University of Georgia School of Law
Digital Commons @ Georgia Law

Continuing Legal Education Presentations March 19, 2012

Mar 19th, 8:30 AM - 9:30 AM

In Real Time: Tips for Fast, Effective Internet Searches

Suzanne R. Graham
University of Georgia School of Law Library, srgraham@uga.edu

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

Part of the Legal Writing and Research Commons

Suzanne R. Graham, "In Real Time: Tips for Fast, Effective Internet Searches" (March 19, 2012). *Continuing Legal Education Presentations.* Paper 2.
http://digitalcommons.law.uga.edu/cle/March_2012/Schedule2012/2

This Event is brought to you for free and open access by the Alexander Campbell King Law Library at Digital Commons @ Georgia Law. It has been accepted for inclusion in Continuing Legal Education Presentations by an authorized administrator of Digital Commons @ Georgia Law. For more information, please contact tstriepe@uga.edu.

In Real Time:
Tips for Fast, Effective Internet Searches

Suzanne R. Graham

University of Georgia School of Law
Alexander Campbell King Law Library
Athens, Georgia

In Real Time:
Tips for Fast, Effective Internet Searches

Suzanne R. Graham

University of Georgia School of Law
Alexander Campbell King Law Library
Athens, Georgia

Table of Contents

Introduction 1
History 1
Google It 2
- Word order matters 3
- Variants 3
- Search as a phrase (quotation marks) 4
- Must include (quotation marks/plus sign) 5
- Boolean 5
- Domain or site specific (site:) 8
- File-type specific (filetype:) 8
- Novelty non-search features 9
- Out-of-the-box help 10

Searching without Search Engines 12
Privacy 13
Conclusion 16

Introduction

Most people would say that they are good at surfing the internet but admit that they might not efficient. The concept of a "search strategy" sounds quaint and out-of-date. A strategy might serve a necessary role while on Lexis, Westlaw or PACER, but search engine searches are free and so fast that trial-and-error or try-assess-and-try-again seem reasonable approaches. However, when a single known result is desired, patience is short or the viewable screen is too small to make scrolling desirable there are a variety of handy tools and techniques to use to get the best out of the Web.

History

Modern search engines with their combination of a simple interface powered behind the scenes by an automated index of Web pages first appeared in 1994. Few of the names from those early efforts (Magellan, Northern Light, Excite, AltaVista, Lycos) are familiar any longer. The pop of the Internet bubble made a diluvian break in the field of contenders.

Curiously, two of the main players today, Google, and Bing (then known as MSNSearch) trace their start to the same year: 1998. Major search engine tracking services, *comScore* and Experian Hitwise give both independently credit Google with 65- 66% of the U.S. market share. Bing and Yahoo!Search alternate at the second and third positions with 13%-15% for the past six months.[1] Although users can choose from at least 66 different search engines, based on those figures, we will focus on Google and

[1] comScore Releases January 2012 U.S. Search Engine Rankings, comScore (28 February 2012) http://www.comscore.com/Press_Events/Press_Releases/2012/2/comScore_Releases_January_2012_U.S._Search_Engine_Rankings. And Search Engines, Experian Hitwise (29 February 2012) http://www.hitwise.com/us/datacenter/main/dashboard-23984.html.

Bing (with popularity on the rise), talk about privacy considerations, and alternatives to using search engines.

Google It

Google's huge advantage in current market share is sustained by its excellent name recognition and because the engine works very well. The simplicity of single box on a mostly white page belies the explosion of data that is about to replace it once the search results roll in—in a fraction of a second. But Google's (and Bing's) power goes beyond the box. It's all about how they index and interpret search strings. Very early on Google had been praised for the intuitive and impressive results, creating an aura of wonder about the variables analyzed and the formula used to produce a results list. While some of the factors that Google considers are known publically (such as number of links to the page, the position and size of the search terms within the page, and the proximity of the search terms to one another on the page), most of the roughly 200 variables are not, which fuels much speculation.

For fun "Google" or "Bing" the search-phrase ["google searching"]. The results highlight how useful crowd-sourcing or web democratization can be. Stringing behind "expert" blog posts and web journal articles on the topic are many comments and suggestions added by regular users with special interests and more up-to-date information. The main takeaway is that over the past (nearly) twenty years of intense competition among these search engines and fitful mergers and reconfigurations of other players, search engines are very powerful but the full extent of their indexing and ranking are tightly-guarded industry secrets.

Word order matters

Each term within a search is not treated as a discrete unit. Google and Bing programmers also consider contextual clues based on all the words in the search, but Google more consistently interprets those clues correctly. Somehow Google is smart enough to know that [private practice] should show items on solo practitioners first and keep the private ice skating lessons below page 30. Even though the same search terms appear in both the ice skater and in the attorney pages, Google has figured out that searchers looking for lessons use other search terms, and it has accounted for this in the rubric.

In addition to contextual clues, the placement of words matters. Reconfiguring the above search as [practice private] does yield hits for pilot training on the first page. Thus, in both Google and Bing, the first words are interpreted to be the most important. Searches for [learning to play the guitar] and [guitar learning to play] retrieve similar lists, but the page rankings are different, and the second (focus on "guitar") has many more hits.

Variants

Unless specifically instructed not to do so, Google and Bing automatically search for variant forms of search words, including plural/singular forms, other verb tenses, and common nicknames for proper nouns. Unfortunately, the English language has many examples of cases in which these *faux amis* are not desired because they take on different meanings, consider the meanings of "classified" and "classifieds."

Google and Bing anticipate and factor out these cases surprisingly well, but to be safe in any interface, tell the search engine to search only for the term as typed by

enclosing the word in quotation marks. For example, ["Internet" "searching"] retrieves pages that have both terms and ranks them higher than pages that more extensively use the variants "Web" and "searches."

Search as a phrase (quotation marks)

In addition to limiting the automatic substitution of similar terms, quotation marks can be used to instruct the search engines to look for only those terms and to look for them in a specific order and without any interceding words. This strategy generally works better for well-established subjects than it does for emerging scholarship simply because terminology has coalesced. A search for the American Civil War should have "civil war" enclosed in quotation marks, otherwise all pages that have "civil" and "war" will be pulled as relevant.

Phrase searching also works better for terms than people because people have so many variations of their names. ["George Washington"] is fairly safe, but ["Suzanne Rockey Graham"] is overly proscriptive and returns only six items related to my activities with one organization. Granted these six hits are all for the correct person, but the search misses the 16,000 other hits that are also relevant.

Beware of Google's "verbatim tool" at the bottom of the left column of options after a search is run. It is not the same as enclosing the entire search in quotation marks. The verbatim tool searches for each term without any variants, but it does not respect the quotation marks as a phrase wrapper. If a search for ["civil rights legislation"] is run through the verbatim tool in Google, it will search for the exact words ["civil" "rights" "legislation"]. The word "law," for example, is not in bold in the results list, but the terms might not be found sequentially on the pages. The search ["civil rights

legislation"] retrieves 1.76 million items, while the search in verbatim collects 6.5 million pages.

Must include (quotation marks/plus sign)

The English language has many homonyms, so as mentioned earlier, giving even these super smart search engine some context is helpful. A user can refine the results by telling the search engine that pages must include or must not include a term. Following up with the example above, a search for ["Suzanne Graham"] uncovers a racy author, a thrill-seeker, a real estate agent, and a librarian in Georgia. To instruct Bing to return pages about the librarian, one can use ["Suzanne Graham" +library], meaning the pages must contain the word "library" (or one of its synonyms).

It is a little more difficult to indicate must-haves in Google. Fairly recently Google abandoned the use of + as a search operator. Too many products, including its own Google+ service, use the + as part of the name. To partially compensate, searchers can enclose important search terms in quotation marks.

Another reason to use a plus sign or enclose in quotation marks is to ensure that a keyword is not ignored. It is common to assume that the search engines are already returning only results with all the terms, but instead they work hard to outsmart the search—even if that entails returning pages with only two of the three search terms.

Boolean

When librarians think search strategy, they think Boolean searching. It's one of those insidious ways that high school math continues to creep into the lives of adults in unsuspected ways. Boolean searching employs the logic operators AND, OR, and NOT. Google and Bing only recognize OR.

Even in all capital letters, search engines treat NOT as just another term. In order to use it to limit a search, the searcher must use the minus sign in both Google and Bing ([civil rights –legislation]). Georgia is good example since it is both a country and a state. Consider the advantage of [Georgia –Tbilisi] when trying to find information concerning the state, not the country.

On their help pages, Google and Bing both claim that they treat keyword searches as AND queries, but they do not.[2] The results list for [civil rights legislation] is quite different from the list generated by the query [civil AND rights AND legislation]. The first search string retrieves only a small fraction of the hits found by the second.

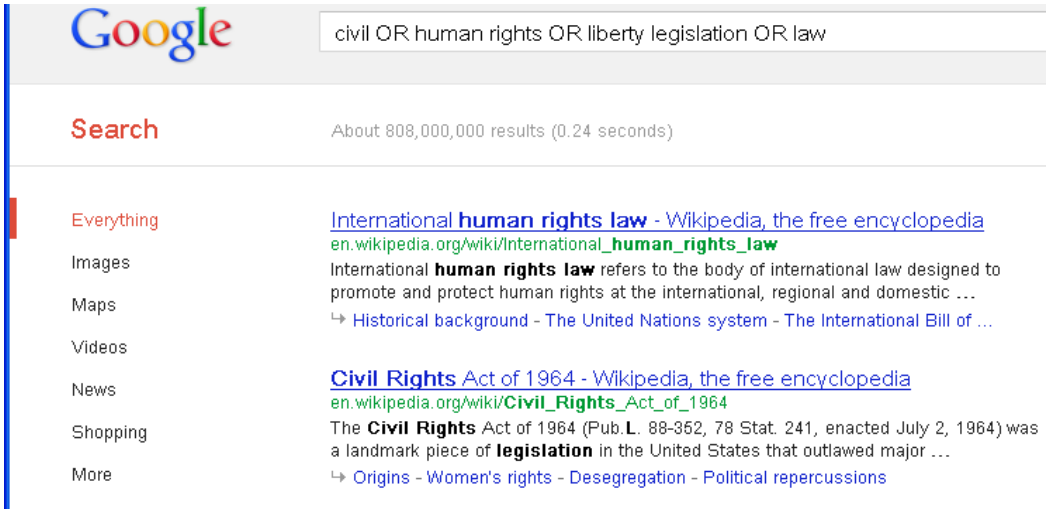
The OR operator is the most expansive search and should be used only to elevating the relevance of one particular variant form over all others or to introduce acceptable synonyms that the user does not think that the search engine's thesaurus recognizes already. The civil rights legislation search yields an overwhelming 25.3 billion hits when combined with OR operator ([civil OR rights OR legislation]). But really such a search does not make sense since these terms are distinct concepts, not variant of each other. A better search query is [civil OR human rights OR liberty legislation]. In this example, the search engine will give higher ranking to pages that contain the words "civil" or "human" and "rights" or "liberty" and "legislation."

To illustrate the impact that this search has on any variants not specified, add "OR law" to the end of the search. In the previous example, many of the hits on the first page referenced the Civil Rights Act of 1968 and the Human Rights Act of 1988. "Act" is

[2] Operators and More Search Help, Google Inside Search (28 February 2012), http://support.google.com/websearch/bin/answer.py?hl=en&answer=136861. And And MSDN (28 February 2012), http://msdn.microsoft.com/en-us/library/ff795645.aspx.

clearly a variant for legislation in Google. By elevating a different synonym, "law," the Google ranking system pushes all those results with "act" lower on the results list. The second hit is still to the Civil Rights Act of 1964, but not because of the word "act." "Act" is no longer highlighted in the results list. It is retrieved because of the prominence of "legislation" on the page.

No variants specified for "legislation"



"Law" specified as preferred variant for "legislation"

Domain or site specific (site:)

Experts also recommend using a site specific search. This advice is helpful if the site does not already have internal search capability, but many do. Going directly to a site and using its navigation and native search functions is often faster than googling. One useful application of a site-specific search is to find professional contact information. Many web sites hide (wittingly or not) staff directories. In these cases, a search that includes a firm's URL and the person's name might be the most efficient way to retrieve current email addresses or business phone numbers: [Suzanne Graham site:law.uga.edu].

This is a good time to mention the "invisible web." Think about the contents of Lexis, Westlaw or Bloomberg. These documents exist on the web, but one cannot do a search in a general search engine to find them. The indexing only goes so deep into a web site, and regardless of architecture, Google only indexes the first 101 KB of a Web page and about 120 KB of PDFs. Search engines also do not include proprietary data or any documents hidden behind a log-in/password. The point here is that one can do a site limit search and still come up empty.

File-type specific (filetype:)

Another tip is to search for specific types of files. If only a PDF (.pdf) or Flash (.swf) video will do, then include those details as part of the initial search, so millions of web pages (.html) do not clutter the results list. For example, to find PDF vesions of the Civil Code of Georgia, type ["filetype:pdf Civil code Georgia]. To find online video clips of Dr. Martin Luther King Jr., try [filetype:swf "Martin Luther King"]. The full list of common file types that Google indexes is available under the Webmaster Tools page at: http://support.google.com/webmasters/bin/answer.py?hl=en&answer=35287

Novelty non-search features

Google and Bing are about more than the search, they are about connecting users to information. To this end, the basic search box also can be a calculator, a clock, a currency converter, a dictionary, a telephone book, weather reporter, and a recipe book. One can type an equation directly into Google and Bing and get an answer, not just a list of links. The operators are fairly predictable and both search engines can now handle common symbols like the basic arithmetic operators and currency symbols.

- Calculator: [(4+5)/2*(3.5-2)=]. The "equals sign" is essential in Bing, but not in Google.
- Clock: [Time in Bismarck]. Defaults to most prominent city, so including the state is helpful. Typing in a state or country with multiple time zones retrieves a nice chart of major cities within the region with their times [Time in Russia]. Bing provides the full date as well. Google provides the day of the week only.
- Currency converter: [$150 in euro]. Rates are updated daily. Citibank N.A. provides exchange information for Google's currency converter.
- Dictionary: [define:digest]. Both Bing and Google have their own dictionaries that power these searches, but both offer links to other dictionaries (Black's Law did not make the cut). As a bonus, the list of synonyms, provided as part of the entry, gives the user the main variants that engines use in searches.
- Phone book: [706-555-5555]. This service is useful as a reverse look-up tool (when the number is known but not the party).

- Weather reporter: [weather in Harrisburg, PA]. Both Google and Bing pull from weather.com, but they also offers links to AccuWeather, the Weather Underground, and iMap Weather.
- Recipe book: [“bell pepper” hamburger “cheddar cheese”]. Both engines are rich with recipes and pull those results to the top. Add “easy” or “quick” to the search string and review the realistic options for after work.

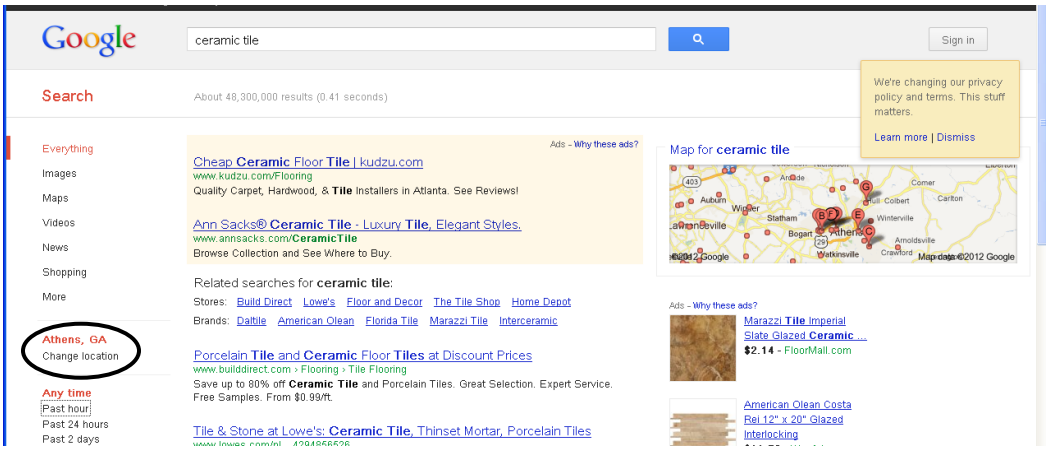
Out-of-the-box help

The search engine developers realize that few people will commit to memory how to do these specialized searches. Users do not need to remember exactly how, only that they can. Google and Bing provide user-friendly icons and labels on the main search screen to help remind users.

When a targeted search is required, most search engines offer more advanced parameters to help a user filter out the chaff and zero in on the best pages. Google recently made the advanced search screen harder to find. It no longer appears beside the search box, but instead is only visible after an initial search is run (down the left side of the screen). To start with an advanced search in Google, the searcher needs to “google” for “Google advance search,” click the link to go directly to the page, and then bookmark it or save it for future reference.

The “I’m Feeling Lucky” button takes one directly to the first result, not to the list. Wikipedia pages frequently appear as the top result. Clearly, some special arrangement has been worked out or the Wikipedia designers have unlocked a key code in the Google webmaster tools

Also be aware that the searches are customized by past search history and current geographic location. One can hope that Google and Bing are doing just what they claim: using the data they collect to improve the user experience. It is easy to find examples of how this is true. Google displays the city are in on the left side of the screen and in searches for commercial services, first page of results will contain several paid advertisements and unpaid listings for companies in the vicinity. Google Map also provides a handy map showing where these businesses are located.



Google knows where your computer is

While much of Google's relevancy ranking magic is not public knowledge, we do know how Google finds web pages. It tracks new web pages in three stages: first is crawling. Googlebots crawl billions of pages and determines how much of the sites to index and notes links on the pages. Next Google compiles an index based on words within the text of the page of tucked into the code. The last part is indexing.[3]

[3] Google Basics, Google Webmaster Tools (27 February 2012), http://support.google.com/webmasters/bin/answer.py?hl=en&answer=70897

Searching without Search Engines

Bookmarking a useful page is one way to bypass the search engines. Once you find an informative web site, don't let it go. Internet browsers, like Internet Explorer, Mozilla Firefox, and Google Chrome, allow users to save pages as favorites (and store them in their log-ins) for a consistent web experience, regardless of the computer being used. If the site is truly the one visited most frequently, consider making it the browser's homepage (or first page that loads when the browser is opened).

Many law-related portals (massive sites that bring together a huge amount of targeted, vetted information) exist on the web. One of the best portals is Legal Information Institute (www.law.cornell.edu). FindLaw for Legal Professionals (lp.findlaw.com), and THOMAS (thomas.loc.gov) are powerful starting points for legal information online. LII is broad in scope with federal and state primary law and regulations, Supreme Court decisions, and some commentary. Findlaw offers free caselaw, newsfeeds and legal articles. THOMAS is a service of the Library of Congress that tracks legislative bills and various other activities of congressional representatives. Rather than starting with Google or Bing, consider starting a search in one of these respected megasites.

A third alternative to a general web search or a single site homepage is a customized personal portal or dashboard. A dashboard is a uniform, tailored page that a user builds and designs just for himself or herself. It's a virtual office with links to useful online resources (a virtual bookshelf), RSS feeds (virtual mail) and electronic mail (virtual telephone). These pages exist only for the user that created them. No one could

"Google" for a personal dashboard and find it. They are log-in protected and not crawled by search engine bots.

Several free web services provide a basic platform and help users create personalized home pages. iGoogle, Netvibes, My Yahoo!, and Windows Live Personalized Homepage are all players in this personalized homepage market and all are worth consideration.

The page aesthetics can be customized, but most importantly, so can the "gadgets" or the contents of the page's various frames. Admittedly, the basic configuration of iGoogle (weather, date/time, YouTube, CNN, Gmail, and Wall Street Journal) appeals to more to the casual surfer than to the legal professional so be prepared to do some work. To begin customization, start with the "Add Gadgets" button at the top left. The long list of pre-fabricated gadgets is searchable, but the search [Law Georgia] does not offer many helpful resources, however, JURIST and Law.com appear. Arguably, at present the real benefit is not in the available gadgets but in the ability to add RSS feeds and to create and share gadgets. The promise of customized internet dashboards is still emerging, and this approach is not for everyone. But if a user is especially tech savvy, then developing great gadgets for online law resources could really boost his or her Internet search efficiency.

Privacy

All of a searcher's online activities leave a virtual fingerprint, and the major search engines collect this data to tailor search results. Google made major waves this winter when it announced a new privacy policy, effective March 1, 2012. On one level

Google earned praise for consolidated its 60-plus privacy-related policies into one unified statement. Less well-received is Google's announcement that it would "combine personal information from one service with information, including personal information, *from other Google services*" [emphasis mine].[4]

Google claims it collects this data to protect Google and users and to offer users more tailored content, results, and advertising. Google Search had been harvesting this information, but the full suite of Google products had not merged this data systemically into such a holistic picture of a user before. Now, Google Search will share with iGoogle and Gmail and Google calendar, merging multiple user profiles.[5]

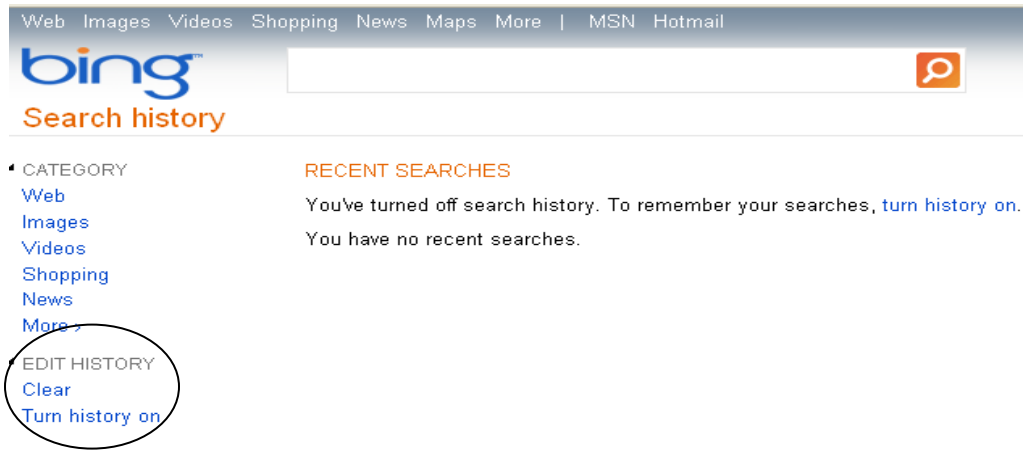
Reactions to the new policy statement range from cries heresy to a matter-of-fact resignation to this is how things have always been—but now more so. All activity on the Internet requires a level of trust: clicking this button will not download a virus, opening this page will not show something disturbing, entering account information will not end in credit card theft, creating a user profile will not increase email spam. As yet, there are no guarantees.

Expect much more on this discussion of online privacy and how we legally define "reasonable expectation of privacy." In two recent concurrences, Justices Alito and Sotomayor consider at length the impact on the Fourth Amendment of a digital age in

[4] Preview: Privacy Policy, Google Policies & Principles (7 February 2012), http://www.google.com/intl/en/policies/privacy/preview.
[5] Ibid.

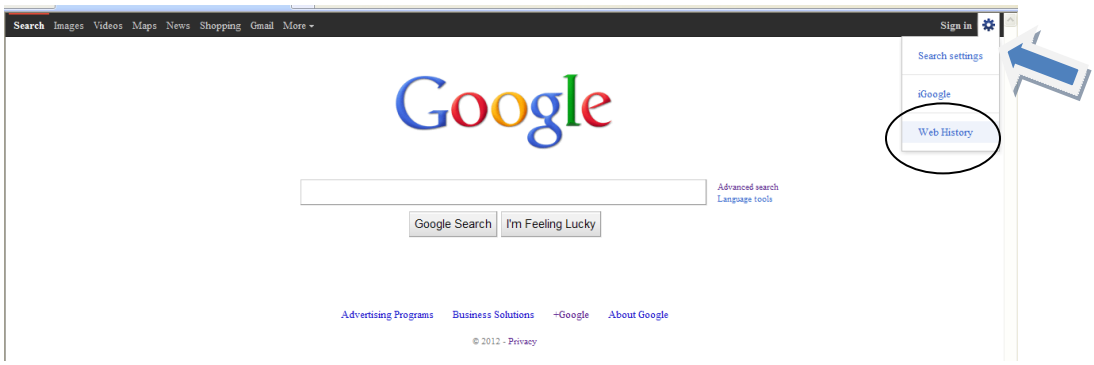
which "people reveal a great deal of information about themselves to third parties in the course of carrying our mundane tasks."[6]

In response to the privacy controversy, the last tip is how to opt out of the tracking of internet searches and storage of search history. In Bing, click on "Search History," which appears prominently in the tool bar. Under "Edit History," select "Clear" and "Turn history off." In Google, the Google Dashboard allows users to configure many settings, privacy is only one. The dashboard is accessible from the top right corner of the Google screen. In Internet Explorer, the icon looks like a wheel, and in Chrome it's a wrench. In Firefox, there is no icon, but users can visit google.com/history/optout . Once on the Dashboard, select "Web History" and then click on "Enable customizations based on search activity" to change the first word to "Disable." Google will not use future activity to tailor ads or search results.[7]



[6] United States v. Jones 565 U.S. ___(2012) (Alito J., concurring). And United States v. Jones 565 U.S. ___ (2012) (Sotomayor, J., concurring).

[7] Turn Off Search History Personalization, Google Inside Search (24 February 2012), http://support.google.com/websearch/bin/answer.py?hl=en&answer=54048.



This step only sets up blinders for Bing or Google. It does not erase the Internet browser's cookies or search history. The browsers also track search queries and page visits. So, don't take the company computer anywhere that Mom would not be proud of.

Conclusion

Driven by a demanding market, modern search engines are wonderful inventions that continue to improve their effectiveness. Often the results of a basic search are fast and good enough, but when patience or a two-inch screen demand extra brevity and precision remember these tips and tidbits. Also keep in mind privacy concerns when setting up online profiles and log-ins, and make sure that the search engine's data collection practices are understood and accepted.