BY CAROL TENOPIR

Evaluating Online Systems

A SPECIAL LIBRARIAN recently commented to me that the most important skill she was looking for in a person to fill a librarian position is the ability to evaluate online systems. She needs someone who can take the time to find out about all of the likely online systems on the market, compare them by criteria established for her organization, and recommend what systems should be added to her search services. Like many of us, this librarian regularly accesses systems that she feels comfortable searching, but she doesn't have time to study whether others might be as good or even better for the needs of her clientele.

Online systems

The importance of these evaluation skills is pointed out by the January 1988 edition of the Cuadra/Elsevier *Directory of Online Databases*, which lists 555 publicly available online services worldwide. This 555 actually represents a slowdown in growth, according to Cuadra Associates. There was a net growth of only 27 new online systems in 1987, which is "only 36 percent of the growth in 1986 and 29 percent of the growth during 1985." For the information professional, 555 systems is still a lot to cope with. Many of the 555 are not of interest to libraries, yet there are dozens of online systems of potential interest in a library setting.

At the 1988 Australian Online Meeting and 1988 National Online Meeting in New York, Martha Williams (University of Illinois-Urbana) presented an overview of online system use by the information center/library market in the United States. She regularly samples 12 percent of this market (550 institutions) by analyzing their online bills. Libraries and information centers are still concentrating their online use on a few systems.

Williams identified 16 systems libraries used, including DIALOG, Mead Data Central (LEXIS/NEXIS), BRS, STN International, Westlaw, NLM Medlars, Orbit, WILSONLINE, Dow Jones News Retrieval, VU/TEXT, Newsnet, Questel, and The Source. The first seven of these systems account for 98 percent of the use and 97 percent of the expenditures in the library and information center market. DIALOG and Mead Data Central together account for 72 percent of the usage and 80 percent of the expenditures, figures that have held relatively steady since Williams began collecting data in 1982.

Evaluating databases

For a more detailed discussion of selecting an online system, see Susanne M. Humphrey and Biagio John Melloni's *Databases: a Primer for Retrieving Information by Computer*, Prentice-Hall, 1986. The *number* of databases is an important evaluation criterion if you are trying to meet the needs of the greatest number of people with one system. One reason to use DIALOG, for example, may be because its many databases will give you the greatest variety.

Perhaps more important than just the number is the *uniqueness* of the databases offered. Even if a system does not have a large number of databases, if the ones it does have are available nowhere else, the system is valuable.

Neither number nor uniqueness matters unless the *subjects* of the databases are of interest to your users. What subjects are covered by the individual databases on the system and are there systemwide specializations? Orbit has concentrations in patents and petroleum information. BRS has many education databases available nowhere else.

Different systems are better for different *types* of databases. If you need statistical information you may be better off with a system that specializes in numeric information and provides you with numeric manipulation capabilities. I.P. Sharp offers many numeric databases with a command language that will let you exploit the unique properties of statistical information. DIALOG's Business Connection will calculate reports that cannot be done with the regular DIALOG software. Full-text databases also have particular retrieval and display needs ("Searching Full-Text Databases," *LJ*, May 1, p. 60-61).

Time coverage of databases may be outside of the control of the online system, but often depends on the system. The databases on WILSONLINE mostly go back only to the beginning of the system (generally 1983). The information on NEXIS mostly starts with 1981, when NEXIS began. *Timeliness* refers to how often databases on a system are updated and whether or not a system usually meets its update schedule. Updating ranges from almost instantaneous for stock market quotes to yearly or less for directories or encyclopedias.

How systems help you search different databases will be important to searchers who plan to use many databases. Cross-file consistency makes multifile searching much easier. Medlars and WILSONLINE offer a fair amount of consistency and vocabulary control across databases, DIALOG has very little. Multifile search capabilities vary from allowing only a single file at a time to be searched on Medlars, to user-specified file groupings on systems (DIALOG and Mead).

Structure

Structure means how the system processes the records to make them searchable. Most of the powerful online systems create secondary inverted indexes that retain field information and word location information so searchers get fast response times even with complex searches using Boolean and proximity operators. Some systems still have no proximity features, making it necessary to "stringsearch" (sequentially scan) the database for free-text phrases, which is extremely limiting. (Medlars requires you to create a 300 or smaller item subject before stringsearching.)



Carol Tenopir is Assistant Professor at the Graduate School of Library Studies, University of Hawaii, Manoa

Systems such as The Source and Compuserve do not provide inverted indexes for every word in most of their databases, making it impossible to do the kind of free-text searching possible in the other textual systems.

There are other stuctural features to look for that impact search capabilities. Can fields be searched separately or all together as the searcher specifies? Can the inverted index be displayed online and can terms be selected directly from the display? To what fields does a subject search default? (If author is included in the default you may get many false drops.) Can individual words in a multiword descriptor be searched?

Search features

Many of the factors mentioned affect the power of the query language. You might use a prioritized list of desired searching features and match it against a system's capabilities. For textual databases these features should include such things as Boolean operations including nesting, set building, proximity operations, ranging, greater than or less than, truncation, and field specification. For full-text searching, all of the above apply plus searching for words in the same sentence and weighting by number of times search words occur in a document. For numeric databases add statistical manipulation and computations.

Interface/help features

In addition to power, the interface to the query language should be evaluated. A good interface depends on the sophistication and experience of the intended users. For novice or infrequent users is there a menu driven interface? For more frequent users is there a command option? Can either be selected? How easy is the system to learn (from the viewpoint of users)? Are online helps available? Are the error messages meaningful?

Output features

Output refers to online display, online printing or downloading, or to offline printing. Are offline prints and online printing both available? How much flexibility does the user have over the formats? Are there system-supplied formats as well as user-designed formats? Are special output formats available, such as label generation (DIALOG) or card catalog output (WILSONLINE)? For full-text files, can you print only those portions of the documents that contain search terms? For numeric databases can you print in columnar or report formats? Can output easily be transferred to a word processing, database, or spreadsheet program for further refinement? Is offline and online sorting available?

Support

How much does the vendor help and support you? Is there a toll-free number, preferably available as many hours as the system is up? If you call do you get immediate and accurate answers to your questions?

Can you learn to search the systems from the manuals? If your library is expanding the number of online services offered, it is especially important to have good information in the manuals because you may be searching some systems infrequently. Are the manuals regularly updated and is there a fee for this service? Are database summary sheets and other search aids available?

Does the system provide a newsletter on a regular schedule? Is training available in your area? What does it cost? Are any tutorials, computer-assisted instruction, practice files, videotapes, or other learning aids available? NEXIS has a new microcomputer tutorial; DIALOG, WILSONLINE, and BRS have all issued videos in the last year.

Finally, does the system sell any front-end software that makes searching easier? LEXIS/NEXIS, DIALOG, WILSONLINE, Medlars, Dow Jones News Retrieval all have microcomputer software. DIALOG and BRS offer simplified versions of their systems available online.

Reliability/technical matters

What is the response time at peak times and at off-peak hours? Of course the most important response time is when you will be doing the most searching, so average times that include responses in the middle of the night are fairly meaningless. Does the software or hardware of the system impose arbitrary limitations on your searching? For example, Medlars allots limited time slices to each user. In complex operations a user will need to respond many times to a time overflow message. Other systems have limits on the number of word variations that can be retrieved in a search. BRS and WILSONLINE will process only 100 stem variations.

Does the system experience frequent downtime (scheduled or unscheduled)? How will this affect your users and search operations? Are several telecommunications networks supported, so if one goes down you will have an alternate path to the system? What transmission speeds are supported? Are there any special hardware requirements? Mead Data Central used to require their own dedicated terminal; now you can use a microcomputer with Mead's software, but dumb terminals won't work.

Costs

What is the basis of pricing? Is it based on connect hour, information retrieved, or a combination? Are there minimum usage charges, per month or per year? Are subscriptions required or are they available to reduce costs for high-volume users? Does the price formula seem fair and easy to estimate? Have costs held relatively steady over time?

Williams calculated some average per hour costs on the major systems. Overall, across all systems, libraries spent an average of $122 per hour online in 1986. Per hour costs for individual systems in 1982 and 1986 were:

System	1982	1986
Mead	$150	$141
DIALOG	93	126
Medlars	28	39
BRS	49	69
Orbit	109	130
Westlaw	30	100
STN	122	203

Oldies aren't the only goodies

Multiple system use is made easier with Online Inc.'s *Online International Command Chart*, which presents commands for 15 U.S. systems as well as many Canadian and European systems. If you know one system well it is easy to find how you would do a similar function on another system. (Online Inc., 11 Tannery Lane, Weston, CT 06883; 203-227-8466.) In addition, the multiple system gateways such as EasyNet allow anyone to access many online systems without having individual contracts with the systems or without knowing all of the command languages. Even before really good front-end software that knows dozens or hundreds of systems (and even without a universal common command language), libraries today do not need to limit their online use to one or two "old favorites."