

## Choosing the Right Tool for Internet Searching: Search Engines vs. Directories

*Teachable Moments for Students ... is designed to provide information that can be used for quick and accessible answers to the basic questions that are frequently asked of librarians and those involved in teaching legal research and writing. These questions present a “teachable moment,” a brief window of opportunity when—because he or she has a specific need to know right now—the student or lawyer asking the question may actually remember the answer you provide. The material presented in this column is not meant to be an in-depth review of the topic, but rather a summary of the main points that everyone should know. It is a companion to the Teachable Moments for Teachers column that gives teachers an opportunity to describe a special moment of epiphany that changed their approach to presenting a particular topic to their students. Readers are invited to submit their own “teachable moments for students” to the editor of the column: Barbara Bintliff, University of Colorado Law Library, Campus Box 402, Boulder, CO 80309, phone: (303) 492-1233, fax: (303) 492-2707.*

**By Joanne Dugan**

*Joanne Dugan is Assistant Director of Public Services for the University of Baltimore School of Law Library in Maryland.*

The conventional wisdom about today’s law students is that they are technologically savvy and sophisticated consumers of Web-based information. Well, yes and no. It certainly is true that they have access to much more sophisticated research tools than students did a decade ago, and that the Web has much more content to search than it did 10 years ago. As a result, it has become unusual for a student to say that she or he “can’t find anything” on a topic. That doesn’t mean that today’s student is necessarily a better researcher or that the information located is of better quality. In fact, one of the ironies of more sophisticated search tools is that it allows researchers to

become less sophisticated in their use of the tools.<sup>1</sup>

Even within the limited scope of Internet searching, there seems to be a divide among the research savvy and research novices. Many of our law students have learned their Internet research skills in an atmosphere where “good enough” was acceptable. One of the first lessons we must teach them is the difference between “good enough” and “best quality”; between information and research. While the former is perfectly acceptable for much of their everyday lives, the rigors of legal scholarship and legal advocacy require the latter.<sup>2</sup>

### **Search Engines vs. Directories: Brawn vs. Brain**

The recent spate of news reports about the relative size of Yahoo!’s database and Google’s database<sup>3</sup> got me thinking: Does anyone use directories to find information on the Internet anymore? More specifically, do law students use directories? Do they know the difference between a search engine and a directory? I suspect that many of them don’t, and yet it’s an important distinction to make when planning an effective Internet search.

<sup>1</sup> “Today’s undergraduates are generally far less prepared to do research than were students of earlier generations, despite their familiarity with powerful information-gathering tools.” Patricia Senn Breivik, *21st Century Learning and Information Literacy*, Change, March/April 2005, at 22.

<sup>2</sup> Several recent articles discuss the tension between these two standards. See, e.g., Steven Bell, *Don’t Surrender Library Values*, *Library Journal*, May 15, 2005, at 79; Todd Miller, *In Defense of Stupid Users*, *Library Journal*, March 15, 2005, at 58; Brian Kenney, *Googlizers vs. Resisters: Library Leaders Debate Our Relationship with Search Engines*, *Library Journal*, December 15, 2004, at 44.

<sup>3</sup> John Battelle, *Search Engines Trade Competing Claims*, NPR Morning Edition, August 15, 2005.

“One of the first lessons we must teach them is the difference between ‘good enough’ and ‘best quality’; between information and research.”

“Online searchers rely on two concepts to evaluate the results of a search: precision and recall. Unfortunately, high precision usually comes at the cost of low recall, and vice versa.”

Boiled down to the essentials, Internet search tools are either full-text search engines, topical directories, or some hybrid or combination of the two. Full-text search engines, such as Google, use sheer computing power to import huge volumes of information from the Web, create a database that indexes every word imported, and find matches to user queries.<sup>4</sup> Topical directories, such as the Yahoo! Directory,<sup>5</sup> are more selective in importing information into their databases. The databases are organized into hierarchical subject categories, and Web sites that are selected for inclusion are then indexed and categorized by human editors.<sup>6</sup>

The dichotomy of full-text vs. topical organization is a familiar one to legal researchers. We have been grappling with it ever since the electronic legal research services introduced full-text databases back in the 1970s,<sup>7</sup> giving us a choice between digests and indexes and full-text searching. Many researchers were tempted by the freedom that full-text searching gave them to customize their own research but found that, in many cases, the results were either overwhelming or not well tailored to their needs. Even with the power that full-text searching provided, there was still value in the analysis and categorization provided by human editors. Even the most skilled legal researcher received different, if overlapping, results when using both online and print resources to research an issue. Similar considerations come into play when deciding what tools to use to search the Internet.

---

<sup>4</sup> Alan M. Schlein, *Find It Online: The Complete Guide to Online Research*, 79–83 (3d ed. 2003).

<sup>5</sup> When Yahoo! first arrived on the scene in 1994, it was a pure directory, and it remains the quintessential subject directory on the Web <[www.yahoo.com/info/misc/history.html](http://www.yahoo.com/info/misc/history.html)>. In 2004, Yahoo! introduced Yahoo! Search, its own proprietary search engine that competes directly with Google. Press Release, Yahoo! Media Relations, *Yahoo! Introduces More Comprehensive and Relevant Search Experience with New Yahoo! Search Technology* (February 18, 2004) <[docs.yahoo.com/docs/pr/release1142.html](http://docs.yahoo.com/docs/pr/release1142.html)>. Yahoo!'s home page now emphasizes the search engine, but the directory is still active and available through a link on the home page.

<sup>6</sup> Schlein, *supra* note 4, at 89–90.

<sup>7</sup> Roy M. Mersky & Donald J. Dunn, *Fundamentals of Legal Research*, 538–539 (8th ed. 2002).

## Precision vs. Recall

The choice of search engines or directories as Internet search tools depends on the type of results that are desired. Online searchers rely on two concepts to evaluate the results of a search: precision and recall. Unfortunately, high precision usually comes at the cost of low recall, and vice versa. Precision measures the proportion of relevant materials in a search result. Let's say you were doing research on the *Miranda* warning for a criminal procedure memo. A measure of the precision of your search would be the proportion of search results that talk about the right to remain silent compared to the proportion of search results that feature 1940s entertainer Carmen Miranda. Recall, on the other hand, measures how complete your search results are compared to the total amount of relevant information available. Let's say you were researching the statute of limitations for medical malpractice in all 50 states. If your search results only included 30 states (because the other states use the term “statute of repose” or “limitation of actions” instead of “statute of limitations”) your recall would be lower than if you had found all 50 states.

Just as with full-text databases on computer-assisted legal research (CALR) services, full-text search engines such as Google excel in quantity of search results, or recall. Unless you are searching for a really esoteric term, you will probably end up with a big results list, and there's likely to be some useful information in the list. Unfortunately, the precision of the results is probably not optimal, so you may end up wading through a lot of irrelevant or low-quality information in order to find the high-value information.<sup>8</sup> Directories such as Yahoo! are more like the West Key Number System\*: a human editor serves as intermediary between the searcher and the information, selecting and categorizing material under a specific subject heading. The result is a

---

<sup>8</sup> Most search engines try to minimize the amount of winnowing the searcher needs to do by using some sort of relevancy ranking. This is a good thing, since studies show that the average searcher scrolls through only 1.8 results pages per search. Pew Internet & American Life Project, *Data Memo on Search Engines*, August 2004 <[www.pewInternet.org/PPF/r/132/report\\_display.asp](http://www.pewInternet.org/PPF/r/132/report_display.asp)>.

smaller set of results than you would get from a search engine, but the search results are likely to be highly relevant.<sup>9</sup>

### How Do You Decide?

Now that we know the basics of how search engines and directories handle Internet searching, the next logical step is to figure out which tool works best in what situation. That answer is going to depend on the purpose of the search, the subject matter being researched, the searcher's familiarity with the subject, and the searcher's skill at online searching.

If the searcher is just beginning to research in an area that is unfamiliar to her, her best bet will probably be to use a subject directory rather than a search engine. There are a couple of reasons for this: first, the searcher needs some background information on the topic and may not know enough about the subject to choose effective keywords. The structural hierarchy of a subject directory allows her to drill down from broader to more specific topics, suggesting vocabulary as she proceeds. Second, while she's in the process of gaining background knowledge, it's more time-effective to rely on the judgment of the directory's editors about the most authoritative and complete Web sites on a particular topic. Because of the selective nature of directories, there's some level of quality control built in.

Once our hypothetical researcher has some background on her topic, she's better equipped to harness the brawn of a full-text search engine. Search engines work best when the searcher has sufficiently defined the limits of her research needs and identified specific keywords to capture the information. Because the brute strength of search engines leads to poor precision, searchers tend to get better results when they use unusual or technical terms. This can pose a problem in law,

since legal concepts are often expressed in ordinary English rather than the technical jargon more commonly found in science or medicine. A savvy researcher can leverage the advanced features of the search engine to combine phrases, search fields, and use Boolean operators to increase the precision of the search results.

A search engine is also the best tool if the researcher is concerned about the completeness of her search results. Directories are selective, which is both their strength and their weakness. Because judgments about what to include are being made by human editors, directories are subject to bias in the selection process. Search engines are also generally more current than directories, since the selection process of directories is more time-intensive than the automated harvesting of information done by the search engines.

Just as with searching on the CALR systems, the best results are returned if the searcher takes advantage of the strengths of both search engines and directories over the course of her research project.

### A Final Warning

Most legal research instructors are deeply skeptical of research conducted solely on the Internet, and for good reason. We know that many of the best gems of information are still only available in print. By the same token, we ought to be very skeptical of Internet research that relies solely on Google or Yahoo!, or even a combination of the two. Many Internet users believe that they are searching the entire Web when they use these services. In fact, they are searching only the portion of the Web that has been captured and indexed by these services. Each service maintains its own database, and each has its own closely guarded techniques for processing queries and ranking results. A wise Internet searcher will never rely exclusively on just one search tool. And a brilliant Internet searcher will search for information outside the world of search engines and directories by learning about the invisible Web and the deep Web. But that's a story for another day ...

© 2006 Joanne Dugan

“ [T]he best results are returned if the searcher takes advantage of the strengths of both search engines and directories over the course of her research project. ”

---

<sup>9</sup> Just as a search engine uses relevance ranking to compensate for the lack of precision inherent in a full-text search, directories compensate for the lack of recall inherent in their smaller search results because the human editors can recognize and include relevant sites even if they use different vocabulary to discuss the same concepts.