RANGEFINDR

rangefindr.ca and tag-based legal research

Matthew Oleynik

> For [legal] digests, where the results are curated where a human has made a judgment about the meaning of a case and placed it in a classification system humans still have an advantage. For citators, where algorithm is battling algorithm to find relevant results, it is a matter of the better algorithm winning. But neither algorithm is doing a very good job of finding all the relevant results; the overlap between the two citator systems is not that large. The lesson for researchers: know how your legal research system was created, what involvement, if any, humans had in the curation of the system, and what a researcher can and cannot expect from the system being used.
>
> - SN Mart, "The Case for Curation: The Relevance of Digest and Citator Results in Westlaw and Lexis" (2013) 32 Legal Reference Services Quarterly 13

Summary

Caselaw research is the art of searching for meaningful and useful results in a set of unsorted legal judgments. Although judgments are usually written with some degree of formality, they are far too unstructured for current technology to perform automated sorting at a non-trivial level. This means sorting the wheat from the chaff requires human interpretation, usually supplied by a researcher.

Tag-based legal research recognizes the need for human curation and interpretation of legal judgments, but performs that step once and records the results in a searchable database. By the time an end-user researcher receives a set of results, all the chaff has been removed. This saves potentially hours of the researcher's time, but also improves the accuracy of the search by ensuring no relevant results are excluded.

This speed and accuracy allow real-time display of metadata about the results of the search, instantly conveying information to the researcher while the search is going on.

The result is a much faster, much more accurate search that delivers real-time information during a search, saving the time and effort of the researcher.

Traditional caselaw research and its drawbacks

Scope of results does not match researcher's scope of interest

The traditional approach to caselaw research is to turn the researcher loose on an unsorted or minimally sorted collection of caselaw with a set of naive searching tools. Researchers search by entering a word or phrase into a boolean search engine and then manually sorting a list of judgments containing that word or phrase into useful and non-useful results.

This is a long, boring process that consumes time, effort, and money, and provides very little confidence in the accuracy or comprehensiveness of the final results since there is no guarantee a particular word or phrase will be used to describe the concept for which the researcher is searching. This problem arises mainly from the informal structure and vernacular of judgments: the same concept may appear in different judgments in different ways and in different places throughout the judgment, and will often appear in non-useful judgments.

For example, a lawyer researching criminal sentencing cases could search for the term "conditional sentence" and miss judgments in which the same legal concept was described as "house arrest," "a sentence served in the community," "a sentence conditional on the following terms," etc. Also, the word or phrase may be used, but in a

negative way: “The defence suggested a conditional sentence, but I disagree.” Likewise, the phrase “conditional sentence” will appear in many judgments where a conditional sentence was not considered; for example, in a recitation of an offender’s criminal record. See Fig. 1.

While this allows for necessary flexibility and creativity in expression by judges, it is at odds with the structure that would be required for automatic or comprehensive machine sorting. Some search tools assist the researcher by filtering, or by programmatically assessing the relevance of the search term in the judgment and adding summaries that help the researcher quickly assess each case, but for the most part the researcher is on their own to find the useful judgments in a set of results with no guarantee they are actually present.

Duplicative work

Another major drawback of traditional research methods is that the work performed by the researcher in sorting useful from non-useful results is limited only to that particular research session and is lost to all future researchers. Every researcher who searches a particular phrase will have to undertake the same exercise of reviewing the results and picking the wheat from the chaff. See Fig. 2.

Tag-based legal research

Human-curated tag-based legal research tools avoid these issues by having a domain expert read each judgment in the database and apply “tags” that describe the concepts that appear usefully in the judgment. The step of sorting useful from non-useful appearances of a particular concept is performed once and the result is captured in the database. See Fig. 3.

To extend the example given above, a researcher using tag-based legal research could click on a tag called "conditional sentence" to instantly receive a list of cases in which the concept usefully appears. This set of results would also include all cases where the concept was described with different wording.

This example only involves the selection of one tag, but real tag-based legal research allows the selection of an arbitrary number of tags. This provides the researcher with results that usefully refer to all the concepts described in the selected tags (i.e., the "intersection" of judgments that contain all selected concepts).

Benefits of tag-based legal research

Tag-based legal research has three main benefits over traditional legal research:

1. **It is much more accurate.** Because the judgments have been read and sorted by a domain expert, the set of results will list all tagged judgments that usefully contain the selected legal concepts, and will exclude all judgments that mention the concepts in a non-useful way.

2. **It is much faster.** The step of sorting useful from non-useful search results is the most time-consuming step of caselaw research. By performing this step once and then allowing researchers to access the pre-sorted results, a tag-based research system saves the researcher the majority of his or her time and effort.

3. **It enables new methods of information display.** Because the tag-based searching is so fast and accurate, display systems can be developed that provide real-time feedback to the researcher while the search is going on; see the example below.

Example of tag-based legal research: rangefindr.ca

rangefindr.ca is an online tag-based tool for searching criminal sentencing judgments. rangefindr.ca uses tag-based legal research to allow researchers to skip the entire step of sorting useful from non-useful results. A researcher using rangefindr.ca can select sets of tags such as *Offence: Assault* + *Record: First offender* + *Judgment: Emphasizes rehabilitation* and see a list of results that includes only judgments that usefully contain all three concepts. This list will include results even if the judge who authored the judgment used non-standard language to describe the concepts (i.e., the list will include judgments that do not contain the word "rehabilitation" if the concept is present).

By selecting these three tags, the researcher has instantly created a list of results that is more accurate and useful than a list that would have taken hours of manual effort to compile using a traditional research tool.

In addition to time-savings and accuracy, rangefindr.ca also uses the inherent speed of the tag-based database to provide information about the results to the researcher in real-time during the search. In the example above, after selecting *Offence: Assault*, the researcher would see the number of tagged cases in the database that are associated with that tag and the sentences imposed in those cases. When the tag *Record: First offender* is selected, the display updates to show the researcher how the types of sentences have changed in the judgments that have both of the selected tags. The same happens when the third tag is selected, delivering a great deal of information to the researcher before the results have even been viewed. See Fig. 4.

Figures

Figure 1

Figure 2

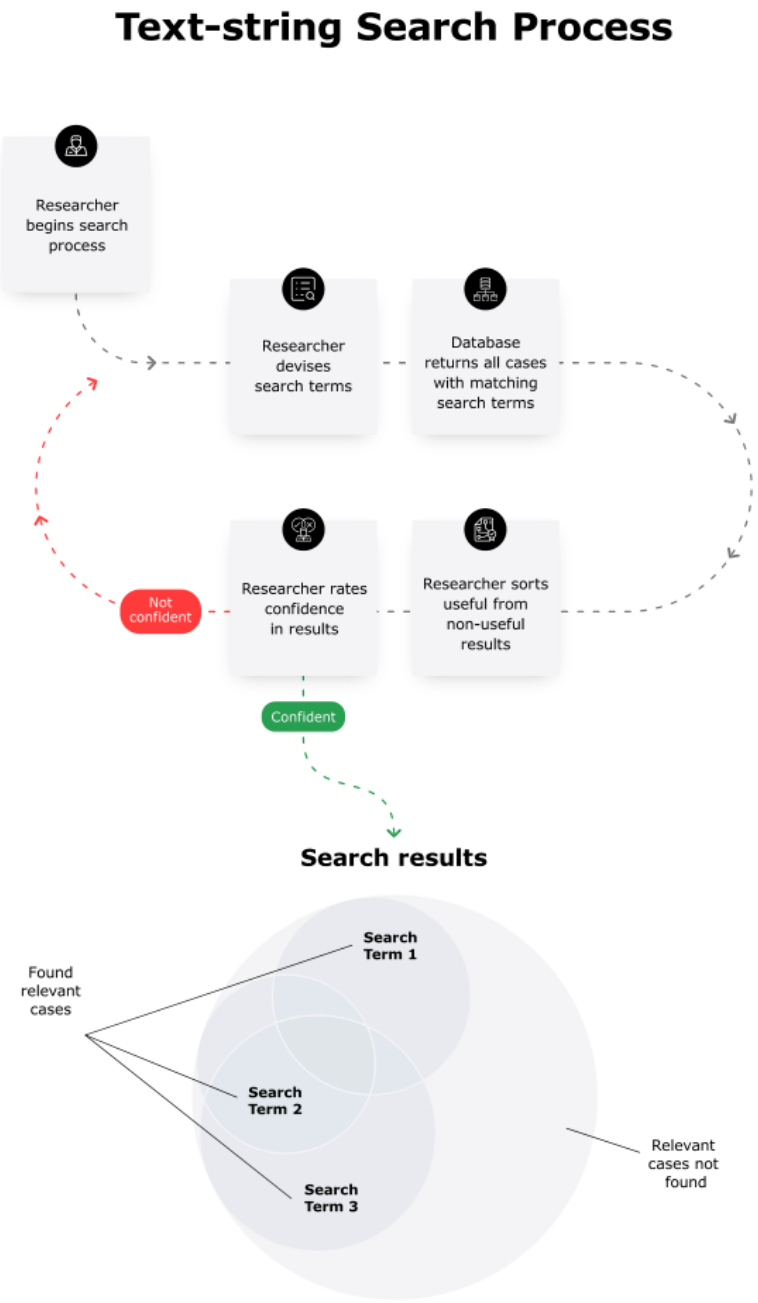

Text-string Search Process
Researcher begins search process
Researcher devises search terms
Database returns all cases with matching search terms
Researcher rates confidence in results
Researcher sorts useful from non-useful results
Not confident
Confident
Search results
Found relevant cases
Search Term 1
Search Term 2
Search Term 3
Relevant cases not found

Figure 3

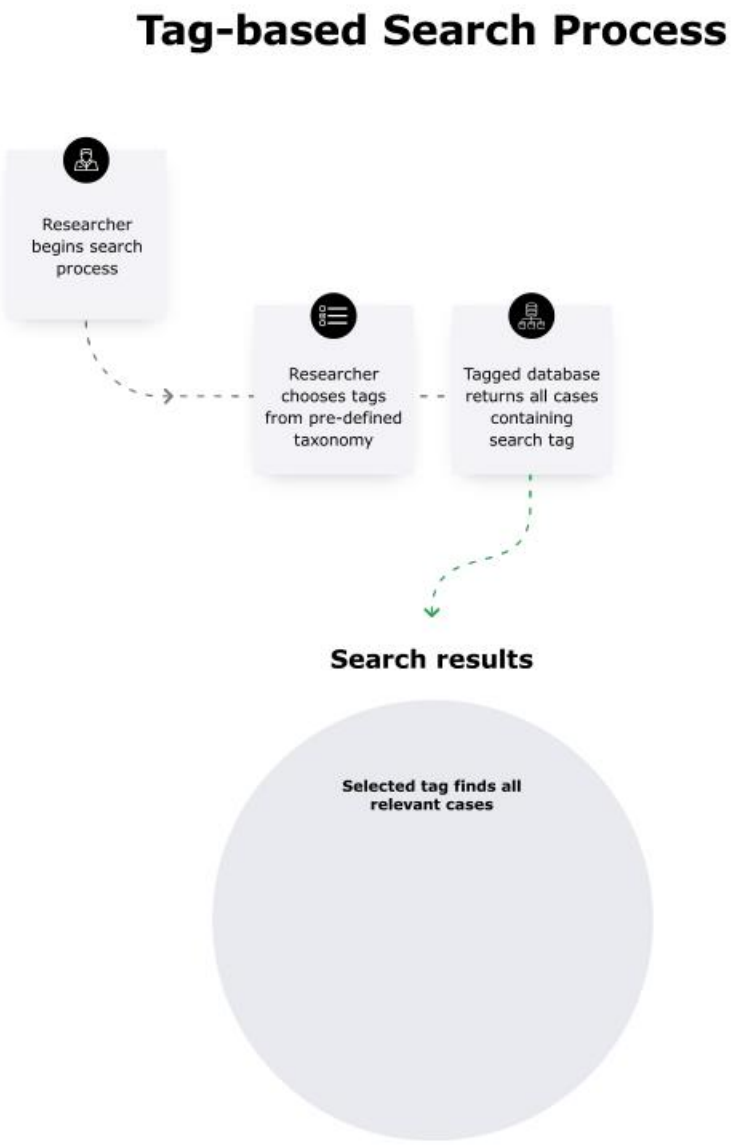
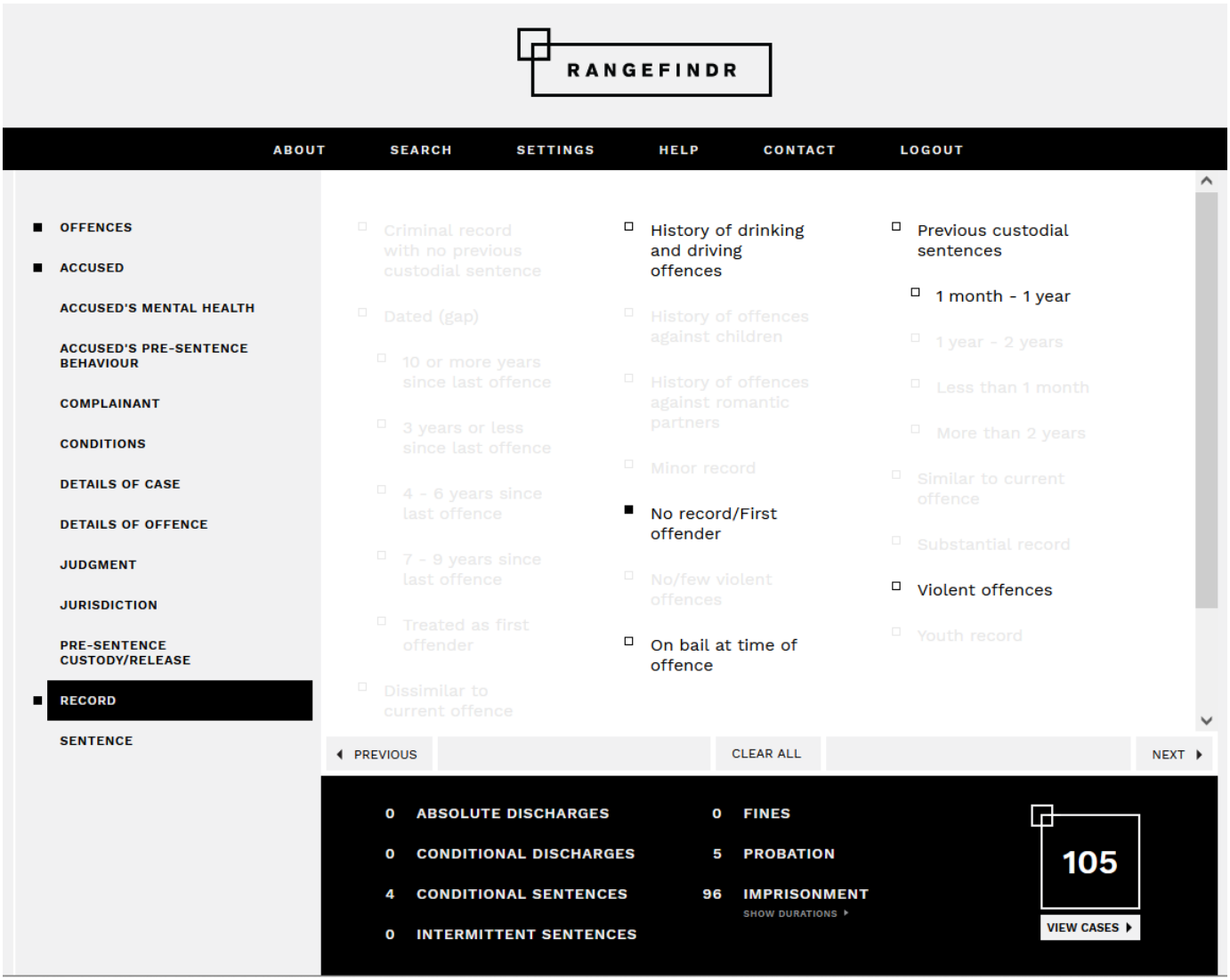

Tag-based Search Process
Researcher begins search process
Researcher chooses tags from pre-defined taxonomy
Tagged database returns all cases containing search tag
Search results
Selected tag finds all relevant cases

Figure 4


RANGEFINDR
ABOUT
SEARCH
SETTINGS
HELP
CONTACT
LOGOUT
OFFENCES
ACCUSED
ACCUSED'S MENTAL HEALTH
ACCUSED'S PRE-SENTENCE BEHAVIOUR
COMPLAINANT
CONDITIONS
DETAILS OF CASE
DETAILS OF OFFENCE
JUDGMENT
JURISDICTION
PRE-SENTENCE CUSTODY/RELEASE
RECORD
SENTENCE
Criminal record with no previous custodial sentence
Dated (gap)
10 or more years since last offence
3 years or less since last offence
4 - 6 years since last offence
7 - 9 years since last offence
Treated as first offender
Dissimilar to current offence
History of drinking and driving offences
History of offences against children
History of offences against romantic partners
Minor record
No record/First offender
No/few violent offences
On bail at time of offence
Previous custodial sentences
1 month - 1 year
1 year - 2 years
Less than 1 month
More than 2 years
Similar to current offence
Substantial record
Violent offences
Youth record
PREVIOUS
CLEAR ALL
NEXT
0 ABSOLUTE DISCHARGES
0 CONDITIONAL DISCHARGES
4 CONDITIONAL SENTENCES
0 INTERMITTENT SENTENCES
0 FINES
5 PROBATION
96 IMPRISONMENT
SHOW DURATIONS
105
VIEW CASES