Hunting and Gathering on the Legal Information Savannah

Susan Nevelow Mart  
*University of Colorado Law School*

Adam Litzler

David Gunderman

Follow this and additional works at: [https://scholar.law.colorado.edu/faculty-articles](https://scholar.law.colorado.edu/faculty-articles)

Part of the Computer Law Commons, and the Legal Writing and Research Commons

**Citation Information**


**Copyright Statement**

Copyright protected. Use of materials from this collection beyond the exceptions provided for in the Fair Use and Educational Use clauses of the U.S. Copyright Law may violate federal law. Permission to publish or reproduce is required.

This Article is brought to you for free and open access by the Colorado Law Faculty Scholarship at Colorado Law Scholarly Commons. It has been accepted for inclusion in Publications by an authorized administrator of Colorado Law Scholarly Commons. For more information, please contact lauren.seney@colorado.edu.
Hunting and Gathering on the Legal Information Savannah

Susan Nevelow Mart,** Adam Litzler,*** and David Gunderman****

No, no . . . you are not thinking; you are just being logical.¹

This article asks, what is it like for novice researchers to research real-world legal problems using four platforms: Bloomberg Law, Fastcase, Lexis Advance, and Westlaw? The study findings produced some surprises, as well as some clear implications for teaching legal research.

Introduction: The Human in the Information Chain ..........................................................6
Navigating the Connected World: Communication ...........................................................7
Hunting and Gathering: On the Road to a Methodology ..............................................13
  Gathering Problems ......................................................................................................13
  Hunting for Computable Search Processes .................................................................14
  Hunting for Relevance .................................................................................................15
  Hunting for a Solution .................................................................................................16
Gathering Knowledge—Unpacking the Coding ..............................................................18
Gathering a Few Thoughts on Legal Research Pedagogy and User Interface Design .................................................................22
Conclusion ....................................................................................................................26


** Professor and Director of the Law Library, University of Colorado Law School, Boulder, Colorado. Thanks to Judith Lihosit, Zach Mullen, and the participants of the Twelfth Annual Boulder Conference on Legal Information: Scholarship and Teaching, July 10, 2020, for their helpful comments. This project would not have been possible without the help of my research assistants: McKenzie E. Brandon, Vanessa R. Cortez, Kimberly Ferrari, Gerard C. Fowke, Kian Pakdel, Eduardo Ramirez, and Jordan T. Vogel.
*** Master’s Student in the Applied Mathematics Department and Statistical Collaborator in the Laboratory for Interdisciplinary Statistical Analysis, University of Colorado, Boulder, Colorado.
**** PhD Candidate in the Applied Mathematics Department and Statistical Collaborator in the Laboratory for Interdisciplinary Statistical Analysis, University of Colorado, Boulder, Colorado.
Introduction: The Human in the Information Chain

¶1 When researchers enter the world of the massive online legal research platform, they are entering a complex, constructed universe where user-interface design and an array of algorithms guide their experience. Those algorithms use a variety of natural language processing methods, deployed in huge vats of content, to provide useful results for researchers. Platform providers might use content mining, context relevancy importance, and user search histories to help refine results. These series of complex mathematic processes are often bundled as “machine learning.” Machine learning is not yet artificial intelligence. There is always some interference between a researcher’s information need as inputted and the output of the system. The researcher can’t see or understand the work that is going on in the back end, which makes understanding how to decrease that interference and improve results difficult.

¶2 The relationship between humans and their intellectual tools has often been discussed as a binary—either a great achievement or an impending doom—since writing emerged in the human toolkit. Online legal research platforms first appeared on the scene in the 1960s, and the preference for online research has gradually come to dominate researchers’ usage. The binary discussion—great achievement or impending
doom—has not diminished with the rise of massive online legal research platforms and
their reliance on complex natural language processing. If we want to empower law
students and attorneys to be in control of the process in complex legal platforms,
instead of having them be the passive recipients of algorithmic results, we need to keep
poking under the hood.

¶3 This article’s narrative shares what we found poking under the hood, using a
small sample of 12 real-world research problems. Some findings will surprise no one.
Researchers did not always find it easy to navigate to relevant sources. Researchers who
were unable to find good secondary sources were not confident in their conclusions.
More surprising was the finding that knowledge did not equal ease in finding results.
One might expect that researchers, after completing a first search using one legal
research platform, which helped them to define the parameters of the legal conclusion,
would have an easier time doing the research in other legal research platforms. Not so.
In 11 out of 12 problems, the time to complete the research in each of the subsequent
platforms did not get much better with knowledge of the legal sources needed to come
to a legal conclusion. It is heartening to know that even novice legal researchers con-
tinue to rely on expert secondary sources to guide their research and confirm their
conclusions. Legal research platform providers need to embed more point-of-need
navigation, to help researchers see how the user-interface is involved in and directing
their research, for course correction. And legal research teachers need to explicitly teach
students how to be thoughtful platform navigators.

Navigating the Connected World: Communication

¶4 We live and work in a connected environment where the interfaces we use to
complete tasks have profoundly changed the way we process and evaluate information.
In the world of the digital interface, how we solve our legal information needs has
evolved since the earliest computer-assisted research platforms. Figure 1 is an early
representation of the online information-seeking flow where “the [human] subject was
envisioned as a smooth space for the transfer of information between the inner and
the outer worlds, between the registers of analysis and stimulus.”

10. More nuanced views of achievement or doom have certainly been the subject of writing on the
current state of AI-assisted programs. For a collection of current views, see Possible Minds: 25 Ways of
Looking at AI (John Brockman ed., 2019). It is fair, however, to say that people tend to take sides.
11. Marcia J. Bates, The Design of Browsing and Berrypicking Techniques for the Online Search Interface,
13 Online Rev. 407, 408 (1989). This chart, even then, did not illustrate the messiness of the research
process. Marcia Bates knew that the search itself changes the search. Id. at 410.
Figure 1. Information Transfer

5 Figure 1 shows the space in the middle as the place where the match happens; that space is also the space for interference, as illustrated in figure 2, which represents the central noise that always interferes with the transmission of information to its destination.13

Figure 2. Mathematical Theory of Communication

6 Communication theory posits that interference will always exist between the information source and the destination; for legal researchers, that interference is between the entry of a research request and the results obtained.14 Our everyday experience bears out the fact that language can be an imperfect carrier of meaning.15 When

---

14. Id.
15. The difficulties of clear communication between humans is an entire subject of its own. For an overview, see, e.g., Richard J. Lanigan, Information Theories, in Theories and Models of Communication 59, 73 (Paul Copley & Peter J. Schultz eds., 2013); see also Juri Lotman, The Sign Mechanism of Culture, 12 Semiotica 301, 302 (1974) (“Non-understanding, incomplete understanding, or misunderstanding are not side-products of the exchange of communication but its very essence.”). Jeanette Winterson, Gut Symmetries 163 (1997), puts it eloquently:

Grandmother and I sat face to face over the sepulchral plastic of the breakfast bar. Common and rare, to sit face to face like this. Common that people do, rare that they understand each other. Each speaks a private language and assumes it to be the lingua franca. Sometimes words dock and there is a cheer at port and cargo to unload and such relief that the voyage was worth it. “You understand
communicating with a computer, we are not even communicating in our native tongue. However, formulating a good query and understanding a little of the thinking behind the translation of a query into results can go a long way toward allowing the researcher to better control the search process and force good search results to the top.

¶7 Figures 1 and 2 do not capture the iterative nature of search. Figure 3 is one attempt of many to document the nuance and complexity of the search process. The figure also illustrates how many entry points there are for “interference” or “noise” to impact results.

Figure 3. Cognitive Information Cycles

¶8 The black box of legal research platforms is the mechanical site of the interference between the information source and the destination. The research process starts with a human, whose level of legal information literacy will impact the actual signal sent to the black box. Legal information literacy implicates issues of ambiguity in the law itself, and lack of legal information literacy will increase the interference in the system.


17. See, e.g., Dennis Kim-Prieto, The Road Not Yet Taken: How Law Student Information Literacy Standards Address Identified Issues in Legal Research Education and Training, 103 LAW LIBR. J. 605, 610–13, 2011 LAW LIBR. J. 37, ¶¶ 8–16, for a discussion of legal information literacy as a way to evaluate and train law students to create a plan and a recursive approach to problem solving that will enable them to interrogate the legal research platforms or other resources they will use to resolve legal problems.
Our understanding of information-seeking behavior has evolved since Bates’s early representation of the relationship between the information need and relevant documents, and a large literature builds on Bates’s perceptions of information transfer and the iterative and redundant nature of search. Elements of Bates’s chart (see fig. 1) still resonate. A need to bridge the gap between the information need and the results remains. That gap is dominated by the black box of algorithms. In the case of legal problem solving, when algorithms do their work and “compromise or adjudicate between mathematical and pragmatic modes of reason,”18 the human researcher always occupies the middle of the gap between the two modes. One way we negotiate this gap “between code and culture”19 is by interrogating the black box. Another way we negotiate the gap is to learn how to communicate with the computer in the language it demands.

From the beginning of online search, the imperatives of algorithmic search forced the researcher to think and process language in a new way, so that the computer could understand. Since natural language processing algorithms have come to play such a large part in online searching, it is easy to lose sight of the fact that a Boolean, or terms and connectors, search is an algorithm. It is just an algorithm created by the researcher and not by the platform provider. Boolean queries are certainly a different form of speech:

Defin! Mean! Interpret! Explain! Constru! Constitut! Typical! Entail! /5 [insert term]20

Or,

(“FAIR DEBT COLLECTION PRACTICES ACT” FDCPA) /P BANKRUPTCY /P CONFLICT /P (TIME-BARRED “STATUTE OF LIMITATIONS” STALE) & DA(AFT 11/16/2016)21

Clearly, the legal researcher needs to think in a new way to translate human speech patterns to Boolean search logic.22 When performing that change in speech, the

---

18. Finn, supra note 6, at 47.
19. Id.
21. This is the search suggested by Westlaw to update Resolution of Conflict Between Bankruptcy Code and Fair Debt Collection Practices Act Where Creditor Seeks to Recover on Time-Barred Debt, 20 A.L.R. Fed. 3d. Art. 5 (2017). Note that at this level of complexity, there are many options for processing the chained operators, and different platform providers’ algorithms will process them slightly differently, leading to disparities in search results. See E-mail from Tito Sierra, LexisNexis, Joe Bred, Bloomberg Law, Khalid Al-Kofahi, Thomson Reuters, and Ed Walters, Fastcase, to Susan Nevelow Mart (Aug. 21, 2019, 3:07 PM) (on file with author).
22. For an overview of how Boolean logic operates, see Bahman Zohuri & Masoud Moghaddam, What is Boolean Logic and How it Works, in BUSINESS RESILIENCE SYSTEM (BRS): DRIVEN THROUGH BOOLEAN, FUZZY LOGICS AND CLOUD COMPUTATION 183–98 (2017).
human algorithm creator is relying mostly on connectors and field limiters. The need to think clearly about what the researcher wants to achieve by a natural language search is a different form of translation: there are rules of “speech” in natural language for effective searching, including word order and number of search terms. Natural language or keyword searching tends to be somewhat reductionist. The researcher needs to think of the exact words or phrases that will retrieve a document that would help resolve the legal question, or part of it, and then enter them into the search box. In the background, the platform is using an array of tools, including Boolean logic, to translate those terms into credible results. Knowledge of how those tools work is helpful in creating a keyword search. As one example, algorithms differ in how they treat proximity—how close words in a document need to be to each other. Just knowing that and thinking about it when formulating a keyword search can help the researcher think through the problem.

23. Common connectors are quotes to search for a phrase and proximity connectors, which include such search terms as AND, OR, NOT, w/, /p, NEAR. There are word variant symbols, such as * and ! Some platforms allow such connectors as “atleast,” so that a researcher can request results where each document has a term or phrase at least a specified number of times.

24. Field limiters or segments are commands to limit the results of the search to a specified portion of the documents being searched. For example, when searching cases, the researcher may be able to limit the search to the synopsis of the case. When searching news articles, the researcher may be able to limit the search to the headline and the lead paragraph. These field limiters allow the researcher to take advantage of the metadata tags assigned to each document in a legal database. The depth and range of field limiters or segments vary from database to database.


26. Id.; see also E-mail from Pablo Arredondo, Casetext, to Susan Nevelow Mart (Sept. 10, 2020, 16:10 MDT) (on file with author), providing a link for a new method of communication being deployed in Casetext's Parallel Search, which allows the researcher to use a full sentence as a search query. Parallel Search uses Google's BERT, or Bidirectional Encoder Representations from Transformers. See WIKIPEDIA, BERT (language model), https://en.wikipedia.org/wiki/BERT_(language_model) [https://perma.cc/AM8T-P8FH] (indicating that BERT is a technique for natural language processing (NLP) pretraining developed by Google; BERT was created and published in 2018 by Jacob Devlin and his colleagues from Google). Even with the enhanced semantic variability this method allows, the researcher still needs to understand how to communicate with the algorithm. On September 15, 2020, I entered the search: An administrative search requires a 4th amendment warrant. The search was limited to the Northern District of California. None of the top 10 search results were about administrative searches, although all of them had semantic variants of the word choice in the search sentence. Finding these variants is certainly a major advance in search capability. Parallel Search does allow the use of quotes, and putting quotation marks around “administrative search” resulted in the top 10 documents being highly relevant. Quotes are, of course, a form of algorithmic communication that needs to be learned, and proper communication with Casetext's algorithms dramatically improved the search results. Variations in sentence structure also changed search results. Lexis+ is also using BERT to improve Lexis Answers. Jean O'Grady, Lexis Rides the "Insight Wave": Launches Lexis+ with New Look, Brief Analyzer, AI Search, Codes Compare and Loads of New Features, DEWEY B STRATEGIC (July 8, 2020), https://www.deweybstrategic.com/2020/07/lexis-rides-the-insight-wave-launches-lexis +with-new-look-brief-analyzer-ai-search-codes-compare-and-loads-of-new-features.html [https://perma.cc/9KKV-N69Y].
¶12 Here is one example of how different algorithms parse the same search. If a researcher is looking for cases about administrative searches and whether that kind of search requires a Fourth Amendment warrant, one search might be:

**administrative search 4th amendment warrant requirement**

Any attorney with even a limited amount of subject-matter expertise can articulate what the underlying legal issue requires: cases on administrative searches and whether the searches require a Fourth Amendment warrant. Natural language algorithms do not yet do as well. When putting the same query into the same reported case dataset in six different legal research platforms, the top results varied enormously, and not all the top results were relevant. Each of the algorithms had a different way of parsing the meaning of a human's translation of a legal concept, and none of the algorithms was successful at turning that translation into 10 relevant documents. The study illustrated that variability in algorithmic implementation has an outsized impact in returning results.

¶13 Searching for cases in a case database is only one element in a researcher's plan to solve a legal research problem. However, in the massive search interfaces legal platform providers offer for fulfilling a research need, interfaces and their intersections with algorithms can augment or impede access to the relevant information that will frame the solution to the legal research question: “Our literal and metaphorical footprints through real and virtual systems of information and exchange are used to shape the horizon ahead through tailored search results, recommendations, and other adaptive systems . . . .”

¶14 What would it look like to interrogate the holistic research process in massive legal research platforms? This article starts the process of answering that question.

---

27. Mart, supra note 25, at 412–15, ¶¶ 45–48. Although the algorithms in use at the time of the study have changed, the general concept has not. For an example of differing search results and relevance in a search made on May 16, 2020, see appendix A, infra.

28. Mart, supra note 25, at 412–15, ¶¶ 45–48; see also Finn, supra note 6, at 138 (“the most effective aesthetic [for AI interfaces] is one of augmentation, complementing human intelligence with computational depth”).

29. Mart, supra note 25, at 412–15, ¶¶ 45–48; see also Gurari, supra note 5:

The reason [the algorithms] are so often wrong relates to a third consideration to keep in mind: there is no requirement that the machine learn features and associations that are intelligent, coherent, or what a person would look for. What these programs are designed to do is learn statistical correlations. That they may be incidental or bizarre is secondary.

Understanding how the algorithms reached a statistical result is complicated by the fact that “machine learning has a transparency problem. We usually don’t know which features these algorithms are learning. This is a major problem for high stakes fields like law and medicine.” Id. Overcoming this limitation is an area of research called “Explainable AI.” Gurari, supra note 5; see also Finn, supra note 6, at 138 (“When designing AI interfaces, the best thing to do is to complement human intelligence with computational depth.”).

30. Finn, supra note 6, at 50.
Hunting and Gathering: On the Road to a Methodology

Gathering Problems

¶15 The current literature analyzing search results has focused on individual processes in legal research platforms, such as the digest function,\(^\text{31}\) the citator function,\(^\text{32}\) how statutes are annotated,\(^\text{33}\) or how case law can be searched.\(^\text{34}\) Those individual search processes do not necessarily solve a legal problem; they provide some information on the way to a larger solution.\(^\text{35}\) One of the many difficulties of trying to compare actual problem solving in different legal research platforms is acquiring real legal research problems, complicated enough to require an iterative search process but not so complicated that the problem cannot be solved in a reasonable amount of time. These types of right-sized problems are often given to summer or new associates to solve. These problems represent actual, fact-based unknowns an attorney needs researched. These problems are manageable enough in scope to give to a novice researcher to research.

¶16 Since 2014, Colorado Law has had a Summer Employment Transition (SET) program. At the end of the first year, willing 1Ls are matched with local attorneys and given one of those short real-world assignments to research and write about in a short time.\(^\text{36}\) I have been collecting those anonymized problems for use in this analysis. I chose 12 of the problems for review in the study, looking for differences in subject matter and jurisdiction, and similarities in complexity.\(^\text{37}\)

¶17 This study is necessarily descriptive. No study this size could hope to do more than raise questions and leave the field open for further research. The study looks at the kind of human-computer translation needed to resolve a research need holistically, in one of the major legal research platforms that support legal problem solving across resources and subject areas. What does the modern research space look like as a whole? What does problem solving look like to a novice user in these platforms: Bloomberg, Fastcase, Lexis Advance, and Westlaw Edge? How would users be guided through the systems? When researchers specify a plan and implement it, those “moments of inclusion mask the many other decisions you are not invited to participate in.”\(^\text{38}\) It is

---

32. Id.
34. Mart, supra note 25; Callister, supra note 2.
35. Looking for a known item would be an exception; an individual search process might well answer the research question.
36. The length of the project has varied from year to year, but it generally takes a day and a half to two days. This allows time for the student and attorney to meet and confer about any issues that arise during the research and writing process.
37. See appendix C for the 12 problems the research assistants used, available at https://scholar.law.colorado.edu/research-data/12/.
38. Finn, supra note 6, at 110.
axiomatic that thinking through the problem, identifying appropriate resources to search, and formulating the search in a way that facilitates the algorithms are still the first steps for a successful search.\footnote{Paul Callister, \textit{The Metacognitive Imperative}, in \textit{The Boulder Statements on Legal Research Education: The Intersection of Intellectual and Practical Skills} 25, 52–53 (Susan Nevelow Mart ed., 2013).} Poor search taxonomy, even in a “natural language” environment, leads to bad results. Thinking is always the first step in the research process.\footnote{See, e.g., Mart et al., \textit{supra} note 3, at 14.} Would the intersection of user-interface design and algorithmic processes in each platform help or hinder the researcher through the search process?

\section*{Hunting for Computable Search Processes}

\textparagraph18 If the first hurdle in setting up a comparison of the problem-solving process in different platforms was finding the problems to solve, the next hurdle was setting up a computable process for comparison of search paths in platforms with different algorithms, search philosophies, content, and user interfaces. Research is an iterative, non-linear process, and the full scope of the search process is not easily quantifiable. Reducing the research process to a series of steps might amplify the “compromises and analogies of algorithmic approximations,” which “tend to efface everything they do not comprehend.”\footnote{Finn, \textit{supra} note 6, at 22.}

\textparagraph19 After multiple attempts, the final choice was to code the search process as a series of access points. There are many first access points, such as putting keywords in the blank search bar, going directly to a secondary source, or locating a case or a statute mentioned in the research assignment. From the first place in each legal research platform that the research assistant navigated to, the coding followed a set process:

- Whether a search string was used (or opening a resource, such as a table of contents)
- Whether the search was natural language or terms and connectors
- The top 10 results (if more than 10)
- The relevance of the top 10 results
- The type of resource being searched
- The total number of results from the search
- The relevance of any sidebar material
- Whether another search string within that access point was needed
- Whether that search was terms and connectors or natural language
- The top 10 results for the second search within that access point
- The relevance of the top 10 results
- The total number of results from the search
- The relevance of any sidebar material
- And so on, until that access point was exhausted
- Repeating the coding process for each access point that was required to complete the research problem
Whether the legal issue(s) was solved
- The time (minus the time to code) to complete the problem
- The total number of access points required to complete the problem
- Whether the search process was successful

The steps enumerated in the coding sheet, for quantitative analysis, are necessarily linear. We all know that research is not linear, and that each step informs and changes the next, revises the research query, and opens new vistas. To try to capture those nuances, research assistants were told, for information that did not fit into the coding, to add their thoughts generally about the process, using a recorder/transcription app or typing their thoughts directly into a Word document. The hope was that adding a narrative component would soften the abstraction imposed in the coding by having researchers add a narrative of their process.

Each research assistant was given one or two problems to research. Each of the problems had to be researched in each of the four legal research platforms. To reduce the distortion that might arise if students started with the platforms they were most familiar with, a random number generator was used to generate 12 sets of randomly sequenced numbers from 1 to 4; that was the order platforms were to be searched for each of the 12 problems.

Hunting for Relevance

In order for the research assistants to determine whether a case or a section of a secondary source was relevant to their research, I had to define standards of relevance. Relevance is a hotly contested issue in law. The lawyers on opposing sides of an issue may determine relevance differently, and any two lawyers may have a different opinion of a case, particularly at the outer edges of relevance. I used a broad and expansive definition of relevance, informing the research assistants that they should think of the legal resources they found as being directly on point, possibly helpful, probably not relevant, or not at all relevant. As Stuart Sutton puts it, “stated simply, a relevant case is one that plays some cognitive role in the structuring of a legal argument or the framing of legal advice.”


43. This was less than successful. The research assistants tended to see the coding as primary, and their comments about the research process itself were brief and somewhat meta; the details of their thinking never made it into the narrative. See Appendix D, available at https://scholar.law.colorado.edu/research-data/14/, for a summary of the narrative comments. Having a human interlocutor watching the research process and prompting reflection is one way to solve the problem, but the time and cost involved in doing that were prohibitive.


23 When lawyers learn about an area of law and focus on specific facts for a specific research problem, they construct a matrix in their minds of resources where the facts are similar or not, the law is similar or not, and the outliers. For experts, that matrix persists, and new facts or new rules and norms can be put into that matrix easily, for a quick analysis of the state of the law on a particular topic. After learning about the mental model framework for determining relevancy, the research assistants were told to code legal resources in their result sets as follows:

1 = most relevant—would definitely go into my research pile

2 = probably relevant—would go in my secondary research pile because it might turn out to be helpful after further thinking or by analogy

3 = probably not relevant—can't think of any way this might relate to my specific research problem

4 = not at all relevant—what is this case about *per stirpes* inheritance doing in the results of my research about the SEC?

24 The research assistants were asked to determine whether a specific resource would help define the contours of the legal issue. This type of research is “creative problem solving under conditions of uncertainty and complexity.” The uncertainty and complexity exist in the indeterminacy of the law, in both its language and its application. Since machine learning bypasses analogic reasoning in favor of statistical probability, when we deploy searches that rely on machine learning in the legal corpus, the relevance of the results is not always optimal.

Hunting for a Solution

25 The aim of the research process was to locate enough relevant resources to reach a legal conclusion about each problem. There is not always a definite answer to a legal problem; the goal is to locate and analyze the relevant law and be able to make a reasoned response to the question being investigated. There were 12 searches, each

---

46. *Id.*


49. *See generally* David Lanius, *Strategic Indeterminacy in the Law* 2 (2019) (“And, most importantly, indeterminacy is ubiquitous in the law—even despite the principles of the *rule of law* and the common perception that the ‘fluidity’ of legal language is a curse.”).


52. For ease of communication, we refer to problems being “solved,” even if the conclusion of the research process was more nuanced.
performed in 4 databases. For each of these 48 searches, the researchers chose to start in secondary sources.\textsuperscript{53} The ease or difficulty of finding relevant secondary sources was a theme in the researchers’ narratives. Failure to locate good secondary sources was frustrating for researchers. In two instances, the research assistant could not find specific resources within the platform that were needed to complete the research.\textsuperscript{54}

\textsuperscript{\textcopyright}26 Without support from secondary sources, the researchers were not comfortable with their conclusions or were unable to come to a legal conclusion; in this case, they coded the problems as “not solved.” In Bloomberg Law, the research assistants identified lack of secondary resources as an issue for Problems 2, 5, 8, 9, and 10. Problems 8, 10, and 11 were coded as “not solved.” In Fastcase, the researchers identified lack of secondary sources as an issue for Problems 4, 8, 11, and 12; Problems 4 and 8 were coded as “not solved.” In Lexis Advance, lack of state-specific resources was an issue for Problem 11. Problem 4 was coded as “not solved.” In Westlaw, Problem 12 was coded as “not solved.” (Table 1 summarizes these results.) Several researchers commented on the need for terms and connectors searching where there were not relevant secondary sources.

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|c|c|}
\hline
Problem # & Bloomberg Law & Fastcase & Lexis Advance & Westlaw \\
\hline
1 & & & & \\
2 & & & & \\
3 & & & & \\
4 & & & & \\
5 & & & & \\
6 & & & & \\
7 & & & & \\
8 & & & & \\
9 & & & & \\
10 & & & & \\
11 & & & & \\
12 & & & & \\
\hline
\end{tabular}
\caption{Solved—Lack of Secondary Sources—Not Solved}
\end{table}

Once the problems were coded and the narratives reviewed, we analyzed the coding to determine the nature of any statistical correlations in the data.

\textsuperscript{53} When no secondary sources were found, two researchers did not code the failed search as the first entry point and the next search strategy as the second entry point. Those two searches were in Problem 5, Bloomberg Law, and Problem 12, Fastcase.

\textsuperscript{54} In two instances, resolution of the legal problem was not complete because the platform was missing specific materials necessary to complete the research. These two instances were in Fastcase, Problem 2 (IBLA decisions not in database), and Problem 9 (DOE regulations not in database). However, within the resources found for those two problems, the missing information was identified. The agency decisions and regulations were simple to locate on agency websites. Because all the information needed to resolve the legal issue was located within the database, we considered the problems solved.
Gathering Knowledge—Unpacking the Coding

\[27\] To determine statistically whether there are any relationships in the data, we considered each attempt at solving a problem to be one experimental unit.\[55\] The problem number, research platform, and order in which the research platforms were attempted were predetermined. The research assistants recorded the total time to reach a legal conclusion, the total number of access points needed to come to a legal conclusion, and the research details within each access point.\[56\] After removing the instances where the problem was not solved, we are left with 40 problem attempts. We evaluated the effect of these categorical variables on the time to reach a legal conclusion, as well as the total number of access points used to complete the research using Analysis of Variance (ANOVA) models, which determine whether there is a statistically significant difference in a response variable in some set of categories. We considered these statistical questions:

- Does the research platform affect the time to complete a particular problem?
- Does the research platform affect the total number of access points used to complete a particular problem?
- Research assistants were asked to research each problem four times: once using each research platform. Regardless of the order of the platforms, does the number of previous attempts affect the total time to complete the problem?
- Does the first type of access point used to research a problem affect the time to complete a particular problem?
- Does the first type of access point used to research a problem affect the total number of access points used to complete a particular problem?

\[28\] We considered the research platforms and their effects on the time to complete. The null hypothesis was that there is no significant difference in time to complete a problem in each research platform. Figure 4 shows the boxplot of the platform versus the time to complete in each platform. In the “time to complete” model, we obtain an F statistic of 0.14 for a p-value of 0.935. This means that, on average, there was no significant difference in the amount of time it took to research a problem in each of the research platforms.

---


56. Although the coding generated data about the research process within each access point, the sample size was not large enough to make any comparisons within access points. Data not included in appendices F and G are on file with the author.
We then looked at each of the research platforms and their effects on the total number of access points used to reach a legal conclusion. Figure 5 shows the boxplot of research platform versus the total number of access points used. In the “total number of access points used” model, we obtain an F statistic of 1.015 for a p-value of 0.397. We fail to find statistical evidence that platform type statistically affects the total number of access points used among legal platforms.
The experiment was designed so that one research assistant solved the same problem in each of the four databases, and the order of the databases for each problem was randomly generated. We considered the possibility that a research assistant might be able to use the knowledge from previous attempts at coming to a legal conclusion on a given platform to come to the same legal conclusion more quickly on subsequent attempts on other platforms. In figure 6, we have the boxplot of these variables. We obtain an F statistic of 0.269 for a p-value of 0.847. Knowledge of the resolution of the legal problem did not really improve the amount of time it took to complete the same research in the second, third, or fourth legal research platform. Only Problem 9 was an outlier.
We also looked at the first access point type to see whether this affected either time to complete or the total number of access points used to solve. We categorized these first access points as:

“File Tree” (11 problem attempts);

“File Tree Search” (23 problem attempts); and

“Main Search Bar” (6 problem attempts).

For a **File Tree** access point, the research assistant navigated through the file structure of the research platform to the table of contents of a specific secondary source and opened relevant sections. For a **File Tree Search**, the research assistant navigated through the file structure to a potentially relevant database (such as “secondary sources” generally or with a subject-specific limitation) and then performed a terms and connectors or keyword search to locate potentially relevant material. For a **Main Search Bar** access point, the research assistant put a terms and connectors or keyword search in the main search bar and then used filters to limit the results to potentially relevant material.

In figure 7, we have side-by-side boxplots for first access point type versus time to complete, and first access point type versus the total number of access points used. In the “time to complete” model, we obtain an F statistic of 0.108 for a p-value of 0.898. The difference in means in the “time to complete by first access point” model is not statistically significant.
¶33 In the “access points to complete by first access point type” model, we obtain an F statistic of 3.228 for a p-value of 0.051, and in this model the p-value is just barely above the traditional threshold for statistical significance. The interpretation of this p-value is that we would expect this much of a difference in sample means for these sample sizes only 5.1 percent of the time if it were the case that queries beginning at these different types of access points were solved on average in the same amount of the total number of access points used. The 0.05 threshold is somewhat arbitrary, but this is much closer to a significant result than any of the other tests conducted in this investigation. A repeat of this study may provide more evidence for or against a relationship between these variables.

Figure 7. Time to Complete by First Access Point and Access Points to Complete by First Access Point Type

¶34 These results are from a small sample, and they could perhaps be strengthened in a future study with more research assistants and with replication. In particular, a larger sample size might help mitigate the effects of outliers, and a more equal balance of first access point types might resolve the indeterminate conclusion of our ANOVA.57

Gathering a Few Thoughts on Legal Research Pedagogy and User Interface Design

¶35 The study is based on investigating 12 real-world legal research problems, and the importance of correlations, or of the lack of correlations, is necessarily limited by the size of the study. The study suggests several important points that have implications for both legal research pedagogy and platform design. The study throws some doubt on the common perception of expert researchers that novice researchers just find a few cases and call their research complete.58 The research assistants were concerned about

57. See Appendix H, infra, for an alternative way of viewing this statistical information.
58. But see Scott Moss, Bad Briefs, Bad Law, Bad Markets: Documenting the Poor Quality of Plaintiffs’ Briefs, Its Impact on the Law, and the Market Failure It Reflects, 63 EMORY L.J. 59 (2013) (focusing on poor
their own conclusions when they could not find relevant secondary sources to confirm their research conclusions. This certainly supports the importance of teaching students the value of secondary sources. The ease or difficulty of locating secondary sources is something legal research platform providers should further consider in interface design. If, because of licensing agreements or other market forces, there simply are no relevant subject-specific secondary sources, embedding help in accessing more general collections of secondary sources, such as law reviews or encyclopedias, would be helpful to novice researchers. There are so many databases of resources in the platforms investigated in this study that it is easy for researchers to get lost. Making sure that researchers are aware of options is a helpful step.

There are existing examples of help embedded in legal research platforms, such as the Must Include function in Lexis+ that highlights terms that are included in a search but missing from the results set, as figure 8 shows.

![Figure 8. Must Include (Lexis+)](image)

This kind of embedded help is a form of making algorithmic choices transparent. Must Include is the same algorithmic function as Search Within Results, but it expressly alerts researchers of the fact that terms are missing.

writing in a subset of employment law briefs, but implicating poor research in the process).

59. Table 1, supra, at 17.


61. When I asked Lexis+ why early results in a result set did not have all the terms in my keyword search, a Lexis+ engineer answered that “there is a lot more that goes into ranking relevance than just the terms that are included. [He said] that terms are of course a big part of it, but we also look to things like frequency of the terms that do appear, how often the case is cited by others, times that the case has been engaged by others, etc.” E-mail from Lynn Pinnecamp, Sr. Acct. Exec., LexisNexis Legal & Professional, to Susan Nevelow Mart (July 20, 2020, 1:33 PM) (on file with the author).
¶37 Fastcase has made the filtering process transparent with guided pop-ups highlighting the navigation on the results page, as figure 9 shows.

![Figure 9. Fastcase](image)

The More button takes the researcher to the next in a series of pop-ups highlighting what is available from this results page.

¶38 Westlaw has highlighted the ability to limit case results by procedural posture with a pop-up, as shown in figure 10.

![Figure 10. Westlaw](image)

¶39 In Bloomberg Law, a pop-up highlights help with the difficult task of how to retrieve a docket by docket number, as figure 11 shows.

![Figure 11. Bloomberg Law](image)

¶40 Legal research platform providers need to continue adding embedded help that highlights next steps in the research process, both for algorithmic transparency and for direction to related and available resources within the platform. Reducing interference

---

caused by the complexity of the user interface design is a form of algorithmic accountability.

¶41 For legal research pedagogy, the study confirms that research courses should include the metacognitive skills necessary to approach new platforms. In my own courses, I teach students to take the time to reflect on the interface of the platform they are approaching for the first time. I suggest they ask questions like these:

- What resources or filtering elements is the platform provider highlighting in the F-shaped pattern on the main page? On internal pages?
- Are those the resources or filtering elements that the student is looking for?
- If not, what are the next steps?

This is another pedagogical example emphasizing that thinking is an important first step in formulating a research plan. To make sure that students internalize the process, so that they have the scaffolding necessary to apply this knowledge to new interfaces, this year I added a mapping exercise comparing different interfaces. A sample mapping exercise is available in appendix B. Although attorneys in law firms may balk at exercises, highlighting the mapping necessary to find typical resources for a practice group when offering training on a new interface may help cement the correct pathways. In subject-specific areas, legal research platform providers could consider adding specific pop-ups for training for typical search patterns within that subject area.

Conclusion

¶42 Communication with and within massive online legal research platforms is a complex matter, as there is always communicative interference between the humans who design the systems and the humans who use the systems. The study raises some interesting issues about how we communicate with and interrogate research universes.

¶43 The study highlights some of the difficulties novice researchers face when using a research interface for the first time. A novice might not be a law student or new associate. A novice could be anyone entering a new online research platform for the first time. As the platforms increase the number of databases and the number of documents in their systems, sophisticated, point-of-need navigation help becomes critical. Although each of the legal research platforms incorporates embedded help in its user interfaces, the study suggests that still more help is needed. For teaching legal research, the study suggests that it is important to teach law students to be thoughtful database navigators.

¶44 The study confirms that human expertise is still important to researchers. The researchers who did not locate relevant secondary sources were uneasy about the

66. Callister, supra note 39; Mart et al., supra note 3.
comprehensiveness of their analysis. Simply researching primary law without guidance has never been the best way to grasp the complexities of a subject-specific legal domain. The legal profession has a rich body of subject-specific treatises, practice guides, and articles, drafted by experts. Experienced lawyers have complex mental models of a subject area, based on the toolkits they have assembled from their chosen secondary sources, relevant primary law, and current awareness tools. The sense-making of experts is not definitive, as the boundaries of the law are always being pushed in new directions. However, the sense-making of experts is grounding and gives researchers a solid basis for taking steps in new directions or answering novel legal issues. The study offers some support to legal research instructors, who have long emphasized the importance of secondary sources: the lessons have been heard.

Appendix A: Comparison of Search Results

<table>
<thead>
<tr>
<th>Bloomberg</th>
<th>#</th>
<th>Fastcase</th>
<th>#</th>
<th>LexisAdvance</th>
<th>#</th>
<th>Westlaw</th>
<th>#</th>
</tr>
</thead>
</table>

¶45 This table shows the results of putting the same search in each of the four legal research platforms and limiting the search to the same pool of reported cases in the Northern District of California.68 White cases are unique and are not relevant.69 Between 20 and 40 percent of cases in the top 10 results are not relevant and are unique. The grey case is not relevant and occurs in two platforms. Light blue cases are relevant and unique. Between 10 and 30 percent of the cases are relevant and unique. Nearly 50 percent of the cases in the top 10 results are unique to each set of returned documents. Yellow cases are relevant and occur in two and, in one instance, three platforms. That means that under half of the cases turn up in more than one platform. Across all four platforms, the average for relevant results is 60 percent. The table illustrates that there is still variation in the uniqueness of cases, and the algorithms still return results that are not relevant in the top 10 results.

68. The searches were performed on May 15, 2020; the source case that generated the language of the search is Camara v. Municipal Court of City & County of San Francisco, 387 U.S. 523 (1967). The search was: administrative search 4th amendment warrant requirement. See Mart, supra note 25, app. B, https://scholar.law.colorado.edu/cgi/viewcontent.cgi?article=1006&context=research-data [https://perma.cc/32XU-P4FT].

69. Relevance here is broadly defined as a case that helps construct a mental model of the contours of an area of the law. See Sutton, supra note 45, at 187.
Appendix B: Sample Research Assignment

Metacognition—Mapping the Research Universe by Comparisons

¶46 I use the following exercises for a class on basic administrative law resources. These types of exercises can also be used to compare massive legal resource platforms, to the same ends. Students find illuminating the task of listing the steps in limiting results in a case database search to a specific jurisdiction and type of case; there are so many pathways! Try asking students to limit the results of a search you give them to Southern District of New York reported cases, and then have students map the process in a few different research platforms. Eye opening!

First Comparison: Locating the Unified Agenda

In small groups or breakout rooms: You are looking for any upcoming meetings on the Disposal of Coal Combustion Residuals from Electric Utilities; Legacy Surface Impoundments (40 C.F.R. 257). [Of course, you should update this with something from a current Unified Agenda.]

Each of you choose one of the following websites to investigate:

- reginfo.gov
- regulations.gov
- govinfo.gov

Answer these questions, and be ready to share your results with the class.

- What was the first tab or link you clicked on? Where was it on the landing page?
- How many clicks did it take you to find the information you needed?
- How long did it take you to find the relevant information?
- On a scale of 1 to 5, with 5 being the easiest, how easy would you say it was to find the meeting information?

In class, poll the students: which was the easiest resource?

Then have a student from each breakout room share their screen, and follow the mapping on the website to their result, emphasizing the ease or difficulty of the interface.

Second Comparison: Finding a Proposed Regulation

You are looking for a proposed regulation titled “Strengthening the H-1B Nonimmigrant Visa Classification Program.” [Update as needed.]

Each of you choose one of the following websites to investigate:

- reginfo.gov
- regulations.gov
- federalregister.gov
- govinfo.gov
Answer these questions, and be ready to share your results with the class.

- What was the first tab or link you clicked on? Where was it on the landing page?
- How many clicks did it take you to find the information you needed?
- How long did it take you to find the relevant information?
- On a scale of 1 to 5, with 5 being the easiest, how easy would you say it was to find the proposed regulation?

Third Comparison: Finding a Final Rule

You are looking for the final rule titled “Exemption of certain cannabis plant material and products made from them, that contain tetrahydrocannabinols.” [Update as needed.]

Each of you choose one of the following websites to investigate:
- federalregister.gov
- govinfo.gov
- LII
- Agency websites

Answer these questions, and be ready to share your results with the class.

- What was the first tab or link you clicked on? Where was it on the landing page?
- How many clicks did it take you to find the information you needed?
- How long did it take you to find the relevant information?
- On a scale of 1 to 5, with 5 being the easiest, how easy would you say it was to find the final rule?

*The polls emphasize that the same resource is not necessarily the best one for every search and answer the student complaint about so many resources being available that they don't know how to choose one. The screen share illustrates the importance of user interface in the process of locating information and creates an awareness of what to look for when navigating new resources.*
Appendix H: An Alternate View of the Statistical Data

<table>
<thead>
<tr>
<th>Experiment Description</th>
<th>Total # of completed problems</th>
<th>P value for anova</th>
<th>File Tree Search</th>
<th>File Tree Search</th>
<th>Main Search Bar</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time to complete (hrs) by First access point type</td>
<td>40</td>
<td>0.898</td>
<td>3.18</td>
<td>5.00</td>
<td>3.33</td>
</tr>
<tr>
<td>Access Points to complete by First access point type</td>
<td>40</td>
<td>0.051</td>
<td>1.43</td>
<td>1.61</td>
<td>1.49</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Experiment Description</th>
<th>Total # of completed problems</th>
<th>P value for anova</th>
<th>First database mean</th>
<th>Second database mean</th>
<th>Third database mean</th>
<th>Fourth database mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time to complete (hrs) by Platform Order</td>
<td>40</td>
<td>0.847</td>
<td>4.1</td>
<td>4.09</td>
<td>4.9</td>
<td>3.89</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Experiment Description</th>
<th>Total # of completed problems</th>
<th>P value for anova</th>
<th>Westlaw mean</th>
<th>Lexis mean</th>
<th>Bloomberg mean</th>
<th>Fastcase mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time to complete (hrs) by Platform</td>
<td>40</td>
<td>0.935</td>
<td>1.48</td>
<td>1.72</td>
<td>1.39</td>
<td>1.58</td>
</tr>
<tr>
<td>Access Points to complete by Platform</td>
<td>40</td>
<td>0.397</td>
<td>4.82</td>
<td>4.45</td>
<td>4.44</td>
<td>3.11</td>
</tr>
</tbody>
</table>

70. Appendixes C, D, E, F, and G are available online; see notes 37, 43, 44, and 55, supra.