

3-2-2006

Google and Beyond: Finding Information Using Search Engines, and Evaluating Your Results

Elizabeth Geesey Holmes

University of Georgia School of Law Library, eholmes@uga.edu

Repository Citation

Holmes, Elizabeth Geesey, "Google and Beyond: Finding Information Using Search Engines, and Evaluating Your Results" (2006). *Presentations*. 2.

<https://digitalcommons.law.uga.edu/speeches/2>

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. [Please share how you have benefited from this access](#) For more information, please contact tstriepe@uga.edu.

**Google and Beyond:
Finding Information Using Search Engines,
and Evaluating Your Results**

Elizabeth Geesey Holmes, M.A., M.S.L.I.S.
Cataloging Services Librarian
University of Georgia School of Law
Alexander Campbell King Law Library
Athens, Georgia

http://www.law.uga.edu/library/research/internet/resource_list2006.htm

Google and Beyond:
Finding Information Using Search Engines,
and Evaluating Your Results

Elizabeth Geesey Holmes, M.A., M.S.L.I.S.
Cataloging Services Librarian
University of Georgia School of Law
Alexander Campbell King Law Library
Athens, Georgia
<mailto:eholmes@uga.edu>

Table of Contents

INTRODUCTION.....	1
BASICS OF SEARCH ENGINES.....	1
Meta Search Engines	3
Dogpile.....	3
Vivisimo (Clusty).....	3
Mamma	4
General Search Engines	5
Google.....	5
Yahoo!.....	6
MSN Search	6
Basic Search Tips	7
The Invisible Web	9
ADVANCED STRATEGIES FOR SEARCH ENGINE SEARCHING	12
Advanced Search pages	12
Searching by format	13
PowerPoint presentations.....	14
Pdfs, Excel Spreadsheets and Word Documents	14
Audio and visual material	15
News	16

Hidden (and not so hidden) search engine tools.....	16
Searching within your results.....	16
Toolbars	17
Cached pages	17
Synonyms, URL info, Who links to a site	18
Yahoo! Shortcuts	19
Context searching or clustering: Y!Q and Clusty.....	19
ASSESSING THE QUALITY OF INFORMATION ON THE INTERNET	20
Authenticity	21
Accuracy	23
Currency	24
Completeness.....	25
SOURCES FOR FURTHER LEARNING	26
CONCLUSION	29

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 task of evaluating information found on the web and choosing between the hundreds of different search tools available – 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 and site evaluation, along with mastering a handful of the top-rated search tools, can make the picture much brighter. This paper will discuss the basics of a few search engines and provide examples of advanced searches that can be done to increase your searching efficiency. It will also address the task of assessing the quality of the information you find on the Internet. In addition it will list and describe places to go for more information on improving you Internet searching and evaluating skills.

Basics of search engines

What are Internet search engines and why should you use them for your legal research needs? It is increasingly important to include Internet resources when conducting legal research: more and more material is being posted directly (and possibly only) to the Internet. There may be no traditional print counterpart to the information source you are seeking, for example: electronic journals and newsletters, lower court/trial level decisions and postings to Internet discussion groups. To find

information today you must look at new developments and sources available only on the Internet.

Search engines are one important tool used to locate these Internet based resources. They are useful for general Internet searching, particularly if you are not familiar with the subject area in which you are searching. They may return hits that are irrelevant, but this flaw can be helpful if the researcher has no idea what he is really looking for. 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.

Directories (such as *Yahoo!’s* and *Google’s* directories) provide links to Web sites organized into an alphabetical hierarchy of topics. They are created by human editors who select quality Web sites and classify them into subject categories and subcategories. Thus, they cover only a tiny portion of the Web (far less than search engines). Access is via browsing categories or searching the directory site by keyword search. They are useful for finding high quality, human-reviewed Web sites on broad, general topics (e.g. affirmative action, capital punishment or civil rights). **Many of the search strategies discussed here can be used to search directory sites.**

Search engine’s data are created by 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 of the database. Users can then search this database by keyword to find Web sites matching their search topic. Because search engines are created by automated computer programs they cover far more Web sites than do manually created directories. I will be focusing on these computer created search engines in this paper.

Meta Search Engines

Before I do so I would first like to mention multiple search engines, also called meta search engines, and their pros and cons as tools to search the Internet. Unlike search engines, meta crawlers do not crawl the web themselves to build listings. Instead, they allow searches to be sent to several search engines all at once. The results are then blended together onto one page.

Dogpile

<http://www.dogpile.com>

This is a popular meta search site that sends a search to a customizable list of search engines, directories and specialty search sites, including the major search engines: *Google, Yahoo!, MSN Search* and *Ask Jeeves*. It then displays results from each search engine individually or in a combined list that notes which search engine or engines it was found on. One can also compare results from one or more different search engines to see which hits are unique. Note that the sponsored sites are mixed in with the regular sites, which is a bit misleading. Dogpile does include an advanced search option, which is easy to find, as it is located directly next to the main search box.

Vivisimo (Clusty)

<http://clusty.com/>

Clusty is the search site owned and operated by Vivisimo. When you enter a search term, Clusty will not only pull back matching responses from major search engines but will also automatically organize the pages into categories – or as it calls them – clusters. Clustered results are especially useful when you are searching for something unfamiliar, and aren't sure which keywords will produce the optimal results. The results are grouped into successively narrower subcategories, allowing you to drill down through a topic without additional searching. Another unique feature on Clusty is the ability to

preview the website on the hit list without opening it in a separate window. This tool is slick and easy to use, and includes an advanced search option, however, they do not at this time include results from *Google* or *Yahoo!*.

Mamma

<http://www.mamma.com/>

Founded in 1996, Mamma.com is one of the oldest meta search engines on the web. Mamma searches against a variety of major crawlers, directories and specialty search sites. It currently includes the search engines: *Yahoo!*, *Gigablast*, *Ask Jeeves*, *Wisnut* and *Entireweb*. It also includes several directories in its searches. It does not have an advanced search page per se, but it does support some Boolean search operators.

The main advantage to using these meta search engines is that one can query a variety of different search engines with one search, but there are drawbacks to these multiple search engines as well, which may outweigh their benefits.

- All have significant limitations as a comprehensive search tool.
- They are subject to time outs, when search processing takes too long.
- Since most only retrieve the top 10-50 hits from each search engine, the total number of hits retrieved may be considerably less than found by doing a direct search on one of the search engines.
- The advanced search features on individual search engines may not be available, and if you use Boolean operators results from search engines that use a different syntax may not be included.

Nonetheless, these are important resources and many of the search strategies discussed later in this paper can be used to search these meta search engine sites.

General Search Engines

What makes one search engine different from another? 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 three 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 changing and growing. 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 out of date.

In addition to contents, the **ranking or relevance algorithm** used by the search engine determines whether the most relevant hits appear towards the top of your results list.

I will be highlighting the search engines: *Google*, *Yahoo!* and *MSN Search* because they meet these criteria, are recommended by advanced Internet searchers, and are the three most searched.

Google

<http://www.google.com>

Google has a well-deserved reputation as the top choice for those searching the Web. The crawler-based service provides both comprehensive coverage of the Web along with great relevancy. It is highly recommended as a first stop in your hunt for whatever you are looking for. In addition you can search for twelve different file formats, cached pages, images, news and Usenet group postings. The advanced searching capabilities are easy to use and their toolbar offers several helpful enhancements.

Yahoo!

<http://www.yahoo.com/> Main *Yahoo!* homepage

<http://search.yahoo.com/> Uncluttered search page.

Yahoo! is one of the best-known and most popular Internet portals. Originally just a subject directory, it now is a search engine, directory, and portal. Its strengths include: a large search engine database, cached copies of pages, links to the *Yahoo! Directory*, and the ability to do full Boolean searching. It does include some pay sites and does not support truncation. *Yahoo!* provides search results in seven categories including a full web search and a search of their directory, which allows one to limit display of results to one area. These results are sorted by relevance algorithm.

MSN Search

<http://search.msn.com/>

MSN Search uses its own Web database and also has separate News, Images, and Local databases along with links into Microsoft's Encarta Encyclopedia content. Its strengths include: this large unique database, its query building "Search Builder" and Boolean searching, cached copies of Web pages including date cached, and automatic

local search options. It has some weaknesses too such as: limited advanced features, inconsistent availability of truncation and no title search, truncation, or stemming. Search results are sorted by relevance and the Search Builder Results Ranking section has three sliders that users can adjust to change the ranking criteria.

Basic Search Tips

Now that we have determined that *Google*, *Yahoo!* and *MSN Search* are good search engine starting points, what are some of the basic search tips to keep in mind when approaching these tools? First, remember that no search engine covers more than approximately twenty percent of the Web. In other words, even when using a large search engine such as *Google*, you are missing at least eighty percent of the sites currently existing on the Web. For this reason, you must **use more than one search engine, and/or meta search engines** if you wish to approach a thorough or comprehensive search. There is remarkably little overlap between 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. Even if the search engines find many of the same sites the hits on the first and second pages – which is as far as most searchers look – may vary greatly. To see visually how results differ from *Google* and *Yahoo*, you can do a search at: <http://www.langreiter.com/exec/yahoo-vs-google.html> (this is an experimental tool designed by Christian Langreiter of Synerge Development Services) or you can do the same search on several different search engines and compare your results. I did a search for “torts” on both *Google* and *Yahoo!* and found that only four of the ten sites on the first page were the same for both *Google* and *Yahoo!*. The results on

page two of each search were completely different! If I had used only one of these search tools and looked only at the first and second pages of each search then I would have missed sixteen, possibly valuable sites – *Yahoo!* and *Google* only agreed on four of the top twenty sites.

Second, search engines offer a variety of features that allow you to construct a precisely targeted search. Examples of these 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 best solution to this problem is to bypass the introductory search page, and instead go to an alternative search page, *Google*, and *Yahoo!* call it “Advanced Search.” *MSN Search* does not have an advanced search page, but does offer something called “Search Builder”. 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, domain, placement of search terms in the document, etc. **Making use of the “Advanced Search” page is the single most efficient way to increase your effectiveness as a searcher.**

Third, all of the above mentioned search engines allow the user to **set their own options** such as turning on or off content filters (the default may be different in each search engine and obviously this could affect the results of a search without the searcher even being aware of it). As well as opening search results in a different browser window (which is very useful for referring back to the search results), selecting a default language, and setting the number of search results to be displayed on each page. These options are called “Preferences” in *Google*, and *Yahoo!* and “Settings” on *MSN Search*. They are accessible via a link directly to the left of the Basic Search box on the search engine’s home page for *Google*, and *Yahoo!*, and immediately below the search box on the *MSN Search* home page.

Fourth, if you really want to get to know a search engine, look for and **read their help or additional information pages**. This information may be slightly buried. On *Google* it is accessible from a link called “About Google” at the top of the Advanced Search Page, or below the search box on the main page. *Yahoo!’s* and *MSN Search’s* help sections are helpfully labeled “Help”. *Yahoo!’s* is accessible from a link on the Advanced Search page called “Help” or from the search home page there is a link at the bottom to “Services and Tools”, and *MSN Search’s* link is on the main page below the search box.

Last, **remember that not all information is indexed on these search engines** – which leads us to a discussion of the Invisible Web.

The Invisible Web

The information, which is not readily indexed by most search engines, is referred to as the Invisible Web. Only a small portion of the vast amount of information posted on the Internet is indexed by search engines. Why is this information invisible?

- The data is in a format that the search engines do not recognize
- The data is contained in a database that must be queried before the data can be retrieved
- The search engine does not know that the Web site containing the data exists because the site has not been submitted to the search engine for indexing, or is too new.
- The search engine has chosen not to index a particular site
- The site has blocked the search engines indexing spiders/robots.
- The site requires a password.

Some examples of Invisible Web sites are: library catalogs, periodical indexes, *LexisNexis* and *Westlaw*, *Expedia* and *Travelocity*, UGA Law's Intranet.

That said, *Google*, *Yahoo!* and *MSN Search* do index many of the Web sites that are invisible to other search engines. In addition to trying these sites there are some other ways to locate information on the Invisible Web. To penetrate the Invisible Web for the useful information that often lurks there you must know where to go by having the URL or using other advanced search strategies discussed later in this paper and in the resources mentioned at the end of this paper.

One example of hidden information is data residing in databases that are accessible via the Web. If you want to find out what a specific stock was worth on July 22, 1970, just typing that date and the stock name or symbol into a search engine will *not* find the answer. The answer is on the Internet, but it is in a database of historical

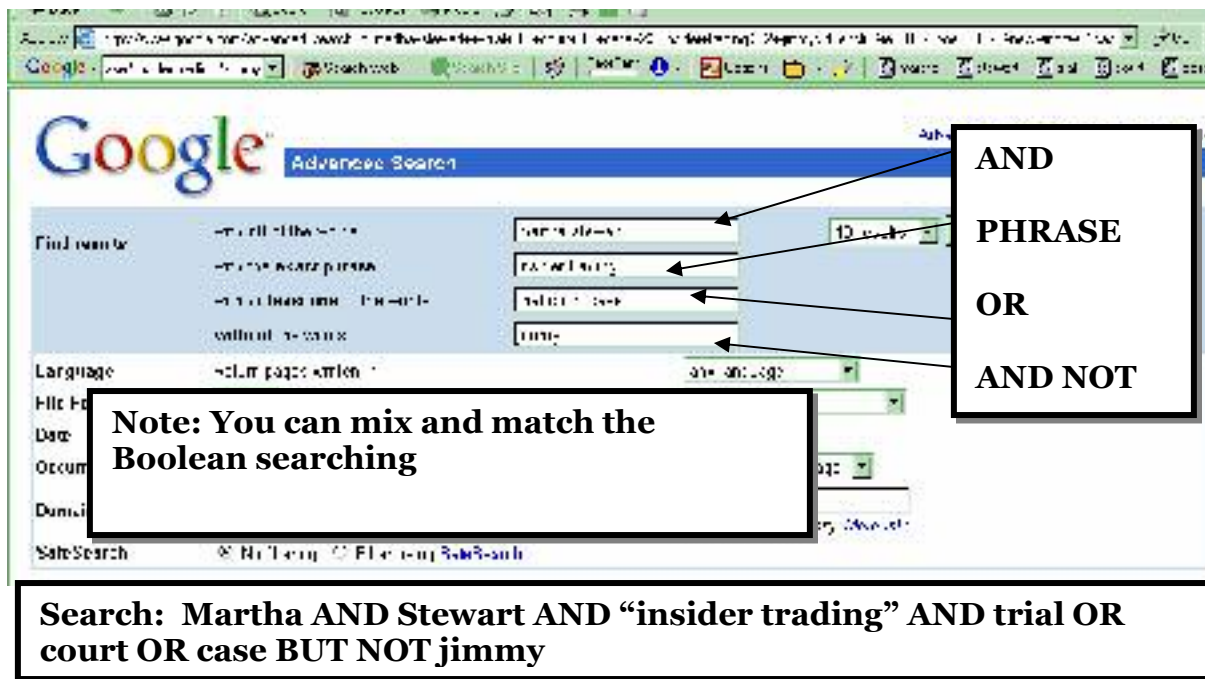
stock quotes. Search engines cannot index the individual data in databases, but the information is there if you know where to go. To find out where to go, do a search in one of the above mentioned search engines and type in the keyword “database” plus any keywords such as “historical stock quotes” that describe the type of database you are looking for. This search will lead you to a discussion of where to find this sort of information, and provide links to the appropriate database for you to query.

Subject-based directories, and Invisible Web portals like *Complete Planet* and *ProFusion* are also tools for finding information that resides on the Invisible Web. I have included more information on these sites in the last section of this paper with other resources for advanced search engine searching. *Google* also has a search tool using its search engine called *Google Scholar*: <http://scholar.google.com/>. As a result of *Google*'s agreements with publishers, associations, universities and others that allow *Google* to index databases that search engine crawlers normally cannot penetrate, this tool covers “peer-reviewed” papers, theses, books, preprints, abstracts and technical reports. Retrieved records for journal articles and books include the title, the author, a few words of text surrounding your search terms, the name of the journal and links to place where it can be found, including pay databases for articles and libraries for books. For the professional scholar *Google Scholar* is not, at least yet, a substitute for searching proprietary subscription databases which provide much more comprehensive coverage; greater clarity about the extent of coverage; are more definitively scholarly; and are much more searchable due to better commands, structure and indexing. However, it is worth giving it a try.

Advanced Strategies for Search Engine Searching

Advanced Search pages

Once you have decided to search the Internet via a search engine, the most important strategy you can use is to go directly to the advanced search page. This page gives you significantly more options for conducting your search. These include sophisticated Boolean connector searching (without knowing the Boolean search symbols), and limiting the search to a certain Internet domain or specific file format. These options are not available via the basic search page.



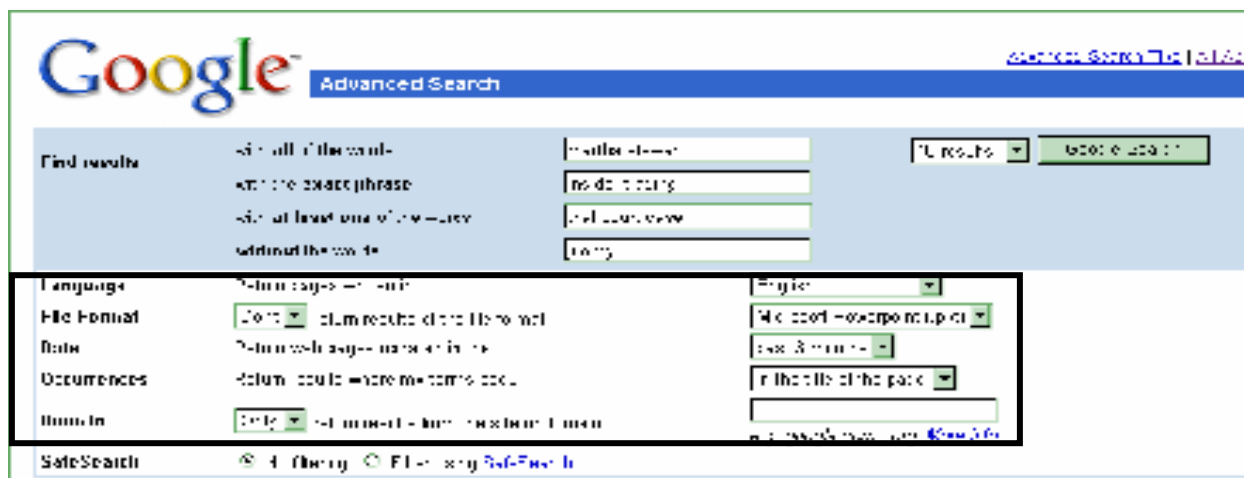
The image shows a screenshot of the Google Advanced Search page. The search bar contains the text: **Martha AND Stewart AND "insider trading" AND trial OR court OR case BUT NOT jimmy**. Callouts point to the search boxes for "AND", "PHRASE" (indicated by quotes), "OR", and "AND NOT". A note box states: **Note: You can mix and match the Boolean searching**. Below the search boxes, there are options for Language, File type, Date, Occurrence, Domain, and Site Search.

Once you've run your search the advanced search page offers ways to further refine that search. From the search results page, click on the advanced search link again and you are taken back to the advanced search page with your search still entered in the boxes. Below the search boxes are additional options for tailoring your search. In the example

in the illustration on the following page I have added several additional criteria to more narrowly target my search:

- The **language** of the page must be English
- No pages with the **file format** of Microsoft Powerpoint will be returned
- The pages must have been **updated** in the past 3 months
- The **words** searched must **occur in the title** of the page.

I could also have limited my search results to only those with a certain **domain name** or without a certain domain name. Obviously you may not want to narrow your search this much, but it can be very helpful if you have a very specific criteria or if your search has returned a huge number of hits.



Searching by format

As we have seen above the advanced search page of *Google* allows the searcher to restrict results to a particular file format, such as PowerPoint Presentations, Word documents, Excel spreadsheets, or a PDF files. This is an excellent way to target your search and return more relevant results. *Yahoo!* indexes all of these as well as RSS/XML and plain text (.txt) files. At present, *MSN Search* does have this feature, but

it is only available through a specific syntax command that must be looked up, and is not a choice on the advanced search page (search builder).

PowerPoint presentations

Use *Google* to restrict your search to a PowerPoint Presentation if you are looking for information such as:

- Your expert witness's (or the opposition's) presentation at a conference – to see if there are any public inconsistencies in an opinion he will be giving at an upcoming trial. Many presentations are created in PowerPoint and posted to the Web, so you might be in luck.
- A “hot” topic that no one has had any time to write about yet. It is possible that someone may have given a recent presentation on the topic and posted his or her PowerPoint slides.

Pdfs, Excel Spreadsheets and Word Documents

Restricting your search to **PDF's** is useful if you are looking for: forms, newsletters, charts, graphs, government documents or brochures. *Google* allows you to display PDF search results as HTML files (regular Web pages). This is very handy if you want to cut and paste from the files, or if there is a problem with the PDF version -- just click on the link marked “View as HTML”.

If you are looking for financial information, charts or graphs try restricting your search to **Excel Spreadsheets**, and if you are searching for someone's resume, a conference paper, an agenda for a meeting or a chart created by a committee try searching for **Word Documents** only.

Audio and visual material

What if you are not looking for a text document at all, but rather a picture, photograph, or video clip. Perhaps you are preparing a client presentation, or multimedia arguments for arbitration, the addition of video news coverage, stock photos, sound effects, and clip art might make your compelling argument more memorable. In addition, photos, diagrams, or videos of a product from a manufacturer's Web site could be useful in a product liability case.

All three of the search engines recommended offer **image** searches. However, *Google* boasts that they are the most comprehensive with over a billion images indexed. To return relevant images *Google* looks at the text on the Web page around the image and the accompanying caption, among other elements, to help determine an image's content. Using the Advanced Search page you can use the above mentioned search boxes to enter keywords, and you can also limit your search to a single domain. This can be very useful if you are looking for photos of an executive or products from a particular company. Using the "Return images from the site or domain" option, you can search only that company's Web site.

Yahoo! also offers a video search. To do a search for a **video clip** – select the "Video" tab and then select advanced search – you are taken to a customized search page to enter your search terms. After entering your keywords you have the option to specify the type of file (MPEG, Avi, Quicktime, etc...), its duration and its source. For example, I could indicate that I only wanted clips longer than 1 minute and from MSNBC.

News

While not a format in the same sense as those mentioned above, all three of these search engines offer a News search accessible via a tab at the top of the main page.

Google's News search collects **recent news stories from sources all over the world**, updating them continuously throughout the day. They retain these news stories in their index for approximately four weeks so this **archive only covers the past 30 days**. Using the advanced search page, you can sort your results list by date or by relevance. This database is terrific for up to date information, as stories note the time they were collected and some in my sample search were collected “less than 30 minutes ago”. *Google* indexes stories from sources such as: the Islamic Republic (of Iran) News Agency, *Washington Post*, News24 (South Africa), *Melbourne Herald Sun* (Australia), *Taipei Times*, MSNBC, Al-Jazeera, *Atlanta Journal Constitution* and the *Christian Science Monitor*. In addition you can browse the latest headlines. *Google* also recently added the ability to **limit news searches by country or state, or to a specific publication**. After doing a search you are given the options to sign up to receive News Alerts on this topic. If your search returned relevant and useful results this allows you to keep up to date on this same topic without running the search periodically yourself. *Yahoo!*'s news search also offers an alert service and *MSN Search*'s results show clearly where the story appears with links to MSNBC.

Hidden (and not so hidden) search engine tools

Searching within your results

A very handy tool for narrowing searches is available on *Google*, but is hidden just below the visible bottom of the search results page. Scroll down and click on the

“search within your results” button. You are then given a search box where you can add additional words to your search. The search engine will search for the word entered within the results of your last search.

Toolbars

Google, *Yahoo!* and *MSN Search* (as well as *Dogpile* and *Clusty*) all offer a Toolbar for downloading to your desktop. I’ve been using *Google*’s toolbar for sometime now and love it. It is easy to download and sits at the bottom of my other toolbars on my Web browser window. It allows me to search *Google* directly without going to their Web page, but it still lets me use some advanced search techniques such as searching within the page or within the site, and searching *Google*’s images files, directory, dictionary, or news site with the click of one button. The feature I like the most is the “highlight search terms” button. This gives me the option to highlight all of my search terms in the result I am viewing or just one of the terms at a time. I also recently discovered – but have not had time to test -- something called *Advanced Search Bar* which claims to be one toolbar access to over 90 search engines including the ones featured in this paper. It works with Internet Explorer and can be found at:

<http://www.advancedsearchbar.com/>

Cached pages

If you are looking for older or “superseded” Web pages then *Google*’s cache feature might be helpful. To access *Google*’s cached copy of any Web site in your results list – click on the “Cache” link next to the result. This takes you to the cached copy that resides in *Google*’s database and dates from the time it was indexed, rather than taking you to today’s current version of the page. These two pages may be exactly the same because *Google* searches the Web and caches pages fairly frequently; however, an added

benefit of using this link is that your search terms are highlighted in the cached copy. If you really need an earlier version of a page – say your client claims to have been libeled on a competitor’s Web site – and you want to find the version of the page with these comments which have since been removed, then first try *Google*. If this doesn’t work then try the specialized Web site called the *Internet Archive* (or the *Wayback machine*) -- located at: <http://www.archive.org>. The *Internet Archive* collaborates with the Library of Congress, the Smithsonian Institution, and others to store and record Web pages. In August 2003 the site boasted more than ten billion Web pages in its archive – while it may not be complete, it does offer a rare opportunity to view Web site content that has been changed or removed from the Internet.

Synonyms, URL info, Who links to a site

In addition to its toolbar, catalog search and cached pages, *Google* offers several other handy “hidden” tools. They have a **synonym feature** that lets you search for not only the word you type in the search box, but also for some common synonyms of the word. To do this use the syntax ~word, so, for example, if you type ~lawyers in the search box, you will retrieve Web pages that have the word law, legal, attorney, attorneys, lawyers, and bar association. Use this tool when you are looking for a broad category of concepts, and be prepared for a few unexpected results.

If you are looking for **information about a particular URL** *Google* offers two ways to enter a search. Enter the search prefix info: into the search box followed by the URL for which you want information (E.g.: info:www.google.com), or simply type the Web page URL directly into a *Google* search box.

Google uses a ranking relevancy based in part on the number and quality of sites linking to a page. If you are interested in finding out **what pages have linked to one**

of the Web sites returned in your results list then scroll down on the advanced search page and look for the heading “Page Specific Search”. Just enter the URL of the Web site you are interested in and click on search. You can also type the prefix link: followed by a URL into the search box. This is a useful way to locate similar sites and also to judge how well known or thought of your Web site is.

Yahoo! Shortcuts

A Yahoo! Shortcut is a quick way to get to the information you want. A Shortcut automatically appears when it is relevant to your search and can contain links to useful content from *Yahoo!*, its partners, or across the web. To trigger these shortcuts yourself you can go to the “Shortcuts” link by clicking on **More** at the right top end of the search box and then selecting the Shortcuts link. You can then scroll down to see what Shortcuts are available. You can enter a search directly on this page next to the type of information you are looking for. There are shortcuts to information on area codes, zip codes, measurement conversions, maps, traffic reports, movie times and many other things. While *Google* offers many of these same limited searches the syntax to use is harder to find and not located easily in one place as it is in *Yahoo!*.

Context searching or clustering: Y!Q and *Clusty*

Yahoo! is offering a beta service called **Y!Q** at: <http://yq.search.yahoo.com> you can use it to show answers to your question in their context. The simplest way to use **Y!Q** is to paste in the text of whatever you are reading that piqued your curiosity. If you were reading a news article about the problems New Orleans will have in disposing of tons of Hurricane Katrina debris, you could highlight the text of the news item, paste it into the **Y!Q** search box, and *Yahoo!* will look for other pages that focus on this topic.

Clusty, mentioned earlier in this paper, is a meta search engine that groups results in context clusters. This can be particularly useful when you are looking for information on something that includes a term used in various contexts. A search on the term “land use” is broken down into twenty clusters, including: Law, Planning and Zoning, Issues, Consultants and Land Use History. This allows the searcher to focus in on his specific topic without wading through lots of sites, thus saving time.

Assessing the Quality of Information on the Internet

Finding a web site with the information you need is only the first step. Next you have to evaluate the quality of this information. Who coined the phrase “*Question Authority!*”? Look at five or six different web sites and you might get six or eight different answers. Honestly. The quote has been attributed to Timothy Leary, Socrates, Unitarian Universalists, Benjamin Franklin and even a bumper sticker.

The Internet is perhaps the ultimate vanity press. Anyone can publish anything. Few costs are incurred and little expertise is required. Unlike traditional print resources, even web resources sponsored by reputable organizations rarely have editors or fact-checkers. Currently, no web standards exist to ensure accuracy, and there is no organized system of peer review for sites in most disciplines. Consequently, each individual user of Internet resources must assume responsibility for ensuring the authenticity, accuracy, currency, and completeness of information obtained from Internet sources. I am not suggesting that you cannot rely on legal information that you find on the Internet. I am also not suggesting that you must spend a substantial amount of time evaluating every web site that you visit. I am suggesting that you must

always remain alert and assess the quality of legal information that you locate on the Internet.

When conducting Internet legal research, one of the first questions you should ask yourself is: Am I in the right place? If you are searching for cases, you should be at the web site for the court that originally published the opinions. If you are searching for statutes, go to the web site for the state or federal legislature. Do not use secondary sources for material that is readily available online from the actual source. If you are not searching a web site that is the authoritative source for the information you are seeking, you must assess the authenticity, accuracy, currency, and completeness of information you have located.

Authenticity

- To judge a web site's authenticity, you must judge its credibility, reliability or believability. In accessing the authenticity of information on an Internet site, you must often employ extensive investigation skills to locate information about the source of information. The first step is to read any site documentation that you can find. To locate site documentation: look under links with titles like "More about us", or "About this site."
- Go to the home page of the site sponsor if the documentation is not evident on the page where you enter the site.
- If you cannot determine the author or publisher of a site, you can always try to examine the html code of a web site to look for clues about its development. This method is similar to REVEAL CODES in WordPerfect. In both Netscape and

Internet Explorer, under the VIEW pull down menu there is an option for web page's SOURCE. Oftentimes an author or web developer's identity will be revealed in the source of the document.

- If you cannot determine the author or publisher of a site, examine the structure of the web address. Many web addresses are readily identifiable by their extensions. For example: gov = government, edu = educational institution, org = nonprofit organization, com = commercial organization. Similarly a web address with a tilde (~) is primary evidence that the web page is an unofficial, unauthorized or personal page. You could also use the info: or link: command search on *Google* to get more information about a web site.

Once you find documentation or clues about a web site's origin, there are a number of useful questions to ask:

- Who are the author and publisher of the site? What are the credentials of the author and publisher of the site? What gives the author the "authority" to write?
- Does the author provide an email address, snail mail address, telephone number, a mailto: link or other contact information?
- Is there a statement of authority included in the site's documentation?
- Does a government entity, a university, a legal institution, a law firm, an attorney, a party to a pending case, a private vendor, or an individual sponsor the site?
- What are the sources of the data provided? Who provided it and in what format?
- Does the site documentation include some sort of guarantee of content authenticity?
- Is there a stated criterion for inclusion of information?

Accuracy

In trying to determine the accuracy of information provided on an Internet site, you will of course apply much of the same scrutiny that helps you ascertain authenticity. You want to assess the authority of the author or creator of the site. Ask yourself:

- Do you recognize the name of the publisher or author? If not, does the publisher provide verifiable evidence of its competency?
- Are there citations to other published works, a corporate profile, and information about editorial standards?
- If you have never heard of the author, does she supply an autobiography or curriculum vita containing verifiable evidence of her authority on the subject?
- Examine the names of individuals or groups responsible for information supplied by the site. A credits and conditions statement might offer this information.

In assessing the accuracy of information, you must always try to determine the **objectivity** of the source. Information pretending to objective but possessing a hidden agenda of persuasion or a hidden bias is among the most common kind of information in our culture. Although lack of objectivity does not necessarily mean a source provides substandard information, you must always beware of partiality. Try to locate and read the source's purpose statement. Take into consideration the appearance of advertisements touting the site or other means of persuasion. Try to judge the extent, if any, of editorial enhancements. Try to identify the reason the web site was created. Is the main purpose to inform, to persuade, or to sell you something? And here is an important, if difficult, question to ask: What is not being said?

If you are familiar with print sources that supply related information, ask yourself how the Internet site compares with those sources. Try to determine if the content of the Internet site is as reliable as its printed counterpart.

Check for errors or omissions in the documents. Look for misspellings and grammatical errors. If you spot more than a few typos on a web site you should be very skeptical. An occasional split infinitive or comma in the wrong place is not unusual, but more than two or three spelling or grammar errors is cause for caution. Visit links provided on the web site to see if they are current, relevant and accurate. If you revisit a site, try to remember to check the stability of the information provided.

Finally, and most important, **verify** all information on which you choose to rely. If you are unsure of the accuracy of data you have located, attempt to find the same data on a reputable web site.

Currency

Information can be timeless (for example Aristotle or Plato) or timely (for example, the latest biogenetics development). You must first decide if timeliness or currency is relevant to your research. Old information is not necessarily bad. However, in general, timeliness is very important for most aspects of legal research.

One particular value of all online information sources is that they can be continuously updated much more easily than traditional research sources. However, just because such updating is possible, does not mean that it is done. When you visit a site, look to ascertain the source's creation and revision dates. Very often each individual page of a site will have a creation/revision date or copyright date beneath

titles or at the bottom of the page. If the site has been revised, see if you can determine what the author/publisher modified. Try to determine if the resources are maintained and how often they are updated. Many sites provide a currency declaration, which details updating schedules or at least articulates the site manager's commitment to ongoing maintenance and stability. Always record the date on which you review information from an Internet source. This is important to include when citing to Internet resources because of their transitory nature.

Completeness

When you obtain information from an Internet source always question its comprehensiveness or coverage. One of the biggest disadvantages of Internet legal research resources is the lack of comprehensiveness. The most comprehensive primary legal research materials available are the U.S. Supreme Court decisions and unannotated federal and state statutes. There is no free comprehensive Internet database of the remaining federal court decisions or any state court decisions.

Many sites provide only "selected" data. Be aware of this and look for statements that give clues of this. In addition, review a site carefully for an indication of the dates it claims to cover. Once again, if you are familiar with print sources for the same information, it is a good idea to compare the coverage of an Internet site with that of the print source.

In conclusion, always bring your skills of critical judgment with you when you visit web sites. Use the above mentioned search engine tools for finding "on point" Internet sites when you don't know where to start, and go to known and recommended

web sites – many of which will be highlighted by my colleagues later today -- whenever you can

Sources for further learning

If you are interested in learning more about Web searching and evaluating your results – I recommend the sites, newsletters, tutorials and book below. There are always new tips to learn and keeping up with the ever growing and changing world of the Internet is a continuing challenge.

Bates Information Service Tip of the Month

<http://www.batesinfo.com/tip.html>

Mary Ellen Bates, a librarian and information broker, is the principal of Bates Information Service and a highly regarded online searcher. She writes this monthly tip newsletter available via e-mail. The tip for October 2005 was “Two cool *Yahoo!* tools.” The current and some back issues are available at the Web site above.

Levitt, Carole A. and Mark E. Rosch The lawyer's guide to fact finding on the Internet. Chicago: American Bar Association Law Practice Management Section, 2004.

Chapter 3 and 4 of this source cover Search Strategies and Search Tools respectively. It is well written and presented and easy to follow with lots of great advice. Other chapters focus on specific legal research topics.

Tutorials

For systematic training in searching the Internet, try a tutorial such as the *Guide to Effective Searching of the Internet* offered free at a Web site called *Bright Planet* at: <http://www.brightplanet.com/deepcontent/tutorials/search/index.asp> or *Finding Information on the Internet: A Tutorial* from the University of California at Berkeley at: <http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/FindInfo.html>

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 Exchange (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. Includes the excellent article *Search the Web More Efficiently* by Daniel Bazac, which includes a section on criteria for evaluating a web resource --

<http://www.llrx.com/features/efficientsearch.htm>

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, materials on evaluating resources, and more.

Profusion

<http://www.profusion.com/>

You can choose to submit your search to from one to nine search engines, or let Profusion pick the best search engines to fit your terms. Results are combined and duplicates are eliminated. Searches may also be done in topics such as news, business, health, investing, and others. This is a directory of over 10,000 databases, archives, and search engines that contain information that traditional search engines have been unable to access. The deep coverage of this site makes it possible to retrieve more sites from the Invisible Web.

Complete Planet

<http://www.completeplanet.com/>

The largest compilation of searchable databases and engines on the Internet -- 40,000 listings and growing -- placed in a 7,000 category directory structure. It offers access to an impressive collection of valuable Web search tools and databases.

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. Effective use of Web search engines ease the problems associated with searching the Web. Learning and applying a few of the mentioned advanced search engine strategies, tools and tips will make your search for information on the Web successful. In addition, applying the techniques for evaluating your Internet sources discussed above will allow you to weed out the wheat from the chaff and assure you of accurate and reliable information.