Digital Commons @ Georgia Law

Presentations

Alexander Campbell King Law Library

5-8-2003

Effective Internet Search Strategies: Internet Search Engines, Meta-Indexes and Web Directories

Wendy E. Moore
University of Georgia School of Law Library, wemoore@uga.edu

Repository Citation
Moore, Wendy E., "Effective Internet Search Strategies: Internet Search Engines, Meta-Indexes and Web Directories" (2003). *Presentations.* Paper 25.
http://digitalcommons.law.uga.edu/speeches/25

This Article 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 Presentations by an authorized administrator of Digital Commons @ Georgia Law. For more information, please contact tstriepe@uga.edu.

Effective Internet Search Strategies:
Internet Search Engines, Meta-Indexes, and Web Directories

Wendy E. Moore, M.S. in L.S.
Acquisitions/Serials Librarian
The University of Georgia School of Law Library
Athens, GA

Effective Internet Search Strategies:
Internet Search Engines, Meta-Indexes, and Web Directories

Table of Contents

INTRODUCTION **1**

CONCEPTS AND TERMS **1**

DIRECTORIES **2**

- **GENERAL DIRECTORIES (COMMERCIAL)** 2
- **GENERAL DIRECTORIES (NON-COMMERCIAL)** 3
- **LEGAL DIRECTORIES** 5

SEARCH ENGINES **8**

- **TIPS FOR USING SEARCH ENGINES EFFECTIVELY** 11

META-SEARCHERS **12**

LEGAL SEARCH ENGINES **13**

SOURCES FOR FURTHER LEARNING **14**

CONCLUSION **15**

INTRODUCTION

Searching the World Wide Web can be a daunting task. The Web has expanded at such a rapid pace that nobody knows exactly how large it is, but it is safe to say that there are many billions of web pages residing on servers all over the world. Add to this scenario the hundreds of different search tools available to choose among – including directories, search engines, meta-searchers, and specialized search engines – and the situation begins to feel overwhelming.

Fortunately, learning a few essential concepts of Web searching, along with mastering a handful of the top-rated search tools, can make the picture much brighter. Simply knowing how to choose the right tool for your information need can make all the difference.

This paper will first discuss basic concepts and terms you must know to be an effective searcher. Next, it will in turn examine each of the major categories of search tools, and recommend the best search engines and directories currently available.

CONCEPTS AND TERMS

The term "search engine" is often used generically to describe both human-powered directories and crawler-based search engines. These two types of search tools gather their listings in radically different ways.

What are the differences between a ***directory*** and a ***search engine***?

Directories

- provide links to web sites organized into an alphabetical hierarchy of topics
- created by human editors who select quality web sites and classify them into subject categories and subcategories
- cover only a tiny portion of the Web (far less than search engines)
- sometimes include reviews and/or ratings of web sites
- provide access via browsing the menu categories or by keyword search

- record only limited information about web sites (title, URL, short description) *not* full-text of web sites
- useful for finding high-quality, human-reviewed web sites on broad, general topics (e.g. affirmative action or capital punishment or civil rights)

Search Engines

- created by computer software programs called by various names such as robots, spiders, crawlers, and worms
- as the spider follows links, moving from one web site to another, it scans each site it visits and records the contents in a database
- users can then search this database by keyword to find web sites matching their search topic
- because search engines are created by these automated computer programs, they cover far more web sites than are listed in manually-created directories
- since there is no human judgment involved, the quality of web sites indexed in a search engine is less consistent than in a directory, where the sites have been selected by editors
- because the spider generally records the full-text (every word) of the pages it visits, search engines are useful when you are seeking web sites about narrow, specific topics or that mention obscure words or phrases

DIRECTORIES

General Directories (Commercial)

Yahoo!
http://www.yahoo.com/

For many years, *Yahoo* has been the most popular directory on the Web. Since its birth, *Yahoo* has used its own human editors to organize web sites into categories. However, recognizing that humans can't index everything because the process is so labor-intensive, *Yahoo* also has for years partnered with a third-party "crawler-based" search engine to provide answers when there are no matches within its own human-powered listings.

In October 2002, *Yahoo* made a giant shift to using *Google*'s crawler-based listings for its main results. However, *Yahoo*'s search results pages still show "Directory Category Matches" as

well. When offered, these will take you to a list of web sites that have been reviewed and approved by a human editor.

By going to the “Advanced Search” page, you can choose between searching the Web (via Google) and searching the *Yahoo!* Directory listings. Also, if you scroll down to *Yahoo*’s “Web Site Directory” and browse through the menu of topics – as opposed to typing in a keyword search -- you are viewing web sites selected by *Yahoo*’s team of editors, rather than matches generated by *Google*’s search engine.

Yahoo, long the best-known Web directory, is today a good example of a "Hybrid Search Tool" which combines aspects of both a human-powered directory and a crawler-based search engine. It is becoming increasingly common for search tools to combine both these aspects, as many directories have partnered with search engines (and vice versa). Generally, though, your searching will be more efficient and effective if you go directly to the source you wish to search, rather than relying on a third-party site to pass your query along to a separate search tool. In other words, if you want to search the Web via *Google*, go to *Google*’s web site where you can take advantage of the many advanced search features offered, rather than letting *Yahoo* interpret your search and pass it on to *Google*.

Yahoo's directory of Law resources <http://dir.yahoo.com/Government/Law/> covers several thousand web sites.

General Directories (Non-Commercial)

As commercial web directories focus on advertising and other ways to make a profit, several general directories put together by non-commercial organizations offer excellent

alternatives for web researchers. These directories strive for quality and authority of the underlying content, rather than trying to create the largest directories available.

Since anyone can place "anything" on the Internet – regardless of accuracy or validity – it's good to know that non-commercial directories such as these are available to help web researchers access high-quality web-based information.

Open Directory Project
http://dmoz.org/

The best known non-commercial directory is the *Open Directory Project*. Its data "powers" the directory services on hundreds of sites, including most of the major search engines such as *Google* and *Lycos*. The *Open Directory Project* is the largest, most comprehensive human-edited directory of the Web. Their global community of volunteer editors (over 56,000) has classified over 3.8 million sites into more than 460,000 categories. "Law" is listed as a sub-category under the heading "Society" at <http://dmoz.org/Society/Law/>.

Librarians' Index to the Internet
http://lii.org/

With the motto "Information You Can Trust" the *Librarians' Index to the Internet* is an annotated subject directory of more than 11,000 Internet resources evaluated and selected by librarians for their usefulness and reliability. Every entry contains a brief annotation, subject heading, and a date when the URL was last checked. The *LII* can be browsed or searched by keyword, and the "Advanced Search" interface allows searching on up to 10 fields. The "Law Topics" page <http://lii.org/search/file/law> lists over 50 different subjects.

Legal Directories

Legal Information Institute
http://www.law.cornell.edu/

An exceptional directory of legal information maintained by the Legal Information Institute at the Cornell Law School. The Institute has pioneered the use of digital information technology in the distribution of legal information, the delivery of legal education, and the practice of law since 1992. The mission of the site is to make the law more accessible not only to U.S. legal professionals, but also to students, teachers, and the general public in the U.S. and abroad. The site first gained notoriety as the source for recent U.S. Supreme Court cases, but now offers links to a wide variety of resources, including, the entire Uniform Commercial Code, opinions of the U.S. Court of Appeals, the Federal Rules of Civil Procedure, Criminal Procedure, and Evidence, the American Legal Ethics Library, and the constitutions and codes for all 50 states. The Institute is a non-profit and dedicated to the concept of free legal content on the web.

FindLaw
http://www.findlaw.com/

Sometimes referred to as the *Yahoo!* of legal web sites, this directory features a huge collection of links to legal materials. A very popular source of information, *FindLaw* recently made it into the top 500 most visited U.S. web sites. Grouped by user category subdivisions and then arranged by topic, it can be a little difficult to know where to begin looking. The subdivision labeled "Legal Professionals" has many topics of interest, including "U.S. Laws: Cases and Codes", "Federal Resources," "Foreign and International," "Forms," and "Research Resources." Because *FindLaw* is a commercial site operated by Thomson – West, it offers

access to some proprietary sources, such as West's Legal Directory. Some of the other subdivisions include links to information on "Adoption," "Elder Law & Aging," "Real Estate," and "Taxes & the IRS."

WashLaw WEB
http://www.washlaw.edu/

Maintained by the Washburn University School of Law as a non-profit service to the legal community, *WashLaw WEB* is an extensive collection of law and law-related links in an easy to browse layout. Arranged alphabetically, there are nearly 200 topic links on the main page. Designed to provide one-stop-shopping for legal professionals, some topics are law-related, such as "9th Circuit," "Federal Caselaw," and "Legal Institutes," while other topics are more practical, such as "Radio," "Street Maps," "Travel," and "Zipcodes." The topic "Subject Index" leads to a subdirectory of legal subjects as arranged by the Association of American Law Schools, such as "Bankruptcy," "DUI/DWI," "Fraud," and "Probate."

Hieros Gamos
http://www.hg.org/

Hieros Gamos is a vast collection of legal web resources with a focus on international materials. The front page is jammed with links, but it is still easy to navigate because of the broad organizational categories used in the left and right columns. Under the area of "Law Practice" there are links for over 70 primary practice areas, with each topic sub-arranged from supra-national to local links. Under "International Law" there are links to 230 counties, as well as links to International Organizations and government sites. Started in 1996 by Lex Mundi, an association of 161 independent law firms worldwide, this is a truly international site with many

links to foreign and international law firms, law schools, and legal materials. Versions of the site are available in French, German, Italian, and Spanish through automated translation via *Google*.

AllLaw
http://alllaw.com/

AllLaw is a small, practice oriented site. The front page has a broad divisions, including "Federal Resources," "Forms," and "Legal Support Services." Under the division "Legal Topics" there are 34 broad topic categories, half of which each offer links to explanatory articles and forms, in addition to related topical web sites. There is advertising on this commercial site, but for the most part it is not too distracting.

Jurist
http://www.jurist.law.pitt.edu/

Maintained by the University of Pittsburgh School of Law, *Jurist* offers a wide breadth of legal information. The format of the site is easy to navigate with the major topics appearing in the left side bar. Under the topic "Legal Research" there are subdivisions such as "Cases & Statutes," "Working Papers," and "Famous Cases." The topic "World Law" is very useful with information on "Constitution, Government & Legislation," "Courts & Judgments," "Human Rights," "Legal Profession", and "Law Schools" arranged by country.

SEARCH ENGINES

What makes one search engine different from another? Why choose one search engine over its competitors?

Remember that when you use a search engine, you are *not* actually searching the Web directly. Rather, you are searching a database that contains information about all the web sites visited by that search engine's spider or crawler. Therefore, the contents of that database are a crucial factor determining whether or not you will succeed in finding the information you need.

Size is one important measure. How many web pages has the spider visited, scanned, and recorded in the database? Some of the larger search engines have databases covering over 3 billion web pages, while the databases of smaller search engines cover half a billion or less.

A second important measure is how **up-to-date** the database is. The Web is constantly growing and changing. New sites appear, old sites vanish, and existing sites modify their content. Unless the spider sent out by the search engine can keep up with these changes, the information recorded in the database will become so out-of-date as to be useless.

Based on these important criteria, the top three search engines at this time are *Google*, *All The Web*, and *AltaVista*.

Google
http://www.google.com/

Google has a well-deserved reputation as the top choice for Web searching. In fact, for many people "Google it" has become a synonym for searching the Web. With a huge database of over 3 billion web pages (refreshed on a monthly basis), the crawler-based service provides

expansive and up-to-date coverage of the Web. It's highly recommended as a first stop in your hunt for whatever you are looking for.

Google is widely acknowledged as a leader in search relevancy, having pioneered relevance ranking based on *link analysis*. “Relevance ranking” determines the order in which your search results are displayed, with the goal being to display the most relevant, useful matches at the top of the list.

Each search engine has its own proprietary computer algorithm, which ranks pages by relevance. Until *Google* came on the scene in 1999, the common practice was to rank pages based largely on word occurrence (i.e. if your search term appeared repeatedly and prominently on a web site, that web site was considered highly relevant to your search topic). *Google* revolutionized this practice by adding the concept of link analysis, which analyzes links between sites as one component of relevance.

Link analysis takes into account the *number* and *quality* of links that point to a particular web site. Sites that the web community has overwhelmingly "voted" for by creating links to them are considered relevant and authoritative. Often misinterpreted simply as a measure of “popularity,” link analysis in fact looks at more than the sheer volume of links (or “votes”) a page receives; it also analyzes the page that casts the vote. Votes cast by pages that are themselves “important” weigh more heavily and help to make other pages “important.”

The bottom line for searchers is that *Google*’s relevance ranking is frequently cited as the best among all search engines, which means that often you’ll find useful sites within the first ten matches, rather than having to wade through hundreds of irrelevant matches before finding what you need. Other search engines have recently moved to incorporate link analysis as part of their ranking algorithm, but none have yet matched *Google*’s success.

Google is also notable for indexing more than just standard HTML web pages. *Google*'s database also includes Web-accessible documents that are not Web pages, such as Adobe Acrobat PDF, Microsoft Word, PostScript, Excel, PowerPoint, WordPerfect, and other files.

AllTheWeb.com (a.k.a. FAST)
http://www.alltheweb.com/

An excellent crawler-based search engine, *All The Web* provides extensive, up-to-date coverage of the Web. *All The Web*'s database of over 2 billion web pages is one of the largest after *Google*'s, and is updated more frequently than *Google*'s. If you tried *Google* and didn't find it, *All The Web* should probably be next on your list.

In addition to web page results, *All The Web* provides the ability to search for pictures, video clips, MP3s, PDF, Microsoft Word, Flash, and FTP files.

The site is operated by the Norwegian company Fast Search and used as a showcase for FAST's search technology.

AltaVista
http://www.altavista.com/

AltaVista is the oldest crawler-based search engine on the web. It opened in 1995 and for several years was the "Google" of its day, in terms of providing relevant results and having a loyal group of users that loved the service. In recent years, *AltaVista* has lost ground to newcomers as the size and freshness of its database has lagged behind its competitors. *AltaVista* remains a useful search engine, but crawlers such as *Google* and *All The Web* provide more comprehensive results. As a result, *AltaVista* is probably a third-choice crawler, one to try if you haven't found what you are looking for at one of its competitors.

Tips for Using Search Engines Effectively

Despite names such as "All The Web," It is important to remember that no search engine covers more than approximately 20% of the Web. In other words, even when using a large search engine such as *Google*, you are missing at least 80% of the sites currently existing on the Web. For this reason, you must use *multiple* search engines if you wish to approach a thorough or comprehensive search. Fortunately, there is remarkably little overlap between the various search engines. Each search engine will find sites not listed in other search engines, and many web sites are found by only one search engine.

Search engines offer a variety of features that allow you to construct a precisely-targeted search. Examples of these features include Boolean logic (allows you to search for terms in particular combinations using AND/OR/NOT), searching particular fields of documents (e.g. title only), restricting your search by date, language, host or domain name, etc.

However, each search engine employs its own syntax for these commands, and learning all the codes and symbols required to execute these search commands can be tedious and time-consuming (especially if you want to use more than one search engine, as the commands vary from one search engine to another).

The simple solution to this problem is to bypass the introductory search page, and instead go to an alternative search page, which can be called by different names – *Google*, *All The Web*, and *AltaVista* call it "Advanced Search," but other search engines may call it by names such as "Detailed Search" or "More Precision." On this page, you are not required to use cryptic symbols and field names to construct a precisely-targeted search. Rather, the page offers a simple fill-in-the-blanks approach where you can search for "all of these words," "this exact phrase," "at least one of the words," etc. You can also check off options to restrict your search

by language, file format, date, domain, placement of search terms in the document, etc. Making use of the "Advanced Search" page is by far the single most efficient way to increase your effectiveness as a searcher.

META-SEARCHERS

Meta-searchers are fundamentally different from search engines because they do not use a crawler and do not create their own databases of Web sites. Rather, they serve as a conduit or intermediary, which allows you to submit a search, and then passes your search along to several individual search engines. In this manner, you can use multiple search engines simultaneously.

Unfortunately, the idea of meta-searching is much better than the reality. Since most retrieve only the top 10-50 matches from each search engine, the total number of sites retrieved may be considerably less than found by doing a direct search on one of the search engines. Meta-searchers do not allow you to take advantage of advanced features in any search engines, and are unpredictable in how they will transmit a complex search. Frequently they are subject to time outs when search processing takes too long.

Overall the problems of meta-searchers outweigh their benefits. All have significant limitations as a comprehensive search tool, and are less effective than learning a few good search engines. Three of the better-known meta-searchers are listed below:

Ixquick
http://www.ixquick.com/

MetaCrawler
http://www.metacrawler.com/

DogPile
http://www.dogpile.com/

LEGAL SEARCH ENGINES

Another category of search tool is the specialized search engine, also known as a topical search engine or vertical search engine. Specialized search engines create their indexes by crawling only a hand-selected set of sites, sites on a certain domain (e.g. .gov or .edu), or sites containing certain types of files. Creators of specialized search engines aim to improve relevancy by creating an index that is focused on a particular topic, region, language, or information type. Today specialized engines exist for many fields including law, medicine, science, government, finance, and more. The primary specialized engine for law is called *LawCrawler*.

LawCrawler
http://lawcrawler.findlaw.com/

Previously powered by the AltaVista search engine, *LawCrawler* is now a Google-powered search engine that returns information only from sites that have been identified as dealing with legal issues. In addition to searching the Web, *LawCrawler* also searches *FindLaw*'s databases of legal materials.

Meta-Index for U.S. Legal Research
http://gsulaw.gsu.edu/metaindex/

This site presents an array of search forms for many law and U.S. government sources, with each form containing sample search criteria. From this page, you can search for opinions of the U.S. Supreme Court and all federal circuit courts. A legislative section allows searching of the U.S. Code, as well as of bills and the full text of the Congressional Record. There are also search forms for federal regulations, people in law, and other legal sources. The site is provided by the Georgia State University College of Law.

SOURCES FOR FURTHER LEARNING

Search Engine Showdown
http://www.searchengineshowdown.com/

Search Engine Showdown is a web site produced by writer and Montana State University Reference Librarian Greg Notess. The site monitors and compares search engines, providing a wealth of information including reviews and tutorials for individual search engines, charts analyzing the size and freshness of search engine databases, and much more.

Search Engine Watch
http://searchenginewatch.com/

Search Engine Watch is another excellent source for detailed, current information about search engines. Features include search engine news, reviews and ratings, web searching tips, and more.

LLRX
http://www.llrx.com/

Law Library Resource Xchange (LLRX) is a unique, free Web journal dedicated to providing legal and library professionals with the most up-to-date information on a wide range of Internet research and technology-related issues, applications, resources, and tools.

The Virtual Chase: Teaching Legal Professionals How To Do Research
http://www.virtualchase.com/

This excellent site is a valuable resource for anyone conducting legal research on the Web. There are annotated guides, online teaching tools, articles on specific research issues and topics, material on evaluating resources, and more.

CONCLUSION

The amount of information available on the Web can make it an overwhelming task to quickly locate the specific web site that is relevant to your needs. Use of directories and search engines, both general and law-related, ease the problems associated with searching the Web. Learning and applying a few of these mentioned search tools will make your search for information on the Web successful.