2017

Research Algorithms Have a Point of View: The Effect of Human Decision Making on Your Search Results

Susan Nevelow Mart
University of Colorado Law School

Research Algorithms Have a Point of View

The Effect of Human Decision Making on Your Search Results

BY SUSAN NEVELOW MART

You might not think that the algorithm returning your search results from a keyword search in a legal database has its own opinions about the results you receive. After all, algorithms don't think, so how can they have opinions? However, algorithms are created by teams of humans. Those teams of humans made choices about how the algorithm would work that became the rules of the game long before you sat down at your computer. When the algorithm follows those rules, as it must, the rules govern what you see in the results. And every team of humans has implemented its own unique opinions about how to filter and sort the search results.

Let's say a lawyer is searching for cases in the Tenth Circuit that discuss the scope of discretion school boards have to remove, retain, or purchase books or other library materials. She opens a legal database, limits the search to the Tenth Circuit, and types in *school board discretion remove retain library material*. No matter what database she uses, she would expect to get relevant cases in the top 10 results. Regardless of the choices those teams of humans made when they set up the rules for the search algorithm, it seems a reasonable assumption that the same relevant cases will show up. That assumption is not correct.

The top 10 results in each database will differ wildly. In a recent study I conducted of 50 jurisdictionally limited searches in Casetext, Fastcase, Google Scholar, Lexis Advance, Ravel, and Westlaw, each database returned about 40% unique cases.[1] And only 7% of the cases showed up in five or six databases.[2] The human factor in algorithm creation means that every algorithm is solving the same problem in a unique way. Each database has algorithms programmed to make unique choices about how to process search terms.

The Human–Computer Interaction: What Happens When Lawyers Search?

The age-old problem computer scientists have been trying to solve is how to make the connection between human input (the search terms lawyers use to reflect the concepts in their legal problems) and output (the documents that will be relevant to the legal problem you are trying to solve). This is a lot more complex than trying to match your need for a local pizza parlor with gluten-free options to the available restaurants in your area. Legal research involves complex cognitive concepts. The toolkit that engineers use to solve the problem for keyword searching includes such decisions as:

- how to treat the number of terms in the search (i.e., if the search phrase has four terms and a document has three of them, will the document show up in the search results?);
- how close the terms have to be to each other (proximity);
- whether terms are stemmed;
- whether legal phrases can be recognized without quotes;
- whether additional concepts are added to the search;
- how numbers are processed;
- whether the algorithm uses citation analysis, a preexisting classification system, or mines other legal content, such as secondary sources;

- how the system prioritizes results by emphasizing some elements over others; and
- how information is included or excluded from results (filtering).

Each database provider weighs these factors differently. The study I conducted illustrates the variability that results from the human decision-making process.

Testing the Algorithms

My study tested 50 unique searches. Each search was limited to the same jurisdictional case database in each of the six legal databases. Each search had to return 10 results so that the top 10 results from each database could be compared. As researchers, we expect the top results to be the most relevant results, and legal database providers all explicitly state that their goal is to return relevant results at the top.

Uniqueness

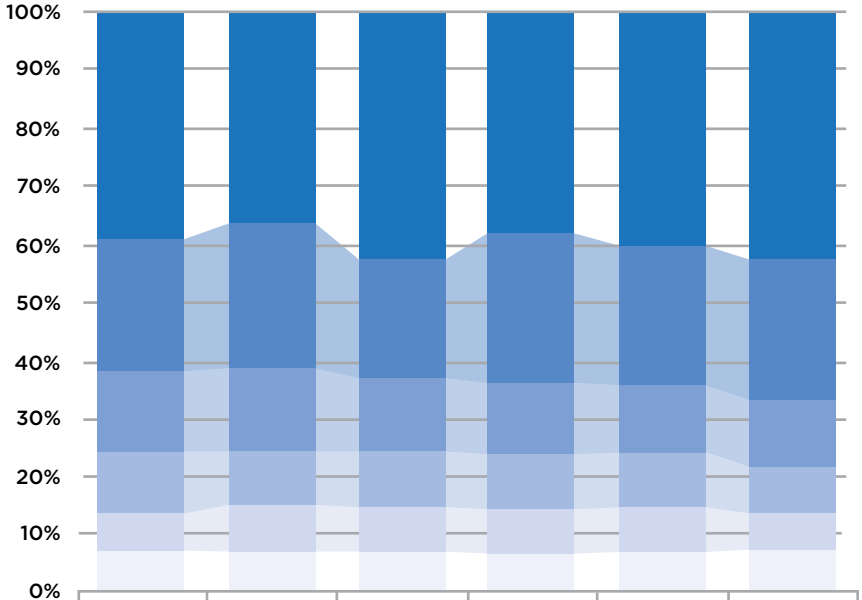
For each search, six algorithms attempted to solve the same problem in the same pool of cases. Chart 1 shows the dissimilarity of the top 10 results.

An average of 40% of the cases in the top 10 results are unique to one database. The remaining cases don't overlap much either. Another 25% of the cases only show up in two of the databases. Only 7% of the cases appear in five or in all six of the databases. If you compare the results for the top 10 in just Westlaw and Lexis Advance, the difference is even starker: only 28% of the cases appear in both databases and 72% of the cases are unique. As a first stop in your research, there will be unique results in each database. Those first results will help frame your research strategy.

Relevance

Lawyers also expect the top 10 results to be relevant. The study defined relevance both subjectively and broadly: any case that might be relevant to framing the legal argument at issue was deemed a case that would be saved for further review. For the sample search *school board discretion remove retain library material*, the researchers were told to look for cases that discuss the scope of discretion of school boards

Chart 1. Percentage of Unique Cases by Database



	Casetext	Fastcase	Google Scholar	Lexis Advance	Ravel	Westlaw
Unique Cases	39.2	36.2	42.8	38.4	40.4	43
Cases in 2 DB	23	25.4	20.4	26.4	23.8	23.8
Cases in 3 DB	13.8	14.2	12.8	12.4	11.6	11.6
Cases in 4 DB	10.4	9.2	9.6	9.4	9.6	8.6
Cases in 5 DB	6.8	8.2	7.4	7.6	7.8	6.2
Cases in 6 DB	6.8	6.8	7	6.8	6.8	6.8

Chart 2. Relevance of the Top 10 Cases in Each Database

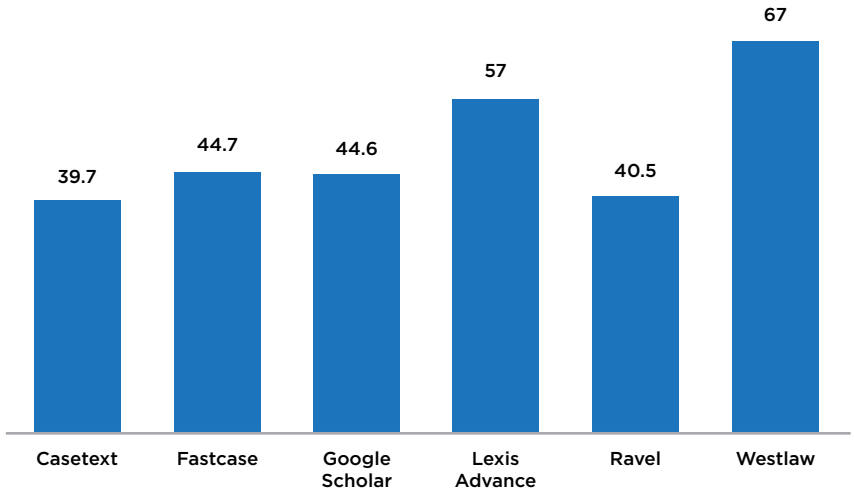
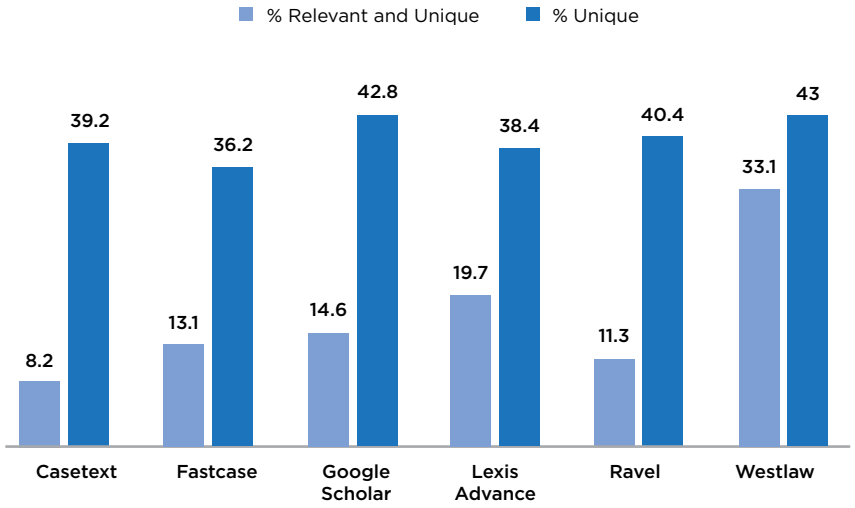


Chart 3. Percentage of Cases that are Relevant and Unique, by Database



in cases involving decisions to remove, retain, or purchase books or other library materials and list every case that might remotely be helpful as "relevant." Chart 2 shows the relevance of the top 10 cases in each database.

The results cluster between the older databases and the newer ones: Westlaw's results were 67% relevant and Lexis Advance's results were 57% relevant. The results from the newer providers (Casetext, Fastcase, Google Scholar, and Ravel) were, on average, 42% relevant.

Relevant and Unique

Each database returned many irrelevant results in the top 10. If all of the top 10 results are not relevant, and about 40% of the top 10 results are unique, how many of those unique results are relevant? Chart 3 displays these results.

Age and Numbers of Cases

The study also tested the age of cases in the results. Some databases, such as Google Scholar, had many older cases. The highest number of more recent cases can be found in Fastcase and Westlaw (67%), and then Casetext (64%), Ravel (56%), and Lexis Advance (56%). How many cases does each database find with a search? At the 50th percentile of results, Casetext, Ravel, and Westlaw are returning slightly more than 100 cases; Lexis Advance is returning over 1,000 cases; Google Scholar is at 180 cases; and Fastcase is at 70. The number of results did not affect the percentage of relevant cases, except when using Lexis Advance. The relevance of its result improved slightly as the number of cases returned by a search went up.

Point of View

The study illustrates that results vary widely depending on the algorithm used. The study also found that the two older databases are better at returning results that are both relevant and unique. Some reasons for this may be that Lexis Advance and Westlaw each leverage complex but dissimilar classification systems, a different set of secondary sources, and large bases of user history. Each of their algorithms is mining those very different points of view, and having different viewpoints in our legal databases can certainly be viewed as positive.

Although the results differ widely in Lexis Advance and Westlaw, they both rely on 19th century classification systems, importing a Langdellian[3] worldview of the law into the algorithm. Searches that look at well-established legal concepts *may* do better in these two databases. There is a fair amount of literature on the slowness of these classification systems to respond to new legal concepts. As newer players, Casetext, Fastcase, Google Scholar, and Ravel might be offering a more modern point of view in their unique results. New opportunities for serendipity in search are being offered by more recent entrants. A few examples are "parentheticals" and "it is well-settled"[4] from Casetext; visualization of citations in Fastcase and Ravel (recently purchased by LexisNexis); personalization of results in Fastcase; brief analysis in Casetext; detailed metadata filtering in Judicata; and new analytics from Ravel. And these are just the beginning of the changes.

Final Thoughts—Algorithmic Accountability

The study is a picture of how the algorithms operated at a specific point in time. Algorithms are constantly being tweaked, and the results would likely differ if the study were replicated today. Furthermore, the study examined algorithms and looked at what happens when a researcher uses keywords as a starting point, not at terms and connector searches. The study shows that every algorithm is unique and every database has a point of view. Any single entry point into the resolution of a legal problem is just that: one entry point. The study also highlights the continued need for redundancy in legal research. Multiple searches, multiple types of searches, multiple resources, and multiple strategies will always be necessary to solve complex legal problems.

Knowing how algorithms work helps researchers do a better job of finding relevant results. The more information legal database providers give researchers about how the algorithms work (i.e., the more *algorithmic accountability* there is), the better the legal databases will perform their job of providing the most relevant results, and the more trust there will be between researchers and providers.

Just bringing this issue into the open has been effective. Since I first began discussing algorithmic accountability with legal database providers, more of them have started publishing helpful information on how to use their databases. Lexis Advance[5] released a few new

fact sheets on the assumptions underlying its search. Fastcase[6] released a whole dashboard of its algorithmic tools, so that any researcher can adjust the relevancy algorithm for a specific search. Judicata[7] released a set of benchmarks to illustrate how it tests for relevance; Judicata hopes its data will both assure researchers of the validity of its results and fuel improvement in all research algorithms. Given the search results' high rate of variability, legal database providers must work harder to regain researchers' trust. CL

Susan Nevelow Mart is associate professor and director of the University of Colorado William A. Wise Law Library—susan.mart@colorado.edu.

Coordinating Editor: Robert Linz, robert.linz@colorado.edu

NOTES

1. For more detailed information about the study and the concepts in this article, *see* Nevelow Mart, "The Algorithm as a Human Artifact: Implications for Legal [Re]Search," 109 *Law Libr. J.*, Issue 3 (2017).

2. *Id.*

3. Christopher Columbus Langdell established the classic law school curriculum and case method at Harvard Law School in the late 1800s. The Langdellian case method is associated with a fairly rigid classification system that reflects a 19th century view of the law. *See, e.g.*, Bast and Pyle, "Legal Research in the Computer Age: A Paradigm Shift," 93 *Law Libr. J.* 285, 301 (2001).

4. *See* www.wellsettled.com, offering summaries afforded by a later court's synthesis of a case in a citation's parenthetical, and any principles that have become "well-settled" by the pronouncements of later courts.

5. *See, e.g.*, Lexis Advance, "Differences that Deliver: The Power of Lexis Advance® Search," www.lexisnexis.com/documents/pdf/20160803090042_large.pdf.

6. *See* the Advanced Search function in Fastcase *7*, www.fastcase.com.

7. Gurari, "Legal Search: Sharing Judicata's Data to Drive Progress," Judicata blog (Aug. 2, 2017), https://blog.judicata.com/legal-search-sharing-judicatas-data-to-drive-progress-811eed64f04b.