Find It on the Web Using the Search Concepts You Already Know

William H. Weare, Jr., Valparaiso University, Valparaiso, IN

Abstract

The author discusses the applicability and effectiveness of phrase searching, the use of natural language, Boolean logic, nesting, proximity, truncation, and wildcards to web search engines.

Keywords

Access to information, electronic information resource searching, information retrieval, information-seeking strategies, internet in education, internet searching, school libraries, web search engines, world wide web

The version of record of this manuscript was published in *Library Media Connection* in 2008.

Weare, W. H., Jr. (2008). Find it on the web using the search concepts you already know. *Library Media Connection 26*(6), 56-58.

When searching the library catalog or a database, most librarians are comfortable with a variety of search concepts, such as the use of phrase searching, truncation, or wildcards. What about when searching the Web? Which search techniques really work? And, with which search engines do such techniques work? It is possible to significantly improve Web search results without using the advanced search screen, learning complex search strategies, or trying to remember complicated commands or special syntax. The following is a review of the applicability and effectiveness of the search concepts you probably already know and use—phrase searching, the use of natural language, Boolean logic, nesting, proximity, truncation, and wildcards—in Google, Yahoo!, Live Search, and Ask, the four largest search entities that operate their own crawling and indexing technology.

PHRASE SEARCHING

The application of phrase searching (enclosing a phrase in quotation marks) rather than treating the words in the phrase as separate keywords can produce significantly fewer results and appreciably more focused search results. The default search in all four major search providers is the Boolean AND; thus, a search for Des Moines River is actually a request to locate pages that include three strings of text: Des AND Moines AND River, anywhere in the document -- and in any order. A search for the phrase "Des Moines River" is a request to locate pages that include one string of text: "Des Moines River"—spaces included. The list of search results is considerably shorter, and the searcher can be confident that these results will include some reference to the channel of water known as the Des Moines River.

NATURAL LANGUAGE

Librarians may not think of the use of natural language searching as one of their strategies for finding information, yet we all use it. Rather than thinking in terms of subject headings, subject terms, subject phrases, or descriptors, a user can enter a question or a phrase—with or without quotation marks—in plain language. Although natural language searching does not work very well in highly-structured databases like library catalogs, it does work remarkably well with the major search engines.

Weare, W. H., Jr. (2008). Find it on the web using the search concepts you already know. *Library Media Connection 26*(6), 56-58.

You might remember that when Ask Jeeves was launched in 1997, the original concept was to "answer your questions in plain English." A user might ask, "Who won the Oscar for best actress in 1980?" The search engine determines which terms in the search phrase are significant (Oscar, best, actress, and 1980) and which are not (who, won, the, for, and in). The significant terms are searched in the database, and those sites deemed most relevant to the search are returned. With a simple query such as, "Who won the Oscar for best actress in 1980?" (without quotation marks), both Ask and Live Search deliver the correct answer (Sissy Spacek, Coal Miner's Daughter) above the organic search results.

Questions that begin with "What is" work surprisingly well. The query, "What is the World Trade Organization?" (again, without quotation marks), locates sites that ask and answer that question. Above the organic search results, Google provides links to both Google Book Search and to a list of definitions found on the Web; Live Search provides a link to the Encarta World English Dictionary. These types of queries need not be questions; try "Gone with the Wind is the best movie ever made" or "Global warming is a hoax."

BOOLEAN OPERATORS—AND, OR, NOT

Boolean operators are used to connect search terms in a way that either deliberately narrows or broadens the results. The use of AND or NOT will narrow the search and thus retrieve more focused results; the use of OR allows the user to connect two or more similar concepts terms (synonyms) and, as a result, broaden the search. Boolean operators must be in upper case.

The default search for the four major search engines is the Boolean AND. A search for chronic child malnutrition should be the same as a search for chronic AND child AND malnutrition; however, the search results are not identical, suggesting that the search algorithms may be considering word order as well. Taking advantage of the default AND by adding more terms to the above search—India, poverty,

Weare, W. H., Jr. (2008). Find it on the web using the search concepts you already know. *Library Media Connection 26*(6), 56-58.

globalization—is an excellent search strategy, as each added term will help the user retrieve fewer, more focused results.

If users have precise search terms and phrases to describe what it is they hope to find, they should use those terms. A search for "global warming" will retrieve millions of results. But if a user really wants specific information about the Kyoto Protocol, the environmental treaty that addresses climate change and the release of greenhouse gases, a search using appropriate and specific search phrases, "Kyoto Protocol" AND "climate change" AND "greenhouse gases" will retrieve fewer, appreciably more focused results.

Many words in English have more than one meaning, so it is a good idea to add one or more terms to a search that clarify the meaning of the words with multiple meanings. When looking for information about the Amazon, add a term such as basin, rainforest, or river that describes the sense of the term Amazon being sought.

The Boolean operator OR can be useful when searching for information on a topic that has more than one term to describe essentially the same concept; lawyer OR attorney, or "biological parents" OR "birth parents." Using OR will retrieve Web sites that contain either of the search terms or phrases you enter. The simple Boolean OR (uppercase) search works in all four major search engines.

The Boolean operator NOT (or the minus sign) allows users to refine their search by excluding specific terms from the results. This can be helpful in cases where a common term has several unrelated meanings, some of which produce large numbers of unwanted results. The use of a minus sign (no space between the minus sign and the search term) is reputed to work in all four search engines; the word NOT (upper case) before the term to be excluded should also work in Yahoo and Live Search. A search for penguins produces results about not only the flightless marine birds but also about the Pittsburgh hockey team. A search for penguins -hockey (or NOT hockey) should exclude sites containing the word hockey, thus excluding sites related to the Pittsburgh Penguins hockey team. It is important to note that

although many hockey sites are excluded, top results from all four search engines include information about the Pittsburgh Penguins, usually from online news sources such as http://sports.yahoo.com.

The use of the Boolean operator NOT to refine a Web search is not a particularly effective method of narrowing a Web search. A search for the term Gettysburg returns millions of results, which contain information about the Gettysburg address, the Battle of Gettysburg, Gettysburg College, and so forth. Suppose a user is interested in the guided missile cruiser, the USS Gettysburg. A search for Gettysburg -address, a search for Gettysburg -battle, or a search for Gettysburg -college will, indeed, all return fewer results. It is also possible to combine unwanted terms: Gettysburg -address -battle -college, which excludes the terms address, battle, and college from the results. However, a better search approach is to forgo the attempt to exclude terms; a simple query for Gettysburg AND ship returns fewer results than any of the query examples above. Better yet, a search for the phrase "USS Gettysburg" returns the shortest list containing the most relevant results.

Although excluding terms is not a particularly effective way to narrow a search, it can be useful in eliminating very specific, unwanted results. A search in Ask for Virginia Woolf -Wikipedia will eliminate results from Wikipedia.

NESTING

Nesting allows a searcher to specify the order in which the search terms are interpreted. Information within the parentheses is read first followed by the information outside the parentheses. In a search for career AND (mortician OR undertaker OR "funeral director"), all of the results retrieved will include the word career, and all will include either mortician or undertaker or funeral director.

A nested search using the Boolean operator OR is a valuable search tool. In Google and Yahoo, it is not necessary to use parentheses: career AND mortician OR undertaker OR "funeral director" is the same as career AND (mortician OR undertaker OR "funeral director"). In Live Search, put the Boolean OR

Weare, W. H., Jr. (2008). Find it on the web using the search concepts you already know. *Library Media Connection 26*(6), 56-58.

operation in parentheses; otherwise, the search performed will search for any one of these: career and mortician, or undertaker, or "funeral director." Nesting does not appear to work in Ask.

PROXIMITY

Proximity searching allows a user to search for two or more words that occur within a specified number of words, or fewer, of each other. The most common expressions for proximity are adjacent (adj), near (n), or within (w), often followed by a number that specifies the numbers of words that may appear between the terms. Unfortunately, the four major search engines do not support proximity operators.

TRUNCATION

Truncation is the search technique used when there are different endings to a particular word and the searcher wants to retrieve documents that contain any version of that word. A search for comput* will return results that include computer, computers, computing, etc. Truncation symbols vary among databases; the symbol might be an asterisk, a dollar sign, an exclamation point, a question mark, or a number sign. Unfortunately, the four major search engines do not support traditional truncation. However, Google uses what it calls "stemming technology"—which operates like truncation in that it will automatically find variants of some of the search terms.

WILDCARDS

The four search engines do not support traditional wildcard searching, such as a search for wom?n to locate documents containing woman or women. However, Google does support the use of a full-word wildcard. An asterisk is used to represent one or more unknown words. Full-word wildcards are useful in finding lyrics, lines of poetry, or famous quotations if the user is unclear about the exact wording, "Stopping by * on a Snowy Evening" (with or without the quotation marks) will retrieve the Robert Frost poem. The full-word wildcard is also useful for finding or confirming factual information. "Bili Clinton was born in *" will retrieve either a place (Hope, Arkansas) or a date (1946). Although the

full-word wildcard is not officially supported in Ask, Live Search, or Yahoo, a search without quotation marks and without a symbol indicating the missing word may easily retrieve the desired results provided that the user has input a sufficient number of other keywords.

WHAT WORKS AND WHAT DOESN'T—A QUICK REVIEW

Phrase searching works very well in all four major search engines and is, perhaps, the most easily taught search strategy to novice users. Natural language searching can also produce excellent results and, again, is easily taught. All four major search engines support the use of the Boolean AND as well as the Boolean OR. The Boolean NOT is of limited usefulness in search engines. A nested search does not appear to work in Ask. Unfortunately, the four major search engines do not support proximity operators, such as adj (adjacent), n (near), or w (within). None of the four search engines support truncation, but Google employs stemming technology, which results in finding terms that share the same root. Single-letter wildcards are not supported, but full-word wildcards are supported in Google.

Weare, W. H., Jr. (2008). Find it on the web using the search concepts you already know. *Library Media Connection 26*(6), 56-58.

Search Concept	*Example*	*Ask*	*Google*	*Live Search*	*Yahoo!*
Phrase Searching	"Des Moines River"	Enclose phrases in quotation marks			
Natural Language	Who won the Oscar for best actress in 1980?	Delivers the answer to simple queries above the search results	User must click on results to find the answer	Delivers the answer to simple queries above the search results	User must click on results to find the answer
Boolean Operators:					
AND	chronic child malnutrition (chronic AND child AND malnutrition)	Keying AND is unnecessary—it's the default operator			
OR	lawyer OR attorney	OR needs to be in upper case between search terms			
NOT	Virginia Woolf – Wikipedia Virginia Woolf NOT Wikipedia	Use "-" before term to be excluded	Use "-" before term to be excluded	Use "-" or NOT (upper case) before term to be excluded	Use "-" or NOT (upper case) before term to be excluded
Nested Search	career AND (mortician OR undertaker OR "funeral director")	Nested search doesn't work (search engine ignores parentheses)	Nested search doesn't work (search engine ignores parentheses)	It works!	Nested search doesn't work (search engine ignores parentheses)
Proximity	whitewater ADJ rafting promotion NEAR tenure online W/2 directory	Proximity operators are not supported			
Truncation	comput*	Truncation is not supported	Uses "stemming technology"—Google will search for words that are similar to the terms entered	Truncation is not supported	Truncation is not supported
Wildcards:					
Single-letter	wom?n	Single-letter wildcards are not supported			
Full-word	"Stopping By * on a Snowy Evening"	Full-word wildcards are not supported	Supports use of full-word wildcard	Full-word wildcards are not supported	Full-word wildcards are not supported

Search Engine Comparison Chart

Weare, W. H., Jr. (2008). Find it on the web using the search concepts you already know. *Library Media Connection 26*(6), 56-58.