Nonbibliographic databases and their use in libraries and information services

Sinikka Koskiala

Helsinki University of Technology


This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.
NONBIBLIOGRAPHIC DATABASES AND THEIR USE IN LIBRARIES
AND INFORMATION SERVICES

Sinikka Koskiala

Helsinki University of Technology Library, Finland

Introduction:

Online searching in bibliographic databases has for many years been a standard feature in many technological libraries and information services. Now that the database vendors are including more and more nonbibliographic databases in their systems, the occasional searching seems to be shifting into regular use of them. To a technical university library heavily into online searching the increase in number and variety of the nonbibliographic databases gives reason to both enthusiasm and concern. It offers more possibilities for direct retrieval of source data rather than just references, and thus, one hopes, improved information access and retrieval. On the other hand the effective utilization of the source databases to the greatest benefit of the end user, whose information needs we want to satisfy, sets us demands which go beyond those encountered in bibliographic database searching. How do we respond to these demands and what will this development require of the technological university library? In the following some of the major concerns will be briefly outlined and discussed from the point of view of the information service.

Types of nonbibliographic or source databases:

In the literature there exists by now several categorisations of the various types of nonbibliographic databases. For the purposes of the discussion in this paper a delineation into three groups, full-text databases, textual-numeric databases and numeric databases, will be used.

Full-text databases include the complete texts of journals, books, legal documents, newspapers, and other such materials. When these texts are available on free-text search systems, where every word of the text is searchable, they are called full-text databases. In the last two years the number of these databases has been growing rapidly. 1983 American Chemical Society put 18 of their scientific journals online, and the commercial publishers are following the suite. 2

Textual-numeric databases include various kinds of directories and handbooks; in general, summary-type information, and facts and figures. Especially in the field of business information there has been an explosive increase in the numbers of these databases. Now the providers of scientific and technical information are expanding in the same direction as witnessed, for instance, by the recent appearance of the online version of the Kirk-Othmer Encyclopedia of Chemical Technology. Many of the tools of the reference librarian are online as well, such as e.g. Ulrichs International Periodicals Directory, GPO Monthly Catalog, and Books in Print, and many more will be soon.
Numeric databases include various kinds of data compilations, such as scientific numeric databases, and survey data files such as time series or cross-sectional data. In general, these databases are often large collections of observational or measurement data on certain phenomena. In addition to a search system which retrieves data computational routines for further manipulation and analysis are needed to make use of the numeric data.

Bibliographic control of electronic data sources:

Though there are good database directories in print form and the Cuadra directory also online, as well as reviews and newsitems in the journals about additions to the various vendors' offerings, the area of nonbibliographic databases is not presently very easy to manage by the information intermediary. The difficulties lie especially in the numeric sector where a large portion of the databases is not offered for public use by the regular online vendors. Though this is the area where the scientists and researchers are most likely to do the searching themselves, knowledge about these databases, their availability, and costs involved in usage is important to the technological university library.

Validity and reliability of the data:

In regard to the source databases the assessment of the validity and reliability of the factual or numeric data is of crucial importance. Errors of course can appear in any database and they always are troublesome. In searching a source database of the textual-numeric or numeric variety though, the knowledge of the origin of the data and the treatment of it by the producer who sets up the database is essential. Attention should be paid to development of methods by which to mark or indicate the degree of reliability that may be associated with a database. 3 When numeric databases become networked or combined for single-source accessing it may well lead to combination of databases of different degrees of reliability or applicability, and also allows easier access by nonspecialist to numerical data that might be used for important decision-making. Data quality indicators can then guide the searcher in appropriate use of the data. Providers of bibliographic databases do not have the degree of responsibility towards their factual content that those providing numeric databases have. 4

Searching of nonbibliographic databases:

An extensive involvement in source database searching would mean some retraining of the information service staff, even those with technical background such as chemical or civil engineering, and experience in bibliographic database searching. Since summary-type databases can be searched similarly to the bibliographic ones, no particular problems exist there, provided that the information intermediary has the subject knowledge to select the relevant facts and figures. Free-text searching of full-text databases requires some new approaches to searching, but since these databases on the commercial systems are generally searched using the same or similar search software features as for bibliographic databases, there are technically only few major changes or additions to cope with. 5 Searching full-text databases though has the aspect of having to make immediate judgement on the relevance of the text to the inquiry, and here the results seem to be noticeably better if the end user is present during the search.

Regarding numeric databases the main problem is the proliferation of different specialized databases, incompatible with one another. Though standardization committees are working towards computer standards, the situation still is that although computerized data in principle are sought more and more, in practice users are frustrated by interface and access problems. 6
Information intermediary may be trained in the twenty ways in which to address a database, but the end user seldom wants to learn that. Recently some gateway systems, such as CIS, Chemical Information System, have been developed to help users gain access in a uniform manner via a single access point to several independent databases. This development of increasingly user-friendly systems is certainly commendable, but the simpler ways to use the system may mean loss in searching power.

In many writings these days the opinion is expressed that the days of the information intermediary are numbered. In a technological university library this will hardly happen. Though the end user will search the main numeric databases in his field directly, and perhaps also the most important full-text journal databases as well, the other relevant databases or occasionally needed ones will be left to the intermediary. We also have to develop the capabilities to search the whole range, because the database that to one sector of engineering is of major importance is marginal to the next. In many cases also there seems to be insufficient use to maintain searching skills, even lack of typing skills may deter some end user from searching.

For the information intermediary the task which is in growing demand is to train and consult the end user as well as to coordinate the information sources and the access to them.

Changing requirements for computer hardware and software:

With source database searching, especially numeric database searching, we often need to send the results electronically to the end user for further manipulation. Many source database systems have also software associated with the databases for transforming data, retrieving and generating reports, modelling, forecasting, performing various statistical operations, etc., but this is usually more expensive way than processing locally on the university computer and using local computational programs. Numeric databases, more than others, often have already sizeable subscription charges, and the statistical analysis of data on the system software, such as e.g. I.P.Sharp's MAGIC, requires a lot of expensive CPU-time.

Downloading data from the source database searches gives us the possibility to reformat the search results, to eliminate duplicates, to include the results within another report, as well as to deliver the results electronically. For this we need intelligent terminals or microcomputers and communications programs with front-end intelligence. The full-text databases appeal to end users who are not interested in specialized vocabulary skills, and do not have experience with existing search systems. Since in this group there is a large number of potential users who are increasing the demand for transparent interface programs, sophisticated file structures and expanded storage capacity, the development towards these changes will be rapid. This will also benefit the information intermediaries engaged in source database searching at the technological university libraries.
References


2. SDI, the Database news section: Dialog signs major agreement with McGraw-Hill. Database, vol. 8 (1985) no. 2, pp. 11.


