Algorithms Are Created by Humans

When searching online, legal researchers must rely on the teams of engineers who created the algorithms that power the searches. It is easy to forget that the computer-generated results returned by a search are determined by the choices that humans made when the system was designed. All algorithms do is follow the rules set by humans who import their own biases and assumptions into the algorithm. With legal research, though, the teams that create the algorithms for legal research databases are trying to solve the same problem: The algorithm should return results that are relevant to a researcher who has entered specific search terms—terms that ought to be related to the legal problem that needs to be solved. Wouldn't that mean the search results would be similar? Does it really matter that the algorithms for each legal research database might be created by different teams of humans?
As it turns out, the human element in algorithms matters a lot. I recently conducted a study comparing the top 10 results of 50 legal searches in six different legal databases. The study looked at Casetext, Fastcase, Google Scholar, Lexis Advance, Ravel, and Westlaw. The study limited the database for each search to reported cases in a specific jurisdiction. Because that pool of information is nearly identical, using jurisdictional limits allows true comparisons of the work each algorithm is performing when it processes the search.

The results of the study certainly indicate that every group of humans will solve the same problem in a very distinctive way. An average of 40 percent of the cases in the top 10 results in each database were unique to that database. Only a few cases turned up in all six databases. Every database has a point of view, offering unique responses to a legal problem that no other database provides. That is because each database makes different choices about how to process terms in a search.

Humans Make Choices About How the Algorithm Will Work
Legal researchers do not know exactly how the algorithm in any given legal database works. It is, however, possible to know some of the variables the engineers work with when they create algorithms for legal research.

Legal researchers do not know exactly how the algorithm in any given legal database works. It is, however, possible to know some of the variables the engineers work with when they create algorithms for legal research.

- **Proximity**: Humans decide how close those words have to be to each other to be returned in the top results.
- **Stemming/Other Search Grammar**: Humans decide which terms are stemmed, (i.e., which legal phrases the algorithm recognizes without quotation marks) and if and when legal phrases are added to the search without researcher input.
- **Network/Citation Analysis**: Does the system rely on citation analysis to enhance the results? Humans decide if so, and how.
- **Classification/Content Analysis**: Does the system boost results by mining its own classification system or mining information in other legal content in the database?
- **Prioritization**: Relevance ranking is one form of prioritizing that emphasizes certain things (like the things in this list) at the expense of others.
- **Filtering**: Including or excluding information according to specific rules or criteria.

Once decisions about how to implement these elements are coded into the algorithm, searches are automatically executed, and researchers have little insight into why certain results are returned.

Algorithmic Accountability
Legal researchers need to know more about how algorithms work.
Understanding how terms are being processed and what content is being privileged would enable researchers to create better searches and would inform legal research pedagogy. There are new entries into the legal research market, and existing interfaces and algorithms change regularly. Just learning to use a specific legal tool will not enable researchers to deal with every new resource or interface unless they are taught to evaluate interfaces and algorithms. The more information legal database providers give researchers, the better those researchers will be.

There is, of course, some information about how algorithms operate. Each legal database provider does publish FAQs, videos, and handouts about how to search in its database. However, the information is not very detailed. For a more comprehensive discussion of what each of the companies in the study says about its search functionality, see my much longer article to be published in Volume 109, Number 3 of *Law Library Journal*, “The Algorithm as a Human Artifact: Implications for Legal [Re]Search.”

Asking for more information often works. Lexis Advance released a fact sheet on jurisdictional filtering when a specific request was made. No trade secrets were revealed, and researchers now have more information about the search process. Please help by requesting more accountability—the more voices that make the request, the more likely it is that a legal database provider will grant it.

Putting the Algorithms to Work
For each of the 50 searches in the study, the research assistants searched in one specific jurisdictional database. Searching with the jurisdiction had to turn up at least 10 results in each of the six legal databases, so that there were 10 cases to compare from each search.
The decision to limit the results to the top 10 was based in part on a desire to limit the number of cases that had to be reviewed to compare 50 searches in six different databases and in part on the assumption that modern researchers expect to find relevant results in the top 10. Studies show that internet users generally focus on the 10 top results. The default in Google, for example, is 10 results. Finally, our intuition tells us that the top results should be the best results. Why else are they at the top?

The study assumed that the goal for a legal research algorithm is to return results the researcher will find relevant. I doubt that any database provider would dispute this goal, and advertisements and announcements from each database provider support this view.

**Uniqueness in Search Results**

No computer scientist would be surprised if six algorithms solved the same problem in different ways. But since each algorithm was attempting to bring back results that matched the expectations of a legal researcher with the same objectives, with the same terms, and the same cases to mine, we might expect to find similarity in the search results. The following chart illustrates the variability in case results.

The top bar shows the percentage of unique cases in each database, and that percentage is high—about 40 percent of the cases in the top 10 results are unique—that is, they only appear in one database. There is not a lot of overlap in the remaining cases. On average, 25 percent of the cases are in only two of the six databases. Only seven percent of the cases show up in five or six of the databases.

If you isolate Lexis Advance and Westlaw, and just compare the cases that appear in those two databases, the results are even more striking: 72 percent of the cases that returned in the top 10 results are unique.

The first conclusion from the study is that, as a first stop on the research process, every database is going to return

### PERCENTAGE OF UNIQUE CASES BY DATABASE

<table>
<thead>
<tr>
<th></th>
<th>CASETEXT</th>
<th>FASTCASE</th>
<th>GOOGLE SCHOLAR</th>
<th>LEXIS ADVANCE</th>
<th>RAVEL</th>
<th>WESTLAW</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>UNIQUE CASES</strong></td>
<td>39.2</td>
<td>36.2</td>
<td>42.8</td>
<td>38.4</td>
<td>40.4</td>
<td>43</td>
</tr>
<tr>
<td><strong>CASES IN 2 DB</strong></td>
<td>23</td>
<td>25.4</td>
<td>20.4</td>
<td>26.4</td>
<td>23.8</td>
<td>23.8</td>
</tr>
<tr>
<td><strong>CASES IN 3 DB</strong></td>
<td>13.8</td>
<td>14.2</td>
<td>12.8</td>
<td>12.4</td>
<td>11.6</td>
<td>11.6</td>
</tr>
<tr>
<td><strong>CASES IN 4 DB</strong></td>
<td>10.4</td>
<td>9.2</td>
<td>9.6</td>
<td>9.4</td>
<td>9.6</td>
<td>8.6</td>
</tr>
<tr>
<td><strong>CASES IN 5 DB</strong></td>
<td>6.8</td>
<td>8.2</td>
<td>7.4</td>
<td>7.6</td>
<td>7.8</td>
<td>6.2</td>
</tr>
<tr>
<td><strong>CASES IN 6 DB</strong></td>
<td>6.8</td>
<td>6.8</td>
<td>7</td>
<td>6.8</td>
<td>6.8</td>
<td>6.8</td>
</tr>
</tbody>
</table>
a lot of unique results that will in turn frame the rest of the research process. Of course, no legal researcher should stop with one search and 10 results as the end of the research process.

Relevance in Legal Research

The second inquiry checked whether or not those top 10 results actually were relevant. Relevance is a highly contested subject, particularly for lawyers, some of whom challenge relevance for a living. So the study needed a definition of relevance that could be understood and shared by all of the coders, and would relate to the way lawyers think about legal issues. Here is an example of a search that student coders were given:

*special relationship constitutional duty protect public from crime (N.D. Cal.)*

Lawyers with any expertise can immediately translate that into an actual legal issue: While state officials normally do not have a constitutional duty to protect the public at large from crime, a duty may (or may not) be imposed by virtue of a special relationship between the state officials and a particular member or class of public. That broader legal issue sets the stage for relevance determinations. Student coders were given that background statement as a framework for their determinations of relevance. If a case they were coding could be helpful to determining the contours of that special relationship in any way, it would go into the pile of cases that are or might be relevant. This is a very broad view of relevance, but it is one that is typical of a researching lawyer’s first cut through a case database. This type of relevance is what author Stuart Sutton (*The Role of Attorney Mental Models in Case Relevance Determinations: An Exploratory Analysis*) calls creating a “mental model” of the law, as these cases might play some cognitive role in the structuring of a legal argument.

The next question is how did the algorithms do when it came to turning those keywords into cases that were relevant to the legal issues the searches reflected?
There is clearly a clustering of results here. The oldest databases provided more relevant results. Lexis Advance had 57 percent relevant results, and Westlaw had 67 percent relevant results. Casextext, Fastcase, Google Scholar, and Ravel had an average of 42 percent relevant cases.

The next logical question is how many of the unique cases were relevant?

Recall that every database returns about 40 percent of unique cases. Here, the four newer entries returned an average of 11 percent unique cases, Lexis Advance returned about 20 percent, and Westlaw returned slightly more than 30 percent of unique cases. Also, recall that the overlap between just Lexis Advance and Westlaw is only 28 percent.

Age, Numbers, and Time

The age of cases each database returns is different. Westlaw and Fastcase had the highest number of new cases (approximately 67 percent), with Casextext right behind at 64 percent. Ravel and Lexis Advance had an average of 56 percent newer cases. Google Scholar had the highest number of older cases. Nearly 20 percent of Google Scholar’s cases were from 1921–1978. Google Scholar relies heavily on citation count, and that privileges older cases.

The number of cases each database returns from a search is quite different. At the fiftieth percentile of the number of cases in the results, Lexis Advance returned more than 1,000 cases. Westlaw, Ravel, and Casextext returned just over 100 results. Google Scholar returned 180 results, and Fastcase returned 70 results. Relevance improved slightly for Lexis Advance as the number of results went up, but the number of results did not affect relevance rankings for the other five legal databases.

Time is critical to this study, which is a snapshot of the results with the algorithms as they were when the searches were performed. Database providers are constantly changing their algorithms. Although you could run the exact same searches in the exact same databases, the cases would be very different. And not just because new cases have been added.

POV

We now know several things about these databases that we did not know before. One is that older and more established databases have an edge in returning cases that are relevant and unique. These databases (Lexis Advance and Westlaw) have a much larger base of user history and they both have complex classification systems and secondary sources to mine from. Each classification system and set of secondary sources is very different. So, each algorithm is relying on a very different point of view in terms of content. As long as we are dealing with algorithms that import viewpoints into the search results regardless of the researcher’s intent, it is good that the viewpoints are different. In the same way libraries want different authorial viewpoints in the treatises others collect on a particular subject (budgets allowing), so too would we want different viewpoints in our legal databases.

Different as the mechanisms are for creating Key Numbers in Westlaw and Topics in Lexis Advance, both of these classification systems rely heavily on the Langdellian worldview of the nineteenth century. Look at the entry for contracts in each of these databases, and, although the order differs, the subject matter is broken down into similar patterns of formation, interpretation, performance, defenses, and breach. These are recognizable to any law student. It is easy to speculate that searches based on legal concepts that have a long history will be the searches that have high success rates in these older databases. There have been many articles written about the slowness of Key Numbers in responding to new legal topics.

The newer entrants into the legal research market—Casextext, Fastcase, Ravel (now part of Lexis Advance), and Google Scholar—may be offering, in their 40 percent of unique cases, something outside of the range of that Langdellian worldview. In the new range of value-added offerings on their results pages, such as parentheticals in Casextext and the citation visualizations in Fastcase and Ravel, these databases are offering new forms of serendipity in search and adding their own unique value to the cognitive universe a researcher is trying to construct.

Final Thoughts

The important takeaways for researchers and teachers are that every algorithm is very different and every database has its own point of view. Researchers need to understand that the variability in results requires multiple searches with multiple terms and differing types of resources. Redundancy in searching is necessary to ensure you are getting a good set of relevant results. Researchers cannot rely on the black box of the algorithm and be satisfied with their initial results.

Since every algorithm and database interface is a completely human construct, and every search is a completely human construct, the researcher should view the online search process as a human interaction, moderated by technology. The goal is actually a very old one in the history of online research: We, the human researchers, need to mediate the information request so that the human engineers who created the algorithm will give us what we want, and vice versa. ■

AALL2go EXTRA


SUSAN NEVELOW MART
ASSOCIATE PROFESSOR AND DIRECTOR OF THE LAW LIBRARY
William A. Wise Law Library
Colorado Law
Boulder, CO
susan.mart@colorado.edu