Feb 21st, 9:30 AM - 10:30 AM

Advanced Internet Research Techniques

Stephen Wolfson
University of Georgia Law Library, stephen.wolfson@uga.edu

9:30   **SESSION 2 – ADVANCED INTERNET RESEARCH TECHNIQUES**
Stephen Wolfson, Research and Copyright Services Librarian, University of Georgia School of Law, Athens

Power Up!: Tips and Tricks to Becoming a Google Power User

Stephen M. Wolfson
Research and Copyright Services Librarian
University of Georgia School of Law
Law Library

Everyday, users run approximately 3.5 billion searches on Google -- 40,000 searches a second -- for a total of 1.2 trillion searches each year.[1] Those numbers are so staggering that it is almost impossible to wrap your head around how important Google is to daily information discovery. Yet, despite Google's near ubiquity -- indeed, because of it -- Google has a bad reputation among information professionals for encouraging poor research habits. Even so, in the hands of a master, Google can find much more than cat pictures[2] and movie times[3], and can be an extremely powerful tool for legal research.

In this paper, I will discuss three keys to unlocking Google's potential for legal research. First, I will identify some of the weaknesses you need to remember when using Google for legal research. Once you know Google's weaknesses, you can begin to take advantage of its strengths. Next, I will discuss a number of advanced search tools that can make your research more tailored and directed. While Google defaults to a keyword, full-text search, this does not always lead to the best results, and you can use its advanced tools to target your research better. Finally, I will review several specialized Google databases that are particularly useful for legal research. Google offers more than google.com, and these alternatives can help you find more relevant results. In the end, you will be well on your way to becoming a Google Power User.

1. Know Why You Are Using Google

Though Google is a powerful tool for legal research when you know how to use it most effectively, the first key to becoming a Google Power User is to understand two of Google's weaknesses: 1. Not everything is on Google, even if it sometimes seems like it is.

---

[1] *Google Search Statistics*, Internet Live Stats, http://www.internetlivestats.com/google-search-statistics/
[2] *Cats and the Internet*, Wikipedia.org, https://en.m.wikipedia.org/wiki/Cats_and_the_Internet
[3] Marziah Karch, *Movie Showtimes with Google*, Lifewire, December 16, 2018, https://www.lifewire.com/movie-showtimes-google-1616497

Google is so effective at finding what you need, even with a simple search, that it can give the false impression that if something is not on Google, it is not available anywhere. 2. Because there is so much information available through Google, but it is not designed for legal research specifically, you can often get many off-topic results (noise) mixed in with on-point results (signal), with little help at separating right from wrong. Knowing how to filter out the noise and home in on the signal is essential to making the most out Google. Ultimately, only after understanding what Google *can't do*, can you truly leverage what it *can do*. I will address both weaknesses in turn.

1.1. Lots but not everything

People often think of Google as a gateway to all the information in the world -- or at least all the information on the World Wide Web.[4] And certainly it is true that you frequently can find what you are looking for with a simple keyword search. Unfortunately, however, this can leave the impression that if something is not on the first page of Google, it is not anywhere online.[5] The truth, however, is that the search engine only provides access to a small percentage of what is online -- somewhere around 4-5%.[6] So, even though Google can offer an incredible amount of information, it is not nearly everything.

This begs the question, what about the other 95%? Much like how most of the known universe is composed of matter that we cannot see -- so-called dark matter and dark

---

[4] *History of the Web*, World Wide Web Foundation, https://webfoundation.org/about/vision/history-of-the-web/. People often use the terms "internet" and "world wide web" interchangeably, but they are not quite synonyms. The Internet refers to a network of networked computers. The web, on the other hand, is how most of us access this network. "So another way to think about it is to say the Internet is composed of the machines, hardware and data; and the World Wide Web is what brings this technology to life." Jessika Toothman, *What's the Difference Between the Internet and the World Wide Web*, How Stuff Works, https://computer.howstuffworks.com/internet/basics/internet-versus-world-wide-web1.htm

[5] Kelly Shelton, *The value of Search Results Rankings*, Forbes, October 30, 2017, https://www.forbes.com/sites/forbesagencycouncil/2017/10/30/the-value-of-search-results-rankings/#bf05c1744d3a
[6] JJ Rosen, *The Internet You Can't Google,* (May 2, 2014), https://www.tennessean.com/story/money/tech/2014/05/02/jj-rosen-popular-search-engines-skim-surface/8636081/

energy[7] -- most of the World Wide Web is similarly "invisible" to Google. We call this the "deep web."

Though its name sounds mysterious, most of the "deep web" is rather mundane. The term merely refers to web pages that are not indexed by common search engines.[8] As such, you can't find the information they hold through something like Google. Nevertheless, you probably interact with deep web websites regularly. This includes commercial databases like Westlaw, Lexis, EBSCO, and ProQuest, library catalogs, private networks, etc.[9]

At least part of what is on websites like these is proprietary information, or at least information that a website wants to have some control over. This is why you are unlikely to find case/statute annotations, headnotes, or library books through Google; the owners of this kind of information often want to control it in one way or another.

Meanwhile, the "dark web" is one part of the deep web, and it is just as ominous as it sounds.[10] These pages are not only not on Google, but you also cannot access them using regular web surfing methods. Instead, you need to use special tools that anonymize their users and allow access to these encrypted networks. Mostly commonly, these tools are Tor browsers.[11] The dark web is probably most famous today for being the home of nefarious black markets like the Silk Road.[12]

Warning: Do not surf the dark web unless you know what you are doing. While there are legitimate reasons to use the dark web, and while not all users on the dark web

---

[7] *Dark Energy, Dark Matter*, NASA Science, https://science.nasa.gov/astrophysics/focus-areas/what-is-dark-energy.

[8] Part of how Google and other search engines work is that the search for new pages -- crawling -- and then they analyze the content on those pages -- indexing -- so searchers can find them. https://support.google.com/webmasters/answer/70897?hl=en. Webpage creators can prevent Google from indexing their sites by including specific code that does this. https://support.google.com/webmasters/answer/93710?hl=en

[9] *How to Safely Access the Deep and Dark Webs*, Symantec, https://us.norton.com/internetsecurity-how-to-how-can-i-access-the-deep-web.html.

[10] Darren Guccione, *What is the Dark Web? How to Access it and What you'll Find* (January 11, 2019), https://www.csoonline.com/article/3249765/data-breach/what-is-the-dark-web-how-to-access-it-and-what-youll-find.html

[11] J.M. Porup, *What is the Tor Browser? How to Works and How it can Help you Protect Your Identity Online* (July 12, 2018), https://www.csoonline.com/article/3287653/privacy/what-is-the-tor-browser-how-it-works-and-how-it-can-help-you-protect-your-identity-online.html

[12] Matt Egan, *What is the Dark Web & How to Access It* (January 30, 2019), https://www.techadvisor.co.uk/how-to/internet/dark-web-3593569/

have malicious intent, the likelihood of coming across dangerous information, files that can corrupt your computer, and/or users who want to steal your personal information is much greater than when searching the surface web.

The important lesson here is this: Before you begin your research, you must understand Google *cannot* find a vast amount of information that is out there. The trick is to try to anticipate what you can and cannot get through Google and then think about where on the deep web you may be able to find an elusive piece of information.

1.2. Finding the signal in the noise

Google's second weakness may seem obvious, but it is equally easily forgotten: Google is so powerful, so flexible, and gives access to so much information, that it is easy to for your signal to be lost in the noise.

Even though Google may only index a small percentage of the surface web[13], this is still represents an immense amount of information. In 2013, Google estimated that it indexed 30 trillion individual webpages.[14] Today that number is probably over ten times greater.[15] Moreover, because Google is free, and because its algorithms are so strong, you can often find relevant results from very broad, unrefined searches. For this reason, basic Google searches can be especially useful in the beginning parts of your research, when you are trying to understand an issue and/or collecting background information. However, for the same reason, these searches are not as useful for finding very specific information. Almost any search on google will return numerous -- often millions -- of results, but you are at the whim of Google's relevance algorithms to decide what is "correct."

One way to improve your results from any electronic search tool -- Google included -- is to understand how it works, and then construct your searches to use it in the most effective manner. This can be difficult, however, with Google, because the exact operation

---

[13] *What is Surface Web, Deep Web and Dark Web?*, Medium (April 9, 2018), https://medium.com/@hackersleague/what-is-surface-web-deep-web-and-dark-web-cdbaf71b30d5
[14] *Billions of times a day in the blink of an eye*, Google Inside Search (March 2, 2013), https://search.googleblog.com/2013/03/billions-of-times-day-in-blink-of-eye.html
[15] Barry Schwartz, *Google's Search Knows about over 130 Trillion Pages*, Search Engine Land (November 14, 2016), https://searchengineland.com/googles-search-indexes-hits-130-trillion-pages-documents-263378

of its algorithms is a closely held trade secret.[16] Still, we have some clues about how it works that we we can use to improve our research.

Google describes its algorithms as a five step process.[17] First is "analyzing your words." Here, the search engine tries to guess what you are looking for, based on common terms, related words, categories, what is trending, what you have searched for before, and different spellings (including interpreting your spelling mistakes). Second is "matching your search." At this point, Google looks for relevant web pages through a complex analysis that is much more in-depth than simply considering whether your search terms appear multiple times on a page. It tries to make sense of what a page is *about,* not just what is says. The third step is "ranking useful pages." *PageRank*[18] is probably what Google searches are most famous for. Here, Google ranks the quality/authority/correctness of each website based on many different factors, including how regularly the content is updated, how many pages link to it, popularity of the page, and others. Fourth is "considering the text." Here, Google looks at your location, past search history, and search settings to further tailor your results. So, for example, if you search for "movie times," Google will give you movie times for local theatres, and not theatres in France. This is because Google assumes you do not want that other information. Finally, "returning the best results." Here, Google puts everything together and tries to give you the best results.

Ultimately, Google is a treasure trove of information, but it does not necessarily differentiate between the gold and the pyrite. You can find cases, statutes, patents, trademarks, legislative history, some public records, news, current awareness, general background information, quickly and often very effectively. But these things can often be lost among untrustworthy information and false positive results. If you do not target your research, you *may* find what you are looking for, but your research will be inefficient, you almost certainly will not find all Google has to offer. The remainder of this paper will discuss ways to home in on the signal of your legal research.

---

[16] Prajwal Nirwan, *Trade Secrets: The Hidden IP Right* (December 2017), https://www.wipo.int/wipo_magazine/en/2017/06/article_0006.html

[17] *How Search Algorithms Work*, Google, https://www.google.com/search/howsearchworks/algorithms/

[18] *PageRank*, Wikipedia, https://en.wikipedia.org/wiki/PageRank

2. More than meets the eye

After considering what Google can/cannot do for you, the second key to becoming a Google Power User is to unlock its potential for legal research by using advanced search techniques, operators, and search tools. When you go to google.com, you are presented with a search box and two options -- "Search" and "I'm feeling lucky" -- without any indication for what those options things mean. Yet even though it is not immediately apparent, Google is more than just keywords and luck, and there are many tools that allow you to to leverage the true power of Google.

2.1 Google Advanced Search

Simply stated, to use Google to its full potential, you need to use its advanced search options. These options provide you with many ways to tailor your search, giving you much more control over your results.

Not long ago, google.com had a link immediately to the right of the search box where you could access its advanced search features.[19] That has since been removed, but there are two other ways to find Google's advanced search. First, you can directly navigate straight to the page by going to https://www.google.com/advanced_search./ Second, if you click on the "settings" tab below the search box after you run an initial search on google.com, you can go to "Advanced search" from there. Both methods will get you to the same place.

Once on the advanced search screen, you will see a number of different search boxes and pull down menus that allow you to target and narrow your searches in different ways. Please note two things, here: 1. You are not limited to using just one of these options, and you can combine them in almost anyway you want. Once you become familiar with these tools, you can create extremely robust and complex searches. 2. Most of these tools have corresponding operators that you can use in the google.com search box. Below, I will describe each tool below an illustrate how you can combine them and use them on google.com.[20]

---

[19] *See, e.g.,* https://web.archive.org/web/20150101040436/http://www.google.com/. This is a capture of google.com from January 1, 2015. Note that immediately to the right of the search box is a link for "advanced search."

[20]This is not a complete list. For a comprehensive list of Google's Search Operators, *see* Joshua Hardwick, *Google Search Operators: The Complete List,* ahrefs blog (May 22, 2018),

All of these words

This search box is a basic google.com search. It looks for websites that contain all of your search terms, as if there was an "and" between them.

This exact word or phrase

This search box works like putting quotation marks around your search terms. As such, it searches for *exactly* what you enter, including any spelling errors or words like "the" that Google might otherwise ignore.

Examples: "colour"; "realise"; "the Rock"

Any of these words

Google will assume an "or" between the terms you enter in this box. Thus, it will return a broader range of results than an "and" search. This can be useful when trying to find something that may go by multiple terms or where you want to include two different types of terms in your search results.

To use this on google.com, "or" must be capitalized (OR). Additionally, you can use the pipe (|) symbol for the same effect.

Examples: "committee report" OR testimony; verdict | settlement

None of these words

When you enter terms in this search box, Google will exclude web pages that include those terms from your results. This operator can be difficult to use, but once you understand it, it can be very powerful. It is probably most useful in situations when you run a few searches, and notice you keep getting some results that are unrelated to what you are looking for but are related to each other. You can use this to remove those unrelated results.

---

https://ahrefs.com/blog/google-advanced-search-operators or *Google Advanced Search: A Comprehensive List of Google Search Operators*, Beyond,
https://bynd.com/news-ideas/google-advanced-search-comprehensive-list-google-search-operators

You can exclude terms in a google.com search by using the minus/hyphen symbol (-) before the term you do not want in your results.

For example, if you wanted to research the economic theory of value but did not want information on the moral concept of values, you might run a search for:

Example: "value theory" | "theory of value" -ethics -moral

Numbers ranging from

Google can search for a range of numbers by entering two search terms in this field. You can do this in a basic google search by separating two numbers by a double period (..)

Example: house committee report 2014..2018

Language

This pull down menu limits your search results to just webpages in a particular language.

Region

This pull down menu limits your search results to web pages from a certain country.

Last update

This pulldown menu limits your results to web pages that were last updated in the past 24 hours, in the past week, in the past month, or in the past year. Since Google tends to prefer pages that were recently updated, this can increase the time range of your search.

Site or domain

The site or domain field instructs google to return only results from a particular web page or domain. By focusing on an individual website, this essentially turns Google into a custom search tool for that website, and is often much more powerful than the website's own search tool. You can add this to a basic search by entering "site:" followed by the webpage or domain without a space. So this search -- "house report" site:.gov -- will look at

web pages that end in .gov for the exact phrase "house report." Similarly this search -- testimony | report site:house.gov -- will search the House of Representatives webpage for the terms "testimony" OR report.

This tool is especially powerful when searching for legislative history documents or information from agency websites.

Examples: site:copyright.gov "orphan works"; site:.org climate change; site:scotusblog.com kelo.

Terms appearing

This pull down menu allows you to specify where on a webpage your search terms should appear. Your options are: "anywhere in the page," "in the title of the page," "in the text of the page," "in the URL of the page," and "in the links to the page.

The operators that let you do these searches on google. Com are: "inurl:", "intitle:", and "intext:".

Example: intitle:"America invents act"

SafeSearch

This field can remove web pages that contain explicit material from your search results.

File type

This pull down menu can limit your search to particular document types, including .doc, .pdf, .xlsx, among others. This is especially useful when looking for official documents, reports, and the like that are usually issued as PDFs, like legislative history documents.

You can use this on google.com with the operator "filetype:".

Examples: "Orphan works" filetype:pdf.; food cost filetype:.xlsx

Usage rights

This pull down menu can help you find items that have been tagged with certain usage rights licenses. You can use this when trying to find rights free images and/or music for presentations/creative projects.

2.2. Combining operators

As aforementioned, you are not limited to using one, or even a few of these operators in a single search. Instead, you can include many, tailoring your search as is necessary. This works especially well when using Google to find legislative and/or regulatory documents. Since things like testimony and official reports usually in PDF format, you can use the filetype operator to focus solely on this document type. Further, you can use the site: operator to narrow down to just one or a range of government cites. So, if I was interested in a report on climate change and national security from the House of Representatives I might search for:

Example: site:house.gov filetype:pdf report and "climate change" and national security

2.3. Additional search operators

In addition to the search options on its advanced search screen, Google also recognizes a number of other operators to modify your searches. I will describe several below.

AROUND (x)

Researchers familiar with tools like Westlaw Edge and Lexis Advance should recognize this tool. This operator searches for terms as they appear within a specified number of words of each other. For example, if you want to search for copyright infringement information, but you are not sure whether an author would write "copyright infringement" or "infringement of copyright" or some other construction, you could search for:

Example: copyright AROUND(5) infringement

Related:

>   This operator will search for web pages that are similar to another webpage.

>   **Example**: related:uspto.gov.

Cache:

>   This operator will return a webpage from what it was the last time Google crawled the page and tell you when Google took that last snapshot. This can be useful when you visit a webpage once, and then the next time you go there, some information has changed. With this, you can try to use this to see what it was earlier in time.[21]

>   Example: cache:law.uga.edu

Define:

>   You can use Google as a dictionary with this operator.

>   **Example**: Define:negligence

3. Google is more than google.com

The third step to becoming a Google Power User is to realize that Google is more than just google.com. Google offers a number of specialized databases that can help you find particular types of information. Here, I will review several of the databases that are relevant to legal research.

Google Scholar (scholar.google.com)

Google Scholar is the most significant alt-Goole database for legal research. As its name suggests, this is a tool for researching scholarly articles across all kinds of disciplines. It can provide access to numerous scholarly articles, papers, books, theses, and dissertations from a variety of locations. Most of the results you will find here are links to

---

[21] Note that this it will only give you a snapshot of the last time google visited the site. It will not go to different points in time. If you need to do that, you should try the Wayback Machine, https://archive.org/web/

abstracts and citations, but they also can help you locate the articles elsewhere.[22] Unfortunately, however, because this database does not differentiate between disciplines, you can often find results unrelated to your research. Google Scholar has an advanced search function that you can access by clicking on the three hash marks on the top left of the screen, but it is not as robust as the one on google.com. Accordingly, it benefits you to construct smart searches when using Google Scholar, to try to minimize the noise.

One great tool Google Scholar offers is its "cited by" feature. After you run a search and find a relevant article, you will see a link that says "cited by" below the article. If you click on this, Google Scholar will take you to a list of articles/books that have cited the first article. From here, you will see a tickbox that says "search within citing articles" if you click on this and run another search, Google Scholar will look through just those articles that cited the first article for your new search terms. Since it is likely that articles which cite an article that is on point for you are also on point, you can use this to target especially relevant articles.

In addition to scholarly articles, Google Scholar also offers an extensive database of federal and state cases. Note, however, that you will not find these if you simply go to scholar.google.com and start searching. To look up case law, you must first select the "case law" button that is below the search box. After that, you can select specific jurisdictions/courts, including federal bankruptcy courts, tax courts, and the court of international trade. Additionally, you will see "cited by" and "how cited" links below each case you find. These work similar to other citators, by looking for other cases that cite to the first case.

The greatest benefit of using Google scholar for case research is that it is free, quick, and easy to use. However, you must once again remember Google's weaknesses: Google Scholar does not necessarily have everything, and it is easy for the singal to get lost in the noise. Consider, for example, the "cited by" and "how cited" functions on Google Scholar's case law database. These are good and useful tools, but *they are not sufficient substitutes* for

---

[22] An increasingly large number of law journals are putting their current issues and archives online for free. If you find a citation to an article, and you cannot find it through Google Scholar, you may be able to find it on the journal's website.

Shepards/Keycite. In using them, you can get a general idea for where a case was cited and whether it is still good law, but you cannot be certain of its currency. You must remember this when using Google Scholar.

Google Books (**https://books.google.com/**)

Google Books is a database of all kinds of books from a variety of sources. Much of this information comes from large academic libraries that participate in the Google Books Library Project.[23]

When you run a search on Google Books, you will look across the full text of the many books in this database. If a book is not in the public domain, you will only be able to see snippets from the book that relate to your search in your results. However, if a book is in the public domain, you will be able to access the entire thing.

Like google.com and Google Scholar, Google Books has an advanced search function that you can access by clicking on the Gear Icon on the search screen. From here, you can use several fields to narrow your search, including by: title, author, publisher, and subject.

Google Dataset Search (https://toolbox.google.com/datasetsearch)

Google dataset search is a relatively new offering from Google. It is a database of published datasets on the web. It also presents usage rights information on the webpage as well as occasionally links to articles that have been published based on these data.

This database can be extremely valuable when trying to find empirical research on a topic. For example, you can find research on how often the Supreme Court overturns cases from the various Courts of Appeals "US Supreme Court" in this database.

Google Patents (patents.google.com)

Google Patents is an excellent search tool for conducting patent research. With "over 120 million patent publications from 100+ patent offices around the world, as well as many more technical documents and books indexed in Google Scholar and Google Books, and

---

[23] *Google Books Library Project -- An Enhanced Card Catalog of the World's Books*, Google, https://www.google.com/googlebooks/library/

documents from the Prior Art Archive," Google Patents is fast, easy to use, and extremely good at find relevant patents.

Once again, Google Patents has an advanced search feature that you can find at patents.google.com/advanced, and like before, this provides you with many ways you can target your research. Additionally, you can tailor and narrow your search after running an initial search by selecting from numerous options on the left of the screen, including: inventor, assignee, status, type, and priority date. After you find and open a relevant patent/application, you can also find links to "find prior art" or to find "similar" documents. Both of these options are very good at helping your find relevant prior art. You can also use Google Patent to find relevant non-patent literature.

Conclusion
Bonus Key: Take a class!

This paper should set you on the way to becoming a Google Power User, and with some practice, you will begin seeing improvements in your results after implementing some of the advice in this paper.

However, if you want to get even better at using Google, consider taking a class. Google offers courses on "Power Searching" and "Advanced Power Searching" that can further strengthen your skills. Indeed, many of the strategies these courses teach work well for any kind of electronic research. You can take these classes from home at your own pace by going to www.powersearchingwithgoogle.com/ and following the links provided.

Ultimately, the most important step you can take to becoming a Google Power User -- and a better researcher in general -- is to be smart. Think about what you are looking for and where you can find it, and take some time to construct a search that uses tools other than simple keywords. Employing just a few of the tips and tricks presented here can dramatically improve your results and your research efficiency. Good Luck!

Stephen M. Wolfson serves as the Research and Copyright Services for the University of Georgia School of Law, Alexander Campbell King Law Library. Before coming to UGA, Stephen worked as the Copyright Librarian for the University of North Texas in Denton and the University of Texas in Austin.

Stephen received his Bachelor of Arts and Juris Doctor from the Ohio State University and his Master of Science in Information Studies from the University of Texas. He has taught classes on Advanced Legal Research, Intellectual Property Legal Research, and Free and Low Cost Legal Research, among others. Additionally, he has published on privacy, intellectual property, and crowdsourcing/crowdfunding law.

Stephen is originally from Chagrin Falls, Ohio, and is a member of the Oregon State Bar Association and the American Association of Law Libraries.