Tools Every Searcher Should Know and Use

BY SUZANNE BELL

A finite number of concepts, techniques, and strategies for searching premium content databases and the Web can make all the difference between aimlessly groping around and actually retrieving useful material. If you spend your whole day searching, then you've probably discovered or developed many more techniques tailored to either your searching style or your clientele. However, for most researchers and reference librarians, the tools and techniques discussed in this article will—most likely—be all you'll ever need. I've dubbed this set of concepts the "Searcher's Toolkit."

The seven tools range from fundamentals—Boolean operators, controlled vocabulary, and field searching—to further refinements or functions—proximity searching, truncation, limits, and the pearl-growing search technique. In addition, I believe there are three "mental tools" that are very useful. The intent is always to enable you to get the information you're after more efficiently and effectively.

For experienced searchers, much of the following may seem too basic (good heavens, everyone knows *that*!), but I encourage you to read on anyway. Try looking at your familiar databases with refreshed eyes. With the proliferation of Web search possibilities, you may discover something new!

BOOLEAN LOGIC

The most basic concept of all is Boolean logic. At the most fundamental level, Boolean logic controls how the search system interprets your search terms to produce a set of results. The Boolean operators used in database searching are AND, OR, and AND NOT (frequently expressed simply as NOT). In Web search parlance, AND is the plus sign (+), OR is spelled out in capital letters, and the AND NOT or NOT command is represented by a minus sign or hyphen (-).

It is probably as ingrained as breathing for a trained searcher to recognize that an AND search represents the intersection of two sets, and will always, in practice, return a set that is *smaller* than either of the original sets, while an OR search represents a union of two sets, so in practice, a Boolean OR almost always retrieves more.
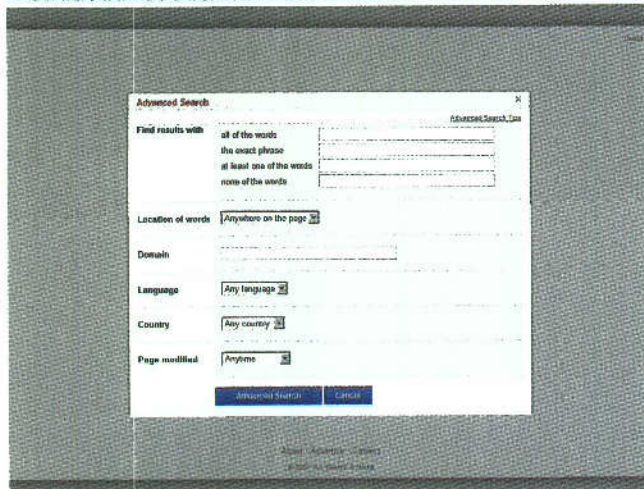
Ask's advanced search page, similar to other Web search engines' advanced search pages, guides you through basic Boolean and some field limitations.

> In my experience, while nonsearchers grasp the Boolean AND concept fairly quickly, it takes much more explaining and persuading to get them to try OR.

It is important to keep in mind that the effect of the Boolean AND and OR operators are contrary to ordinary usage of the words "and" and "or." For a layperson, the regular use of the word "and" is additive—it produces more ("two scoops, and sprinkles, and some whipped cream, and a cherry on top"). I have worked with any number of intelligent end users who have suggested ANDing in another term when the results set was too small, thinking it would help. Then I have to explain that the common "and" isn't like a Boolean AND; the latter retrieves less rather than more.

BOOLEAN OR NOT

A similar situation exists with "or." In common parlance, we usually use the word "or" to mean "either one or the other"—"I'd like the banana *or* the orange." We don't expect to be handed both fruits in response to that request. The Boolean idea that OR means "either the one, or the other, or both" is usually very foreign to nonsearchers. (In my experience, while nonsearchers grasp the Boolean AND concept fairly quickly, it takes much more explaining and persuading to get them to try OR.) Finally, and thankfully, the Boolean NOT means the same thing to both searchers and nonsearchers: It excludes records containing whatever term has been "NOTed" out and should reduce the number of records retrieved.

In ordinary searching, you've probably found that you don't often use the NOT operator in commercial databases. The possibility of missing useful records just because they happen to include the NOTed-out term is too risky. If too many results are coming back, the better strategy is almost always to AND in another term, rather than to NOT out a term. NOT works best when using iterative systems, such as Dialog, where you can NOT out records you've already viewed. The use of NOT in a Web search also seems like a safe bet: In a pool of material that big, I worry less about missing a few potentially useful results.

In the days of command line searching, searchers were probably more aware of how to use these operators, since they were typing them in deliberately. Many current Web interfaces present the Boolean operators as built-in to the search screen, in the form of drop-down menus whose default is almost always AND. This is generally fine, but I sometimes wonder if our OR and NOT skills are atrophying. This is a particular concern for OR, which can be very helpful in low or zero results situations. Having the operators "built-in" like that also tends to make us forget about things like the order of operations that govern Boolean operators and how that can be controlled easily and simply with the use of parentheses. Oddly enough, sometimes the vendors' advanced search screens seem too controlling; I find I receive more accurate results by typing a more advanced Boolean statement into the one text box on the basic screen.

CONTROLLED VOCABULARY

You can essentially assume that any premium content database will support traditional Boolean logic. The next most useful thing to look for is the availability of controlled vocabulary, which could be termed subject headings, a thesaurus, descriptors, or authority control. Use of controlled vocabulary *should* help you get all the database has to offer on a topic, without having to think of all the possible synonyms. It should also ensure that your results are more precise and not littered with false drops, because they disambiguate among words with several meanings. Finally, it provides a "safe" and helpful entry point into an unfamiliar subject area. Even if you know nothing about the subject, you have the assurance that the terms in the subject list are correct and appropriate. By browsing in the list and getting a sense of the terms (especially if there are "see" or "see also" references), you can often get ideas and develop or refine a search strategy. As one who has to delve into unfamiliar subject areas on a regular basis, I am always grateful for databases that offer a subject list of some kind and the ability to browse it.

TOP SEVEN SEARCH TOOLS

1 Boolean Logic
Commands the system to interpret search terms as the searcher wants them to be related to one another

2 Controlled Vocabulary
Terms that encompass multiple synonyms for a search concept

3 Field Searching
Restricts your search to one element of a database, such as author or date

4 Proximity Searching
Increases precision

5 Truncation
Increases recall

6 Limits, or Preset Options
Constrains your search by descriptive elements about the articles you want to retrieve

7 Pearl Growing
Identifies controlled vocabulary to perform the next iteration of a search strategy

The exact terms in the controlled vocabulary will vary from one database to another and from one search service to another. Controlled vocabulary is one of the added extras, one of the contributions of the companies that put together the databases—and one of the reasons they charge a subscription fee to access them. You are paying for a bit of human analysis, or automated taxonomy creation, on each entry in the database. If done well, it is worth the price of admission for the efficiency it provides.

Web search engines lack the controlled vocabulary of the premium content databases. Web page metadata is a very poor, often vilified, substitute.

FIELD SEARCHING

Field searching means the ability to restrict your search to a specific field, or structural element, of the database record. Examples would be author, date, journal name, company name, patent number, descriptor, or industry/ product code. Most databases offer some kind of default set of fields that get searched, so if you're unsure, in a hurry, or just getting a sense of what the database might contain, you can always throw a word or phrase into the first available search box and hit search, just like you usually do on the Web. However, taking a few moments to determine what fields are available for searching can be very valuable. Web search interfaces usually make this

fairly easy to see: Either the available fields are presented in the form of drop-down menus associated with text input boxes or the interface is simply laid out as a series of labeled input boxes, representing the various fields. In databases with very elaborate record structures, the myriad field choices may be listed in a separate screen.

Field searching focuses your search and usually makes it more efficient. For example, say you wanted to search a database of English literature for works by an author named—English! Just searching on the keyword "English" without limiting to the author field would result in hundreds of irrelevant results, because a great many of the records undoubtedly mention the word in a title or abstract, or English might appear in a "language" field that gets included in a "default fields" search. The database designers spent all that time deciding what fields to have. Definitely exploit this feature if it is available.

Information professionals accustomed to the robust field searching offered by premium content vendors are disappointed in the anemic versions of field searching by Web search engines.

Proximity Searching

In addition to Boolean operators, many subscription database systems offer proximity operators that dictate how close to each other, and sometimes in what order, terms must appear in the text to qualify for retrieval. When your search topic falls below the radar of subject headings (controlled vocabulary), and you have some text to work with (at least an abstract, if not full text), proximity searching is a wonderful way to get greater precision in your results. You will still get false drops, but not nearly as many as you would using only Boolean operators. Indeed, the more searchable full text a database has, the more important the ability to do proximity searching becomes, as it is the only useful way to really mine all that text for everything it has to offer.

Proximity searching, other than for exact phrases, is almost universally

lacking in Web search engines, with the exception of Exalead.

Proximity searches in premium content databases are expressed with special operators. Unfortunately, unlike the universal and easily recognized AND, OR, NOT (AND NOT) used for Boolean expressions, proximity operators vary from system to system, so it's harder to produce a nice neat list to memorize. Even the syntax can be mysterious, using simply N (for "near") or W (for "within"), and a number to indicate the number of possible intervening words. Some systems even offer two flavors of proximity operator—one for specifying just proximity, the other for dictating both proximity and word order (termA must occur within so many words and *before* termB). Also unlike Boolean operators, proximity operators usually don't appear as dropdown menu choices in the database interface. While the use of N or near and W or within is fairly standard, the way they are interpreted and the use (or not) of additional characters such as a slash mark make each database just a bit different. Given that there are several subtle variations that might be used, it's best not to just guess. Explore files such as Help or Examples to find out whether a database provides proximity searching and what syntax to use.

Really Close Proximity: Phrase Searching

Searching on exact phrases can be extremely important in some cases, and the inability of some databases to do this (easily) can really inhibit how effectively you can search. In the commercial database world, the way you indicate to the system "this is a phrase search" varies, although there seems to be a trend toward the use of double quotation marks to indicate, "This is a phrase," following the common usage in Web search engines (surprise, surprise).

Phrase searching is useful any time you're searching for things such as the name of a place or an organization (especially if the name is made up of common words), a multiword concept or topic (latch key children, gourmet pickles, missile defense shield), and—especially on the Web—for tracking down more complete information from incomplete fragments. Problems such as the rest of the lyrics, or indeed the real title, of a song whose only line you can remember is "is the moon out tonight?" can quickly be resolved and put in context by plunking the quote-bound phrase into Google.

In a more academic scenario, say you have a bad or incomplete reference to a thesis or a journal article. (My reference desk rule No. 1 is that citations almost always have a mistake in them somewhere.) If available, head for a database such as Dissertation Abstracts or an appropriate subject database, and try a partial phrase from the title or the author's name and a word from the title, depending on what information you (seem) to have. If that doesn't work, try only the most distinct words as a title field search. Still no match? Go to the Web. You may not find the actual item, but you're likely to find someone else's bibliography with a more accurate, complete reference.

TRUNCATION

Truncation is an efficient way of extending and broadening your search to pick up many variations on a word without having to either think of all the possible variants or input them with endless ORs. Truncation allows you to search on a word stem and retrieve any word beginning with those letters,
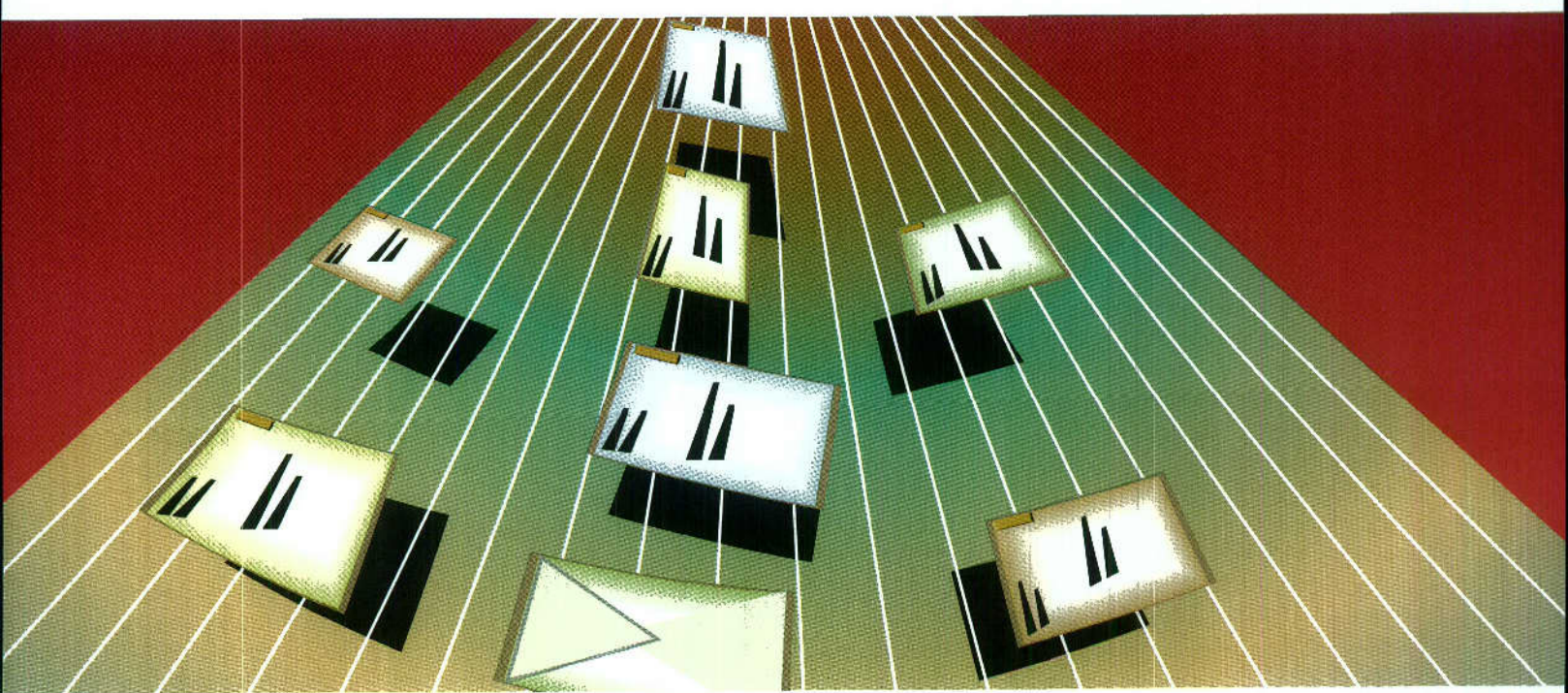
including the word itself (if the stem is a word in its own right). Truncation is another tool largely lacking in general Web search engines.

Truncation is equally useful for both fielded and full-text searches in commercial databases. In a field search, for example, truncation is a wonderfully efficient way to pick up several related subject headings at once (searching `poet*` to pick up poet, poetics, and poetry) or variations on author names (with and without a middle initial, for instance, or even with or without a first name spelled out: `Adams, J !`). In a full-text search, obviously, truncation greatly increases the number of documents that will be eligible for retrieval. When you are fishing around for a concept or topic that you think might be rather rare and that isn't expressed with any set phrases or words, the combination of truncation and a proximity search can be invaluable.

Like proximity operators, truncation symbols vary somewhat from database to database, and there is likely to be nothing in the initial search interface to indicate whether truncation is supported and if so, what symbol to use. There are databases that don't offer a true truncation function but simply search on a limited set of variants, such as plural forms, automatically. Rather than trying to memorize a great deal of arcane knowledge, my advice is always to look for links to "Help" or "Examples" to determine how the database at hand handles truncation.

Wildcards

Closely related to truncation are "wildcard" symbols, in the sense that a symbol is used in place of a letter. Where truncation symbols represent any number of characters, wildcards substitute for characters on a one-to-one basis. A search situation in which single wildcarding is really helpful is for picking up U.S./U.K. alternate spellings, such as labor and labour. If you were searching a database that included British publications and you wanted to be sure you picked up relevant material from them, this use of wildcarding is very important.

Be prepared for confusion: The symbols used for wildcards are the same as those used for truncation, but with the effect changing depending on the vendor. That is, one vendor may use "!" for truncation, and "*" for a wildcard, while another exactly reverses those two meanings. As with truncation, consult the database's Help file for the most up-to-date information on what functionalities the database supports and the appropriate symbols to use.

LIMITS OR PRESET OPTIONS

Limits, or limiters, are preset options in the search interface that can be used to define your search further. They are described here as "preset options" to distinguish them from the free-form text input fields, where you type in terms of your choice. *Limits* make use of fields in the database record that are used to store attributes of the record rather than conceptual content: You could say limit fields are about the article, not what the article is about.

Limiters usually appear as check boxes or drop-down menus and are usually also visually set off from the "text input" part of the search interface, almost always appearing below. Limit choices naturally depend on the content of the database, but they frequently include publication or article type (book, conference proceeding, review, editorial), language, full text (in a database that offers some records with full text and some without, this limit will constrain the search to retrieve only matching records offering full text), and date. If an institution has enabled the technology that links to full text in other databases or identifies library print holdings (as with SFX), limiting by full text just in the current database can eliminate many potentially useful results.

The Date Limit

The date limit is a bit of an anomaly. In the context of a known citation, you would consider it a "content" field—the date would be part of the unique information identifying that citation (Joe Blogg's 1996 article is not Joe Blogg's 1998 article). However, if you are searching for material published *before* or *after* a certain date, or within a particular date range, the date information becomes an attribute used to limit your search results. The way the date option is displayed in the search interface is frequently hybrid as well, offering both

a drop-down of preset choices, as well as fields where you can enter a specific date (or date range).

Some databases let you perform a search using *only* limit values, without any textual input at all. Depending on the database content (and what you are trying to do), this can be very important and useful. Systems requiring some text input when using limits can be frustrating when you are trying to assess and evaluate the database, as when you simply want to see for yourself, "How many records in French does this database contain?" or "How many records bearing the document type 'interview' are there?"

PEARL GROWING

This charming expression refers to the process of doing a very simple search first, with the intent of achieving high recall, and then examining the results to find appropriate subject headings or to discover further terms to search on, from the most on-target hits. You then either replace your original terms or add one or more of these terms to your original search strategy to produce a more precise list of results. You can continue iterating in this fashion until you aren't seeing anything new or compelling, at which point you have probably exhausted the resources of that database. This is very useful when you are venturing into a new database or unfamiliar subject matter or when you simply don't have the time or inclination to do formal preparatory work by hunting around in the subject indexes.

A classic article describing the pearl-growing search technique, along with the building block and successive fractions techniques, is "Online Bibliographic Search Strategy Development," by Robert Wagers and Donald T. Hawkins in *ONLINE*, Vol. 6, No. 3 (1982): pp. 12–19.

YOUR MENTAL TOOLKIT

Understanding and being able to use concepts such as Boolean logic, controlled vocabulary, proximity searching, and limits will definitely go a long way toward making you a more effective searcher. In addition, there are certain mental attitudes that will help you a great deal as well. Some aspects of mind or personality you either have or you don't— general curiosity, interest and enjoyment in puzzles, and an ability to think "out of the box," to make connections or have ideas (light bulb moments) *beyond* the research request as explicitly stated. But there are three "mental tools" you don't have to be born with—rather, you can make a conscious effort to develop them. These are tools to employ in any search, and they may be just as important to your success as a searcher as your knowledge of search functions.

The mental toolkit includes the following:

• A healthy skepticism: Do not trust anything people tell you that "they remember" or even anything that is printed in a bibliography.

• Willingness to let go: A patron may offer a great deal of information, but if you keep getting zero or wrong results, *let go*. Drop pieces of information, one at a time.

• Maintain mental clarity and patience: Be systematic about your searching; don't just thrash around rapidly trying this and that. It may seem longer to stop and think and try one change at a time, but in the end, it saves time.

To add emphasis to the second point: One of the biggest pitfalls in searching is not being willing to *not* look for a part of the information provided. In general: Be flexible, not fixated.

And that's your toolkit. These are some concepts and tools that you probably already use over and over in various combinations. I've codified some attitudinal tools as well. My advice for how to employ this information to the best advantage is fairly simple:

1. Master the concepts.

2. Do not attempt to memorize exactly which databases offer which capabilities.

3. Train your eyes!

Learn to scan an interface quickly. Look for "Help" or "Search Guides." (And actually *read* them, although be prepared: Sometimes the "Help" is not updated as quickly as changes are made to the database.) You now know *what* to look for, so just look for it. Nothing on that screen should be "noise" or ignored. This is the most important thing you can do: *Look* with your trained eyes. Why? Because things can change at any time—and they will change!

Now that database vendors have moved their products to the Web, their interfaces have become more fluid. They are more stable than most Web pages, but the lure and the ability to change things so easily is hard to resist. You need to be flexible, and able to relearn continually as the interface designers move things around and change their terminology. You are probably already used to doing this, but you need to be alert and ready for change ("Oh, that tab is now orange and they've changed the name Advanced Search to More Options. Same thing. OK."). I cannot emphasize this enough: Use your eyes—they are the best tool you have, but do connect them to your brain for validation.

(Adapted from *The Librarian's Guide to Online Searching*, Libraries Unlimited, 2006.)

Suzanne Bell *(sbell@library.rochester.edu) is economics/data librarian, UR research projects coordinator at the University of Rochester.*
Comments? Email letters to the editor to marydee@xmission.com.

COPYRIGHT INFORMATION

TITLE: Tools Every Searcher Should Know and Use
SOURCE: Online 31 no5 S/O 2007

The magazine publisher is the copyright holder of this article and it
is reproduced with permission. Further reproduction of this article in
violation of the copyright is prohibited.