



11-1-1988

Database Selection Tools

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, "Database Selection Tools" (1988). *School of Information Sciences -- Faculty Publications and Other Works*.

https://trace.tennessee.edu/utk_infosciepubs/307

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.

ONLINE DATABASES

BY CAROL TENOPIR

Database Selection Tools

WHICH DATABASE should I search to answer this question or for that patron? Is more than one database necessary or appropriate? Which database will give the most unique relevant citations? What databases include a particular journal or several journals of interest? Online searchers face such questions every day. The answer too often ends up being "the database or databases I am accustomed to searching or feel most comfortable with." Studies have shown that online intermediaries search the same handful of databases over and over again. Is it because this handful of databases contains the best and most relevant information or is it because database selection techniques are based mainly on intuition and custom? The answer is most probably some of both.

Although there are no precise answers to any of these questions, there are some tools to help make database selection less reliant on intuition alone: 1) database or periodical directories, 2) online dictionary files, and 3) automatic database selection.

Directories

I reviewed database directories available in print and online in two previous columns (*LJ*, August 1985, p. 64-65 and September 15, 1985, p. 56-57). The most popular of the database directories is still the quarterly *Directory of Online Databases* published by Cuadra/Elsevier. This (or a similar directory) should be available in every library that offers online search services.

A database directory gives basic facts about thousands of databases available on hundreds of online systems. With their subject indexes in

the print versions and free-text searching in the online versions, they provide a broad subject approach to database selection. They are a good first step for locating potential databases in a subject area you may not be familiar with or do not search regularly. They keep searchers aware of new databases and databases on online systems they don't use much.

Another type of directory, probably less familiar to searchers, approaches database selection by periodical title. Two such directories are *Directory of Periodicals Online: Indexed, Abstracted, and Full-Text, News, Law and Business* (3d ed. edited by Maria S. Sims, Federal Document Retrieval, Inc., 1987, \$125) and *Books and Periodicals Online: A Guide to Publication Contents of Business and Legal Databases* (edited by Nuchine S. Nobari, Learned Information, 1987, \$125).

Directory of Periodicals Online

An annual publication, each edition of the *Directory of Periodicals Online* is meant to replace the last. The 1987 edition includes over 7700 periodical titles that are available in over 200 different databases through 125 different online vendors.

The main part of the directory is an alphabetical listing of periodical titles. Brief information is given for each periodical, including: full title, former or alternate title, name of publisher, ISSN, frequency of publication, subjects, name of database(s) that cover the periodical. Cross references are provided from former titles to current titles.

Section 2 provides "Database Online Availability" information. For each database referred to in Part 1, Part 2 gives more information on how to access it online. Information includes: the database name, producer, timespan, update frequency, whether it is a full-text, abstract, or citation database, and the vendors that offer it. Section 3 is a subject index and Section 4 gives addresses for all online vendors that are included in the directory.

To use the directory you would first check Section 1 to see what databases include the periodical you are interested in. Next a check of Section 2 for each of the databases you found will show what online services have the databases. Information in Section 2 varies in detail and completeness, ranging from a listing with timespans of every single publication indexed in a database to just a mention of the online services that carry a database with no further information on individual periodicals. No information is given on how much of a particular periodical is covered by a database.

It is a big task to compile a directory of this sort. The editors explain that complete information was not always available, but sometimes the omissions in Section 1 seem as if they could be taken care of with better editing. Some entries leave out what databases carry the periodicals; in some entries the periodical title heading itself is left off; there are duplicate entries for some titles and sometimes the duplicates list different databases; and subject indexing is inconsistent (*TMA Leaf Bulletin* has the subject heading "Tobacco"; the *TMA Leaf Bulletin Summary* has the subject heading "Tobacco Industry and Trade").

Books and Periodicals Online

Books and Periodicals Online includes over 6800 periodical titles that are available in publicly available source, numeric, bibliographic, or full-text databases. First available in fall 1987, an update included in the purchase price is planned for 1988. (The promise in the book's introduction of twice-a-year updates will not be fulfilled.) Like the *Directory of Periodicals Online*, the information in this volume was verified with database producers. The introductory material describes in detail the fairly rigorous verification process.

The main section of *Books and Periodicals Online* is the alphabetic title index. For each periodical the following information is given: present title, country of publication, for-



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

mer title, publisher, database, producer of the database, scope of coverage in the database, start date of coverage in the database, names of vendors that carry the database.

Each entry is for one journal title and one database. If a journal is covered in more than one database there is a separate entry for each. *Harvard Business Review*, for example, has 12 entries in the alphabetical listing: ABI/INFORM, Accountants, AMI, F&S Indexes, FINIS, HBR/ONLINE (full-text version), INFOBK, Insurance Abstracts, Magazine Index, Management Contents, NEXIS, and Trade and Industry Index.

The disadvantage of this policy is that it adds space to the directory and some information is repeated. The main advantage is that the many differences in coverage among databases can be clearly seen. Inclusion policies of database producers are given under each periodical title.

In addition to the alphabetical listing, there are a publishers' name and address listing, a list of producers and vendors' names and addresses, and a listing of databases including all of the journals they cover and the vendors that make them available.

Online dictionary files

Database and periodical directories provide broad approaches to database selection, but neither helps a searcher decide what database would be best for a specific topic or question. In the mid-1970s Martha E. Williams, director of the Information Retrieval Research Laboratory at the University of Illinois, experimented with a Database Selector that merged the dictionary (inverted) subject indexes of several different bibliographic databases. Searchers would enter their queries in the selector to see which databases had the greatest number of postings on their search terms.

The Database Selector idea was later implemented by several of the major online systems. Orbit's Database Index, Dialog's DIALINDEX, and BRS's file CROSS each contain the inverted indexes of all of the databases on the respective systems. No actual records are in the files, only the terms and posting information that will allow searchers to select databases for searching that have the most postings for a query. Unlike other database selection tools, these online files only cover one system's databases, but for that one system

they allow database selection at the specific query level.

Database Index, DIALINDEX, or CROSS offer a relatively inexpensive way to narrow down a search to the databases that have the most records on the topic. They allow a searcher to prioritize database selection by number of postings and remind searchers to try files that may not otherwise be searched. On most systems, searches can be saved in these files for subsequent running in the databases with the most postings.

Online dictionary files have some limitations. Their help is quantitative only—they provide no information about quality or appropriateness of the databases. They still rely on human judgment to select the best files for a particular patron or query. Overlap is not considered, so a searcher does not know which of the databases that have the most postings on a topic will provide the greatest number of unique hits. Although the costs are modest for these files, there are connect time charges for using them.

Automatic database selection

Automatic database selection is not yet common, but is becoming an important component of gateway systems. Because gateways offer end user searching of databases from several different online systems, the person searching cannot always be expected to know what databases are even available, much less are appropriate. The EasyNet gateway system (called INFOMASTER and Einstein) selects a database for a user after the user chooses from among a series of subject-related and format-related menu choices. EasyNet's database selection is at the broad topic level, not the individual query level, because database selection is made before the searcher enters his or her actual search.

Chengren Hu, a doctoral student at the University of Illinois, recently evaluated automatic database selection on the INFOMASTER version of EasyNet ("An Evaluation of a Gateway System for Automated Online Database Selection," *Proceedings of the Ninth National Online Meeting, New York, May 10-12, 1988*, Learned Information, 1988, p. 107-114). Students at the Graduate School of Library and Information Science used INFOMASTER to search questions given them by Hu. The system selected the database to

search after students selected from among menu choices about the subject of their query. Hu then asked experienced online searchers to select databases for the same queries.

Hu found that "INFOMASTER could select databases as well as human intermediaries when the gateway user properly selected the subject area of a particular query."

The major technique for "automated database selection" by INFOMASTER is through narrowing down the subject selections for a particular query using menu choice by the human searcher, then, selecting a database seemingly at random from among a group of databases for the queries falling in a particular subject field.

The machine does not replace human judgment; instead the searcher's interpretation and then choice of subject area from the menu is the most important factor in database selection by INFOMASTER. For queries where broad subject categorization is not so obvious, INFOMASTER does not perform as well. Although it could be called semi-automatic database selection, the gateway still relies heavily on how human beings judge where a query fits into a subject area. When several databases cover a broad topic, the final choice of one by INFOMASTER seems to be based on some other indefinable criteria programmed into the system. Hu speculates that the other criteria might be economic—certain databases produced by certain database producers seem to be favored by the EasyNet system creators.

Future aids

The limitations of all database selection aids and Hu's research suggest future directions for database selection aids. The best automatic aids must combine general subject help with information relating to the specific query. Machine-aided subject categorization and quantitative word frequency analysis should be combined with such qualitative factors as a human expert's opinions about databases or the ability of the machine to "learn" from its users regarding previous choices.

In the future, expert systems techniques should follow improvements in database selection tools. For now, searchers must rely on their own experience, expertise (and intuition) plus a variety of print and online aids.