers, part of an address, data item may be mistaken for publication year
Commands

- + sign before a search term to force an exact match
  - +norne

- intitle: to search for a single word in the title of the document (rather than the web page)
  - intitle:zeolites environmental remediation

- allintitle: to search for all of your terms in the title of the document
  - allintitle:zeolites environmental remediation
Commands

• author: to search on an author’s name
  • zeolites environmental remediation
    author: rhodes

• Commands can be combined for a precise search
  • author: wolford site: psu.edu
    allintitle: “marcellus shale”
ALTERNATIVES TO GOOGLE

1. Bing/Yahoo
2. DuckDuckGo
Bing/Yahoo

• Yahoo now uses Bing’s database, commands and ranking algorithms
  • Yahoo Finance still available

• No advanced search screen on Bing - use commands
    • filetype:
    • site:
    • inbody:
    • intitle:

• AND, NOT, OR parentheses for complex Boolean searches

• near:n where n is a number, specifies that the terms must be within that number of words of each other and in any order
Bing/Yahoo

- Results seem to be more consumer/retail focused
  - More ‘shopping’ than research
  - Results improve as soon as you start using the advanced search commands
- Sometimes more up to date than Google
  - Updates sites more frequently
  - Adds new sites more quickly
  - Useful if you are looking for information on a new company or organisation
- Many features and options available to US users only
  - Changing location and version of Bing does not always work
  - Using anonymous proxy does not always work
Is Bing dropping search terms?

- Bing may be dropping terms from your search if you are getting very few or no results

- Solution:
  - use AND between your terms (AND must be in capital letters)
  - prefix an essential term with inbody:
    - `inbody:aardvark`
Bing It On

www.bingiton.com

Statistics - Occupational asthma - HSE: Information about health ...
www.hse.gov.uk/statistics/causdis/asthma/index.htm
Occupational Asthma. During the last decade, our data sources suggest there has been an overall decrease in occupational asthma. However, new cases continue to occur...

Occupational Asthma in Great Britain in 2013
www.hse.gov.uk/statistics/causdis/asthma/asthma.pdf - PDF file
Health and Safety Executive: Occupational Asthma in Great Britain 2013. Contents . Summary 2 Introduction 3 Overall scale of occupational asthma 3

Oasis and Occupational Asthma
www.occupationalasthma.com/default.aspx
The Shield scheme reports some statistics for occupational asthma. ... BOHRF shows the evidence based occupational asthma guidelines for the UK

Occupational asthma: measures of frequency from four countries
www.ncbi.nlm.nih.gov/pmc/articles/PMC1090684
Department of Epidemiology and Medical Statistics, London Hospital Medical College at QMW, London, UK

Does the introduction of exposure limits reduce the incidence of...
www.occupationalasthma.com/occupational_asthma_pageview.aspx?id=5012
The OM/OMR-046 working scheme for occupational lung disease in the UK has...
DuckDuckGo

http://duckduckgo.com/
- Does not track, does not personalise

- Results are a compilation of about 50 sources including Wikipedia, Wolfram Alpha, Bing, Blekko and its own Web crawler DuckDuckBot.

- Advanced search commands include:
  - site: [though it seems to be ignoring it now]
  - inbody:
  - intitle:
  - filetype:
  - sort:date to sort by date (uses results from Blekko)
  - region:cc (e.g. de) to boost a country

- Does not support use of * for truncation

- DuckDuckGo Syntax
  http://help.duckduckgo.com/customer/portal/articles/300304
Questions?

Check with your subject librarian for their subject specific advice: 
http://www.bodleian.ox.ac.uk/subjects-and-libraries/subjects/librarians