

tional Law” the topic number 221. Working with the online listing in Westlaw, under this topic there are 13 key numbers, of which only five are broken down further into sub-keynumbers. The researcher can scan to see that international organizations and tribunals is a separate topic with a sub-keynumber for privileges and immunities. This gives the legal researcher a sense of arrangement of the law in the United States.

The ability to search across all key numbers, rather than browse them, is a useful way to find relevant key numbers. For example, a search for “international law” across all Topic and Key Numbers will yield a results list that includes: Courts (topic 106), Sentencing and Punishment (topic 350), and Treaties (topic 385). Indeed, much of what would otherwise be classified under public international law is included under domestic U.S. law based on the application and relation of international law in the U.S. system, including the treaty implementation process discussed in Chapter 2. This is demonstrated by the fact that international law appears not only as a topic but also as a subtopic under other topics. For example, International Law appears as a sub-keynumber under Statutes.

- 361 STATUTES (Up to 10,000)
- 361III Construction (Up to 10,000)
 - 361III(G) Other Law, Construction with Reference to (Up to 10,000)
 - 361 $\rightarrow$ 1209 International law.(36)

A legal researcher does well to recognize that all attempts to arrange public international law are influenced by the domestic jurisdiction’s legal tradition and history. Bearing this simple fact in mind helps the legal researcher to remain open-minded and creative when searching in foreign jurisdictions.

F. Digests

Digests are a set of volumes that contain short summaries of the legal points of law considered by the courts followed by citation information. In essence, digests are an indexing tool for case law. As such, researchers use digests to find case law by topic. Digest sets generally cover a specific jurisdiction. The digest set itself will have a descriptive word index that is used as an entry point for finding information.

Originally a print finding aid, some digests can now be found online. West’s Topic and Key Number System is one example of an online digest and the World

l-361255.pdf [https://perma.cc/Q3RR-D4H5] at 1 (contains a listing of topic and key numbers).

Court Digest is another.¹⁴ Researchers will find digests for some but not all international courts and tribunals. Some IGOs capture their work or practice in digests called repertoires.¹⁵ Digests are covered in greater detail in Chapter 7.

IV. Natural Language Search and Boolean Search

At the start of any project, researchers should compile a list of potentially useful key words¹⁶ and add to that list as the research unfolds. It is essential to select appropriate key words to achieve a relevant results list. There are two options for searching full-text using key words: (1) natural language and (2) terms and connectors, which is also known as Boolean search.

i. Natural Language Search

Many researchers use natural language search based on their positive experiences with search engines like Google. It involves typing key words or phrases into a search bar and relying on the relevancy ranking algorithm to include the most relevant results toward the top of the list. A natural language search will offer a results list that contains documents that use not only the key words that were entered into the search bar, but also associated words based on the algorithm working behind the scenes. For example, if a researcher types in the key word *labor*, the algorithm may also search *labour*. Similarly, if a researcher types in the key word *convention*, the algorithm may also search *treaty*. While this is beneficial to those who do not have familiarity with an area of law, it may frustrate others who want a precise results list. At a minimum, it prevents a legal researcher from knowing exactly how and with which particular search terms the algorithm created the results set.

Generally, relevancy ranking algorithms incorporate a number of factors in making a determination of relevancy. On a very basic level, relevancy ranking algorithms include how frequently key words are used in a document as well as the location in which the key words appear in the document. Those documents that contain the key words more frequently and those that contain the key words in the title or the headings (as opposed to a footnote, for example)

14. World Court Digest (1986–2000), http://www.mpil.de/en/pub/publications/archive/wcd.cfm?fuseaction_wcd=aktat&aktat=100000000006.cfm [https://perma.cc/WDL4-QDK8].

15. See *Repertory of Practice of United Nations Organs*, <http://legal.un.org/repertory/> [https://perma.cc/6AH2-YXUC].

16. Developing key words is discussed in Chapter 10.

will be considered more relevant. Of course, strong relevancy ranking algorithms take into consideration many other factors, as well, including court hierarchy and citation information. Additionally, as mentioned above, the algorithm usually incorporates synonyms to the key words and will assign a differing weight for each synonym and search for those words as well as the entered key words. The relevancy ranking algorithm is but one of the "value-added" features to any subscription database and as such is a closely-guarded trade secret. In addition, relevancy ranking algorithms are routinely updated or refined which makes comparisons between them difficult.

It is best to use a natural language search and rely on the relevancy ranking algorithm when dealing with an unfamiliar area of law or if one's Boolean search skills are not yet well-developed. After reviewing a number of results and gathering a stronger understanding of the area of law, the researcher may elect to follow-up with Boolean searching, as discussed below. A Boolean search at the end of the process, places one in a better position to evaluate the results set to decide whether the Boolean search was properly constructed. Also, because well-constructed Boolean searches are more precise, they will reveal anything inadvertently overlooked from the larger natural language result set. Eventually, when a legal researcher's ability to construct effective Boolean search queries is sufficiently developed, the Boolean query is generally considered the preferable and superior approach.

Sometimes search engines will offer options when a researcher is typing key words into the search bar. For example, researchers may observe a synonym list or "did you mean" or clustering results (depending on the search engine). These are opportunities to peek behind the relevancy ranking algorithm to gather a sense of the factors used for ranking relevancy.

ii. Boolean Search

Boolean search involves generating key words and connectors to direct how the key words should appear in relation to each other. A Boolean result set offers all documents that satisfy the query. This means that every document in the results set is relevant. There is no need for an associated relevancy ranking algorithm. Researchers must evaluate the entire results set, not only those entries that are listed at the top of the list.

When constructing a Boolean search, researchers may connect their terms or phrases using the search syntax provided by the subscription database or platform. Most researchers are familiar with the basic AND, OR, and NOT connectors. Legal subscription databases offer a variety of other connectors that allow the legal researcher to be much more precise. Key words may be connected

by paragraph (/p), sentence (/s), or within a given numeric proximity to another word (/15). The legal researcher may direct the algorithm to search for a word preceding another word (pre/s) or restricting the search to exclude a particular term (and not). Each database provider tends to have its own search syntax (examples provided in preceding parentheses are merely illustrative). Although many legal databases use similar syntax, researchers should seek the syntax legend or use the help features to learn the terms and connectors search syntax.

When crafting a Boolean search, terms and phrases of a related concept should be connected first and enclosed in parenthesis; then, that concept should be connected to other concepts, which themselves have been connected according to terms and phrases enclosed in parenthesis, using the appropriate search syntax. This is represented abstractly below and more concretely by example:

((concept one term OR concept one term) OR (concept one phrase))
/p (concept two phrase) AND (concept three phrase OR concept three
phrase)) /s (procedural concept)

((jus cogens OR ius cogens) OR (peremptory /5 norm!)) /p "sovereign immunity" AND ("international law" OR "human rights") /s jurisdiction

This can be an especially efficient and economical way to search when the subscription databases offer the option to search within (or locate or find within) a result set at no additional cost. In essence, these 'search within' features are merely a way to **reorder** the results set so that only those documents with the desired terms are included in the list, cutting down on the time it takes to review the complete results set.

iii. Multiple Language Synonyms

When searching online, researchers should determine whether the search algorithm incorporates U.S./U.K. spelling differences. If not, and certainly when conducting Boolean search, it is necessary to account for differences in spelling for certain words. A short sampling is included below:

Analyse/analyze
Centre/center
Colonise/colonize
Encyclopaedia/encyclopedia
Favourite/favorite
Harbour/harbor
Honour/honor

Kilometre/kilometer
 Labour/Labor
 Licence/license
 Offence/offense
 Practise/practice

Generally, these can be reduced to the differences listed below, with U.K. usage on the left and U.S. usage on the right. Thinking in terms of these differences can help the legal researcher identify certain less-frequently used words:

-our/-or
 -re/-er
 -ce/-se
 -ise/-ize
 -yse/-yze
 -ogue/-og
 -ae/-e

iv. Search by Field in Online Subscription Databases

Databases are nothing more than a collection of documents, whether articles, treaties, case law, statutes, or otherwise (possibly images or audio files). Each of the documents contained within a database are tagged with metadata. Metadata is data about the data within a document. For example, in a case law database, each case opinion is tagged with descriptive metadata related to the parties' names, the judge(s) who authored the opinion, the date of the opinion, the attorneys of record, the court's holding, the disposition, and the full-text of the opinion. The metadata for different types of documents will vary according to the document. For example, while there may be judge metadata for a case opinion, there would not be judge metadata for legislation. Including metadata increases discoverability and retrievability of the documents within a database. On the front end of systems, researchers tap into this metadata when they search using the advanced search template. Typically, a case law database will have individual fields for each of the discrete pieces of descriptive metadata collected for the cases.

To demonstrate the value of metadata and field searching, consider the example of a researcher attempting to find all opinions written by Judge Hamilton. An inexperienced legal researcher might type *Judge Hamilton* into a natural language search bar. This query would return all documents that contained the word judge (this term alone would return virtually every document in a case law database) and that contained the word Hamilton. All cases that con-

tained the word Hamilton whether a party name, the name of a company, the name of a product, or a location would be included in the results list, along with the desired cases authored by Judge Hamilton. Because a natural language search is searching the database full-text, it will return many more irrelevant results than an inexperienced researcher might otherwise expect.

A more advanced legal researcher might develop the Boolean search *judge +5 Hamilton* to obtain a more precise result set. By connecting *Judge* and *Hamilton* so they appear within five words of each other, the legal researcher is accounting for those circumstances where Judge Hamilton's full name may be used. Results will include all those documents that contain Judge Hamilton as well as Judge David F. Hamilton. This search results list will include all opinions by Judge Hamilton as well as all cases authored by another judge who referenced Judge Hamilton. These additional results would be considered irrelevant and not what the legal researcher intended by searching *judge +5 Hamilton*.

For this reason, an experienced legal researcher will seek to search by field. By restricting the research to the field for Judge and searching by the term *Hamilton* within that field, the legal researcher will obtain a result list of only those documents authored by Judge Hamilton. This field-restricted search offers a complete results set of all documents authored by Hamilton. The legal researcher must review all of the documents in the list because all of the documents satisfy the query. The results set will not contain any of the irrelevant documents that a natural language search for *judge Hamilton* or a Boolean search for *judge +5 Hamilton* would return.

Combining, or searching more than one field at a time, is another strategy that provides the legal researcher with a smaller and more relevant results set. This decreases the amount of time researchers must spend ruling out irrelevant results. Legal researchers may create a Boolean search to combine the judge field with a date field to obtain only those opinions authored by Judge Hamilton within the past ten years, for instance. Certain fields lend themselves to being searched together. For example, before appearing in court, an attorney may search judge and date fields or judge and holding fields to gather a sense of how that judge may have decided similar types of motions or cases during a specific time frame. Similarly, by searching the attorney and holding field, a researcher may gather a sense of how many cases an attorney has handled, as well as the outcome of those cases.

Because the nature of a document will determine the types of metadata collected and fields offered, researchers must determine which fields are associated with any particular document to utilize this strategy effectively. For example, legal researchers can select the advanced search link from the home page on either Lexis or Westlaw to obtain a list of field options. This advanced

search home page allows the legal researcher to search full-text using AND, OR, NOT, and date, citation, and title fields. However, by first selecting cases from the home page under Browse > All Content listing (Westlaw) or Explore Content (Lexis), and then clicking on the advanced search link, the legal researcher will obtain a different listing of fields specifically tailored to case law documents. This advanced search page, specifically for case law, offers more fields (e.g., judges, synopsis, attorney, docket number) than the advanced search home page. Similarly, by selecting statutes from the Browse All Content (Westlaw) or Explore Content (Lexis) listing, and then advanced search, the legal researcher obtains a template that offers options related to statutes, including restricting a search to the caption, historical notes, or annotations.

v. Search by Field Online

For researchers accustomed to using field searches on Westlaw or Lexis, searching the open web becomes an exercise in frustration due to a lack of precision. The ability to mimic some of the enhanced search functionality that researchers have come to expect with the legal subscription databases may be achieved on the open web using some search engines, like Google.

For instance, a loose equivalent of field searching online may be achieved by using the extensions from URLs. Many legal researchers are aware that *.int* is the extension for international organizations. By including *inurl:.int* into the search query, the legal researcher restricts the results list to only those websites that contain the *.int* in their URLs. It is recommended that when using the *inurl* command, the legal researcher add this command to the beginning of the search query rather than the end. This is because *inurl:.int* "human rights" and "human rights" *inurl:.int* will return a different results list. By including *inurl:.int* command first followed by the search terms 'human' and 'rights,' only those results with *.int* in the URL that also contain human rights in the documents will be included on the results list. Whereas "human rights" *inurl:.int* returns those sites that contain *.int* as well as *.org* in the URLs with human rights in the documents. This is because the relevancy ranking algorithm takes into consideration the order in which the legal researcher lists the search terms in the search bar. Those entered first are either given greater weight or are processed first, or a combination of both.¹⁷ Indeed, entering 'rights human' will provide a different results list than 'human rights.' As mentioned previously, legal researchers who know how relevancy ranking algorithms generally

17. Because the relevancy ranking algorithms are proprietary it is impossible to know the details.

work understand the need for carrying out a few searches to serve as a control strategy.

vi. Search by Command

Searching by command, where available, allows a legal researcher to tap into field searching without having to resort to a template or advanced search feature. The command abbreviations followed by a colon and the search terms is the standard format for search by command.

The Westlaw and Lexis platforms offer the ability to search by command. However, at the point of publication, most of these commands are not listed on homepage. Researchers familiar with these platforms proceed to the advanced search link and fill in the templates to discover the commands that are being utilized behind the scenes and then use those commands directly. For example, if a legal researcher is interested in finding a case where one of the parties was Colorado Coal, the following command could be used: advanced: TI("colorado coal"). It is important to remember that the commands are part of the Boolean search functions for the search engine. As such, if a legal researcher knows both parties names rather than type in *Roe v. Wade* in order to use the command, the legal researcher would connect *Roe* and *Wade* with a narrow connector, either /s or /15. Or, simply type into the search bar: TI(roe /s wade).

The same approach holds true for dates. Once the legal researcher knows the date command it may be used at the outset of query development. For all documents after January 1, 1995, the following date command could be used: DA(aft 01-01-1995). Other alternatives for dates include the 'between' restriction: DA(aft 12-31-1959 & bef 01-01-2017). However, on some level, it may be just as efficient to fill out the advanced search template as it is to ensure that the command format is correct. These advanced strategies are for the legal researcher who uses a particular subscription database exclusively and frequently.

Another way to investigate the commands for any particular database is to check whether, after conducting an advanced search using a template, the database supplies the search query (either in the search box or at the top of the results list). In circumstances where it does, the commands may be apparent. For example, suppose the legal researcher was interested in finding an article by an author with the last name Smith related to law reform. After conducting an author search in the HeinOnline Law Journals database, using the template, the resultant search query shows: ("law reform" AND creator:(smith)). The HeinOnline database, by making this information transparent, allows the user to see the syntax that was generated through the template. When researchers can see the syntax generated, they can incorporate it directly into

their subsequent searches. Legal researches are cautioned to avoid adding a space following the colon when conducting command searches.

V. Comparing Search Approaches and Tools

By utilizing a natural language search, a legal researcher will obtain a results list that includes all documents that contain the key word or phrase, whereas searching the index or table of contents will offer only those documents where the text contains a discussion related to the key word or phrase. What makes searching full-text less efficient is the need to spend additional time evaluating a lengthy list that may very well contain a number of irrelevant documents. While reliance on the relevancy ranking algorithms may allow the legal researcher to focus only on those documents near the top of the list, at some point the legal researcher must make a judgment call on when to stop checking additional results on the list. It is at this point that the legal researcher is declaring that the rest of the list probably offers results that contain the key words used only once or in a manner irrelevant to the particular research needs. This may result in a nagging feeling that something has been overlooked.

Carrying out a search of the indexes and table of contents first, to find only those documents that contain a discussion related to the key words or phrase, is usually productive. An index or table of contents search will deliver a result set that includes only those documents that are about the key words. Whereas, a full-text search will contain all documents that contain the key words, whether the document is about those words. Online indexes and tables of contents increase efficiency even though they require the legal researcher to carry out an additional search.

For example, a researcher may have a fact pattern related to lights being used to torture an individual. If this researcher were to enter the query: *light /s torture* into the search bar, the results list should contain all documents where the word *light* is within the same sentence of *torture*. The legal researcher is expecting to find documents that describe the torture suffered by a plaintiff. The legal researcher can envision the following types of sentences: "plaintiff was tortured by the use of flashing extremely bright lights." Or, "strobe lights were used to torture the plaintiff." And indeed, the results list contains a number of documents that contain the words *torture* and *light*. However, by scanning the results set, the legal researcher discovers that the word *torture* is sometimes used in a sentence containing the phrase "... in light of the fact that..." As it turns out, this full-text search yields a results list that contains a

number of irrelevant results. This could be avoided by searching an index for *torture* and then scanning for a subentry for *lights*, the legal researcher obtains only those results that cover using lights as a method of torture.¹⁸

It is the nature of the English language, where words have multiple meanings and can be used in phrases or contexts that do not immediately occur to the researcher while generating a search query, that requires researchers to carefully evaluate the results list. The legal researcher must develop a sense for when to re-query rather than to begin with a search of the index or table of contents. A re-query to eliminate the phrase "... in light of ..." may be used; however, it might inadvertently exclude a highly relevant result—one where the terms are used as originally anticipated but that same document also uses the phrase '... in light of ...'. The legal researcher who understands the strengths and weaknesses of the various approaches does not hesitate to run multiple and pointed searches, despite the cost, when comprehensive research is required.

VI. Control Strategies for Various Search Techniques

Experienced legal researchers employ at least one of the previous strategies and usually a combination of these strategies. At times, the legal researcher may be utilizing an additional strategy or approach to confirm what is already uncovered. This is considered utilizing a control strategy. It allows the legal researcher to verify that the original search achieved the desired objective.

For instance, a legal researcher may first conduct a search of abstracts and then follow it with a full-text search. Starting with abstracts followed by full-text allows the researcher to evaluate the full-text results set more efficiently because all of the relevant results from searching the abstracts should appear in the full-text search (although not necessarily near the top of the results list depending on the strength of the relevancy ranking algorithm). The full-text results that are unfamiliar to the researcher who has already carried out an abstracts search may then be skimmed for relevancy. This lends to efficiency because the searcher, when skimming, understands that the result may not be

18. Additionally, it should be pointed out that by searching an index for *torture* and then scanning for *lights* as a subentry, the researcher can readily find other means of torture that are listed as subentries. Such an approach encourages a legal researcher to develop arguments by analogy.

about the search terms but rather may only contain references to the search terms. This allows the legal researcher to more confidently focus on the portion of the document that might be relevant. This approach also accounts for any articles or documents that do not have an associated abstract, or that would be overlooked if the researcher had carried out only the abstract search.

Control strategies may also be used to help the legal researcher understand how a particular database functions. If there are no associated costs with a database, the legal researcher may elect to run a few searches to assess the strength of the relevancy ranking algorithm (presuming the researcher has a sense of what should appear toward the top of the results list) or to better understand the search syntax.

VII. Conclusion

The international legal system has no single over-arching architecture or arrangement of the law. However, by surveying several available arrangements, legal researchers can begin to develop a framework for carrying out international legal research by topic. Researchers should remain flexible yet persistent and utilize the variety of finding aids available to them. Due to the complexity of international legal research, researchers should develop and practice their natural language and Boolean search skills while employing control strategies to ensure nothing is inadvertently overlooked.

Chapter 10

Planning and Recording Research

I. Introduction

A research plan serves as a road map for the legal research process. A good research plan includes all the sources of law that the researcher must consult together with a comment about how to find and use each source. The complexity of international legal issues, the nuances related to handling the sources of law, and the abundance of databases and legal publishers worldwide contribute to the need for a detailed research plan. Ultimately, taking time to generate a research plan at the outset of a project will save the researcher time and frustration.

When crafting a plan it is important to recognize that there is typically more than one good route for gathering all relevant material. The route taken will depend on the availability of resources as well as what that researcher already knows about the subject area. However, the methods for handling the international law sources, described in detail in earlier chapters, must be included in every good research plan. By adhering to a planned and methodical approach, the legal researcher can be sure that nothing is inadvertently overlooked.

When handling public international law sources, researchers must account for the fact that certain publications may express the law itself as well as serve as evidence of customary international law or general principles of law. For example, a researcher may cite to the *Vienna Convention on the Law of Treaties* as treaty law in one instance and as evidence of customary international law