

DIVINING THE LEGAL MYSTERIES OF THE ALL-POWERFUL SEARCH BOX

In this white paper, we explore the history of online legal research—and look forward to where legal search technology will take us next.



... the English or American lawyer resembles the hierophants of Egypt, for, like them, he is the sole interpreter of an occult science.

—Alexis de Tocqueville, *Democracy in America*

Online Legal Research—A Brief Look Back

Instead of mystical tomes, modern lawyers have the search box—their one true source of arcane knowledge and wisdom.

Searching legal content for use in cases is one of the most important, yet time-consuming tasks a lawyer, paralegal or law librarian can have. It's important because a case may live and die on the strength of the lawyer's argument. It's time consuming because even simple cases may require large numbers of citations and very thorough research to ensure that the arguments are strong. While the legal information industry has innovated by improving search accuracy and efficiency, the effectiveness of search largely relies on the research expertise of lawyers, paralegals and law librarians. The practice of law hangs on words—and their interpretation. When it comes to search, what you get out still depends on what you put in.

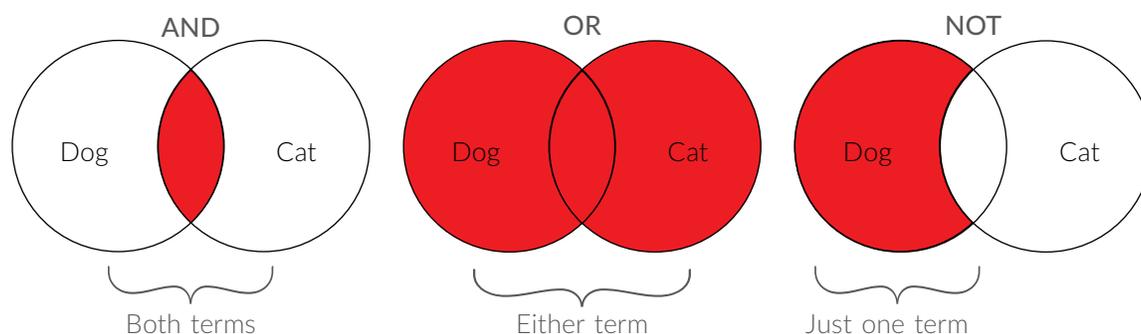
Search = Online Search

Today, legal case research is overwhelmingly done online. In fact, recent studies have shown that law firms consider online search the most important thing a new attorney needs to learn. Law schools will need to respond and adapt their curricula. While new attorneys are very Web-savvy, they often don't know how to evaluate sources and tools. One respondent said, "They rely too much on Google™ or other Internet search engines these days. They do not have a plan in mind when researching." Part of that plan is knowing the advantages and disadvantages of different types of tools.

Boolean Search: First Steps for Search

Looking at the tools we have available today, it's impressive how quickly electronic text search has arrived. Vannevar Bush, one of the first scientists to even propose a computer-accessible, networked source of knowledge (he called it a "memex"), thought the task would be so difficult that a special language would need to be created so computers could understand what we wanted.

Luckily, that wasn't necessary. By the 1960s, as computers became common, one of the very first legal text searching methods started with what's called Boolean search—exact matching of given terms. This type of search was a step ahead of Bush's computer-specific language, but it can also look a bit like programming. If you want a document from a specific source or source type, from a specific date range, or that includes your search terms in a particular way, this is the way to do it.



While this was all that search technology was initially capable of (the first Boolean-based legal text search projects started in the '60s), it became quickly clear that this was not enough. [Research in 1985](#) showed that Boolean text searches might return only about 20 percent of matching material, even though they gave the impression of being thorough. Having too much faith in your search results can be dangerous, but smart filters and search modifiers go a long way to making Boolean search a strong way to find something—if you know what you are looking for.

Making Search More Natural

A decade later, scientists began to develop natural language search methods. This technology went beyond exact text matches to produce better results, taking into account both a more flexible range of inputs and a better understanding of the source content. This included tallying the number of documents a search returned, and ranking documents by their relevance based on how often exact and related terms appear.

Natural language allows us to research general issues rather than very specific topics. This is extremely helpful when we don't know much about an issue or are researching something broad and complex. It's also a way to be more thorough by complementing a specific, technical search with one that is more conceptual.

Bringing in Citations

When it came to searching legal text, one challenge with the Boolean and basic natural language models was that it treated all the text equally. This was a problem when it came to citations, which were treated just like another string of words. Citation-based retrieval is a crucial part of legal research, and lawyers rely heavily on cites to find specific cases. With early search tools, it was up to the user to come up with variations in the search box that would give them what they wanted.

Fortunately, legal information providers were able to adapt their tools to meet the unique requirements of legal research. Twenty years ago, tools like LEXCITE® and LEXSEE® incorporated retrieval by citation, as a search or as a link from the case doing the citing. By recognizing and normalizing embedded citations combined with creating an authority of citable cases, lawyers had a new approach for retrieving important documents during legal research that nicely complemented traditional Boolean and Natural Language searches.

The Costs of “Free”

In more recent years, Google has consistently been the leader in advancing general search. Their initial innovation was in analyzing the linkage-based relationships among documents and exploiting those relationships to improve search results. Today they are leaders in applying linkage and user analytics to improve search results ranking, recommendations, ad click-thru rates and other areas. But Google has limitations by its nature. One limitation is content.

[Recent research at Stanford University](#) revealed that free legal research tools were simply not reliable, because they had no access to unpublished cases, and could not connect cases with higher court reversals or overrules. Another limitation is the business model itself: [if you're not paying for it, you become the product](#).

Online Legal Research—Adding Smarts to Search

There's no one perfect algorithmic solution to searching law. It's too complex. To balance this, companies have been adding supplementary specialized intelligence to legal search tools to support nuanced and in-depth legal analysis. For example, one reason for this complexity is that the words used in U.S. law are themselves confusing. Legal terms and concepts vary in meaning and validity depending on the jurisdiction you are working in. Many parts of the law can be filled with open-ended terms, the meanings of which change over time. A lawyer from California might be bewildered by the degree to which [Louisiana law](#) diverges from what he or she considers “standard.” Searching for “code” will give you very different results in patent law than property law.

Solving ambiguities in legal terminology requires specialized software that can parse and match terms to jurisdictions and areas of law. For example, a [Lexis Advance®](#) search recognizes over 15,000 legal phrases. Search algorithms are now able to understand lawyer research requirements and parse legal information with tools like phrase and case recognition, as well as implied phrases and the relationships between specific search terms.



The relationships between courts can also make a difference. Different judges have different levels of expertise, and will get recognized in the legal case data. Courts in Delaware may get more attention (in the form of citations) for corporate cases, just as East Texas will for IP cases, or New York for finance. Just look at the [complexity of the network of Supreme Court citations over time](#). But which ones do you want to read first? Case decisions from senior or superior courts may be more relevant, but the lower court's decisions often have more discussion of the relevant facts. Research into [lower court compliance with Supreme Court decisions](#) has revealed similarly complex patterns of links and relationships. Understanding these networks will help improve the accuracy and reliability of legal content search systems.

The Human Component and Beyond

Though technology has come far, law content is about people as much as numbers. It's still the human element that really makes legal search work. When it comes to paid legal information services, that is where much of the money goes. Thousands of experts are needed to customize and enhance legal information.

The unique Lexis Advance editorial process applies extensive quality control measures together with experienced attorney-editors to provide results intended to give you absolute confidence in your legal research.

Headnotes are a good example. In Lexis Advance, attorney-editors synthesize and prioritize the elements of a case, showing what's interesting and where it belongs in the larger context of law so you can be sure that court decisions are not misrepresented.

Humans still have the advantage in digesting legal information, even if their efforts are not scalable in the same way servers are. The architects of legal information systems will be looking to balance not only the ability of technology to set the pace of change, but also the way real people can contribute to making the change lasting and valuable.

The legal information landscape will continue to change. The continuing growth of digital legal information will make it tougher to identify the most critical and relevant information to a case or client matter. More courts are publishing information digitally, and some are going paperless. The predominance of online access might even change the way cases are written.

This white paper is presented by LexisNexis on behalf of the author. The opinions may not represent the opinions of LexisNexis. This document is for educational purposes only and does not guarantee the functionality or features of LexisNexis products identified.

About LexisNexis Legal & Professional:

LexisNexis Legal & Professional is a leading global provider of content and technology solutions that enable professionals in legal, corporate, tax, government, academic and non-profit organizations to make informed decisions and achieve better business outcomes. As a digital pioneer, the company was the first to bring legal and business information online with its Lexis® and Nexis® services. Today, LexisNexis Legal & Professional harnesses leading-edge technology and world-class content to help professionals work in faster, easier and more effective ways. Through close collaboration with its customers, the company ensures organizations can leverage its solutions to reduce risk, improve productivity, increase profitability and grow their business. LexisNexis Legal & Professional, which serves customers in more than 175 countries with 10,000 employees worldwide, is part of RELX Group plc, a world-leading provider of information solutions for professional customers across industries.

