

12-12-2007

Find It Free and Fast on the Net: Strategies for Legal Research on the Web (2007)

Michelle Rigual

University of New Mexico - School of Law

Follow this and additional works at: https://digitalrepository.unm.edu/law_facultyscholarship



Part of the [Law Commons](#)

Recommended Citation

Michelle Rigual, *Find It Free and Fast on the Net: Strategies for Legal Research on the Web (2007)*, National Business Institute Continuing Legal Education Seminar (2007).

Available at: https://digitalrepository.unm.edu/law_facultyscholarship/153

This Presentation is brought to you for free and open access by the UNM School of Law at UNM Digital Repository. It has been accepted for inclusion in Faculty Scholarship by an authorized administrator of UNM Digital Repository. For more information, please contact amywinter@unm.edu, lsloane@salud.unm.edu, sarahrk@unm.edu.

I. HOW TO SEARCH LIKE A PRO

- A. What's the Difference Between a Search Engine and a Directory?**
- B. How do Search Engines Really Work?**
- C. Learn Boolean Logic to Create the Best Search Strategies for Finding Relevant Documents**
- D. Which Search Engines Are the Best?**
- E. A Close Look at Google and Search Engines That "Think"**
- F. Learn Browser Tricks – The Power of the "Find" function to Scan Web Pages Quickly**
- G. Managing Bookmarks**
- H. Effectively Using RSS Feeds, Blogs and Podcasts**

A. What's the Difference Between a Search Engine and a Directory?

In a nutshell: search engines are fully automated while directories have human editors.

Search engines. Search engines are huge databases of web pages and other files that have been assembled automatically by machine. The databases are compiled by employing "spiders" or "robots" to crawl through web space from link to link. Once the spiders get to a web site, they typically index most of the words on the publicly available pages at the site. Web page owners also may submit their URLs to search engines for "crawling" and eventual inclusion in their databases.

Whenever you search the web using a search engine, you're asking the engine to scan its index of sites and match your keywords and phrases with those in the texts of documents within the engine's database. You are not searching the entire web as it exists at this moment. You are actually searching a portion of the web, captured in a fixed index created at an earlier date. Spiders regularly return to the web pages they index to look for changes. When changes occur, the index is updated to reflect the new information. The process of updating can take a while, depending upon how often the spiders make their rounds and then, how promptly the information they gather is added to the index. Until a page has been both "spidered" AND "indexed," you won't be able to access the new information.

While most search engine indexes are not "up to the minute" current, they have partnered with specialized news databases that are. For late breaking news, look for a "news" tab somewhere on the search engine or directory page.

Search engines are best at finding unique keywords, phrases, quotes, and information buried in the full-text of web pages. Because they index word by word, search engines are also useful in retrieving tons of documents. If you want a wide range of responses to specific queries, use a search engine.

Search engines provide access to a fairly large portion of the publicly available pages on the Web, which itself is growing exponentially. They are the best means devised yet for searching the web. Stranded in the middle of this global electronic library of information without any recognizable structure, how else are you going to find what you're looking for?

On the down side, the sheer number of words indexed by search engines increases the likelihood that they will return hundreds of thousands of responses to simple search requests. Remember, they will return lengthy documents in which your keyword appears only once. Additionally, many of these responses will be irrelevant to your search.

With the ascendancy of Google, the number of sites with unique search engines is dropping.

Subject directories. Subject directories are created and maintained by human editors who review and select sites for inclusion in their directories. The resources they list are usually annotated. There is some question as to whether directories can remain viable as the number of web pages continues to grow exponentially.

When you initiate a keyword search of a directory's contents, the directory attempts to match your keywords and phrases with those in its written descriptions. Subject directories come in assorted flavors: general, academic, commercial, and subject-specific.

Like the yellow pages of a telephone book, subject directories are best for browsing and for searches of a more general nature. They are good sources for information on popular topics, organizations, commercial sites and products. When you'd like to see what kind of information is available on the Web in a particular field or area of interest, go to a directory and browse through the subject categories.

Directory editors typically organize directories hierarchically into browseable subject categories and sub-categories. When you're clicking through several subject layers to get

to an actual Web page, this kind of organization may appear cumbersome, but it is also the directory's strength. Because of the human oversight, they have the capability of delivering a higher quality of content. They may also provide fewer results out of context than search engines.

Unlike search engines, most directories do not compile databases of their own. Instead of storing pages, they point to them. This situation sometimes creates problems because, once accepted for inclusion in a directory, the Web page could change content and the editors might not realize it. The directory might continue to point to a page that has been moved or that no longer exists. Dead links are a real problem for subject directories.

Today, the line between subject directories and search engines is blurred. The most popular search engine is Google. However, Google also has a directory, which you can reach by clicking on the directory link. The directory is simply imported from the Open Directory Project. Similarly, Yahoo uses Google's search engine results but also has its own directory too.

B. How Do Search Engines Really Work?

Spiders. Spider-based search engines have three major elements. First is the spider that visits a web page, reads it, and then follows links to other pages within the site. This is what it means when someone refers to a site being "spidered" or "crawled." The spider returns to the site on a regular basis, such as every month or two, to look for changes.

Index. Everything the spider finds goes into the second part of the search engine, the index. The index, sometimes called the catalog, is like a giant book containing a copy of every web page that the spider finds. If a web page changes, then this book is updated with new information.

It can take some time for new pages or changes that the spider finds to be added to the index. Thus, a web page may have been "spidered" but not yet "indexed." Until it is indexed -- added to the index -- it is not available to those searching with the search engine.

Software. Search engine software is the third part of a search engine. This is the program that sifts through the millions of pages recorded in the index to find matches to a search and rank them in order of what it believes is most relevant.

All crawler-based search engines have the basic parts described above, but there are differences in how these parts are tuned. That is why the same search on different search engines often produces different results.

Search engines have a variety of ways for you to refine and control your searches. Some of them offer menu systems for this. Others require you to use special commands. The more specific your search is, the more likely you will find what you want. Don't be afraid to tell a search engine exactly what you are looking for.

C. Learn Boolean Logic to Create the Best Search Strategies for Finding Relevant Documents

Boolean logic takes its name from 19th century British mathematician George Boole who wrote about a system of logic designed to produce better search results by formulating precise queries. He called it the "calculus of thought." From his writings, we have derived Boolean logic and its operators: AND, OR, NOT, and others which we use to link words and phrases for more precise queries.

Boolean logic is a powerful tool for efficiently obtaining precise results from database search engines, whether it the database is Google, Lexis, CaseMaker, or any other.

AND. The Boolean AND narrows your search by retrieving only documents that contain every one of the keywords you enter. The more terms you enter, the narrower your search becomes.

truth AND justice

truth AND justice AND ethics AND congress

OR. The Boolean OR expands your search by returning documents in which either or both keywords appear. Since the OR operator is usually used for keywords that are similar or synonymous, the more keywords you enter, the more documents you will retrieve.

college OR university

college OR university OR institution OR campus

NOT/AND NOT. The Boolean NOT or AND NOT (sometimes typed as ANDNOT) limits your search by returning only your first keyword but not the second, even if the first word appears in that document, too.

saturn AND NOT car

rico AND NOT puerto

Nesting. Nesting, i.e., using parentheses, is an effective way to combine several search statements into one search statement. Use parentheses to separate keywords when you are using more than one operator and three or more keywords. For best results, always enclose OR statements in parentheses.

(hybrid OR electric) AND (Toyota OR Honda)

Boolean logic is not always simple or easy. Different search engines handle Boolean operators differently. For example, some accept NOT, while one accepts ANDNOT as

one word, others AND NOT as two words. Some require the operators to be typed in capital letters while others do not.

Proximity. Proximity, or positional, operators (NEAR and ADJ) are not really part of Boolean logic, but they serve a similar function in formulating search statements.

Not all search engines accept proximity operators, but a few accept NEAR in their advanced search option. The NEAR operator allows you to search for terms situated within a specified distance of each other in any order. The closer they are, the higher the document appears in the results list. Using NEAR, when possible, in place of the Boolean AND usually returns more relevant results.

Advanced Searching. Nearly all search engines and professional databases offer "basic" and "advanced" searching. Advanced search refining options differ from one search engine to another, but some of the possibilities include the ability to use Boolean operators, to search on more than one word, to give more weight to one search term than you give to another, and to exclude words that might be likely to muddy the results. You might also be able to search on proper names, on phrases, and on words that are found within a certain proximity to other search terms. Some search engines also allow you to specify what form you'd like your results to appear in, and whether you wish to restrict your search to certain types of pages such as those from a university or from the government or to specific parts of Web documents such as the title or URL. Look at the home page for links that say "advanced search."

Using Keywords Creatively Before searching, try to imagine what the ideal page you would like to access would look like. Think about the words its title would contain. Think about what words would be in the first couple of sentences of a webpage that you would consider useful. Use those words, or that phrase, when you enter your query.

A search engine's ability to understand what you want is very limited. It will obediently look for occurrences of your keywords all over the Web, but it doesn't understand what

your keywords mean or why they're important to you. To a search engine, a keyword is just a string of characters. It doesn't know the difference between cancer the crab and cancer the disease and it doesn't care.

Boolean operators and advanced search functions give meaning to your keywords. Here are a few other useful search strategies:

- Put your most important keywords first in the string.
- Use quotation marks around phrases to ensure they are searched exactly as is, with the words side by side in the same order.
- Type keywords and phrases in lower case to find both lower and upper case versions. Typing capital letters will usually return only an exact match.
- Use plus and minus signs in front of words to force their inclusion and/or exclusion in searches.
- Combine phrases with keywords, using the double quotes and the plus and/or minus signs.
- Use truncation and wildcards to look for variations in spelling and word form.
- Know the default settings your search engine uses (OR or AND).
- Know whether or not the search engine you are using maintains a stop word list. If it does, don't use known stop words in your search statement.
- In Boolean searches, always enclose OR statements in parentheses.
- Always use capital letters when typing Boolean operators in your search statements. Some engines require that the operators be capitalized; others will accept either so you're on safe ground if you stick to all CAPS.

D. Which Search Engines are the Best?

Many of my favorite databases are gone because, though they had some strong attributes, they could not compete with Google's size and ability to retrieve highly relevant material.

Google. Google has become the preeminent Web search engine. It made its mark with its relevance ranking based on link analysis, cached pages, and aggressive growth.

Strengths:

- Size and scope: It is the largest, and includes PDF, DOC, and many other file types
- Relevance based on sites' linkages and authority
- Cached archive of Web pages as they looked when indexed
- Additional databases: Images, News, Directory, etc.

Weaknesses:

- Limited search features: no nesting, no truncation, does not support full Boolean
- Only indexes first 101 KB of a Web page and about 120 KB of PDFs
- May search for plural/singular, synonyms, and grammatical variants without telling you
- Not as comprehensive as is popularly believed to be.

Yahoo. 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.

Strengths:

- A very large, unique search engine database
- Includes cached copies of pages
- Also includes links to the Yahoo Directory
- Supports full Boolean searching
- Wildcard word in phrase

Weaknesses:

- Lack of some advanced search features such as truncation
- Only indexes first 500 KB of a Web page (still more than Google's 101KB)
- Link searches require the inclusion of the http://
- Includes some pay for inclusion sites

Live Search. Live Search is the successor to MSN Search. Sometimes called just Live.com or Windows Live Search, this is the Microsoft Web search engine. Launched in September 2006, it uses its own, unique database

Strengths:

- Large, fresh, unique database
- Query building Advanced Search and full Boolean searching
- Cached copies of Web pages including date cached
- Automatic local search options.

Weaknesses:

- No truncation, stemming, or wild card word in a phrase
- Limited to 10 words in a query
- Advanced search not on front page, but available after running a search.

Ask.com. The search engine formerly known as Ask Jeeves has changed greatly from its early days as a question answer matching service to being a real Web search engine using a database originally developed by Teoma. The question-answering tradition continues in Ask's shortcuts that give a direct answer for many popular query types right at the top of regular search results.

Strengths:

- Identifying metasites

- Refine feature to focus on Web communities

Weaknesses:

- Smaller database
- No free URL submission
- No ability to uncluster results to easily see more than two hits per site
- No cached copies of pages

Search engine attributes are monitored and measured by the Search Engine Showdown website. Some of the attributes measured include: total size, relative size, freshness, change over time, unique hits, and dead links. Overall, Google outperformed all other databases on these attributes. See <http://www.searchengineshowdown.com> for more information.

Jeteye. Jeteye users can create custom content packages (Jetpaks) consisting of links, images, notes, and even other Jetpaks. A researcher might use one to chronicle sites visited while researching a project. Building a Jetpak is as simple as performing a search using the four or five supported Web-search engines, including Google. Results appear in a list, search engine–style; click a link to preview the site in an embedded window or click the accompanying check box to add the link to your Jetpak. A similar process adds images, though Jeteye currently limits image searches to Ditto and Flickr (Google's image search isn't available). Jetpaks can be made private or public, including publicly modifiable, like a wiki.

Blinkx. A video search engine fed by automatic spiders that crawl the web for audio video content and by content partnerships with leading content and media companies. Users can search for content and create personal TV channels that automatically splice relevant content together.

Rollyo. Create personalized search engines by picking the sites you want to search. The tool offers the option of saving, sharing, and personalize these.

E. A Close Look at Google and Search Engines that “Think”

Google is one of the smartest search engines. It was created in the winter of 1998 by graduate students at Stanford University. From the beginning, Google has endeavored to present a clean, attractive interface. This is a straightforward, easy-to-use engine that is noted for its highly relevant results.

The heart of Google’s software relies on the uniquely democratic nature of the web by using its vast link structure as an indicator of an individual page's value. In essence, Google interprets a link from page A to page B as a vote, by page A, for page B. But, Google looks at considerably more than the sheer volume of votes, or links a page receives; for example, 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." Using these and other factors, Google provides its views on pages' relative importance.

Of course, important pages mean nothing to you if they don't match your query. So, Google combines its ranking system with text-matching techniques to find pages that are both important and relevant to your search. Google goes far beyond the number of times a term appears on a page and examines dozens of aspects of the page's content (and the content of the pages linking to it) to determine if it's a good match for your query.

A search in Google can encompass its own database of indexed files, paid advertisements (these are shown on the right side under a "Sponsored Links" heading), an image database, the past 30 days of Web-based news sites, the Open Directory Project, and a shopping and product and catalog search.

Default Operation. Multiple search terms are processed as an AND operation by default. Phrase matches are ranked higher.

Boolean Searching. Google uses an automatic Boolean AND between terms and has slowly been moving towards more Boolean support; however, it does not yet support the AND operator, NOT operator, or full Boolean searching with the ability to nest operators. Since its inception, Google has added the - symbol to perform a NOT function and the ability to use an OR (which must be in upper case) to do some Boolean OR operations.

The + used to be able to be used to require a term, but since the default operation was AND, the + was never really needed. However, the + can be used for forcing a search on stop words and for requiring Google to search for only that exact term without any possible plural/singular, synonyms, and grammatical variants.

Proximity Searching. Google allows phrase searching designated in the usual manner by enclosing the phrase in "double quotes." Google also detects phrase matches even when the quotes are not used and usually ranks phrase matches higher. No other proximity searching is directly available.

Truncation. No truncation is available. Some automatic plural searching and word stemming occurs for English words and can be turned off by using the plus sign in front of each term that should not be stemmed. However, within phrases, there is a trick that can be used for a wildcard word. Use an asterisk * within a phrase search to match any word in that position. So, for example, to find "a little neglect may breed mischief" when you are not sure of the second to last word, search "a little neglect may * mischief". Multiple asterisks can be used as in "a little * * * mischief". This is the only way Google supports a wildcard symbol.

While not exactly truncation, the synonym operator of a tilde ~ before a search term, with no space, directs Google to look for synonyms. So a search on Yosemite ~trails will find pages that have terms like 'hiking,' 'rides,' and 'maps.' This synonym finder will sometimes include plural, singular, or other grammatical variants as well. So the ~ can be used to get something a bit closer to truncation but not very. The ~ only works in Google's Web database and only for English language terms.

Case Sensitivity. Google has no case sensitive searching. Using either lower or upper case results in the same hits.

Limits. Google has language, domain, date, filetype, and adult content limits. Most of these are only available on the Advanced Search page.

Stop Words. Google does ignore frequent words. Its documentation mentions terms such as 'the', 'of', 'and', and 'or'. However, it also notes that these can be searched by putting + in front of them. Be sure to only place the + in front of stop words. If a + is placed in front of a non-stop word in the same query, all + signs will be ignored. Stop words within a phrase no longer require a + sign and will automatically be searched. Also, if only stop words are entered even without phrase markings, they will be searched.

Sorting. Results are sorted by relevance using its PageRank analysis, determined by links from other pages with a greater weight given to authoritative sites. Pages are also clustered by site. Only two pages per site will be displayed, with the second indented. Clicking the "repeat the search" option will bring up more pages, some of which are near or exact duplicates of pages already found while others are pages that were clustered under a site listing.

Display. The display includes the title, URL, a brief extract showing text near the search terms, the file size, and for many hits, a link to a cached copy of the page. The cached copy will display highlighted search terms. If more than one search term is used, each has a different color highlighting. The default output is 10 hits per screen, but the searcher can also choose 20, 30, 50, or 100 hits at a time on the preferences page.

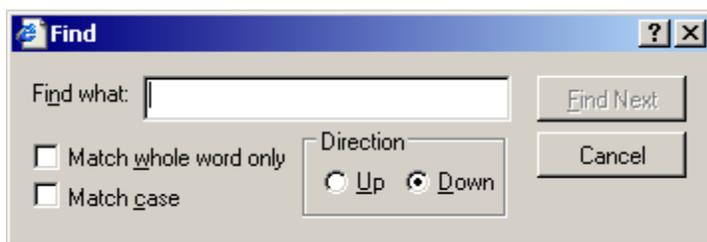
Google refreshes the indexing of certain pages (those with daily updates) more frequently than the rest of the database. These are marked with the indexing date. As of Feb. 2002, 3 million pages were being refreshed on an almost daily basis.

Cool Google Features. Google has a number of regular and beta features that can help the searcher and that are just fun to use. The Google website currently lists the following web search features:

- Book Search - Use Google to search the full text of books.
- Cached Links - View a snapshot of each page as it looked when indexed.
- Calculator - Use Google to evaluate mathematical expressions.
- Currency Conversion - Easily perform any currency conversion.
- Definitions - Use Google to get glossary definitions gathered from various online sources.
- File Types - Search for non-HTML file formats including PDF documents and others.
- Froogle - To find a product for sale online.
- Groups - See relevant postings from Google Groups in your regular web search results.
- I'm Feeling Lucky - Go to the first web page returned for your query.
- Images - See relevant images in your regular web search results.
- Local Search - Search for local businesses and services in the U.S., the U.K., and Canada.
- Movies - Use Google to find reviews and showtimes for movies playing in your area.
- Music Search - Use Google to get quick access to a wide range of music information.
- News Headlines - Enhances your search results with the latest related news stories.
- PhoneBook - Look up U.S. street address and phone number information.
- Q&A - Use Google to get quick answers to straightforward questions.
- Refine Your Search - Add instant info and topic-specific links to your search in order to focus and improve your results.

- Search By Number - Use Google to access package tracking information, US patents, and a variety of online databases.
- Similar Pages - Display pages that are related to a particular result.
- Site Search - Restrict your search to a specific site.
- Spell Checker - Offers alternative spelling for queries.
- Stock and Fund Quotes - Use Google to get up-to-date stock and mutual fund quotes and information.
- Street Maps - Use Google to find U.S. street maps.
- Travel Information - Check the status of an airline flight in the U.S. or view airport delays and weather conditions.
- Weather - Check the current weather conditions and forecast for any location in the U.S.
- Web Page Translation - Provides you access to web pages in other languages.
- Who Links To You? - Find pages that point to a specific URL.

F. Learn Browser Tricks - The Power of the “Find” Function to Scan Web Pages Quickly



Once you have found the Web page you want, it is easy to quickly scan the page.

Look to the top left corner of the webpage, click on the Edit menu, then click on Find. A handy "find" box will appear in the middle of your screen. This box lets you type in a specific word. After you type in your search term, hit the Find Next button and your

browser will look for that word on the page and take you right to it. Once you see the word, click Cancel.

Even easier, is holding down the CTRL and F buttons simultaneously. This shortcut will also make the find box appear.

Other Browser Commands. With just a little practice, you will quickly find you can enter and invoke multiple browser commands with the old-fashioned keyboard faster than with any other method, often faster than the program can react. Now that's fast.

Not all web browser commands have equivalent keyboard commands, but most of the common commands do. Some of the most well-known and useful commands are listed below:

Keyboard Command	Description	Internet Explorer	Mozilla Firefox / Netscape
<Alt><LeftArrow>	Page back.	X	X
<Alt><RightArrow>	Page forward.	X	X
<Ctrl>-[Decrease text font		X
<Ctrl>-]	Enlarge text font		X
<Ctrl>-a	Select all of a page, form, or field.	X	X
<Ctrl>-b	Edit bookmarks	X	X
<Ctrl>-c	Copy the selected text.	X	X
<Ctrl>-d	Add bookmark	X	X

<Ctrl>-f	Find a text string.	X	X
<Ctrl>-h	Open history file	X	X
<Ctrl>-i	Page info.		X
<Ctrl>-n	New window	X	X
<Ctrl>-r	Reload / Refresh	X	X
<Ctrl>-v	Paste.	X	X
<Esc>	Stop page load.	X	X

G. Managing Bookmarks

Traditional Internet Bookmarks. These bookmarks (called "Favorites" in Internet Explorer) record an Internet URL so you can revisit it later without having to remember its address.

Bookmarks are one of the early innovations of Mosaic, the first web browser. Early on, bookmarks were simple lists of URLs stored in a "Bookmark" menu. They couldn't be sorted, and all new sites went to the bottom of the bookmark list. Today, modern browsers and third party bookmark managers provide a range of advanced management features.

A knowledge of Bookmarks enables you to easily revisit the sites you like without having to remember their exact URLs. You can collect lists of good sites, and build up a library of your favorite and most useful destinations.

Use. Once you add a bookmark to your list, you simply select it from your browser's bookmark list to revisit the page. You add a page to your bookmarks by visiting it and then adding it with your browser's add bookmark feature.

Managers. A range of third party tools have been developed to help you manage bookmarks. Many provide import and export capability with the leading browsers.

Different bookmark managers provide different capabilities, such as advanced organization and searching features, automatic checking for dead links, upload to the web, and other functions. You can find more information at the following sites:

- Google Bookmark Managers
- Open Directory Bookmark Managers
- Yahoo Bookmark Managers

Categories. Managing bookmarks consists of deciding on a set of categories in which to store your bookmarks. Most browsers come with a built-in set of categories and bookmarks, but you should create your own so they aren't cluttered by sites you don't use.

Creation - Create a set of categories for your major subject area interests.

Order - Move the categories you use frequently to the beginning, and arrange the rest in alphabetical order. - Drag items up and down in right-half of window.

Subfolders - Add subfolders for subcategories when necessary; a general rule of thumb is to create subfolders when you get more than twenty bookmarks in a category. Keep in mind that you maximize a bookmark's visibility by keeping it at a high level, and you double the difficulty of finding a bookmark when you move it down one level.

Separators - Put in separators to divide different sets of bookmarks.

Editing - Leading browsers allow bookmark editing

Export - Internet Explorer favorites files can be exported to an HTML page.

Social Bookmarks. Social bookmark sites allow you to store, tag and share links across the internet. In a social bookmarking system, users save links to web pages that they want to remember or share. These bookmarks are usually public, but depending on the service's features, may be saved privately, shared only with specific people or groups, shared only inside certain networks, or another combination of public and private. The allowed people can usually view these bookmarks chronologically, by category or tags, via a search engine, or even randomly.

Most social bookmark services encourage users to organize their bookmarks with informal tags instead of the traditional browser-based system of folders, although some services feature categories/folders or a combination of folders and tags. They also enable viewing bookmarks associated with a chosen tag, and include information about the number of users who have bookmarked them. Some social bookmarking services also draw inferences from the relationship of tags to create clusters of tags or bookmarks.

Many social bookmarking services provide web feeds for their lists of bookmarks, including lists organized by tags. This allows subscribers to become aware of new bookmarks as they are saved, shared, and tagged by other users. As these services have matured and grown more popular, they have added extra features such as ratings and comments on bookmarks, the ability to import and export bookmarks from browsers, emailing of bookmarks, web annotation, and groups or other social network features.

Del.icio.us . Founded in late 2003, they pioneered tagging and coined the term "social bookmarking". To register, go to <http://del.icio.us/register>.

Related Sites. A related type of web service that provides a system for social news:

Digg - <http://digg.com>

Reddit - <http://reddit.com>

Newsvine - <http://www.newsvine.com>

H. Effectively Using RSS Feeds, Blogs and Podcasts

RSS Newsfeeds. RSS, the abbreviation for Rich Site Summary, is a means of distributing news headlines on the Web that allows readers to sign-up and to receive news, headlines and short summaries from websites, weblogs, and online newsletters.

The first step is to find and download a news aggregator to your desktop in order to receive feeds. A news aggregator is a dedicated software program capable of contacting and gathering pre-selected news items posted at specific addresses on the Internet. What the news aggregator does is simply to go out to the URL/addresses you have specified, collect the information stored inside an auto-generating file that sits invisibly next to certain Web pages (and which contains all of the most recent updates to that page) and bring that information back inside the newsreader/aggregator while storing that information in a specific folder you have specified. Many exist today on the market, with names like Newzcrawler, AmphetaDesk, FeedDemon, NetNewsWire, Newsmonster, etc. Once the aggregator has been downloaded and installed, news feeds need to be placed inside. You may have seen sites with **XML** or RSS placed inconspicuously on the page. This means that this site is set up with RSS feed capabilities and can be placed in an aggregator for viewing and constant updating.

Blogs. A blog is a frequently updated, publicly accessible journal. While the “blogosphere” started off as a medium for mostly personal musings, it has evolved into a tool that offers some of the most insightful information on the Web.

In brief, blogs are reverse chronologically arranged collections of articles or stories that are generally updated more frequently than regular web pages. Just like any other information on the net, there is no guarantee of authority, accuracy, or lack of bias. In fact, personal blogs are frequently biased and can be good sources of opinion and information from the “man on the street.” Because blogs can be updated on the fly, they frequently have unfiltered information faster from war zones and sites of natural disasters than the mainstream media outlets. Blogs are also good sources of unfiltered information on either faulty or very useful products.

Blog searching breaks down into at least two categories. First are searches for addresses of feeds from blogs so that you may subscribe in your aggregator. Waypath provides this service. The second category is searches of information from within blogs/across blogs.

Use Feedster for a very precise search and Google for general searches. To search blogs in Google, enter your keywords then **~blog inurl:archives**. Legal issue blogs include Blawg.org and The Blawg Ring.

Podcasts. At its core, a podcast is nothing more than a recorded audio file, just as a blog is nothing more than a website. But like blogging, the value of podcasting comes from the content that is written and how widely it is distributed. Legal podcasters, like legal bloggers, share knowledge, opinions, and commentary on topics that can advance the profession while keeping us up-to-date on legal topics or new legislation. In many cases, they are produced by the same person creating content in both mediums, written and spoken.

You do not need an iPod to listen to podcasts. You can listen to them on any MP3 player or computer. To do this you will need to download a podcast aggregator. If you already use a news aggregator, you can use it for this as well. Once the aggregator is secured, you will need to "subscribe" to content. There are many directories that list available podcasts to which you can subscribe. Finally, connect your MP3 player of choice to your computer and transfer the audio files to your portable player. Once loaded onto your portable music player, you can listen to the podcast from anywhere. If you download the audio file to your computer, you can burn a CD and listen to it on the go as well.

Popular podcast directories include iTunes, Odeo, iPodder, and Podcast Alley. Each provides listings to many legal podcasts, though none have a specific legal category at this time. Podcasts specific to the legal industry include:

Lawyer Podcasts

- Bag and Baggage – Denise Howell, appellate and intellectual property lawyer with Reed Smith.
- May it Please the Court Podcast – J. Craig Williams, a founder of The Williams Law Firm.

- Coast to Coast – J. Craig Williams and Robert Ambrogi
- The Legal Talk Network – a network of lawyers nationwide providing the latest legal news.
- The Legal Underground Podcast – Evan Shaeffer, a trial lawyer in the midwest.
- The Rethink (IP) Aloud Podcasts – IP lawyers, Douglas Sorocco, J. Matthews Buchanan, and Stephen Nipper

Legal Information Providers

- FIOS Podcasts on Demand – FIOS podcasts featuring hour long seminars covering a wide range of E-Discovery topics.
- Merrill – Legal solutions provider has several on-demand seminars for download in mp3 format.
- Ten Minute Mentor – The State Bar of Texas' podcast on mentoring and practical information to lawyers.
- The Westcast – Thomson West podcasts on legal news.

II. DISCOVER THE INVISIBLE WEB

- A. What is the Invisible Web?**
- B. How Does it Differ From the Visible Web?**
- C. Learn to Find “Invisible Documents”**
- D. Where do Old Web Pages Go? Learn to Locate Them**

A. What is the Invisible Web?

The invisible web is what you cannot retrieve in the search results and other links contained in these types of tools.

B. How Does it Differ From the Visible Web?

The visible web is what you see in the results pages from general web search engines and subject directories. Search engines include millions of pages in their databases, but none of them come close to indexing the entire Web. Here is a list of some of what is missing:

- The content in sites requiring a log in
- CGI output such as data requested by a form
- Intranets; pages not linked from anywhere else
- Commercial resources with domain limitations
- Sites that use a robots.txt file to keep files and/or directories off limits
- Non-Web resources

There are two reasons a search engine does not contain a page: technical barriers that prohibit access and choices or decisions to exclude.

If the only way to access web pages requires you to type something or scan a page and select a combination of options, search engines are unable to proceed. Search engine spiders crawl the Web by following the links in the web pages that are already in the database of their parent search engine. If there is no link to a page, a spider cannot "see" it. Thus intranet sites which are not linked from outside become part of the invisible web. They also cannot scan a set of options and choose which one to select. They not only lack fingers for typing, but also lack a brain capable of judgment.

Pages created as the result of a search are called "dynamically generated" pages. The answer to your query is encased in a web page designed to carry the answer and sent to your computer. Often the page is not stored anywhere afterward, because its unique content (the answer to your specific query) is probably not of use to others. It's easier for the database to regenerate the page when needed than to keep it around.

The opposite of a "dynamic" page is a "static" page. Static pages reside on servers, each identified by a unique URL, and waiting to be retrieved when their URL is invoked. Spiders can find a static page if it is linked to in any other page they know about. They follow links to it and retrieve it much as you would by clicking if you knew the link. Static pages are not invisible.

The content of a lot web pages is both searchable and browsable by clicking on links. To the extent that the content found by searching is replicated in web pages with links somewhere, part (or all) of the content might be found in a general Web search engine (unless the pages are excluded from a search engine for policy reasons, discussed below). Any content not contained in static pages linked to somewhere else remains invisible. You have to search a database directly to find these.

All sites requiring a password or login are closed to search engine spiders because they require typing something spiders cannot "know." The contents of these sites are very unlikely to be in any general Web search engine. This includes all resources that require a password.

A script is a type of programming language that can be used to fetch and display web pages. There are many kinds and uses of scripts on the Web. They can be used to create all or part of a web page, and to communicate with searchable databases. Many database queries and responses use scripts. When you find a question mark (?) in the URL of a page, some kind of script command is used in that page. Most search engines are instructed not to crawl sites or include pages that use script technology, although it is often technically possible for them to do so. If spiders encounter a ? in a URL or link,

they are programmed to back off. They could encounter poorly written script or intentional "spider traps" designed to ensnare spiders, sometimes bogging them down in infinite loops that run up the cost and time it takes for spiders to do their work. So search engine companies instruct their spiders not to retrieve pages with URLs containing ?. This may result in the contents of an entire site using scripts being excluded from a search engine, or a search engine may crawl to a safe part of a site and omit others. A spider doesn't have the freedom and creativity to jump around a site intelligently.

C. Learn to Find “Invisible Documents”

You can find searchable databases and other invisible web material in the course of routine searching. Use Google and other search engines to locate searchable databases by searching a subject term and the word "database". If the database uses the word “database” in its own pages, you are likely to find it in Google. The word "database" is also useful in searching a topic in the Yahoo directory, because Yahoo sometimes uses the term to describe searchable databases in its listings.

for Google & Yahoo:

plane crash database

languages database

toxic chemicals database

There used to be a series of directories that purported to lead you to the invisible web, but most of them have deteriorated into a mixture of visible and invisible sites.

Current Invisible Web sites:

- Invisible Web Directory (<http://www.invisible-web.net>)
- Direct Search (<http://www.freepint.com/gary/direct.htm>)
- Profusion (<http://www.profusion.com>)

- CompletePlanet (<http://www.completeplanet.com>)
- The Super Searchers webpage (<http://www.infotoday.com/supersearchers>)

D. Where Do Old Web Pages Go? - Learn to Locate Them.

The web is constantly changing and web pages disappear or are changed without warning creating dead links and frustration. Even after a page has been modified or taken off the web there are ways to find it, access its content, and even link to it.

The Internet Archive is working to counter the transitory nature of information on the Internet through its Wayback Machine (<http://www.archive.org>). The Wayback Machine is the largest database in the world, containing multiple copies of the entire publicly available web. The database is created by using the same spiders employed by the commercial search engines. Each time the spiders crawl through the web they identify pages that have changed since the previous visit. Visitors to the site type in a URL and are presented with a list of archived versions of the site going back to 1996. Clicking on a date reveals the archived site. The archived pages can be linked to simply by copying the URL. Many of the websites that are part of the invisible web are also missing from this site. A site may not be included in the database if it is password protected, includes a command that won't allow the spider to search, or if the owner has requested it be removed.

The CyberCemetery (<http://govinfo.library.unt.edu>) provides permanent public access to the Web sites and publications of defunct U.S. government agencies and commissions. The old sites are listed alphabetically, by subject, and by region. The CyberCemetery was created and is maintained by the University of North Texas Libraries and the U.S. Government Printing Office, as part of the Federal Depository Library Program.

Search Engines also provide access to old web sites. Google takes a snapshot of each page examined as it crawls the web and caches these as a back-up in case the original

page is unavailable. If you click on the "Cached" link, you will see the web page as it looked the previous time it was indexed. The cached content is the content Google uses to judge whether this page is a relevant match for your query.

When the cached page is displayed, it will have a header at the top to serve as a reminder that it is not necessarily the most recent version of the page. Terms that match your query are highlighted on the cached version to make it easier for you to see why your page is relevant.

The "Cached" link will be missing for sites that have not been indexed, as well as for sites whose owners have requested their content not be cached. Yahoo provides a similar service, and Gigablast has a link to the Wayback Machine's archived versions of the pages.

III. FINDING FREE LEGAL RESEARCH SITES AND FREE CASE LAW

A. Finding the Shortcuts to Legal Research: Free Legal Portals and Meta-Sites

B. Finding and Using Free, Full-Text Case Law Sites

C. Learn About the Free (Poor Man's) "Shepards"

A. Finding the Shortcuts to Legal Research: Free Legal Portals and Meta-Sites

Portals, or meta-sites, in the general sense, are the one-stop-shops of the Web. The idea behind their design is to give you all the resources you are likely to want in one spot so that you spend your time there and don't bother to go elsewhere. The more time a user spends within a site, the more income the site generates from advertising. The payoff should be that the user saves time and effort. Typical offerings at portals include a search feature, subject links by topic, reference information, news, shopping links and sometimes customization options. Some portals also offer e-mail accounts and web hosting. Legal portals are similar sites focused on legal resources. While a general portal – Yahoo, for instance – would probably have some legal resources, a good topic-specific portal should quickly direct its user a more exhaustive range of resources, and those resources will mainly be targeted at a special audience.

Major Portals. The major portals, or old standards, include FindLaw, Hieros Gamos, Legal Information Institute from Cornell Law School, and Washburn University's WashLaw.

Minor and Newer Portals. There are many more legal portals available on the Web. These include the general sites of AllLaw.com, LawyerExpress, Law Guru, and The Virtual Chase. Others have a more narrow focus such as Employment Law Information Network, GigaLaw.com (a news-oriented site for Internet professionals and attorneys specializing in Internet and technology law), ILW.COM (Immigration Law Portal), PageBid (litigation pricing research). Also, don't forget that your local law libraries provide useful portals.

B. Finding and Using Free, Full-Text Case Law Sites

The legal portals discussed above provide access to many sites that contain free, full text case law. Especially useful here is the Legal Information Institute from Cornell Law School.

Federal Case Law. The United States District Court and the United States Bankruptcy Court for the District of New Mexico have a free ‘Opinions Database’ that was created in January 1989. The United States Court of Appeals for the Tenth Circuit’s opinions are available back to 1995. Both the District and Appellate courts’ databases are keyword searchable.

State Case Law. Full text opinions from New Mexico District, Magistrate, and Municipal courts are not available. However, you can use the New Mexico State Judiciary’s ‘Case Lookup’ to obtain case details. Municipal court data is limited to Domestic Violence and DWI historic convictions from September 1, 1991 onwards. Bernalillo County Metropolitan Court DWI data is available on the website; however, to search other Metropolitan Court case data you must go directly to the Bernalillo County Metropolitan Court’s own website. Cases filed prior to the completion of the statewide court automation project in June 1997, may exist in a database, depending on the individual court. If you are looking for cases filed prior to 1997 and do not find information on the web page, you should call the court directly.

Full text opinions of the New Mexico Court of Appeals and New Mexico Supreme Court are available on the Conway Greene website back to 1995. The site is keyword searchable.

Other. CaseMaker and other government internet resources for finding court opinions will be discussed later in this material.

C. Learn About the Free (Poor Man's) "Shepards"

Verifying that the law you are citing is good law, is absolutely crucial to legal research. The up to the minute nature of electronic publishing makes it the ideal means for ensuring that you have good law.

Free. A free alternative to Shepards or KeyCite is to use your web browser's search capability. For example, if you are planning to rely on the opinion in Rio Grande Chapter of the Sierra Club v. N.M. Mining Comm'n, 2003-NMSC-005, 133 N.M. 97 , 61 P.3d 806 and want to find out if the case is good law, you can search the New Mexico Supreme Court and Appellate Court cases using the free Conway Greene website. In the search box, enter the name of the case or its citation. The results of your search will include every citing case. Unfortunately, the results list won't give you any idea of the treatment your case received, you must determine that based upon your reading of the case. After all, you get what you pay for.

Free, but....The University of New Mexico Law Library offers Shepard's online for free to library users. The only catch is that the terms of the license agreement state that it cannot be made available remotely. That means you must physically go to the Law Library.

Pay Per Use. Both Shepard's and KeyCite are available by Credit Card. The service is purchased by the day for \$32, week for \$52, or month for \$107.

IV. FINDING “PRACTICE” TOOLS

- A. Sample Briefs, Motions, Complaints and Settlements**
- B. Sample Forms and Contracts**
- C. Uncover Jury Verdicts**

A. Sample Briefs, Motions, Complaints and Settlements

New Mexico has a relative dearth of free, online sample briefs, motions, complaints and settlements. This is largely due to the fact that electronic filing has not yet been adopted by the New Mexico courts. While the Caselookup database gives docketing information for New Mexico courts, it does not have full-text versions of the documents, but rather includes only the titles of pleadings filed. The Supreme Court and Supreme Court Law Library, however, both have forms sections on their web pages (<http://www.supremecourtlawlibrary.org/Forms.htm>) that include a number of pleadings. These forms can be downloaded into Wordperfect or as an html file (the “TV” hyperlink) and tend to be quite extensive within the limited areas that the forms cover.

For federal briefs and pleadings, the best source is probably the PACER database (<http://pacer.psc.uscourts.gov/>). Since the federal courts require electronic filing, you can sample a wide array of documents from similar cases, perhaps even those against the same plaintiff or defendant, when drafting a pleading or brief. Some attorneys are beginning to post samples of their briefs on their websites (see, e.g., Dana Curhan’s web site from Boston, Massachusetts at <http://www.criminalappeals.com/SampleBriefList.html>).

If you’re unable to find good motions, briefs, or pleadings on the Internet, please visit either the Supreme Court Law Library or the UNM Law Library. Print-based sets, such as Am. Jur. Pleading and Practice Forms or Causes of Action remain the best tool for drafting.

B. Sample Forms and Contracts

A number of web sites exist that sell contracts and other legal documents to non-attorneys. Caveat emptor is the phrase of the day when dealing with such an operation.

While this practice may not quite amount to the unauthorized practice of law, the more serious issue is whether the forms are tinkered to comply with either the client's individual needs or the laws of the particular state. Without commenting upon the quality of the company or its forms, here is an example of an internet legal forms business:

The wiser course of action remains print-based legal research or the equivalent on Westlaw or LexisNexis. Various specialized treatises, West's Legal Forms, Am. Jur. Legal Forms, etc., all have highly refined contracts and other documents that can be tinkered to comply with any peculiarities in New Mexico law.

C. Uncover Jury Verdicts

Jury verdict research can be difficult to conduct. One methodology is to use news resources to find verdicts against the same defendant for similar tortuous conduct. Unfortunately, only the extreme cases tend to be news worthy, thus skewing the results of the research. The best jury verdict print sources remain Jury Verdict Research (JVR) and What's It Worth?. Jury Verdict Research also has a fee-based web site with a free trial (<http://www.juryverdictresearch.com/>). The National Association of State Jury Verdict Publishers has a website, but there are no New Mexico publishers, hence no New Mexico-specific information (<http://www.juryverdicts.com/>).