

## TEACHABLE MOMENTS...

EDITED BY BARBARA BINTLIFF

*Barbara Bintliff is Library Director and Professor of Law at the University of Colorado School of Law in Boulder. She is a member of the Perspectives Editorial Board.*

**Teachable Moments ...** is designed to provide information that can be used for quick and accessible answers to the basic questions that are frequently asked of librarians and those involved in teaching legal research and writing. These questions present a “teachable moment,” a brief window of opportunity when—because he or she has a specific need to know **right now**—the student or lawyer asking the question may actually **remember** the answer you provide. The material presented in this column is not meant to be an in-depth review of the topic, but rather a summary of the main points that everyone should know. Readers are invited to submit their own “teachable moments” to Barbara Bintliff, University of Colorado Law Library, Campus Box 402, Boulder, CO 80309, (303) 492-1233, fax: (303) 492-2707.

*In this issue, Jessica R. Hogan, Electronic Resources Specialist at the University of Denver Westminster Law Library in Denver, offers advice on what to tell the individual who presents an ideal “teachable moment” by asking:*

### “Why won’t my Westlaw search work on Lycos?”

Legal researchers who use both Westlaw and LEXIS-NEXIS have learned that the two computer-assisted legal research (CALR) systems accept basically the same search commands and are remarkably forgiving if an incorrect connector is entered. Depending on the circumstance, the CALR services will either automatically correct an incorrect command or respond with an error message, highlighting the problem and suggesting how it can be corrected. However, the many search engines on the Internet each have their own command language, or syntax, and do not necessarily accept commands unknown to their syntax. Further, the Internet search engines won’t tell you if there is a problem with your search commands—they just produce unresponsive (or no) results.

Understanding the syntax of a search engine and the order in which search commands are processed is

crucial to effective online searching. This article explores the search connectors used by the two major commercial CALR systems, Westlaw and LEXIS-NEXIS, and on the Internet. Boolean search connectors, the order in which the search engine processes connectors, the use of parentheses to change the order of processing, and other search basics are reviewed. This is contrasted with the “New Boolean” language of the Internet. The goal is to explain how the CALR services produce their results, and highlight the differences between them and the Internet services.

### Boolean Search Connectors

Traditionally, users of Westlaw and LEXIS-NEXIS have used Boolean<sup>1</sup> operators to structure their searches. The basic operators—AND, OR, and NOT—are the most commonly used connectors. They are also used by most Internet search engines.

Powerful commercial search engines such as Westlaw and LEXIS-NEXIS add numerical and grammatical Boolean connectors, such as “within a certain number of words” (W/N), “within the same paragraph” (W/P), and “within the same sentence” (W/S). Some search engines on the Internet add the Boolean connector NEAR. These terms serve to refine the capabilities of the more general AND, OR, and NOT.

### Grammatical Versus Numerical Connectors

Basic Boolean connectors are used to search for key terms in the documents contained in a database. One advantage of using these connectors is that the order of the search terms in the document is irrelevant. For example, a search on Westlaw or LEXIS-NEXIS of

marijuana and smell and “probable cause”

will return documents that contain those three terms, regardless of their position in the document or their relationship to each other in the document. This is a very general type of search, and will usually return a large number of documents of varying relevance.

<sup>1</sup> George Boole (1815–1864), a 19th-century mathematician, first suggested that logical concepts could be expressed with algebraic symbols. The method he developed to illustrate his theory is used today to perform many computer searches, using words like “and” and “or” to direct the computer to identify relationships of words and phrases.

“Understanding the syntax of a search engine and the order in which search commands are processed is crucial to effective online searching.”

“Numerical connectors work best when the searcher is very familiar with the subject at hand and knows the relationship between the words.”

Grammatical connectors specify a relationship between the key terms. A search using grammatical connectors is more refined than one using just the basic Boolean operators. It allows the searcher to ask for a logical relationship between words. For example, the search

marijuana /p smell /p “probable cause”

containing the same search terms but linking them with paragraph connectors, is much more likely to retrieve documents in which the smell of marijuana gives rise to probable cause, since all of those terms appear clustered together in the same paragraph. Again, the order of the words in the paragraph is not important; all documents with the three terms in one paragraph will be retrieved. This eliminates many irrelevant documents, and returns a more responsive search result.

Numerical connectors require that the search terms be within a definite relationship to each other. Searching for

marijuana /10 smell /15 “probable cause”

will retrieve documents that have the word *marijuana* within 10 terms of the word *smell*. The word *smell* will be within 15 terms of the phrase “*probable cause*.” Results will not necessarily show a proximate connection between the terms *marijuana* and *probable cause*. If you guess the wrong order of the search terms, or if you use a number that is too small when you specify the maximum distance between the words, you may miss important documents. This paragraph clearly shows the risk of using numerical connectors improperly:

There was sufficient **probable cause** in this case. The **marijuana**, locked in the defendant’s trunk, was not in plain view. However, the officer who first arrived at the scene had his police dog with him. The officer had the dog **smell** the car, and the dog started barking.

Both the search using basic Boolean connectors and the search using grammatical connectors would return this document, but the numerical connector search would not. Numerical connectors work best when the searcher is very familiar with the subject at hand and knows the relationship between the words. Otherwise, numerical connectors should be used with caution.

### Standard Order of Processing on CALR Systems

Once a search is transmitted to the CALR service, the search software starts to work on the parts of the request. Each online service has a standard order of

processing the Boolean search connectors.<sup>2</sup> They are essentially the same in that they move on a continuum from narrow to broad. If more than one numerical connector is used, they are processed from the smallest number to the largest.

On Westlaw, terms joined by Boolean connectors will be searched for in the following order:<sup>3</sup>

“”, SPACE (OR), +N<sup>4</sup>, /N, +S, /S, +P, /P, & (AND), % (BUT NOT)

On LEXIS-NEXIS, the order of processing Boolean connectors is:<sup>5</sup>

OR, W/N, PRE/N, NOT W/N, W/S, NOT W/S, W/P, NOT W/P, W/SEG (segment, or field search), AND, AND NOT

For example, in the search below, Westlaw will first process the OR connector (represented by a space between *harassment* and *discrimination*), and then will search for all documents with either *harassment* or *discrimination* within three words of *sexual*. From among the documents now identified, Westlaw will then retrieve those documents with the word *hostile* in the same paragraph:

sexual /3 harassment discrimination /p hostile

### Changing the Order of Processing

Advanced online researchers have learned that they can change the order of processing Boolean connectors with parentheses in order to retrieve documents with more specificity. As an illustration, suppose you want documents to contain the terms *probable cause* or *reasonable suspicion*. The search

probable w/3 cause or reasonable w/3 suspicion

will first look for documents that contain *cause* or *reasonable* in the same document, then look for *cause* or *reasonable* within three words of *probable* and *cause* or *reasonable* within three words of *suspicion*. The results of this search will be quite different from what is actually intended. However, the searcher can correct the unintended grouping and retrieve the desired documents by using parentheses around logical groupings of words:

(probable w/3 cause) or (reasonable w/3 suspicion)

<sup>2</sup> Now most of the standard Westlaw and LEXIS-NEXIS connectors are interchangeable and using W/N will work on Westlaw and /N will work on LEXIS-NEXIS.

<sup>3</sup> See *Discovering Westlaw* 37 (8th ed. 1998).

<sup>4</sup> “N” represents the number chosen by the researcher to specify an adjacency. For example, the search would be written *jessica +3 hogan*.

<sup>5</sup> See *Emmanuel’s LEXIS-NEXIS for Law Students* 2–11.

## Results Ranking and Results Clustering

When a researcher uses Boolean commands on LEXIS-NEXIS and Westlaw, the results of the search are displayed in reverse chronological order; the most recent documents are listed first. These services retrieve only documents that fit the parameters of the search request. If only five cases have the terms *marijuana*, *smell*, and *probable cause* in the same paragraph, then only five documents will appear in the results list.

In contrast, if the researcher switches the search to Natural Language on Westlaw or Freestyle on LEXIS-NEXIS, the results ranking is not by date, but by the number of times the search terms appear in the text of the document. The document in which the terms appear most often will be first in the results ranking. In Natural Language searches, if a search term does not exist in any document, then the term is discarded and the remaining terms are searched.

## The “New Math” on the Internet

It’s typical for Internet researchers to use the most basic of search strategies, entering a single word or two, often with no connectors. However, most of the major Internet search engines feature basic and advanced Boolean searching and even “nesting” (using parentheses to build a more complex search). A new syntax, variously called “search engine math” or “implied Boolean” has emerged on the Web. It uses the + and – symbols. This is in addition to the general availability of the standard Boolean AND and OR connectors often found on a Boolean template that is completed by the researcher.

Adding the + symbol before a word requires that the term appear in the document. Similarly, adding a – symbol before a word excludes that word from the document. So, the search

+attorney +jobs

will retrieve Web pages containing both words.

However, if you want to exclude a particular practice area, then you could search for

+attorney +jobs –criminal

On the Internet, the ranking of search results is like that of a natural language search. Even with advanced Boolean searches on the Internet, most search engines retrieve documents with the greatest number of search terms on the Web pages. Because

of this, Danny Sullivan, editor of the *Search Engine Watch* Web site, suggests that Boolean commands are “overkill for the average web user.”<sup>6</sup>

Results “clustering” is a feature on the Internet that retrieves only one page per site instead of multiple pages from one site. That is, the search result won’t list separately several pages that are really part of the same Web site. Some search engines allow the user to uncluster the results, showing the various pages available at one site. Other search engines provide a feature that allows the user to retrieve additional, similar Web pages. (On LEXIS-NEXIS, the two features “More Like This” and “More Like Selected Text” also allow the user to retrieve additional documents based on the original search.)

## Phrase and Field or Segment Searching on the Internet

Although some search engines have automatic phrase detection, Internet searchers should enclose phrases in quotation marks to retrieve the exact phrase. The search

“occupational outlook handbook”

will retrieve Web pages with an exact match to that phrase. The searcher can even combine quotation marks with the + and – symbols as follows, requiring that the term *1998-99* appear:

“occupational outlook handbook” +1998-99

Like LEXIS-NEXIS, some Internet search engines have automatic phrase detection, even when the words are not surrounded by quotation marks. For example, on many Internet services the search *probable cause* retrieves documents in which these two words appear as a phrase.

Some of the Internet search engines provide for special field or segment searches, such as a title search. Using that capability, a researcher can search for the titles of Web pages.

## Nesting and Wildcards on the Internet

“Nesting” on the Internet uses parentheses to create more advanced Boolean searches, such as jobs and (attorney or lawyer)

<sup>6</sup> See Danny Sullivan, *Search Engine Watch: Boolean Searching* (visited June 2, 1999) <<http://searchenginewatch.internet.com/facts/boolean.html>> (suggesting that even advanced researchers are better served if they abandon Boolean searching on the Internet).

“Results ‘clustering’ is a feature on the Internet that retrieves only one page per site instead of multiple pages from one site.”

.....

The searcher must first click the “advanced” tab to run a Boolean search on the Internet. Once in the Boolean search mode, the user can then read the help menus that explain the syntax of that particular search engine.

A wildcard on the Internet operates similarly to wildcards on Westlaw and LEXIS-NEXIS. There are two wildcards on Westlaw and LEXIS-NEXIS, the ! symbol and the \* symbol. The ! symbol acts as a universal root expander and will retrieve any variant of a root word. For example, the search *bank!* will retrieve documents containing the words *bank*, *banks*, *banker*, *banking*, *bankrupt*, and *bankruptcy*. The \* symbol acts as a placeholder, like the blank tile in the word game Scrabble®.

The effect of these symbols is just the opposite on the Internet. The \* symbol acts more like the ! symbol does in the CALR services because it retrieves various endings of the original search terms.<sup>7</sup> For example, *play\** may retrieve *plays*, *player*, and *playing* depending on the search engine. Some search engines will automatically generate various word endings, like a Natural Language search does on Westlaw.

### Conclusion

Researchers familiar with traditional Boolean search operators on Westlaw and LEXIS-NEXIS have transferable skills for Internet searching. However, unlike CALR, there is no standardization of Boolean operators on the Internet. Each search engine is slightly different in its approach to Boolean commands. Therefore, the Internet searcher must become familiar with the syntax of a search engine before relying on it. To do otherwise is to waste time and effort.

© 1999 Jessica R. Hogan

<sup>7</sup> This may be referred to as stemming.