



THE LEGAL RESEARCH CORNER

Searching Smarter: Finding Legal Resources on the Invisible Web

by Bonnie Shucha

By now, most attorneys have discovered that the Internet can be a powerful tool for legal research. Increasingly, Web search engines like Google have moved up in the ranks of computer-assisted legal research tools alongside more expensive resources such as LexisNexis™ and Westlaw®. Even some judges are using the Web to check facts and statistics presented by attorneys and reporting their findings in written opinions.¹

As the Web makes its way into the courtroom, it is important that legal practitioners know how to search it effectively. Unfortunately, this is not always the case. A new study reveals that while professionals are spending increasingly larger amounts of time doing computer-based searching, most are dissatisfied with their search experience.²

It is estimated that most searchers locate only 0.03 percent—or 1 in 3,000—of the Web pages available to them.³ Although such results may be due, in part, to a poorly constructed search, a large portion of the blame also falls upon the search engine itself. Even the most experienced searcher, using the largest search engines, can access only about 16 percent of all Web content. Why? Because 84 percent of the information available on the Internet is found only on the “invisible Web,” also known as

the “deep Web,” and is not searchable using a general search engine such as Google.⁴

By recognizing how the invisible Web differs from other Web content, you will understand how to alter your search strategies to find this information in a time-efficient manner. This article investigates the nature of the invisible Web and offers strategies for locating invisible Web content.

What is the Invisible Web?

In order to understand the concept of the “invisible Web,” it may be helpful to first explore the nature of the “visible Web.” A visible Web page is one that exists in “static” or unchanging form until its creator alters it. In this way, it is similar to a document that you might create in a word processor. Both types “physically” exist as files on a computer: the word-processed document might be saved as a .doc or .wpd file on your hard drive, whereas a visible Web page might be stored as a .htm or .html file on a Web server.

These static Web pages are considered *visible* because standard search engines are able to index them, or display them as search results. Most search engines index new documents in one of two ways: (1) by using automated “spiders” or “crawlers” to follow links from other documents that are already indexed, or (2) when a Webmaster registers a Web page. Because this method of indexing documents is well established and relatively inexpensive, most search engines draw primarily upon visible Web content.

As you now know, there is another type of Web content known as the “invisible Web.” Most invisible Web content is considered “dynamic” because it consists of bits of information that are stored in a database and pulled together on-the-fly into a Web page at your request.⁵ Invisible Web pages don’t actually exist until you submit a query to the database containing the

This department, published quarterly, is sponsored by the Colorado Association of Law Librarians (“CoALL”) to assist attorneys with common problems in legal research. Readers interested in submitting research questions may send them to: CoALL, The Legal Research Corner, at <http://www.aallnet.org/chapter/coall>; or to: “The Legal Research Corner,” c/o Arlene Abady, Managing Editor, The Colorado Lawyer, 1900 Grant St., Suite 900, Denver, CO 80203-4336 or e-mail: aabady@cobar.org.

Members of CoALL will attempt to answer as many questions as possible, either individually or as part of this department. The information provided in this space is for educational purposes only and is not intended as legal advice. No endorsement or recommendation is made of any product named in this department. Department contributors are CoALL members and include Andrea Hamilton, Wanda McDavid, Mariann Storck, and Patty Wellinger. For more information about CoALL, see <http://www.aallnet.org/chapter/coall>.

Bonnie Shucha is the reference and electronic services librarian at the University of Wisconsin Law Library—bjshucha@wisc.edu. She is past president of the Law Librarians Association of Wisconsin. Excluding Colorado material, which was added by CoAll, this article first appeared at 77 Wisconsin Lawyer (Sept. 2004). Reprinted with permission.

information and matching information is drawn together into a Web page. Usually, an invisible Web search is conducted via a specialized search interface, or search box, provided by the creator of the database.

This concept is somewhat similar to the mail merge feature in most word processors. In a mail merge, content is drawn from an outside data source, such as an Excel file, and inserted into a new, customized document. Like an invisible Web page, the mail-merged document did not previously “physically” exist as a stored file on a computer. Rather, both types are created at the point of need.

Because it is dynamic, or “physically” non-existent, most conventional search engines are unable to retrieve invisible Web content. Traditional methods of indexing, which are based on following links from other documents or Webmaster registration, are inadequate because they rely on the existence of a static file. Once most search engine spiders hit a database’s search form, they are forced to stop because user input is required. Conventional search engines are simply not capable of automatically generating that input in the form of a search. As one author notes, “It’s not that the information is really hidden or invisible. It’s there, freely available and waiting to be found. The problem is that general search engines are built in such a way that they just can’t go into [a] database and search the information contained [there].”⁶

Besides dynamic Web pages, other types of content are also considered “invisible.” Very recently created static Web pages are effectively invisible because search engine spiders have not yet had a chance to index them. It is estimated that it takes three to four months before a static Web page is indexed by a search engine.⁷ Password-protected information may also be considered part of the invisible Web because search engines are unable to access and index this content without the proper authorization. Such content might include information within subscription databases, such as LexisNexis and Westlaw or confidential business databases.

What Type of Content is (and is not) Freely Available on the Invisible Web?

Fortunately, the vast majority of invisible Web content, 95 percent, is publicly accessible, free information. And studies reveal that the quality of documents found on the invisible Web often exceeds those that are accessible via conventional search engines.⁸ This includes a vast amount of legal and governmental documents such as case law, statutes, bills, regulations, patents, briefs, census data, government reports, treaties, and

much, much more. A great deal of business and corporate data are also available on the invisible Web, including SEC filings, stock quotes, company profiles, etc. More general types of information can also be found on the invisible Web, such as address and phone number directories, flight schedules, dictionary definitions, maps, and more.

Searchers should be aware that much of the law-related information that is freely available on the invisible (and visible) Web are materials that are in the public domain. Despite assertions by some novice researchers that “everything is free on the Web,” there are certain types of content that are *unlikely* to be found at no cost on the Web. These include books and articles that are published for profit, public domain documents that have editorial enhancements, or other authoritative materials that are considered to be someone’s intellectual property.⁹

Tips for Finding Invisible Web Content

The first step in locating any type of information is considering where an authoritative source of that information might be found. It may be a print source, the Web (visible or invisible), a subscription database, a phone call, etc. If it is available from more than one source, you will need to consider what will be the quickest, most cost-effective way to obtain it.

If you do determine that the information you need might be available on the invisible Web, how do you find it? Fortunately, there is nothing magical about finding content on the invisible Web. It’s simply a matter of knowing where to look. Consider that:

1. A great deal of excellent legal and business information is freely available on the Internet.
2. Unfortunately, much of it is contained within databases and is, therefore, invisible or inaccessible to most conventional search engines.
3. The most effective way to access this information is using the database’s own search interface, or search box.
4. Fortunately, the search box is usually found on a static, visible Web page, which *is accessible* using a conventional search engine.

The following scenario may help illustrate: You have been asked to locate Wisconsin Statutes concerning livestock. You don’t have a copy of the Wisconsin Statutes in print, but you think that they might be available on the Internet. You go to Google and do a search containing the keywords “livestock statutes Wisconsin.” You find some interesting information about your topic and references to the statutes, but not the statutes themselves.

How to Find Information on the Invisible Web

1. If you believe that the information you need is available on the Web, consider whether it is likely to be “visible” (a static htm or html page) or “invisible” (dynamic content located inside a Web database). Invisible content can include, but is not limited to: case law, legislation, patents, census data, government reports, SEC filings, stock quotes, directories, flight schedules, and much more.
2. Locate a database that may contain the information you seek by searching your favorite search engine. Look broadly for the database itself, rather than for the specific information contained within it. Remember that the search interface page for a database is often visible, while the data contained within it is invisible. Or, select a database from a directory of invisible Web resources. See “Invisible Web Resources for Attorneys” in the accompanying box.
3. Once you’ve located a database that might contain the needed information, construct a search within it, using specific keywords that describe your topic.

Take a moment to reconsider the search. If you were doing the research using print sources, you would first locate a copy of the Wisconsin Statutes, then search the index for your keyword, "livestock." The same strategy applies when doing research on the Web. Because providers of legal and business information often publish their collections within "invisible" databases, it is more effective to first limit your search to find the provider's "visible" search interface page. Once there, you could query the database using your specific keywords. As one author notes, "often the key to the answer is not locating the answer itself as the first step, but locating the right database in which to search for it."¹⁰

Back at Google, you decide to try your search again, this time using just the keywords "Wisconsin Statutes." The very first item in your search results is the freely available Revisor of Statutes Bureau's ("RSB") Wisconsin Statutes database. In the RSB's search box, you proceed with your specific search for "livestock" and successfully locate the specific statutes that you need. Chances are that you have already used invisible Web content without realizing it. Maybe a librarian directed you to the RSB's Wisconsin Statutes page or perhaps you saw a link on WisBar. With a better understanding of why some Web content is considered "invisible" and a knowledge of the strategies used to locate it, you will be able to *search smarter* and maximize the time you spend researching on the Web.

A similar search in Google for Colorado, that is, Colorado livestock statutes, results in 33,200 hits. The first two are links to the Office of Economic Development. To search more logically, you would begin a statutory search as "Colorado Statutes." The first link of under 10,000 hits will automatically link and redirect you to <http://198.187.128.12/colorado/lpext.dll?f=templates&fn=fs-main.htm&2.0> (copy and paste to a browser), the official site of the Colorado Statutes where a search using the

term "livestock" results in 109 hits. Other Colorado-specific resources that parallel what is cited for Wisconsin are listed in the accompanying box, "Invisible Web Resources for Attorneys."

NOTES

1. Pena, "Google's Domain Even Takes in Law Offices," *East Bay Bus. Times* (May 7, 2004), available at <http://eastbay.bizjournals.com/eastbay/stories/2004/05/10/story7.html>; McCullagh, "Search Engines Take the Stand," *CNET News.com* (May 13, 2004), available at http://news.com.com/Search+engines+take+the+stand/2100-1032_3-5211658.html.
2. "Delphi Research Asks: Does Search Contribute to Productivity?" *DelphiWeb.com Newsflash* (May 5, 2004), available at http://www.delphiweb.com/knowledgebase/newsflash_guest.htm?nid=953.
3. Bergman, "The Deep Web: Surfacing Hidden Value," *7 J. Elec. Pub.* 1 (Aug. 2001), available at <http://www.press.umich.edu/jep/07-01/bergman.html>.
4. *Id.*
5. One way to recognize if a Web page is static or dynamic is to look at its URL, or Web address. Static pages often contain the extension .htm or .html. Dynamic pages usually include a question mark, indicating that the resulting Web page is based on a database query.
6. Botluk, "Mining Deeper Into the Invisible Web," *LLRX.com* (Nov. 15, 2000), available at <http://www.llrx.com/features/mining.htm>.
7. Bergman, *supra*, note 3.
8. *Id.*
9. There are a few exceptions to this rule. A number of news sources, such as *The New York Times*, offer full-text content on their websites. Usually, however, this includes current content only, and the reader is subjected to numerous advertisements. Additionally, some authors and publishers have chosen to offer their content free on the Web for several reasons: to facilitate the distribution of scholarly information, to market themselves, or to put forth their own point of view. Beware the latter.
10. *Botluk, supra*, note 6 ■

Invisible Web Resources for Attorneys

Colorado:

- Colorado Bar Association's "For the Public" "Law-related Websites": <http://www.cobar.org/hotlinks.cfm>.
- Colorado State Legislative Branch: <http://www.leg.state.co.us> for statutes, session laws, and bills.
- Colorado Online Court Records (Register of Actions Sheets), with the exception of Denver District Courts, available through fee-based services only, such as: <http://www.cocourts.com> and <http://www.cojustice.com>; LexisNexis File and Serve provides Register of Action Sheets for free to parties to the case [see Colorado Association of Law Librarians, "Online Background Searching in Colorado—Part III," 32 *The Colorado Lawyer* (Nov. 2003) at 61].
- Denver County Court provides records free of charge at: <http://198.202.202.66/court/courtselect.asp>.

Federal:

- American Factfinder: <http://factfinder.census.gov>—population, housing, economic, and geographic data from the U.S. Census.
- FedStats: <http://www.fedstats.gov>—statistics from United States government agencies.
- FindLaw's Cases & Codes: http://www.findlaw.com/case_code—links to databases of federal and state cases and legislation.

- FindLaw's Supreme Court Center: http://supreme.lp.findlaw.com/supreme_court/resources.html—recent U.S. Supreme Court opinions, orders, briefs, docket, etc.
- GPO Access: http://www.access.gpo.gov/su_docs/help/hints/index.html—U.S. Code, CFR, etc., from the Government Printing Office.
- Thomas: <http://thomas.loc.gov>—U.S. legislation and other congressional information.

Wisconsin:

- WisBar State and Federal Legal Resources: <http://www.wisbar.org/legalres>—links to Wisconsin and federal resources.
- Wisconsin Legislature Infobases: <http://folio.legis.state.wi.us>—Wisconsin Statutes, Acts, Bills, etc.
- Wisconsin Online Court Records: http://www.courts.state.wi.us/global/court_records.html—status information for Wisconsin cases (WSCCA.i & CCAP).

General Invisible Web Directories:

- CompletePlanet: <http://www.completeplanet.com>.
- Direct Search: <http://www.freepint.com/gary/direct.htm>.
- ProFusion: <http://www.profusion.com>.
- Librarian's Index to the Internet: <http://lii.org>.