Google for academic research

Catherine Scutt and Tom Dale
The Bodleian Libraries

Google cannot search the deep or dark web, it can get to about 4-6% - that is still a lot!

Adapted from Karen Blakeman’s Google and Beyond http://www.rba.co.uk/as/
https://creativecommons.org/licenses/by-sa/2.5/

This presentation also available via a Creative Commons Attribution-ShareAlike License
Search strategies

Start simple or go for a more complex strategy such as you would use in a library literature database?

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

• Google does not use Boolean logic and offers different options to the library literature databases you may be used to

• To get Google to show you what you want, you will need to try to understand how it works – but give up now if you hope to fully understand what Google is up to!
How Google ranks your search

• Hundreds of criteria used, including:
  • Where your search terms appear in a document and how often
  • Popularity, i.e. the number and type of links to a page AND what has been clicked on before
  • The “authority” of a page (as decided by Google?)
  • What you *meant* by your query – including using *machine learning* to figure out what you meant (RankBrain)
  • Personalisation is an important part of how results are ranked and presented (more on this in a moment)

• 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
Your search is personalised

- Your search will be personalised based on:
  - Location
  - Past search history (from your Google account)
  - Past browsing history (from local cookies)
  - Activity in other areas of Google (e.g. YouTube)
  - The device you are using (e.g. phone v. laptop)
    - Although web pages not suitable for mobile are not showing in search results even on laptops
  - Search history of your contacts
How to “un-personalise” your search

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

Or just **Use private/incognito browsing**
• Firefox ; Internet Explorer: Ctrl+Shift+P
• Chrome: Ctrl+Shift+N
• Safari: Cmnd+Shift+N

• Private browsing will not remove country personalisation and will not hide your IP address.

Further reading: [https://www.wired.co.uk/article/google-chrome-incognito-mode-privacy](https://www.wired.co.uk/article/google-chrome-incognito-mode-privacy)  
[https://www.wired.co.uk/article/how-to-delete-google-search-history-tracking](https://www.wired.co.uk/article/how-to-delete-google-search-history-tracking)
Country versions and local information

• Country versions of Google give priority to local content – so this could potentially be useful if you are researching a person/company/sector in another country.

• Google not keen on you doing this but you can try:
  • Settings > Advanced search > Narrow your results by… language
  • Settings > Advanced search > Narrow your results by… region
  • site: command, e.g. site:.no ; site:.fr
  • using something like TOR or a VPN service to lose the country you are actually in and see the results as if you were in another country

• note the EU Court of Justice ‘right to be forgotten’ ruling
  http://googlepolicyeurope.blogspot.co.uk/2016/03/adapting-our-approach-to-european-right.html

• GDPR – access to some sites from outside Europe blocked or restricted whilst they catch up with our data protection laws
“Google knows best”

Although it is a good idea to apply some of the search techniques from library literature databases to your Google search be aware that Google will not always respect your search strategy.

For example, adding in further terms to narrow down your topic is a good idea, but:

- Google will ignore terms if it thinks you do not have enough results.
  - Most of the time it will tell you
Missing terms

Osney Mill - Wikipedia
Osney Mill was a flour mill on a branch of the River Thames in Oxford, England, located south of the Botley Road, down Mill Street. While the mill was destroyed ...

Osney Lock Hydro
www.osneylockhydro.co.uk/
Welcome! Osney Lock Hydro is the first community owned hydro scheme to be built on the Thames. The hydro is now fully operational, generating clean green ...

Images for osney lock flour mill

[Images of Osney Lock Hydro location]
“Google knows best”

Exact phrase searching is just as useful in Google as it is in the library literature databases
- “…” around terms, phrases, names, titles of documents, but 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, and it gives you different numbers of results and differently ranked results depending which synonym you use
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 search**
  - **Tools > All Results**
    - Occasionally doesn’t quite work, nothing in Google is definite!
Google *doesn’t* always know best

- Google does not recognise NOT, but you can use the minus sign “-” to exclude terms
  - `quantitative`
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. Artificial intelligence developing & improving Google’s algorithm all the time

• Google tests new ideas for improving search on live searchers by sending you to different versions of Google without you realising
1. Commands
2. Options

USING GOOGLE SEARCH WITH YOUR OWN RESEARCH QUESTION
Commands: General

• " " around phrases
  • "Metadata mega mess in Google Scholar"

• - to exclude a term
  • "research methods" qualitative -quantitative
Commands: General

- * to stand in for one or more words
  - E.g. solar * panels picks up solar PV panels, solar photovoltaic panels
  - E.g. european * policy picks up european education policy, european health policy, european privacy policy etc.

  * DOES NOT TRUNCATE!

- intitle: to search the title of a webpage
  - intitle:“quantitative methods”
  - Similar to searching titles or abstracts in bibliographic databases, although much less reliable! The title of a web page is what you see in the tab header, or if you bookmark the page

- inurl: to search for your terms in the url
Commands: File types

Think about file formats

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

• ppt or pptx
  • For presentations, tracking down an expert on a topic

• xls or xlsx
  • For spreadsheets containing data

• filetype:
  • “inclusive education” research filetype: pdf
  • “inclusive education” research filetype: ppt
  • “inclusive education” research 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
  • agricultural occupational asthma UK site:www.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 work with web archives such as http://www.nationalarchives.gov.uk/webarchive/
Date searching – 2 options

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

Choose to search for results before or after a certain date. Use format **YEAR/MONTH/DATE**

- **Does** work with Verbatim
OTHER SEARCH ENGINES WORTH A LOOK

1. Bing
2. DuckDuckGo
3. Swisscows.ch
Bing

• Similar commands to Google
  • filetype: (automatically searches for .pptx as well as .ppt, unlike Google)
  • site:
  • inbody: \[doesn’t always work, sometimes searches for ‘body’!\]
  • intitle:
  • List at Advanced Operator Reference

• AND, NOT, OR parentheses for complex Boolean searches
  • Therefore Bing does not have a Verbatim option
  • \textit{However, Boolean searches do NOT always work in Bing}

• Cached versions highlight your search terms (Google does not highlight any longer)
Bing

- Results seem to be more consumer/retail focused
  - More ‘shopping’ than research initially
  - 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
Bing

- Ordnance survey is an option on Bing maps

- Country versions
  - Settings > COUNTRY/REGION
  - Settings > LANGUAGE > Search Language > Limit my searches to pages written in specific languages:

- Date limit similar to Google

- Images and videos a good alternative (includes YouTube)
- Bing has close links with Facebook, LinkedIn – may be better for people searches

- Many features and options available to US users only
What is Bing actually doing?

- Bing may be dropping terms from your search if you are getting very few or no results.
- Possible solution:
  - use AND between your terms (AND must be in capital letters).
- The order of your search matters not just to how your results are ranked but also to the amount of results it finds.
  - e.g. (evaluation OR assessment) AND (learning OR teaching)
    - 56,100,000
  - e.g. (evaluation OR assessment) AND (teaching OR learning)
    - 54,900,000
DuckDuckGo

• Does not track, does not personalise

• Recognises some Boolean AND and NOT (OR doesn’t work!)

• Advanced search commands include:
  • site:
  • inbody:
  • intitle:
  • filetype: (good as it automatically searches for .xls as well as .xlsx)
  • region:## (e.g. de) to boost a country
  • !tw searches Twitter for you. Other site searches possible: https://duckduckgo.com/bang
• A different kind of search command you may enjoy is:
  • https://duckduckgo.com/bang

• !tw searches Twitter for you
• !solo to bring a book up on SOLO
• !mmol to search for a medieval manuscript.

  • Let’s try !solo Lord of the Rings
  • Let’s try !mmol MS Bodley 860

• Lots of other site searches possible, e.g. Oxford English Dictionary etc.
DuckDuckGo

- DuckDuckGo now makes its money from advertising (based on what search terms you have just typed in, not tracking/personalising) but they try to keep it to a minimum and you have the option to disable ads: https://duckduckgo.com/settings
Swisscows.ch

- Swisscows doesn’t store IP addresses
- Swisscows doesn’t store tracking cookies
- Swisscows doesn’t store user queries

- uses artificial intelligence and machine learning to analyse your search
- select a region to get more relevant search results

- sourcing its results from Bing (the Microsoft search engine)
Links to keep you up to date

- Karen Blakemans’ blog http://www.rba.co.uk/wordpress/
- Phil Bradley’s website http://www.philb.com
- http://searchengineland.com
- https://tacticaltech.org/

Book suggestions