eralized data base management systems commonly used in business applications. The methods range from structured menu selection (in which the user is offered a series of choices that narrow down the subject) to English-like query languages and sophisticated but restricted language-understanding systems. The special challenge to producers of bibliographic and other textual search systems lies in the fact that the text portion of the documents is written in English or some other versatile but redundant and ambiguous natural language. There is little doubt that the scientist also prefers using the same natural language in expressing search topics of interest. Experience shows that most scientists are unwilling to learn the intricacies and subtleties of access protocols, command languages, Boolean search strategy formulation, and controlled vocabularies.

A citation is unique and unambiguous, but numerical data or analytic statements are dependent on context for their utility. Moreover, if we hope to achieve simplified retrieval systems that are based on the use of natural language, the ambiguities of syntax and grammar will have to be dealt with. Recent work at the NLM (*13*) resulted in a prototype of an English language interface to MEDLINE, TOXLINE, and the Hepatitis Knowledge Base. No special training is required to use this system, named Current Information Transfer in English (CITE). Questions may be posed in English; the software then searches for documents that contain all or most of the key terms in the query. By using a special algorithm based on combinatorial term weights, the system then ranks the retrieved records according to their relevancy. The searcher can select documents of interest and can command the system to expand the search by finding other items that contain terminology similar to the selected citations. Figure 3 shows an actual search on MEDLINE in which CITE was used. Although the system does not perform syntactic and semantic analysis (it cannot "think"), its comparative simplicity and performance offer a genuine potential for vastly increased interactive access to the literature of science by the scientists themselves.

Conclusions

Although this article attempts to identify the many advantages of full use of existing automated data bases, it also points out some of their shortcomings. There are many problems attendant on information retrieval systems in science and technology, and we have not discussed all of them. For example, what criteria should be used to determine whether a datum or statement is a valid entry to a data base? The NLM tends to rely on the published literature. Other systems rely on patient records or expert opinions. But regardless of which is used, there immediately arises a second set of questions: What published literature? Which patient records? Whose expert opinion? These problems become even more complex when information in data bases is updated. As the number of data bases grows, more sophisticated computational methods will be required to ensure complete updating of the many data bases. The criteria need not be the same for all data bases, but they must be clearly defined in each case.

There are those who believe that the increasing amount of scientific and technical research will create a volume of information so large as to frustrate the very purpose for which it was created. If this prediction is not going to become a reality, then a larger percentage of the resources now expended on generating scientific and technical information must clearly be invested in research on how to handle the mass of information being generated.

References and Notes

1. L. Karel, C. J. Austin, M. M. Cummings, *Science* **148**, 766 (1965).
2. D. B. McCarn and R. V. Katter, in *Interactive Bibliographic Search: The User/Computer Interface*, D. E. Walker, Ed. (American Federation of Information Processing Societies, Montvale, N.J., 1971), pp. 121–141.
3. D. B. McCarn and J. Leiter, *Science* **181**, 318 (1973).
4. M. E. Williams and J. H. Rouse, *Computer-Readable Bibliographic Data Bases: A Directory and Data Sourcebook* (American Society for Information Science, Washington, D.C., 1979).
5. *Med. Inf.* **3**, 3 (1978).
6. L. M. Bernstein, E. R. Siegel, W. H. Ford, *Proc. 2nd Annu. Symp. Comput. Appl. Med. Care* (1978), p. 366; A. D. Merritt, E. R. Siegel, C. M. Goldstein, L. M. Bernstein *Proc. 3rd Annu. Symp. Comput. Appl. Med. Care* (1979), p. 752.
7. E. R. Siegel, *Proc. Am. Soc. Inf. Sci.* **16**, 124 (1979).
8. M. E. Williams, *On-line Rev.* **2**, 353 (1978).
9. _____ and J. E. Preece, *Proc. Am. Soc. Inf. Sci.* **14**, 34 (1977).
10. T. E. Doszkocs, *ibid.* **15**, 105 (1978).
11. R. T. Niehoff, *J. Am. Soc. Inf. Sci.* **27**, 3 (1976).
12. R. S. Marcus and F. J. Reintjes, *Experiments and Analysis on a Computer Interface to an Information Retrieval Network* (Rep. LIDS-R-900, Massachusetts Institute of Technology, Cambridge, 1979).
13. T. E. Doszkocs and B. Rapp, *Proc. Am. Soc. Inf. Sci.* **16**, 131 (1979).