

□ ONLINE DATABASES □

BY CAROL TENOPIR

Value-Added Searching

ADDING VALUE to information products and services is not a new concept to librarians. Bibliographers use collection development criteria to find the most valuable materials for their acquisition dollars, catalogers provide valuable finding information to bibliographic records, and reference librarians add the valuable service of instruction or interpretation. In the information industry, *value-added* has become a buzzword, most often applied to databases. It describes something added by human intellectual effort to the basic bibliographic or textual information found in the records.

Database searchers can use the value-added information from database producers (e.g., subject descriptors, codes, classification schemes, abstracts) to refine or improve their search. Another value that is added to database searching is the editing and analyzing searchers do with the search results before presenting them to a patron. The positive points are obvious to professional searchers, but both kinds of value present trade-offs to the database producer and to the searcher. Time, money, and effort must be weighed against the advantages that any value-added process brings.

The value of descriptors

In both bibliographic and full-text databases, the amount of intellectually created information added to records varies from database to database. What to add is a database producer's decision, within the constraints of the online or CD-ROM search software. Most of the major bibliographic databases add controlled vocabulary descriptors. Since controlled vocabulary descriptors are assigned by a human indexer to represent the subject concepts in a docu-

ment, they allow searching for meaning rather than just for a string of characters. If the descriptor PLANT is chosen to mean an industrial building, it will never mean a botanical object in the descriptor field.

Experienced searchers use descriptors to improve the precision of a search, because indexers in most databases assign the most specific term possible to represent the most important concepts in a document. For example, in a medical database an article about the negative psychological impacts of the anabolic steroid nandrolone would be indexed under the term nandrolone, not the broader term anabolic steroids.

In many cases descriptors can also be used to improve recall because they provide a consistent name for a subject instead of the variety of words that different authors may use to mean the same thing. If a descriptor term AIRCRAFT exists, for example, it may be applied for airliners, airplanes, planes, etc. Searchers are relieved of the burden of thinking of all possible words that describe a concept. Arthur Elias, president of BIOSIS, describes the value of descriptors as "standardized handles" that let us grab and hold on to the documents we need.

Descriptors are most often words, but subject codes can be used as well in certain limited fields, e.g., genus/species, type of business, etc. Codes enhance the process and allow cost-effective, precise searching.

Trade-offs with descriptors

Searchers recognize the value of descriptors and codes. The trade-off to searchers for descriptors is the cost of purchasing relevant thesauri or the cost of checking the thesaurus online. For codes, a manual and list of pertinent codes is needed. New editions of both thesauri and code lists must be purchased on occasion and searchers have to keep up-to-date with how they are applied. More search preplanning may be necessary (but almost certainly less time will be spent online than when free-text searching). Since it takes time to in-

dex, we may have to wait to have documents available through the indexed database.

The trade-offs to publishers are more dramatic. Controlled vocabularies are expensive to create, maintain, and keep up-to-date. Many publishers have a person, a department, or a committee charged with the task of reviewing new candidate words for the thesaurus or changes to old ones. In addition, good indexers who can accurately apply the vocabulary to documents must be hired, trained, and retained. Probably no group is more aware of these trade-offs than the National Federation of Abstracting and Information Services (NFAIS), an organization comprising many of the database publishers that provide value-added products. Recognizing the importance of the indexing question, NFAIS offers seminars on what works in indexing, hiring good indexers, and the possibilities of allowing indexers to telecommute.

Classification

Classification schemes are less often used than descriptors, but in combination the two value-added fields offer the best in retrieval. Classification provides a context for ideas, putting similar subjects together. It allows broader and narrower concepts to be searched easily, when, for example, the searcher truncates a classification number for anabolic steroids, that search will also retrieve all specific types and names of anabolic steroids without having to explicitly name them. If on the other hand, only a specific one was wanted, the full classification number for that anabolic steroid could be specified.

Abstracts

Abstracts are offered by a smaller percentage of secondary database producers. The searcher may pay extra to view an abstract, so abstracts therefore have a direct cost trade-off. The benefits, of course, are additional retrieval points in bibliographic databases and a concise summary of an article's contents. If not charged extra, most patrons would prefer a



Carol Tenopir
is Associate
Professor at the
School of Library
and Information
Studies, University
of Hawaii at
Manoa, Honolulu

ONLINE DATABASES

bibliographic search to include abstracts because abstracts make it easier to judge relevance and can sometimes substitute for the original article.

Full-text databases

Even though the value-added process is expensive and time-consuming, a vast majority of bibliographic database producers feel the value of at least indexing is worth the costs. Some companies may even advertise their indexing quality or the presence of other value-added fields as a selling point over their competition. A debate over the value of controlled vocabulary indexing started with the recent proliferation of full-text databases.

At first, full-text database producers hoped that full text would render indexing obsolete. Experience and research has now convinced many this is not true. Controlled vocabulary indexing adds value to full text just as it does to bibliographic databases. The trade-offs here are less clear, however. Are searchers willing to have delayed access to text just so we can have controlled vocabulary handles? Are we willing to pay more for having the option of controlled vocabulary in addition to full-text searching? Are full-text database producers willing to make the commitment to vocabulary development and indexing quality that will improve their products?

What seems to be emerging is a difference in approach depending on the type of information and the needs of the users. For very timely full texts such as wire service stories, speed of access is more important than improved searching. With scholarly journal articles or business magazines, a choice of retrieval options is more important so users can refine the search process according to their needs. Ironically, after spending money on indexing, full-text database producers may see searchers spend less time online in their files because a controlled vocabulary search can usually be done more quickly and efficiently than a free-text one.

Software solutions

One solution to the value-added trade-off that has not been pursued much in commercial online systems is the software solution. Software can be used to help the value-added process at creation and at searching. At the

creation (indexing) end, machine-aided indexing can help indexers create better products. At the searching end, good software solutions might make indexing unnecessary with full text.

Machine-aided indexing

In spite of much research over the years, completely automatic indexing for secondary publications has never worked well. Today's indexing procedures still depend very much on the human indexer. Machine-aided indexing is helping indexers to be more productive and accurate, not replacing them, by providing good software tools to help indexers do their jobs better.

Machine-aided indexing can be defined as using the computer for any part of the indexing process. At the simplest level, that might be just the editing flexibility and speed that good word processing brings, but word processing alone does little to aid the intellectual processes. Word processing tied with a good data entry program offers more to speed productivity including such features as spell-check, validation by type of data (numbers, dates, etc.), and menu prompts that offer a limited set of choices for certain fields such as document type or illustration codes.

Machine-aided indexing can go further. When indexers deal directly with the system or when authority files are a part of the data entry system, such things as feedback about the prior use of a term in the database, automatic checking of the thesaurus, online suggestions of other appropriate terms, and speedy updating of the vocabulary become possible.

Software to replace indexing

Machine-aided indexing is still very labor-intensive. Every article in a database is read, indexed, and reviewed—even if the likelihood of that article ever being retrieved or used is small. Some search software solutions applied to full-text databases may one day make indexing less important, but today's commercial online systems don't do enough. We may begin to see the results of ongoing research, but not very soon on a large scale.

In addition to the standard software features we expect of online systems, software solutions at the searching end must be more "knowledgeable" about words, document structures, and meaning. Simple things like

automatic word equivalency lists are necessary, but do not go far enough. Thesaurus development becomes even more important as a knowledge base for a given document collection. An extensive hierarchically structured thesaurus with many cross references might improve retrieval at the time of searching. When users input their search words the system first checks its thesaurus and matches the words and appropriate levels in the documents.

Even this probably does not go far enough, because of the difference in the meaning of text words according to the context, grammar, and co-occurring words. (PLANT co-occurring with flower and soil probably has a different meaning from PLANT co-occurring with manufacture). Software has a long way to go in recognizing and using the grammatic structure of documents and word patterns.

Value added by searchers

All of the things discussed so far are the value added by database producers or online systems and exploited by searchers. Once a search is completed, the online searcher can add value by post-processing search results. Like any part of the value-added process, there are trade-offs in this as well.

In the old days (five to ten years ago), most online searchers simply did a search, had the results printed out, and passed on the printout to the patron. With faster searching speeds, greater storage capacity on our microcomputer/terminals, and better search software, post-processing is easy to do. Searchers can download results and transfer the results to a word processing or information retrieval software package. Citations can be edited, rearranged, or the results from several searches can be merged. Searchers can add their own value-added information such as local call numbers or interpretation of the significance of a document. Citations can be incorporated into texts and reports can be generated.

Just like the database producers, searchers have decisions to make about how much value should be added to the basic information. The possibilities are there but do we have the time or commitment? Are patrons willing to pay the costs of time and perhaps an increased fee to get additional value? Every time you do a search you weigh these value-added trade-offs.

Copyright of Library Journal is the property of Library Journals, LLC and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.