



5-1-2002

Sorting Through Online Systems

Carol Tenopir
University of Tennessee - Knoxville

Follow this and additional works at: https://trace.tennessee.edu/utk_infosciepubs



Part of the [Library and Information Science Commons](#)

Recommended Citation

Tenopir, Carol, "Sorting Through Online Systems" (2002). *School of Information Sciences -- Faculty Publications and Other Works*.
https://trace.tennessee.edu/utk_infosciepubs/425

This Article is brought to you for free and open access by the School of Information Sciences at TRACE: Tennessee Research and Creative Exchange. It has been accepted for inclusion in School of Information Sciences -- Faculty Publications and Other Works by an authorized administrator of TRACE: Tennessee Research and Creative Exchange. For more information, please contact trace@utk.edu.

Sorting Through Online Systems

RECENTLY, A GRADUATING student, hoping to work as a corporate information specialist or academic reference librarian, came to me in a state of near-panic. Her distress went something like this: "Factiva, LexisNexis, Westlaw, Dialog, ProQuest, CSA...they are all running together! I just can't keep track of them!"

This student is specializing in online reference and spends many hours on these systems. So my first reaction was to pity more the poor end user who faces all of these choices with little knowledge, or even the librarian who spends only part of the day online.

Librarians often ask if there's a simple way to keep online systems straight in order to help users and conduct efficient searches. Although there is danger in oversimplifying the complexities and contrasts among the hundreds of online systems and databases to which libraries provide access, all but full-time online searchers should practice a simple coping mechanism. First recognize the similarities rather than the differences (today's commercial online services are more alike than not) and then focus on the few important differences that make each system stand out.

Similarities in interfaces

Ten years ago we thought that the problem of multiple interfaces would be solved with Z39.50 compatibility, which allows varying content to be presented via a single interface, ideally that of the library's online catalog. Although Z39.50 remains a factor, the web and web browsers now dominate the interface scene and have somewhat simplified the problem for users.

Most online services in libraries are



Carol Tenopir
(ctenopir@utk.edu)
is Professor at the
School of Information
Sciences, University
of Tennessee,
Knoxville

web versions and fit within the framework of a web browser. Certainly within that framework there is incredible variation in interface design, from the simple dialog box in FirstSearch Basic to the big-button menus of Easy Search in Web of Science. Most, like Factiva and FirstSearch Advanced, combine templates and dialog boxes. Some, like Ovid, still have lots of specific icons, but I suspect most users just go directly to the search box. If the system is based on the web, the surface correlations outnumber the disparities.

At the basic level, most systems can be expected to share web browser features like the back button. Studies have shown repeatedly that the back button is the most commonly used strategy when searching the web. Most web-based online systems allow users to resort to it if they get lost. It also isn't a bad strategy for librarians checking on how a user got into trouble in a search.

Look for a system that has a simple interface as the default, with more advanced features readily available. The best simple interfaces will allow at least a quick-and-dirty search, which for some purposes may be enough. Save advanced interfaces for when the search requires it or for systems you know.

Similarities in search features

Whether in simple search or advanced mode, online systems for research almost always have certain standard search powers, including Boolean logic, set building, proximity operators, truncation, and record/field structure. Almost all offer Boolean logic searching, particularly combining concepts in an AND relationship as well as specifying synonyms in an OR relationship. Excluding concepts with NOT is less common and also less useful.

You don't need to memorize all of the ways systems can implement AND and OR, just know how to find it. Most systems like SilverPlatter, LexisNexis, and Dialog explicitly use the terms AND and OR, but others hide the function behind templates (typically terms listed across are ORed, terms listed down are

ANDed). They still work the same way. Faced with a blank between words, online catalogs may default to the Boolean AND, but in research systems that provide access to full text in addition to bibliographic records that is less common.

Set building

Set building should be a part of Boolean logic systems—and it is in most—but unfortunately it is not available in all. Early versions of Factiva did not allow sets to be built for combining and recombining. LexisNexis offers rudimentary set building at best. Most systems, however, such as SilverPlatter, Ovid, and CSA, build a set for the terms a user specifies and allow users to recombine the set by referring to it by set number. It is best to check on set building in a new system and, since set building sophistication varies, then choose to enter one concept at a time.

Proximity operators are also commonly available in all but the simplest web search engines or OPACs. Again, the exact way they are implemented varies, but that can usually be found initially fairly quickly in help screens. Don't be ashamed to check help screens, because if you don't search a specific system very often, you'll soon forget its syntax for proximity operators (and standardization isn't likely to happen anytime soon).

Most systems, like LexisNexis, SilverPlatter, and Factiva, default a blank to an adjacency operator; that may be all that is needed for simple searches. The more subtle operators that allow searching within a specified number of words or within a grammatical sentence or paragraph can be found in help screens or advanced help.

The option of truncation or stemming also can be taken for granted, but, again, its implementation cannot be assumed. If you can't remember which systems automatically truncate for plurals or for word form variations, it is always safer to specify truncation explicitly. Only hardcore searchers will remember all of the system truncation symbols—whether you place a colon, ex-

ONLINE DATABASES

clamation point, plus sign, question mark, or asterisk at the end of a stem—but this information is often located on a main menu screen or one click into Help. In FirstSearch Basic and Advanced levels it is hidden, but in the Expert mode it appears directly on the search screen.

Output

Although standard in web search engines, relevance ranking shouldn't be assumed in all research systems. A Boolean search with a default output order of reverse chronological order is the most common, although many systems, such as FirstSearch and Factiva, now offer relevance ranking as a sort option. If this option is selected, the final Boolean-created set is sorted in relevance order. The sorting is done slightly differently in each system, but all consider such things as how often each search word occurs and how many of the concepts are present in the retrieved records. Although some might think so, don't assume that relevance ranking will be the default. It is still most commonly an output feature, not a search feature. The relevance ranking happens after the restrictive Boolean search.

Users will want output formatted in a way they specify, available in PDF, sent to e-mail, integrated into a bibliography, or in other ways. Unfortunately, only the basic output features can be assumed—the bells and whistles vary considerably from system to system. A user is typically at the mercy of the system and its licensing agreements as to whether full text is available and, if so, what form it takes. Users have a choice of PDF or ASCII text for some journal articles in systems like ProQuest or Ingenta, but usually there's less choice. Bibliographic output in many systems, such as CSA and ISI, can be incorporated into bibliographic database software like Reference Manager or directed to an e-mail address. This is becoming more standard.

Structure

The structure of research databases remains more alike than not. Unlike web sites, even web versions of databases retain their standard structure—a collection of records, each of which is made up of fields, each field of which comprises words and/or phrases. It was once safe to go one level higher and say online systems such as FirstSearch or Dialog are made up of many individual

databases (each of which is made up of records, etc., in turn). This is no longer a given, as systems such as ProQuest or LexisNexis Universe seem to appear as one huge conglomerate.

Although the fields available in each database vary, certain core fields are common. Bibliographic and full-text databases typically have author/byline, title, journal name, corporate source, and date fields (and usually descriptor and abstract or full text). Directories usually have company name and a variety of address fields.

Most systems allow users to search

It is natural for users to think of specific databases rather than the online systems behind them

for terms only in those fields or to restrict a search by values within a field. Even most OPACs start the search process by allowing users to specify where they want to search—title words, exact title, subject descriptors, keywords, etc. Field specification is common enough to be taken for granted. Most systems today use check boxes to limit a search to a field or to use a field value to restrict a search.

Often, as with Web of Science Advanced and Factiva, these check boxes are part of the main search screen, allowing the searcher at the start to specify a date range, subset of the system, or language. Fortunately, the searcher rarely has to memorize all the field possibilities in each system, as most systems list them all. If a choice isn't listed (e.g., language), most often that field is not supported in that particular database. (Beware: even if a field is listed, it may not be a part of every record in a database.)

Content differences

Although search features and structure are more alike than different at the gross level, you can ignore variations in content. Some are easy to identify, such as that between Ovid (medical sources) and Westlaw (legal and government documents). Others are more subtle, such as the singularities among

EBSCOhost, ProQuest, and InfoTrac full-text coverage.

Much bibliographic instruction in academic libraries still involves pointing students to the best sources for their topic. It is natural for users to think in terms of specific databases rather than the online systems behind them. Many libraries separate databases from the system that offers them in the library database menu. This matches the way users think—"I need a zoology database," not "I need a database on the CSA system"—but it makes it harder for the librarian who thinks in terms of system features. Although subject menus are probably the best compromise, they do make it more difficult for troubleshooting or help with search strategies.

Full text is another important area in which content deviates. Scholarly texts from systems like Ovid, FirstSearch Electronic Journals, or Ingenta are, of course, quite a departure from business and general texts from Factiva, InfoTrac, or LexisNexis, but undergraduate students often don't recognize the variance.

In the news

A recent study by Cornell University librarian Philip Davis discovered a significant decrease in the frequency of citing scholarly resources in undergraduate term papers in the 1990s ("The Effect of the Web on Undergraduate Citation Behavior: A 2000 Update," *College & Research Libraries*, 63(1), p. 53-60). Scholarly materials are cited less often and newspaper articles (and web sites) are cited more. The newspaper citations can almost certainly be attributed to the widespread availability of full-text newspaper databases.

Although the problem is likely to be solved soon with new standards, we still can't ignore linking distinctions. Without standards fully in place, some bibliographic systems with full-text links take users only to a journal's homepage (thus necessitating another search), while others link directly to the article. Sometimes users get the advertised full text only if the library subscribes to the print or electronic journal. It is still important to explain why a link to full text doesn't always behave as expected.

There are many choices of systems and much complexity in the library's online environment, but if you recognize the similarities when possible, you can put more effort into being mindful of the most important differences.