

Transcript: Google for academic research

Full presentation

Part 1: Introduction

Google is often dismissed as being irrelevant to the academic literature search. However, changes to the nature of the dissemination of scholarly research and official information in recent years mean that a wealth of relevant information can be accessed via standard Internet search engines such as Google. Too often however such information is lost amongst thousands of irrelevant, spurious and misleading results.

Based on an understanding of how the Google search engine works this session aims to show you some techniques to quickly filter your results for high quality academically relevant material.

A couple of other search engines will be considered briefly, including DuckDuckGo which has become a popular alternative to Google in the light of concerns about the privacy costs of using search engines such as Google or Bing.

Much social science research, and indeed research undertaken in many other disciplines, draws on literature beyond the traditionally published journal article, (although these are of course, key!). Certain types of documents such as government policy, or statistics collected by charitable organisations, or research projects done with schools, need to be sourced from the free web rather than retrieved from library literature databases.

So, yes, I'm afraid that means yet more places to look and more search skills to master.

When we talk about internet searching these days we generally mean Google. We'll talk a little more about alternative search engines later on.

You may be surprised just how much of the internet Google does NOT search. You may have heard of the dark or deep web. One estimation is 4% of the internet can

be searched by Google as opposed to about 90% which is password protected or only available by a paid subscription, leaving about 6% of the internet for the more dubious or illegal content.

The point I wish to make is that not everything is freely available online.

Note that we can and do access some of the deep web content is contained in library subscription databases.

The other thing I really want to stress is that this session is definitely not about using Google instead of SOLO. However, for some types of literature, it is a really good and necessary tool.

For now we need to accept that Google is the most popular search engine, most people's 'go to' choice, and I think that is for a good reason – it is a very clever search engine.

However, to use Google in your academic research, it is important to understand how it works, as far as it is possible to see this. Google plays its cards close to its chest, much of how it works is proprietary information so they are not very transparent about how it all works.

Much of our knowledge in this area comes from a UK guru on Google, Karen Blakeman, who has kindly shared her wisdom with us. Google changes and develops all the time so we rely on checking in with her up to date expertise to keep clued up to how Google is working at the moment. There is a link to her website on slide 2 so you can keep up to date yourself.

Part 2: Getting to know Google

Can you treat Google as you would a library literature database?

To some extent, yes, developing your search strategy will draw on the techniques and ways of thinking you have already developed for successful library searches. However, you really need to understand what Google is actually up to before you can get what you need from it.

For a start, Google does not use Boolean (the standard search logic used by databases and search engines) in the strictest sense, but you will see that there are commands you can use in Google allowing you to, sort of, get similar results.

So (somewhat differently from the advice we give when discussing search strategies for library literature databases) with Google you start with a simple search string then, rather than add in further search strings to create an advanced search, instead, with Google, you stick with a simple search string, but add in some of the tricks & techniques we will be covering today.

Because Google will not use a complex search string of clever Boolean and nested terms etc. you cannot cover all your bases with one clever search- you will probably have to do a few different searches within Google to make sure you've covered everything.

Indeed, you probably don't even want to limit your searches just to Google. Different search engines work in different ways and therefore although one is not necessarily better than the other they can bring you different sets of results.. more on this later...

BUT, give up on any hope of fully understanding what your Google searches are doing

And also give up on any hope of claiming a *comprehensive* or *systematic* search. With Google the best you can hope for is a *practical* search. You'll see what I mean ..

Let's have a look together now to try to tease out some of the issues:

www.google.co.uk

Search for: **Research methods**

Simple keyword searching like this will find too many irrelevant poor quality results.

In many cases not from academically relevant sources..

We actually will not see many adverts on these university computers, but do a search at home or on café wi-fi and I am sure you will.

Google is a business, it exists to make a profit and is very successful at doing so through targeted advertising.

Part 3: Personalisation in Google

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) [more than half of Google's searches happen on mobile and it is not always presenting the same content as it would on a laptop or desktop.]

Also note that Google ranks mobile friendly sites higher in results whichever device you are using for your search

Search history of your contacts

Why? Ostensibly to present you with a more personally relevant set of results but it isn't just for your benefit – it is so they can target advertising more successfully. And this works, they make incredible amounts of money and most (90%) of this profit comes from their successful use of advertising.

As we are on academic network Google recognises this and actually strips out some of the ads so you'll see less ads here than on a home or in a café.

You can see this very clearly when you do a non-academic search, for instance

Cinema

You can see it is showing me a map of my local area and listing local cinemas.

Could be useful for ordinary searching, up to you how much you use and appreciate the personalisation, or find it creepy and avoid it.

It is unsettling though to think that academics in different places and with different hobbies could do a literature search for a project they are working on together and get different results, or at least differently ranked results – which given the sheer numbers we are talking about pretty much amount to the same thing.

Even if you have no Google account, if someone has sent you an email ever from their gmail, then Google will know about you...

You can un-personalise your search:

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

Or simply 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.

One approach that I see many academic colleagues adopting, is to deliberately separate your normal and academic searches and to open a separate window for academic searching which has no personalisation going on.

Chrome helpfully explains what is going on:

You've gone incognito. Pages you view in incognito tabs won't stick around in your browser's history, cookie store, or search history after you've closed **all** of your incognito tabs. Any files you download or bookmarks you create will be kept.

However, you aren't invisible. Going incognito doesn't hide your browsing from your employer, your internet service provider, or the websites you visit.

i.e. Incognito browsing doesn't use cookies or search history (so is good because it gives you a neutral search and doesn't save any history) However it will not remove country personalisation and will not hide your IP address from anyone (the university/employer will still know!)

Part 4: Locations in Google

Country versions of Google give priority to local content - this could be useful if you are researching a topic in a country other than the country you are based in.

However, Google does not make it easy to use country versions.

Notice that if you navigate to google.com or google.fr, for instance, you still see 'United Kingdom' in the footer.

Your results will not be the same as if you were actually in the USA or France using Google.

To emulate using a country version of Google, 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 mask the country you are actually in and see the results as if you were in another country

Tor's method of randomly sending the connection through many servers throughout the world means it is slower than you are used to, and it will bounce you around various country versions of Google, so interesting in terms of playing with country bias results.

A **VPN** is a service that keeps your web browsing secure and private over public Wi-Fi and it is also used to create a secure connection to another location, thereby allowing you to appear as if you are in another place. E.g. the university's VPN

makes it seem as if your laptop is on campus so allows you to access drives on university computers and to access the online library without having to log in so much. The Opera or Vivaldi browsers have an optional VPN (limited country options) and there are many free and cheap VPN services available - you can choose where to be so note, if you use a VPN then the version of Google you end up using will be the country it thinks you are in.

These options can be useful if you want a country bias, just not the country you are physically located in. **It is not possible, as far as I know, to get rid of the country bias altogether.**

Using an alternate search engine such as DuckDuckGo is the best way to ditch Google's country bias, but even DuckDuckGo has regional choices in a drop down.

We noticed once that with Verbatim selected, Google was showing "unknown location" at the foot of the screen. This does not work every time though – certainly not on a library staff computer – so yet another of Google's mysteries.

Part 5: Some search techniques

Google knows best. Or does it?

Google does not always respect the search terms you are putting in.

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 by noting the missing word/s underneath each result (but it doesn't always!)

Exact phrase searching is useful in Google, just as it is in library literature databases. Use double quotation marks around terms, phrases, names, or titles of documents to force the search engine to give you results with those terms directly next to each other (but note that as with everything in Google, although it works most of the time, it doesn't always work!)

As when developing a search strategy, consider alternative terms and variations on the search terms you are using such as singular and plural. Google will try to help you out here by:

- making automatic spelling corrections
- looking for both British and American spellings for you
- searching for various variations of your search terms including singular/plural versions of the word you used

Note that the asterisk * has a very different function in Google, so do not try to use this for truncation!!

The ranking is affected by which of your alternative terms you choose to use, switch to a more academic sounding synonym and the results should re-order to show more academic results nearer the top. Indeed, some results may even be missing.

Google initially achieved this by referring to an English language thesaurus but this has got more sophisticated using artificial intelligence to develop its own thesaurus. It now uses trial & error and sees what people end up clicking on. For example, it could figure out that locomotive was an alternative term for steam engine by learning from people's searches and clicks.

A recent search I did, for instance, I could see that Google was showing me variations on the term 'assessment' as I could see some results with 'assess' or 'analysis' or 'appraisal'.

You may be used to using * to truncate your search terms.

The asterisk * is used very differently by Google (because it is already truncating for you).

Google uses * to stand in for one or more words. e.g. european * policy picks up: european education policy, european health policy, european privacy policy etc.

Part 6: More search techniques

Google doesn't always know best.

If you want to stop Google being so clever with its synonym, plural, and alternative terms searching, use the **intext:** command

Intext: will work with phrases.

Unlike in a library database, you cannot be 100% sure that Google is paying attention to your request for an exact phrase search and your efforts to narrow your search. Google thinks it knows best, and is always keen to bring you back as many results as it possibly can. So, if it doesn't think there are enough results coming from what you've typed in, it will start ignoring quotation marks and dropping some of your search terms (sometimes shows you this underneath, not always), seemingly at random and you won't always realise its doing this. It sometimes also searches for variations of words within your quotation marks. If you have gone to the time and trouble of developing a good search strategy, you will also want to make sure that Google is actually paying attention to this.

To ensure that only your search string is being using to generate results turn 'verbatim' on.

You can find this by looking underneath Google's search box for:

Tools – all results – verbatim.

NB. With verbatim on, to see amount of results simply click on 'search tools' –this toggles the amount on and off.

Verbatim is not completely reliable – it sometimes still tries to search for alternative spellings or remove search terms if there aren't enough results.

Google does not recognise **NOT**, which is a piece of search logic you may have used successfully in library literature database, but you can use the **minus** sign to exclude terms.

Simply place the minus sign directly at the start of the word you wish to exclude.

It can be really useful to get Google to show you only PDFs or only PowerPoint presentations as results. If you think about it, many of the academically relevant literature you will find useful for your research may very well be in a PDF format, or shared at conference as a PowerPoint perhaps.

You can find this option in the advanced search, or simply add file type colon (**file type:**) directly before your search terms.

For searching large websites, or groups of sites by type, for example, government, NHS, or university websites, use the '**site:**' command.

You can also exclude sites using a **minus** sign.

N.B. **site:** does not work with web archives such as <http://www.nationalarchives.gov.uk/webarchive/>.

These are part of the deeper web as the contents sit behind a search.

There are a couple of options for date searching. The first is to look in Tools and to use the **Anytime** menu to restrict your results to information that has published within the last hour, day, week, month, year; or you can also set your own custom date range.

This technique does NOT work with verbatim switched on, you have to choose one or the other.

The second date search option is another command which you can include in the search box. This DOES work with verbatim.

Part 7: Replicating Google searches

There is no guarantee that the same search will present the same results – either by different people on the same day, or by the same person a few days apart etc. Why?

Google is constantly updating the way it searches. New algorithms are released a couple of times a week on average.

Google is also always testing these new ideas on live searchers, and you'd never know whether you are currently being tested on or not - you may be searching in one of Google's 'sandbox' environments without realising it.

Also, Google is using artificial intelligence or machine learning to develop and improve all the time. This is taking over from the sandbox approach to development.

This means that even if you do exactly the same search yourself, the Google you are searching in is not the same as it was when you last did the search.

Part 8: Using Google with your own research question

In the course of explaining how Google works, I have mentioned various tricks to try. Now is a good opportunity to work with your own research question to try out various tips and tricks within Google to help you narrow your results to those that will be academically useful for you.

Download the accompanying workbook and use this alongside re-watching the previous parts to try out these Google techniques to find academically relevant literature for your own research.

Part 9: Alternatives to Google – Bing and DuckDuckGo

Google often seems ubiquitous and has even entered the dictionary as a verb, but there are still alternatives out there. When searching the internet for academic

literature it is a good idea to go beyond Google as you may be presented with different results.

So, I would see these as 'additions' to Google, rather than 'alternatives'.

I am only going to talk about 2 other search engines, but there are plenty more out there for you to explore.

In some ways Bing is actually more akin to the library literature database – it does recognise Boolean and can cope with parentheses being used to nest terms.

So, it is worth trying out your search that you have constructed to work in a library literature database, and see what Bing shows you.

Something I used to appreciate in Google is that it would offer you a link to a cached version underneath most results and these cached versions had your search terms highlighted so it was easy to see where they appeared. Bing still does this.

I suspect there is a point to be made here about Bing being less advanced in its development than Google. Quite possibly we will see Bing become more like Google over time as more is invested in it.

But for now, you can make the most of these features, if they prove useful to you.

You can use many of the same techniques, or search commands, in Bing that you can in Google, such as file type and site search.

Intitle: also works in Bing as does a similar command, **inbody:**

Some differences that commentators have noticed between Google and Bing are that Bing's results tend to be more consumer/retail focused – though you can get round this initial bias by adding some of the commands to a simple search.

Bing may also be more up to date than Google – it updates sites more frequently, adds new sites more quickly so if you are researching in area where this is key, e.g. politics, perhaps refugee studies, this could be really handy.

My favourite thing and the only thing I use Bing for is that their maps include proper Ordnance Survey maps.

The country versions and date options are similar to Google.

Annoyingly though, many features and options are available to US users only – there are more and better commands available when you use Bing in the USA.

There are similar problems to Google in that you are not quite sure what it is up to – it may drop search terms, for instance if there are too few results, like Google does.

And the order of terms seems to matter to how many results you get.

Inbody: doesn't seem to work all the time; sometimes it just searches for the word "body"

How much is Bing tracking and personalising?

Bing is also a business, indeed a massive, profitable business – Microsoft.

Bing uses click data from its users to influence their rankings – just one of thousands of parts of their algorithm – as Google does.

There seems to be some feeling that Bing might be slightly better with privacy concerns than Google, but they are running the same kind of business. Bing is tracking and personalising as much as Google – or at least it is trying to. It probably doesn't do it as well, especially if you haven't used it as much as Google.

So Bing is not a search engine to choose for reasons of privacy concerns. It is one to choose for, perhaps, comprehensiveness, a change of results, the ability to use a Boolean search – something to give you alternative or additional results that you might not be seeing in your Google search.

Given the privacy concerns with search engines such as Google and Bing, an alternative search engine which has become widely popular with those keen to avoid all tracking and targeted advertising is DuckDuckGo.

DuckDuckGo doesn't use cookies to identify you, and it discards information including IP addresses from its server logs. It simply does not have any information on you, not even relating to searches you do one after the other from the same computer.

They do store your search terms, not in a personally identifiable way. The privacy policy states, "We use aggregate, non-personal search data to improve things like misspellings."

Because it does not store or track the searches that you do so obviously it doesn't (cannot) personalise your results. There is a log in option which lets you change things such as the 'look n feel' – this is only what you have set up, it is not tracking your searches, so quite different.

It recognises similar commands to Google and Bing, though as with the other two you can't be quite sure that it is paying attention and exactly which of its sources it is looking at.

You have to be more precise with your searches as, unlike Google, it does not know what you are thinking! It is also worth looking at the 2nd or 3rd pages of results.

The Asterisk does not truncate in DuckDuckGo either. We aren't sure how to do truncation in fact. It seems as if it may be searching for variations for you like Google does, but I am not sure.

The **Region:** command boosts country results in your search. Although, obviously if you search in English it won't really work to ask it to focus on a non-English speaking country.

DuckDuckGo can also search other sites for you, e.g. it will search Twitter for you.

DuckDuckGo makes some 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>.

The CEO of DuckDuckGo once said: “if the FBI comes to us, we have nothing to tie back to you”.

However, they are linked with Yahoo (now owned by Verizon, a big American telecoms company) so I could not say for sure that they are an entirely independent company despite their strongly promoted ethical stance on privacy.

I hope some of those tricks, both in Bing and DuckDuckGo prove useful and that you find exploring two other search engines adds to your literature search.

Part 10: A Brief word on Google Scholar

This session is not really about Google Scholar as that searches for journal articles mainly, and we offer other workshops about searching for that sort of key academic literature.

But, a brief note to explain that as well as the focus on what Google classes as ‘academic literature’, the main difference from ordinary Google is that Google cannot search behind paywalls, whereas Google Scholar has made deals with publishers to let it search their full text.

Ordinary Google does not use the publishers’ metadata, whereas Scholar does.

Google Scholar provides a very similar service to SOLO's Articles search and much of the content is the same. The Bodleian Libraries subscribe to a curated service (Primo) bringing together content from many publishers and databases. Google Scholar gathers its content from many of the same publishers in an automated way using search algorithms. The Bodleian Libraries also supplement the index with additional local content (e.g. online articles received via legal deposit, items from the Oxford Research Archive) to create SOLO's Article search.

Scholar is particularly useful to those without access via a university library so is great for practitioners and members of the public with research interests.

Whilst you are here at the University with us I strongly encourage you to make the most of SOLO and the integrated subscription sources which we in the Bodleian Libraries pay a lot of money for on your behalf.