



Internet Searching vs. Researching; Legal-Specific Search Sites

Prepared by:
Mark Giangrande
DePaul University



October 2016

Internet Searching vs. Researching; Legal-Specific Search Sites, ©2016 Lorman Education Services. All Rights Reserved.

INTRODUCING

Lorman's New Approach to Continuing Education

ALL-ACCESS PASS

The All-Access Pass grants you **UNLIMITED** access to Lorman's ever-growing library of training resources:

- ☑ Unlimited Live Webinars - 120 live webinars added every month
- ☑ Unlimited OnDemand and MP3 Downloads - Over 1,500 courses available
- ☑ Videos - More than 700 available
- ☑ Slide Decks - More than 1700 available
- ☑ White Papers
- ☑ Reports
- ☑ Articles
- ☑ ... and much more!

Join the thousands of other pass-holders that have already trusted us for their professional development by choosing the All-Access Pass.



Get Your All-Access Pass Today!

SAVE 20%

Learn more: www.lorman.com/pass/?s=special20

Use Discount Code Q7014393 and Priority Code 18536 to receive the 20% AAP discount.

*Discount cannot be combined with any other discounts.

Searching vs. researching

The Internet is a great place to find the law, but it is not necessarily a great place to research “the law.” Search engines help immensely in locating specific, useful materials for a research problem. If a researcher looks for a specific document using related keywords and it appears on an official or other site, a search engine will almost always return a result for that document within the first page of results. On the other hand, searching a concept, such as the use of parole evidence¹ in interpreting a contract, and the results will be more random and general. Case law and statutes appear on the web, though these relevant legal sources hardly ever appear in results from concept searches in any search engine. Rather, they are typically found in searches for specific opinions and legislation on sites dedicated to those collections. There are also better strategies for locating statutes as discussed below.

Researchers can use the local search feature at specific legislative and court sites, though the quality of local search varies widely from site to site. Another method is to use a general search engine’s advanced search feature and limit search to specific web sites (more on this below). Sometimes a search engine does a better job at indexing a site and organizing search results compared to the local search feature. All the major search engines have an advanced search option. Even with advanced search, however, the results are likely to appear somewhat general for a researcher, limited by date range, and most likely not directly citable. Lexis and Westlaw have the advantage of organizing legal material jurisdictionally and topically; by including all available publication dates and citations; accessible using precise search syntax; and linked to an extensive library of scholarly secondary material. The free web has never duplicated these features with one or two major exceptions. More details on that will appear below.

A researcher is usually successful at searching and less so at researching in these circumstances. For example, if a researcher wanted a copy of the United States Sentencing Guidelines Manual, placing those words in a search engine will return links to the manual from the U.S. Sentencing Commission. The specific document will almost always appear at the top of the first page of results. Search the more generalized phrase “sentencing guidelines” in a search engine will return results for the U.S. Sentencing Commission and related sentencing manuals within the first few results.

¹ This is the definition: A rule that governs the extent to which parties to a case may introduce into court evidence of a prior or contemporaneous agreement in order to modify, explain, or supplement the contract at issue. The rule states that where the parties to a contract intended for their written agreement to be the full and final expression of their bargain (i.e., the writing is an integration), other written or oral agreements that were made prior to or simultaneous with the writing are inadmissible for the purpose of changing the terms of the original agreement. Source: Cornell Legal Information Institute. https://www.law.cornell.edu/wex/parole_evidence_rule. While this is a great help in understanding the concept, it has limited value in applying the rule to specific situations or locating materials that show application.

However, the results can also include pages related to juvenile sentencing, a Wikipedia entry, sentencing guidelines randomly from Michigan, Minnesota, and other states, and a paper on the evolution of sentencing from the Department of Justice.

Add the word “robbery” to the search and the results drill down to materials that include the term, but are just as generalized in the type and range of documents. Add a jurisdictional element, for example “Illinois,” and the results include advertisements for Illinois attorneys, materials from the United Kingdom, and other random results that match the search terms. These technical but useless hits are essentially noise and they take time to sort through to more useful pages. The added terms do not focus the search. Rather, they often broaden the results.

One of the immediate hits from this broader search will point to a page from the Illinois Legislative Research Unit (an agency of the legislature) that has links to a survey of sentences for crimes in Illinois. The actual statement of the law which defines the mandated penalty for crimes is in the Illinois Compiled Statutes, [available online](#) at the [Illinois General Assembly](#) web site. As noted, case law and statutes interpreting or affecting sentencing guidelines for robbery in Illinois are not represented in the first several pages of results, if at all. These may be the kinds of materials sought. The document from the Illinois Legislative Research Unit may be useful for the information it contains, but it is not the law.

A researcher will have to have a detailed understanding of how law is organized and how it’s represented on the web to be efficient and effective. It’s a matter of anticipating sources for certain kinds of information and how that information is represented. Knowing, for example, that multiple government web sites host versions of the Code of Federal Regulations bypasses the need to use a search engine to find a copy of a particular regulation from a particular date. Knowing how a document appears in print will inform on which site will be a more useful source. One site may present the document in text while another may present it as a PDF. This understanding comes with experience.

A search engine is most valuable to locate very specific documents or to find out information about them. If the task is to find a specific case, searching for known yet incomplete information may yield more details such as a citation or a full case name that can be used to find the case. This strategy in searching for specific documents is to use search engines to cross-reference information about an item. For example, if a researcher is looking for a specific report by a federal agency and doesn’t have a lot of detailed information about the report, he or she could run a series of successive searches that accumulates information, such as the specific name of the issuing agency, the specific name of the report, the date it was issued, and so on. Once armed with that information a researcher can locate a copy of the report with confidence either online or in print.

One way to make using a search engine more efficient is to anticipate where specific information is likely to be found. A document produced by the Department of Justice is most likely going to be available at the DOJ web site. That DOJ listed as a search hit should be easy enough to pick out from the other results. It's possible that the document could be available at other sites acting as a mirror. However, it's best to get the document from a site maintained by the entity responsible for generating it. In any event, the more accurate information a researcher knows about a document, the easier it will be to construct a search that will locate it online.

Note that search engines do not index every page on the web or index pages as they are created or changed in real time. As good as search engines may be, their crawlers and robots that contribute to the index are not monitoring the entire web at every possible moment. Web site may deliberately blocked a search engine's indexing mechanisms through a code placed on a web page ([robots.txt](#)). Web content is not be necessarily available in a form that lends itself to easy indexing or searching. Information contained in a database is an example of that. Some sites use databases to create pages on the fly using links or local search. Search engines can easily miss this content in presenting results. It is important to strategically check all likely sources for an item. Do not assume that if an item is not in a search engine result list that it is not available. The next step would be to browse likely or predictable sites.

Search Engines

There are many search engines that index the web. Among major search engines by market share are [Google](#), [Yahoo](#), [Bing](#), [AOL](#), [Ask.com](#), [Wolfram Alpha](#), [AltaVista](#) (now part of Yahoo but separately branded), [Dogpile](#) (which searches Google, Yahoo and Bing simultaneously), and others. Duckduckgo is popular with some as it specifically does not track searches or leave a history of searches. Lists and descriptions of general and specialized search engines is at [The Search Engine List](#), [Search Engine Watch](#), and [The Ohio State University Library](#).

Information on the Internet is constantly changing--new sites appear and old sites revise their content or disappear entirely. Google [announced](#) in July of 2008 that it indexed 1 billion unique URLs. The volume of pages has grown to an estimated 50 billion pages since then.² While this figure represents an exceptional amount of pages referenced by a search engine, some estimates are that around a quarter to a third of all pages on the Internet are still not indexed. Most public legal information that is available via government and free sites in one form or another is generally represented in most major search engines.

² Visit <http://www.worldwidewebsize.com/> for more up to date estimates on indexed pages in each of the major search engines. Google, for now, appears to be far and above its competitors in total number of indexed pages.

A typical keyword search of one or two words or a phrase will yield an exceptional amount of results. Some result sets may number in millions of hits. Even in this circumstance the most relevant results will typically show up within the first several pages. General keyword searching, however, is not the most efficient way to retrieve information, even when using terms of art. Most search engines offer an advanced search that gives more control over the search terms and options. Google, for example, is capable of adding filters to search strings.

Search term filters:

Find pages with...

- all these words
- with the exact word or phrase
- any of these words
- none of these words
- numbers ranging from

Then narrow your results by:

- Language
- Region
- Last Update
- Site or domain
- Terms appearing
- SafeSearch
- File type
- Usage Rights (Pages with difference types of licenses.)

Google's Advanced Search is not obviously accessible from most screens. Go to https://www.google.com/advanced_search to use this feature and to get explanations of the various filters.

Google is the example here though all search engines have an Advanced Search feature on their pages in one form or another. These filtered results may still be in the millions of hits, but as noted, the most relevant material will be present in those first few pages. One key to using search filters is to understand something about the material and to match that understanding to the appropriate option. For example, more a recent document can be found by using date limitations, making it easier to eliminate older documents that may be returned as hits.

There is another search option that implements the type of search filters available through Advanced Search, and that is through undocumented search operators. These

operators may be placed in the generic search entry box and Google will observe the limitation.

For example the search:

filetype:pdf motion to compel discovery

will return results that are only PDF documents that have the words **motion to compel discovery** contained in the document (not necessarily as a phrase). Typically, in this circumstance, the results are more likely to contain formatted motions that may have been submitted in court or sample motion documents. The search automatically excludes general web pages that contain those words. This example is useful to find documents that can be used as models for creating similar documents.

Note that Google modifies these operators from time to time. A reliable list of supported operators is available from Google's Search Help page at <https://support.google.com/websearch/answer/2466433?hl=en> and Google's Search Education page at <https://sites.google.com/site/gwebsearcheducation/advanced-operators>. Note that there are other sites on the web that give other undocumented search operators. These may or may not work depending on how up to date these sites may be. Use the search [Google search operators] without the brackets to find them. Yahoo and Bing and other search engines have similar features.

Legal-specific search sites

There are a growing number of free sites that are devoted to organizing and indexing law on the Internet. Among these are Justia.com, the [Legal Information Institute \(LII\)](http://Legal Information Institute (LII)) from Cornell University, and FindLaw.

Justia is organized by jurisdiction and subject. It's a searchable portal that includes links to primary legal documents, legal news, and some commentary organized by subject. Justia also features a collection of federal judicial opinions organized by circuit, year, or [series of the Federal Reporter](http://Justia) back to 179 F.2d, or 1950. None of the reported case law in Justia has headnotes or other editorial enhancements such as those which are published in the Federal Reporter by West. Nonetheless, Justia is the only site that has attempted to replicate a collection of federal appellate cases browsable by citation to the Federal Reporter and for such a timeline. Justia will link to citations in opinions that appear in its database.

Another key feature of Justia is the part of the site [containing docket information](http://Justia) from the federal courts. The site is a free alternative to PACER, though it is not at all comprehensive. The docket archive tends to feature important cases that have received some notoriety. A case appearing in a Justia docket links to available PDF copies of documents in the court file maintained on Justia servers. Two limitations in Justia's

presentation of these materials are that the collection is incomplete and keyword search is very basic. Federal Court libraries offer free public access to PACER as an alternative to subscribing and buying documents from the federal courts.

Cornell began the LII in the 1992 and links jurisdictionally to content located contained in official state and federal law sites. Cornell also hosts a wiki like overview of various subject areas of law. The site is known for its editorial accuracy in the way it presents its collection, which would be expected from one of the nation's leading law schools. Some foreign universities have created their own version of the LII for their countries. See, for example, the [Australian Legal Information Institute](#).

[FindLaw](#) is owned by Thomson Reuters, the same vendor that produces Westlaw. The materials on FindLaw are organized by topic and jurisdiction. The site features court and business forms. The forms library can include actual though redacted versions contracts and other business agreements. The site has two versions, a consumer version which is the default and a "[professional version](#)" aimed at legal professionals. The difference is that the latter is organized with information geared to practitioners while the consumer site is aimed at individuals without a legal background. The professional site is organized by court and topic. Case name search is possible but citation search is not. None of the case law or statutes on FindLaw have any value added metadata that appears on Westlaw.

[Google Scholar Legal Opinions and Journals](#) (GSLOJ) site was released to the public on November 16th of 2009. It represents a leap over any other collection of free legal opinions on the web because of the time range available (approximately 60 years for state cases and 80 years for federal cases), a statement of parallel citations, the hyperlinking of citations within an opinion to other cases in the archive, star paging, a simplistic citator, and a sophisticated search algorithm that understands jurisdiction, case names, and other technical details of ranking relevant search results.

Google engineers have disclosed that the database is licensed from a major legal publisher. That suggests that the text is accurate as having been edited by the unnamed publisher's editorial staff. The contract between Google and the publisher calls for updates, keeping the archive current. Searches have demonstrated that updates and additions are not made in real time. There can be a month or more before recent opinions appear on the site.

The contract calls for some limitations, keeping Google from creating a full-featured citator present in the online citators Shepards and KeyCite. Google links to secondary materials that may be present in pay-for-access databases from other publishers. This is through the articles side of the site. For example, Google will link to law review articles are in the vast Hein Online law review database (which requires a commercial subscription) rather than to any free sources which may co-exist. There are no statutes,

administrative opinions, or other forms of legal documents in the database aside from cases, articles, and patents.

A researcher can access legal materials by going to the main Google Scholar web page. There is a radio button just below the search entry box that selects legal materials. Enter key words or phrases, and a jurisdiction and Google will display relevant results that link to the full text of the opinion. The “text display” is one of two tabs on an individual result. The other tab, “how cited,” offers snippets of other cases that have cited the main case with links to those cases. There is a link on the page that presents a list of other cases that cite the main case. However, the display does not offer any indication of the treatment by other courts of the cited opinion. That ability is the hallmark of Shepards Citations and KeyCite.

Google SLOJ represents the only substantial opportunity so far to search a comprehensive (meaning deep) free case law database that’s formatted consistently and with high quality results that are immediately useful. The advanced search options can limit search to specific jurisdictions or add other filters similar to the general advanced search feature. As the results are limited to case law, a researcher will have to ferret out some of the references in the opinions, such as statutory cites which are likely online but not linked by Google. Other cited but unlinked material may not be available online for free, or even online at all. A combination of print and online materials may be best to retrieve cited materials. For example, a print copy of annotated statutes can identify useful annotations and secondary materials which are retrievable by citation from Google SLOJ. Print collections are alternative sources for online materials that require payment for access.

Google is committed to improving the utility of the database to the extent that their contract allows. Expect improvement over time as Google and its users become more acquainted with managing and searching legal information. As an example, entering a citation as a search now tends to bring the case with that citation as the first result in the list. This was not the situation when Google SLOJ first appeared.

Leagle is another online portal to free case law with a substantial archive of opinions. Leagle describes itself this way:

Leagle, Inc. is a leading provider of copies of primary caselaw from all Federal courts and all State higher courts. Our collection is up to date within 24 hours of release of opinions from the courts and is also complete historically for all time for Federal courts and back to 1950 for state appellate and supreme courts. We add Slip Opinions daily, and Advance Sheets and Bound Volume copies as they become available. Our materials are fully copyrighted by Leagle, Inc.

Published cases on Leagle are star paged. The site features links to the most recent released opinions. It includes options for advanced search and a citator similar to that of Google Scholar except that it does not offer snippets of citing cases. A feature that is somewhat unusual is that Leagle allows individuals to comment on cases. Not many opinions have comments, but they are there.

