Continuing Legal Education Presentations   March 24, 2014

Mar 24th, 8:45 AM - 9:45 AM

Savvy Web Searching with Advanced Tools and Techniques

Anne Burnett
University of Georgia School of Law Library, aburnett@uga.edu

Follow this and additional works at: https://digitalcommons.law.uga.edu/cle

Part of the Legal Writing and Research Commons

Savvy Web Searching with Advanced Tools & Techniques

Anne Burnett, J.D., M.L.I.S.
University of Georgia Law Library
Athens, Georgia

Savvy Web Searching with Advanced Tools and Techniques

Anne Burnett, J.D., M.L.I.S.
University of Georgia Law Library
Athens, Georgia

TABLE OF CONTENTS

Contents

I. Introduction

With the word "Google" now being used as a verb meaning to search the internet (to the dismay of the underlying corporate entity),[1] the namesake search engine is the first step in many researchers' exploration of available web resources. Google is not the *only* search engine game in town, however, and savvy searchers are familiar with competing products, including Bing and ever-evolving alternatives. This paper discusses several web-searching sites, including Google, as well as sites designed to provide answers to questions.

Power internet searchers also employ advanced tips and strategies for crafting effective search queries in whichever search tool they utilize. The emphasis herein is on Google's and Bing's advanced searching tools, but many other search engines, such as Yahoo and DuckDuckGo, allow for similar advanced search terms and operators to increase effectiveness and efficiency in retrieving relevant results. Web searchers will also learn about selected special services and enhancements in Google and Bing. Finally, with an estimated 90% of internet content unsearchable by traditional search engines, a burgeoning area of development is for tools to search the Deep Web.

---

[1] See Joseph Philipson, *How Google Became a Verb*, The Lingua File (Feb. 1, 2013), http://www.thelinguafile.com/2013/02/how-google-became-verb.html. Merriam-Webster Dictionary continues to specify that "Google" means to search the Web using the Google search engine.

II. Keep A Couple of Search Engines in Your Back Pocket

If you cannot find what you seek using Google, do you assume it either does not exist or is unfindable? Do not give up! Savvy searchers are familiar with one or two additional search engines and know that a quick run of the same search query through an alternative tool might find the exact site needed. The following are worth keeping in mind.

Google
www.google.com

Google still leads the U.S. search marketplace with a 67.6 percent market share, according to comScore's analysis of the web searching activity in January 2014. comScore reports on the search results from Google, Bing, Yahoo, Ask and AOL. See Section III below for strategies to use Google more efficiently and effectively.

Bing
www.bing.com

Microsoft's search engine, Bing, is a distant second to Google in the search engine market, but it ended 2013 with a market share of 18.2 percent, up from its 16.3 percent share at the end of 2012. Bing is a good alternative and worth exploring. PC World magazine conducted a showdown between these two search engine giants in April, 2012, and found that neither search engine outright won the competition. While Google scored higher in some categories, the tests also showed that "sometimes, 'binging it' might be the better way to go."[2]

---

[2] Liane Cassavoy, "Bing Versus Google: Search Engine Showdown," PC World (April 24, 2012), available at http://www.pcworld.com/article/254405/bing_versus_google_search_engine_showdown.html

Yahoo!
www.yahoo.com

Yahoo!'s rank in the search engine market has declined in the past year, but at number three it may still be worth keeping in the web researcher's tool box. If a search in Google or Bing returns unsatisfactory results, try it in Yahoo! which offers many of the same advanced searching options discussed in more detail for Google and Bing in Section III below.

DuckDuckGo
duckduckgo.com

This "anti-Google" search engine promises to offer far better privacy than the big name search engines. This small start-up is receiving a good bit of attention as a viable alternative to Google and Bing, especially for searchers who do not want the search engine to save their usernames, email addresses, social media logins and other identifying information.[3] DuckDuckGo promises a "better search and real privacy" for searchers concerned about tracking by search engines. DuckDuckGo also promises that it does not store searches nor does it transmit information to third parties.

The DuckDuckGo search interface is simple and clean, reminiscent of Google at its inception. For image, video and map searching, use the dropdown menu to the right of the DuckDuckGo search box to pass a search query through to other sites, including Bing, Google and YouTube (but note that could negate the privacy-related advantages of basic DuckDuckGo). Use the dropdown box to sort the search results by date, category or alphabetically.

---

[3] Kim Lachance Shandrow, "6 Things You Should Know about the Anti-Google Search Engine DuckDuckGo" Entrepreneur (February 27, 2014), available at http://www.entrepreneur.com/article/231852.

Computational and Question/Answer Sites

Rather than searching an index of "crawled" web pages, sites like Wolfram|Alpha and Quora attempt to answer questions or provide computational functions.

Wolfram|Alpha
www.wolframalpha.com/

Wolfram|Alpha is a Computational Knowledge Engine, or a question and answer site where the answers are computer-generated. Its stated goal is "to make all systematic knowledge computable and broadly accessible." Users submit a textual question in the input box which is invitingly labeled "enter what you want to calculate or know about." The results box includes a statement of how the system interpreted the question. The results also suggest additional queries that the system determines may be relevant.

Wolfram|Alpha is, of course, limited by the data and algorithms of which it is comprised. A search for **how many lawsuits were filed in 2010?** returned a disappointing result of "Wolfram|Alpha doesn't understand your query." A search for **litigation statistics** resulted in information about the ABA periodical *Litigation*. A search for **Georgia crime,Alabama crime** brought a much more satisfying result of a side-by-side chart comparing Georgia and Alabama crime statistics for 2009 including overall numbers and breakdowns by broad categories (larceny, robbery, etc.).

A Wolfram|Alpha app is available for both Android and iPhone.

Quora
www.quora.com

Quora is a question and answer site where the answers are provided by humans. The answers are rated by readers, who also rate the questions by indicating interest in an answer. Questions are tagged into topical areas, including law. Examples of recent law-related questions on Quora include:

- Will people filing insurance claims due to the $400 million dollar Mt. Gox Bitcoin theft be able to claim losses at the Bitcoin price before the robbery when Bitcoin prices were higher — or after the robbery when Bitcoin prices had fallen?

- What are some odd or interesting Supreme Court cases?

- What was it like to face your first case in court as an attorney?

- What are the best college majors to prepare for LSAT?

Questions on a variety of topics include:

- What meal can you buy in your home town for $5?

- Given our current technology and with the proper training, would it be possible for someone to become Batman?

- What do truck drivers eat?

- I'm a student creating a start-up. What do I do?

Quora is edited and maintained by its users. The site is free, but registration is required. Quora apps are available for both Android and iPhone.

III. Crafting Advanced Searched Queries

Spend time up front in crafting a precise query and you'll spend far less time meandering through page after page of results. Following a few conventions and learning some key symbols to include in the string of characters entered in the search box can greatly improve your results. Many of these strategies will work with numerous search engines, but look for a link to "search help" or "search tools" or "advanced searching tips" or something similar on any search site to get exact symbols and strategies:

- Google: http://support.google.com/websearch/?hl=en links to several search help pages within Google, including Search Operators and Search Tools & Filters.
- Bing: http://onlinehelp.microsoft.com/en-us/bing/default.aspx provides some searching tips and techniques
- Yahoo: http://help.yahoo.com/kb/index?page=content&y=PROD_SRCH&locale=en_US&id=SLN2242&impressions=true
- DuckDuckGo: https://duck.co/help/results/syntax

The following tips and tricks focus on Google due to its ubiquitous use; however, a few different conventions for Bing and Yahoo are also noted.

Use quotation marks

Surrounding a phrase with quotation marks is the most common way to accomplish phrase searching. Google, Bing and Yahoo all use quotation marks to search for an exact phrase.

Limit your results to a specified time

Google and Bing both allow filtering of results to a specified time within which pages were last updated (past 24 hours, past week, past month, etc.). Google also allow includes a "last update" field in its Advanced Search box to impose the same limits on the results.

Exclude words from your results

 Use the minus sign or NOT directly in front of words you do not want to appear in your search results. **Vikings –football –team** should produce results about the seafaring raiders rather than the sports team and a search for **RICO – Puerto** should exclude many irrelevant hits from a search for sites about the racketeering statute. (Note: Yahoo also uses the minus sign. In Bing, use the word NOT in front of words you do not want included in your results.) Caveat: use the minus sign or NOT limiter carefully as you may unwittingly eliminate relevant sites from your results.

Find definitions

Type "define:" followed by the word you want defined, and Google will provide a list of web sites providing a definition. Try **define:infeft**. Preceding a search term with "what is" often produces the same results.

Search for synonyms using ~ and alternate terms using OR

Google routinely searches for synonyms, but inserting the tilde symbol ~ directly in front of a word will search for even more synonyms. For example, **~police** will also retrieve pages containing "law enforcement," "cops" and "military." To search for results containing related terms that might not be considered by Google to be synonyms, use a capitalized OR between the two terms. Google gives the example of **"Olympics 2014" OR "World Cup 2014"** to have one search retrieve sites containing either phrase.

Use wildcard operator

Use an asterisk (*) to formulate a search query asking Google to fill in the blank. **Christopher Columbus died in *** retrieves numerous sites providing the year of Columbus' death (it was 1506, by the way).

Limit by language

Google's advanced search form http://www.google.com/advanced_search?hl=en allows researchers to limit search results by one of over 40 languages. Simply choose the language in the drop down box.

Search (or exclude) a specific domain

Google allows searchers to limit their results to pages within a specific domain by typing "site:" before the desired domain abbreviation. For example, a researcher may desire only pages assigned a government, or .gov, domain. A search of **improving school lunch site:.gov** will only retrieve pages with the .gov domain. Placing a minus sign in front of the same search will retrieve pages from all domains but the typed one, so that **improving school lunch –site:.com** will find sites with .edu, .org, and .gov domains while excluding those with a ".com" domain.

When searching for foreign materials, specifying both a language and a country-specific domain often retrieves the best results. For example, a search for **bananas industry site:.sv** limited by English in the drop down box retrieves a relevant list of pages about El Salvador's banana industry in English. A search for **"climate change" site:.mx** limited to English might produce sites in English with the Mexican perspective on climate change. The Internet Assigned Numbers Authority (IANA) provides a complete list of domain extensions at http://www.iana.org/domains/root/db/.

Searching a specific web site

The "site" limiter also allows the researcher to further limit the search by limiting the results to pages within a specific web site. Type "site:" followed by the web site. A search for **constitutional law site:www.law.uga.edu** will only

retrieve sites with that term on the UGA School of Law web site. (Note: the site: limiter also works in Bing and Yahoo.)

Limit your results by a number range

Google allows you to specify a specific number range by placing two periods between the desired numbers. This is useful for searching for ranges of things like dates, prices and measurements. A search for diabetes research 2008..2009 theoretically retrieves results limited to diabetes research from those two years. The number range limiter works best when the search engine can ascertain from the context of the query exactly what is being limited.

Force search for exact word

Google assists searches by automatically including results for certain synonyms and variant spellings and abbreviations. To bypass this automatic expansion of your search query and force Google to return results including only the exact word as typed, place quotation marks around the word. For example, a search for "canine" diseases will only return results including the word "canine" whereas the same search without the quotation marks would also return results with the word "dog" even if the word "canine" was not included. (Note: in Yahoo, place a plus sign directly in front of the word you want to appear exactly as entered; for example, **+childcare** will only retrieve results containing the word "childcare" while those sites using "child care" will be excluded.)

Pay attention to word order

The order in which words are entered in a search query usually do not make a difference in the total overall results list; the difference may be in which sites make it to the top of the results list. Search engines like Google index words as *bigrams*, or words next to each other in a specific order, and the closer your query is to the order in which the words appear in relevant sites, the more relevant your top results will be. A short and entertaining but informative video provides several examples of searches where word order in the query returned very different results – see blog entry at http://searchresearch1.blogspot.com/2013/02/1mm-8-why-search-query-word-order.html.

Proximity searching

In Google, use a capitalized AROUND(n) between keywords, replacing the "n" in parentheses with a number signifying the maximum number of terms between the two keywords. For example, a query of **John AROUND(2) Smith** searches for all the possible combinations that a proximity search such as 'John within 2 of Smith" would create, thereby allowing the searcher to simultaneously search for John Smith, John X. Smith, John Xxxx Smith or Smith, John.

In Bing, use NEAR:n to search for words within a specified number of words from each other. Try **attorney NEAR:5 ethics** or **television NEAR:3 violence.**

Use the Advanced Search Box

Searchers can use a search engine's "advanced search" function to build a search query incorporating many of the above tools without having to remember all of them. Google no longer links its advanced search page from its home page, but a quick search query of "Google Advanced Search" retrieves the page quickly. At present, use the following to access advanced search in the major search engines:

- Google – either search for **Google advanced search** or go to http://www.google.com/advanced_search?hl=en

- Bing – enter search in basic search box; on results page, click in search box then on link to Advanced Search at bottom of search query suggestions.

- Yahoo – either search for **Yahoo advanced search** or go to http://us.yhs4.search.yahoo.com/web/advanced?ei=UTF-8&fr=altavista&p=

IV. Google Enhancements & Services

This section discusses just a few of the numerous useful options and tools provided by Google (similar ones are found in Bing and other search engines). Google's "even more" page at www.google.com/intl/en/about/products/ lists its many different products, from specialized search tools (alerts, Scholar) to social networking (Google+, Blogger) to collaboration tools (Calendar, Drive) and much more.

Google Alerts
www.google.com/alerts)

Google Alerts sends email updates of latest results of a specified query. Searchers specify whether the email alert should be sent once a week, once a day, or as-it-happens. Setting up an alert is easy and quick, and searchers can modify existing alerts.

Google Search History
www.google.com/history

Google Search History allows searchers who are signed into Google to review their recent search queries as well as results over a period of time. Searchers can pause Search History to temporarily turn off the tracking.

Google Translate
http://translate.google.com/?hl=en

Google Translate will automatically translate text, web pages, and files between over 50 languages. As with any automated translator, Google Translate is useful for getting the gist of a site but cannot be relied upon to give a sophisticated translation.

Google Scholar
scholar.google.com

Google Scholar searches for articles, theses, books, abstracts and court opinions, from academic publishers, professional societies, online repositories, universities and other web sites. If searching for court opinions in Google Scholar, check the box next to Legal Documents under the search box. The search form then allows the researcher to limit the results to federal court decisions or to those from

specified states. Google Scholar also serves as a citator by listing in the sidebar the cases which subsequently cite the displayed case.

Google's inclusion of the following disclaimer on their Legal Opinions page

> Legal opinions in Google Scholar are provided for informational purposes only and should not be relied on as a substitute for legal advice from a licensed lawyer. Google does not warrant that the information is complete or accurate.

leads to a similar disclaimer here: do not assume that the results include all relevant case law on a specific legal issue! This collection should serve solely to give the researcher an idea of the law or to help locate a specific case.

Google Reader
support.google.com/reader/answer/113517?hl=en

Google Reader provides a centralized location to read updates to blogs of interest to the researcher. Using RSS (Really Simple Syndication) technology, a researcher can easily add new blogs to Google Reader and then read updates to those blogs in one spot when convenient. This eliminates the need to visit each blog site individually to check for updated content.

V. What If It Is "Ungoogleable?"[4]

Not all content on the web is accessible through search engines like Google or Bing. In fact, the vast majority of web content is missed by traditional search engine spiders and crawlers, which cannot "see" the content buried in subscription sites or generated on-the-fly by databases. Referred to as the "Deep Web," "Invisible Web" and "Hidden Web," this content is estimated by the Deep Web Source Repository to be at least 400 to 500 times the size of the "free web" (also referred to as the "surface" or "visible" web).[5]

How, then, is a researcher to find this content? Search engines that attempt to retrieve content from the Deep Web are popping up daily, and any list of these resources is likely to be outdated by the time the list is published.[6] With that caveat, here is a sampling of current deep web search tools that are worth a try:

CompletePlanet
aip.completeplanet.com

CompletePlanet, which appears on many lists of Deep Web search resources, currently indexes over 70,000 databases. It uses a Bright Planet technology

---

[4] In 2012, Sweden's Language Council added to its list of Swedish words the term "ogooglebar" which translates as "ungoogleable" and was defined as "something you cannot find on the web with the use of a search engine." The council removed the term in 2013 after a kerfuffle with Google's lawyers who had asked the council to amend the definition to specify that the term meant only searches made using Google and not their competitors. Rather than revise the definition, the council removed the term from the list. "'Ungoogleable' Removed from List of Swedish Words after Row with Google," The Independent (March 26, 2013), http://www.independent.co.uk/news/world/europe/ungoogleable-removed-from-list-of-swedish-words-after-row-over-definition-with-google-8550096.html.

[5] "Deep Web Source Repository: Stop Searching Site by Site" (January 10, 2013), available at www.brightplanet.com/2013/01/deep-web-source-repository-stop-searching-site-by-site/. See also Alex Wright, "Exploring a 'Deep Web' that Google Can't Grasp," New York Times (February 22, 2009) available at www.nytimes.com/2009/02/23/technology/internet/23search.html?th&emc=th&_r=0 (part of web that is searchable by Google "represents only a fraction of the entire web").

[6] An earlier draft of this article referenced Scirus as a reliable tool for finding science literature that a general search engine might not retrieve; unfortunately, Scirus "retired" in early 2014.

called Deep Query Manager (DQM) to search across numerous databases

simultaneously. Researchers choose a content type (ranging from agriculture to

health to politics to weather) and CompletePlanet will return a list of databases

identified as containing relevant content. The researcher then enters a search

query which will be run across the multiple databases using the DQM technology.

SurfWax
www.surfwax.com

SurfWax is a meta-search engine, meaning it simultaneously runs a search

through multiple other search engines, theoretically giving the searcher the best

of all search engine worlds. Other meta-search engines include Dogpile, ZapMeta,

IxQuick and Yippy (formerly Clusty).

USA.gov
www.usa.gov

Use this as a search engine and as a portal to federal, state, local and tribal

government documents and databases.

Pipl
pipl.com

Pipl searches personal profiles, public records and other documents stored in

databases and is especially useful for locating people-related information.

Infomine
infomine.ucr.edu/

InfoMine was designed for the academic researcher. It was created by and is

maintained by librarians at the University of California, Wake Forest University,

California State University, the University of Detroit - Mercy, and other

universities and colleges. InfoMine is designed to be a virtual library of databases, electronic journals, electronic books, bulletin boards, mailing lists, online library card catalogs, articles, directories of researchers, and many other types of information.

CiteSeerX
http://citeseerx.ist.psu.edu/index

CiteSeerX describes itself as 'an evolving scientific literature digital library and search engine that has focused primarily on the literature in computer and information science." Also take a look at **DeepDyve** (www.deepdyve.com) if searching for scientific literature.

Wayback Machine
archive.org

The Wayback Machine at the Internet Archive includes over 397 billion pages of archived web sites from 1996. This is especially useful for finding a snapshot of content that has been removed from the web.