November 19, 2005

The Trials and Tribulations of Internet Research

Lee F Peoples, Oklahoma City University

Available at: https://works.bepress.com/lee_peoples/12/
The Trials and Tribulations of Internet Research

By Lee F. Peoples

Internet use is pervasive in all areas of modern life and the law is no exception. Stories abound of judges and their clerks using the Internet to verify facts, delve into the backgrounds of parties before them and challenge information presented by attorneys. Some scholars argue that lawyers must search the Internet to satisfy ethical obligations of competence and due diligence. At least one jurisdiction requires attorneys who rely on information obtained from the Internet to “take care to assure that the information obtained is reliable.” Are you confident that your Internet searching skills are up to par?

ASSESSING THE QUALITY OF INFORMATION

Information quality is something we often take for granted when working with known authoritative sources. More care is required when venturing away from familiar legal publishers and Web sites. Genie Tybruski, a law firm librarian at Ballard Spahr Andrews & Ingersoll LLP in Philadelphia, has put together an excellent tutorial detailing how to evaluate the quality of online information, available at www.virtualchase.com/quality. Some of her tips are fairly obvious and include evaluating the expertise of an author and assessing the objectivity and accuracy of information. Other tips involve more skill and are worth examining in detail.

WHO IS PROVIDING THE INFORMATION?

Anyone can publish a Web site on any subject. Some obvious techniques for establishing who is behind a Web site including reading any “About Us” or “Disclaimer” pages you find at a Web site and investigating the people behind it using Martindale Hubbell or other relevant directories.

A more savvy technique is to pay attention to the Web site’s address or URL (uniform resource locator). For example, take the bar association’s URL, www.okbar.org. The “.org” ending indicates that the site belongs to an organization as opposed to “.com” (usually a for-profit entity), “.gov” (government), “.au” or “.uk” (examples of Web sites originating in other countries, Australia and the United Kingdom in this example). To find out who is behind a Web site, submit the URL to a domain name lookup service like WhoIs, www.whois.net, or CallingID, www.callingid.com. These free services tell you who the URL is registered to. This information comes in handy when detecting spoof or fake sites like www.gatt.org, which looks exactly like the real
World Trade Organization site www.wto.org. Submit gatt.org to Whois and you get the name and address of a private individual in Brooklyn, New York, not the WTO secretariat in Geneva.

ESTABLISH THE DATE OF PUBLICATION

Timely information is invaluable in legal and factual research. LexisNexis and Westlaw tell you how often the information in their databases is updated. When conducting general Internet research, there is no reliable way to verify when Web site content was last updated. Copyright dates are of little help as they are often just a range of dates. Some Web sites include a “page last updated” notation but these are only as reliable as the person editing the content. Google’s advanced “search by date” function is actually searching by the date Google last indexed the site and not the date the information within the site was updated. For more on this conundrum, see the article “It’s Tough to Get a Good Date with a Search Engine” available at http://searchenginewatch.com/searchday/article.php/2160061.

ADVANCED INTERNET SEARCHING TIPS

Google, the ubiquitous search engine that is at once both noun and verb, will be the focus of this section. Most of the advanced searching you can do with Google is explained on their Advanced Search screen, accessible by clicking the “advanced search” link immediately to the right of the search field on Google’s home page. To go beyond the tips discussed in this article check out the book Google Hacks by Tara Calishain and Rael Dornfest.

BASIC BOOLEAN

Google puts an “and” between search terms by default. If you want an “or” between terms you must type it yourself, the | symbol also works. A useful and often neglected tip is to enclose phrases, citations or anything else you want searched as a group in quotation marks. This trick has saved the day for me on a number of occasions. Try searching for old English legislation using the following citation: 6 Geo. I, c. 18., first without quotes and then with them. The difference is dramatic. Quotations are also an excellent way to fill in an incomplete citation that you can’t complete using traditional tools. Back to our English legislation example, 6 Geo. I, c. 18. This citation is incomplete and should include a year. When we surround the cite with quotation marks, our search results contain several hits including a year. It is good practice to then take the citation you found on the Internet and verify it using reliable print or online sources.

EXCLUDING TERMS

Legal terms often have other meanings in everyday discourse. Try searching for information on RICO and you will get tons of hits on Puerto Rico. Google lets you narrow your results by excluding terms using a – symbol placed directly before the term to be excluded. A search for: “RICO –Puerto” will return Web sites about the RICO laws but not Web sites about Puerto Rico.

SYNONYMS

Including synonyms is always good practice when searching. Placing the ~ sign before a term tells Google to look for your term and all synonyms of it. Instead of just searching for “labor law” try searching for “~labor law.” Your results will be more complete and will include Web sites on employment law in addition to Web sites on labor law.

NUMERIC RANGE

Google allows you to search for pages containing a range of numbers. If you are looking
for information about bankruptcy crimes, searching for “18 USC 151...157” returns pages related to the range of crimes defined by these statutory sections and is more efficient than a search for each individual section.

GOOGLE AS A DICTIONARY, PACKAGE TRACKER AND MORE

Enter “define: your term here” and Google will return definitions of your term. Some online glossaries of legal terms are searched but unfortunately Blacks is not included in this function yet. Enter the following numerical data into Google, and it will search the relevant database for your results: airline name and flight number, UPS and FedEx tracking numbers, VINs, UPC codes, patent numbers, FAA registration numbers, stock ticker symbols and FCC equipment IDs. More information about these features is available at: www.google.com/intl/en/help/features.html.

MORE ADVANCED TRICKS

From the Google Advanced Search page, you can limit your searches to a specific file type. For example search only for PowerPoint presentations involving personal injury damages. Another very useful advanced feature is the search only in the domain or site function. You can limit your search to specific domains, .edu or .gov, or to specific Web sites. This feature is particularly useful for searching specific Web sites that lack search functionality or have very poor search features. For example, searching the irs.gov site using this feature returns more relevant results than searching the IRS site using their internal search function.

EVEN MORE

Click the “More” icon just above and to the right of the search field on the Google homepage to delve even deeper. Google Images searches over two billion images, many of which could be useful in litigation. Click Google Labs, labs.google.com, and you can experiment with cutting edge applications still under development. My Search History, www.google.com/searchhistory, allows you to recreate your past Google searches similar to Westlaw’s “Research Trail” function and the LexisNexis “History” utility. Google Earth, www.earth.google.com, is an amazing download that provides satellite imagery of the world allowing you to “fly from space to your neighborhood” and navigate with your mouse or search by specific address.

WHAT’S WRONG WITH INTERNET SEARCHING?

The savvy Internet searcher must also be aware of the shortcomings of Internet research and search engines. Search engines aren’t so great at returning meaningful results for commonly used or procedural terms. Try an Internet search (or LexisNexis or Westlaw search) for “burden of proof” and you will be up to your eyeballs in a mass of information that is mostly irrelevant to your particular situation. Searching for factual information about things that can be purchased is tricky. Most Internet search engines accept sponsorship and display sponsored products prominently at the top of search results. The Internet is only about 20 years old, and a great deal of information prior to the mid-1990s is simply not available online. Internet searches miss an entire universe of knowledge contained in books and journals that is not, and may never be, available online. Google has two projects underway aimed at remedying this problem. Google Print, www.print.google.com, is a collaborative effort of five of the world’s largest libraries to make millions of books avail-
able online. Works related to law have been slow to appear in the collection. Another drawback is that generally only works published before 1923 will be available in full text because of copyright concerns. Google Scholar, www.scholar.google.com, includes scholarly articles from publishers' Web sites and databases. A traditional Google search also searches Google Print and Google Scholar and results are returned at the top of your Google search results.

SEARCHING THE DEEP WEB

A common misconception is that search engines search everything that is available online. There are over 600 billion pages of information on the Internet but even the best search engines like Google only search about 8 billion of those pages, only one percent of available information on the Internet. The problem is twofold. First, search engines only search for words contained within pages that have been indexed and not all available pages. Indexes are created by "crawling" through pages, following hyperlinks, and indexing the pages they lead to. This random process excludes documents not hyperlinked to anything within existing search engine indexes. The second reason is that search engines simply can't keep up with the explosive growth of the Internet.

The solution is to search what has been referred to as the hidden, deep or invisible Web. These terms are misnomers as the information is in fact not hidden, deep or secret at all, it is simply irretrievable with conventional search engines. A strategy for searching the hidden Web is to search a directory or portal of information that has not yet been indexed by traditional search engines. Try www.completeplanet.com, a directory of over 70,000 specialty databases grouped by subject, or www.turbo10.com and www.profusion.com, sites that let you create your own search engine to search hidden Web databases. Two great articles on searching the hidden Web are Genie Tyburski's "The Invisible Web: A Brief Note and Bibliography," available at: www.virtualchase.com/iweb/handout.html and Marcus P. Zillman's "Deep Web Research 2005," available at: www.llrx.com/features/deepweb2005.htm.

THE INTERNET ARCHIVE — SEARCHING FOR WHAT ONCE WAS

Web pages are dynamic creatures. With the click of a webmaster's mouse what is online today can be gone in an instant. The solution is the Internet Archive, www.archive.org, a collection of over 30 billion archived Web pages. To access the archive, enter the URL of a Web site into the Internet Archive's Wayback Machine search engine. The results allow you to view different versions of the page you entered from 1996 to the present day.

The Internet Archive is particularly useful to lawyers. The Bureau of Indian Affairs Web site, www.doi.gov/bureau-indian-affairs.html, has been down for several years because of litigation. Savvy searchers can use the Internet Archive to access the multitude of materials available at archived versions of the site. The Internet Archive can also be useful in litigation, for example to determine what a Web site contained at specific points in time. It is commonly used in trademark infringement cases to see how a mark has been used over time.

Blogs about law are frequently called blawgs.
SEARCHING BLOGS

Blog is short for Web Log, which is an online journal that is frequently updated. Blogs about law are frequently called blawgs. As of July 2005, there were over 70 million blogs in existence and somewhere between 30,000 and 40,000 new blogs are created each day. Lawyers have begun using blogs as marketing tools, to put their spin on trial publicity, to obtain information about adverse parties and to stay on the cutting edge of legal developments. Advanced Internet searchers can’t ignore the wealth of information available in today’s blogs. Specialized search engines for discovering information contained in blogs include Technorati, www.technorati.com, which searches over 20 million blogs and Google Blog Search, blogsearch.google.com. To find out more about how lawyers are using blogs, read Jim Calloway’s articles “Was 2004 the Year of the Blog?” and “Of Blogs, Bloggers, and Blawgs,” both available at www.okbar.org.

CONCLUSION

Advanced Internet searchers are able to evaluate the quality of online information, search the internet efficiently and locate material that eludes others. Becoming an advanced Internet searcher requires just a small amount of time, but it yields big results for you and your clients.

5. Google recently announced that it would temporarily stop scanning books for the project until an agreement over copyright issues was reached with publishers. Margaret Kane, “Google Pauses Library Project,” CNET, Aug. 12, 2005, http://news.com.com/Google+pauses+library-project/2100-1025_3-5830035.html. Two of Google’s rivals, MSN and Yahoo, recently announced they have similar projects underway.
6. A law firm that used the Wayback Machine for just this purpose in a trademark case is being sued by the defendant in the case who claims that accessing the defendant’s old Web pages using the Wayback Machine was unauthorized and illegal. The defendant claims violations of the Digital Millennium Copyright Act and the Computer Fraud and Abuse Act and has also named the Internet Archive as a party. The case is Healthcare Advocates, Inc. v. Harding, Earley, Follmer & Frailey, No. 2:05-CV-03294-RK (E.D. Penn., filed July 8, 2005); See also Tom Zeller Jr., “Keeper of Expired Web Pages Is Sued Because Archive Was Used in Another Suit”, N.Y. TIMES, July 13, 2005, at C9.

LEGAL AID SERVICES OF OKLAHOMA

Legal Aid Services of Oklahoma is hiring an attorney for the Family Law Unit in its Tulsa Law Office. The attorney will be responsible for handling an active family law caseload. Legal Aid is a nonprofit provider of civil legal services for low-income people.

This is a staff attorney position for entry level attorneys (0-5 years) with a minimum salary of $30,000 DOE plus generous fringe benefits including health insurance, pension plan, student loan repayment program, etc. Application forms are available at www.legalaidok.org or LASO Operations, 2901 Classen Blvd., Suite 110, Oklahoma City, OK 73106. Please send LASO application form, resume, and writing sample to the above address. Applications will be accepted until November 28, 2005 and thereafter until filled. EOE