

Retrieval from Full-Text Medical Literature: The Dream & The Reality

MaryEllen C. Sievert, Ph.D.*, Emma Jean McKinin, M.A.*,
E. Diane Johnson, M.A.** , Joyce A. Mitchell, Ph.D.***

*Department of Information Science
**J. Otto Lottes Health Sciences Library
***Medical Informatics Group
University of Missouri, Columbia, Missouri

Abstract

While the retrieval of the full-text of a document might seem to end all the hassle of using traditional retrieval systems, the results of the MEDLINE/Full-Text Project indicate that retrieval from the current full-text databases of biomedical journal literature does not match the dream. During the first phase of the project the researchers learned that searching the full-text databases resulted in the retrieval of a significantly larger number of relevant documents than MEDLINE. However, the full-text databases also resulted in a large number of non-relevant documents. Currently the researchers are focusing on how to search these databases to continue retrieving the large number of relevant documents but without so many non-relevant items.

The Dream

Full-text -- the concept is so promising. A researcher or clinician could sit at his/her workstation and browse through the current issue of a medical journal of interest by reading from the computer monitor. Some, who still like to read from paper rather than a screen, could, while sitting at their desks, find an article, give it a quick skimming and print it for serious reading. Either scenario, however, depends on the dream of full-text retrieval.

The dream of full-text retrieval is not just about the ease of having the article available to read without leaving the office. Many who think about full-text dream about using natural language to access literature. No longer would a user need to worry about how something is indexed in MEDLINE. The terms now would be terms used in everyday practice. The thinking is that somewhere in articles of interest the author(s) would use the same term as the researcher/clinician sitting at his/her desk. Or, so the dream goes.

But, what is the reality of full-text retrieval? Currently, in the literature of information retrieval, there is a controversy about the merits of full-text databases. Further, there has been a major study of retrieval of full-text retrieval from clinical medical journal databases currently available from commercial vendors.[1] A summary of the controversy and a discussion of the findings of the MEDLINE/Full-Text Research Project may shed some light on the reality of full-text retrieval in the biomedical literature.

The Full-Text Controversy

Though there are many kinds of databases which are called "full-text" some do not fit the dream described above. For our purposes we will use the definition from the Cuadra/ Elsevier Directory of Online Databases [2] where a full-text database is defined as one which: "contains records of the complete text of an item, e.g. a newspaper article, a specification, a court decision, or a newsletter."

Since the Cranfield experiments in the 1960s, two measures have become the standards for the evaluation of information retrieval tests [3,4]. The first of these measures, recall, measures the ability of a system to retrieve all that it should. It is formally defined as: the number of relevant items retrieved / the total number of relevant items in the database. This measure is not always easy to calculate since it depends on knowledge of the number of relevant items in the database, a figure seldom known when using large, commercial databases. Given this difficulty, several methods have been proposed for estimating the number of relevant items in a database [5].

The second retrieval effectiveness measure is precision. It is formally defined as the number of relevant items retrieved / the total number of items retrieved. Unlike recall, precision can be calculated

whenever at least one relevant item is retrieved in a search.

Since the Cranfield tests the belief has always been that recall and precision are inversely related. That is, when a search results in high recall, it has a correspondingly low precision and vice versa. At least two studies have called this belief into question and have posited that some actions can raise precision without lowering recall [6,7]. Whether recall and precision are indeed inversely related, they are still the most commonly used measures in the evaluation of retrieval.

The controversy surrounding full-text retrieval asks whether full-text retrieval is more likely to result in high recall or high precision. On one side of the controversy are those, like Blair and Maron, who feel that retrieval from full-text databases will result in high precision but amazingly low recall [8,9]. On the other side of the controversy are those who, like Tenopir, find full-text searching results in high recall but corresponding low precision [10]. Confounding the problem is that the studies on full-text retrieval have been done on materials in different disciplines, such as chemistry [11], law [12], or business [13,14]. In addition, the software used to perform the retrieval in these studies has not been consistent. Finally, the kinds of documents in the databases have varied from in-house documents used in support of litigation to journal articles. Undoubtedly, each of these factors contributes to some degree to the different results.

The MEDLINE/Full-Text Research Project

For more than four years a research team at Missouri has been investigating searching the commercially available full-text databases in clinical medicine. MEDIS is available from Mead Data Central and contains two separate full-text journal databases: a current journals file, JNLS, and an archived journal file, ARCJNL. The archived journals have not been updated since 1987. Currently MEDIS contains the full-text of 15 journals being regularly updated and 11 journals in the archived file. BRS Information Technologies offers CCML (the Comprehensive Core Medical Library) both on its command system and on its medical system, Colleague. CCML contains information from a set of medical textbooks, the Year Book series of abstracts and the full-text of journal articles. The journal articles and the Year Book abstracts are available as a separate file, JOUR. Currently there are 67 current journals and 13 closed

journals in the JOUR subset of CCML. As an indication of the size of the file, BRS personnel reported that there were over 225,000 items in JOUR as of July 1990.

During the first two years of the project the researchers compared the retrieval from these full-text journal databases (CCML and MEDIS) with that from MEDLINE. A detailed discussion of the process and results of this part of the project appears elsewhere [15]. A brief summary, however, will show that their findings place the medical literature clearly on one side of the full-text controversy.

Searching the Full-Text Databases

Three searchers performed searches for 100 topics. (For this project, MEDLINE was searched using the ELHILL software from the National Library of Medicine.) For MEDLINE a search using MeSH was always performed. If the searcher felt it appropriate, she could also do a second search which used natural language, either in conjunction with MeSH terms or by itself. Since neither of the two full-text databases was indexed, the searchers constructed natural language searches for these systems.

Constructing a natural language strategy which will retrieve with any reasonable recall is a time-consuming process. Each concept in a request will have to be searched by using all the synonyms, antonyms or near-synonyms the searcher can think of. An example here may clarify the difficulty of creating a natural language strategy. The physician requested information on autologous bone marrow transplants with high dose chemotherapy for breast cancer patients. The request contains three concepts: autologous bone marrow transplant, high dose chemotherapy and breast cancer. Following is a representation of how the first and third concepts might be searched using either MEDLINE or a natural language strategy. (This natural language example is expressed with BRS software in which ADJ indicates that the two terms must be adjacent to each other in the order specified, WITH indicates that the two terms must be in the same sentence, and \$ represents truncation.)

Concept 1:

MEDLINE

transplantation, autologous or bone marrow transplantation

Natural Language
 autologous OR abmt OR abmr OR
 abms OR aubmt OR (bone ADJ
 marrow WITH (autotransplant\$ OR
 purg\$ OR reinfus\$ OR infus\$ OR
 rescue\$ OR harvest\$ OR reconstitut\$))

Concept 3:

MEDLINE
 EXP antineoplastic agents

Natural Language
 high ADJ dose WITH (therapy OR
 treat\$ or manag\$) OR hdc OR ict OR
 chemotherap\$ OR anticancer OR
 antineoplastic OR anti ADJ (cancer
 OR neoplastic) OR cyclophosphamide
 OR thiotepa OR thio ADJ tepa OR
 etoposide OR 4hc OR
 hydroperoxycyclophosphamide, etc.

The EXPLODE command on MEDLINE collects all the drugs used in cancer treatment. With natural language the user must gather all the likely drugs as well as the broader terms such as chemotherapy. When searching with natural language the use of strategies such as these will overcome many of the problems presented by the lack of indexing control.

Results

The comparative effectiveness of the full-text and MEDLINE searches was compared using two measures, precision and comprehensiveness. Comprehensiveness, which is similar to a relative recall measure, points toward recall. It is: the number of items retrieved by a database / the number of items which could have been retrieved by that database, given those retrieved by all three databases. Since the journals covered by the two full-text databases were not the same it was necessary to keep records of which database could have retrieved any item.

The results of this part of the project showed that both MEDIS and CCML retrieved significantly more relevant items than MEDLINE. Further, statistical tests showed that the two full-text databases did not differ significantly from each other but each differed significantly from MEDLINE. Likewise, while the MEDLINE MeSH and the MEDLINE natural language approaches did not differ significantly from each other, the results of both MEDLINE strategies

differed significantly from the full-text files. In retrieval terminology, then, the full-text databases had significantly higher recall than MEDLINE.

On the other hand, MEDLINE, searched strictly with MeSH or with a combination of MeSH and natural language, achieved significantly higher precision than either of the full-text databases. Again, the two full-text databases did not differ from each other statistically. Tables 1 and 2 show the specific figures for comprehensiveness and precision.

Table 1

Comprehensiveness		
System	Mean	Median
CCML	.76	.85
MEDIS	.78	.89
MEDLINE/MeSH	.42	.41
MEDLINE/other	.41	.60

Note. N = 89 Evaluated Searches

Table 2

Precision		
System	Mean	Median
CCML	.37	.33
MEDIS	.37	.36
MEDLINE/MeSH	.55	.53
MEDLINE/other	.62	.67

Note. N = 89 Evaluated Searches

The lack of a statistical difference between the two full-text databases is of interest because the softwares for searching the two systems are markedly different. One system, BRS, has an operator which exploits the structure of written English. Two concepts can be retrieved if they occur in the same paragraph with the operator SAME or even in the same sentence with the operator WITH. Mead Data, on the other hand, does not rely on grammatical structure but allows for numerical word proximity. That is, on Mead one can specify that two concepts must occur within 5 or 50 or even 200 words of each other. The mere fact that the full-text was available seems to have been of more importance in the retrieval than this kind of difference in software.

With this study, then, the same kind of literature was searched on three systems covering this literature but with different software. The results clearly suggest that at least for the biomedical journal literature full-text retrieval will result in high recall and low precision. In layman's terms this means that for every ten documents retrieved from one of the full-text medical journal databases, on the average only three or four will be relevant.

Current Research

Once these results were determined, the researchers turned their attention to a different aspect of the problems of full-text retrieval. Since precision was so low, they attempted to learn if there were any heuristics which could improve precision without a corresponding loss of recall. Ten physicians were part of the research team and, thus, the searchers were able to create fairly broad searches on a topic without undue concern for the time needed to evaluate output containing such a high percentage of irrelevant records. A number of different techniques were tested and results of each technique compared against the broad search, which was treated as the optimum recall set. For each technique, then, it was possible to calculate both a precision figure and a relative recall measure.

The results are preliminary since the testing is still underway. These preliminary results, however, suggest that some of the conventional wisdom about searching full-text databases may not necessarily be true. For example, current wisdom on full-text searching was that one method of improving precision was to use a more restrictive operator [16]. On BRS, therefore, this wisdom, would suggest that if the retrieval has low precision, the operator combining the concepts should be changed from SAME to WITH. The preliminary results from the research indicate that this move is not always the most effective. The relative recall figures are low, but, most surprising, the precision does not improve dramatically. In fact, often the precision is lower than that resulting from the use of SAME. In slightly over half of the topics for which we changed the operator from SAME to WITH, for example, the precision of the set retrieved by WITH was lower than that generated by SAME.

A second finding, supported by the preliminary results, is that the repetition of a concept can be a very effective way to insure precision. That is, it may be more important that the concepts are mentioned at least twice in a document than that they occur in any kind of

proximity (either grammatical or numerical) to each other.

Another common heuristic to improve precision is to restrict the search to certain fields. The title and abstract are the most often mentioned but to search the full-text file for title and abstract is to retrieve no more than the same search could retrieve on MEDLINE. One fact which has emerged from our study so far is that the strategy used to search the entire text will fail to retrieve many items which the title and abstract suggest are relevant [17,18]. Currently we are investigating using the natural language and combining the concepts with AND in a search of the title or abstract. The results show that this strategy will improve the precision of the full-text databases but with a corresponding drop in recall. The research team is also investigating the use of other fields, such as the references, and the combination of fields. It appears that the references by themselves will, for most topics, result in both low recall and only slight improvements in precision but combined with the title will result in very high precision.

The Reality

A look at some of the current research in retrieval from the biomedical literature shows that some are working to create new retrieval programs which might result in better access to both the bibliographic and full-text databases [19,20]. And, there is one aspect of the dream which is currently available, albeit in a limited form. BRS Information Technologies has developed a mechanism whereby those searching MEDLINE or EMED (Excerpta Medica as loaded on that system) will be able to retrieve the full-text of articles for which there is a counterpart on CCML by a simple command. If the full-text article is available, the record in the bibliographic database will carry a note, FULL-TEXT AVAILABLE. The users of the system can then enter the LINK command to retrieve the full-text of the article. This feature is available only on the one system and then only for a limited number of articles, but it is a step toward the dream.

In general, however, retrieval from current operational commercial full-text databases does not fulfill the dream. First, it seems clear that searching the operational commercial systems will result in high recall but low precision. It is likely that the conventional wisdom about retrieval from full-text databases may be faulty in that the number of occurrences of the concepts may be as or more

important than their proximity to each other. Finally, restricting the search to certain fields may help to improve either precision or recall, depending both on the fields and how they are searched. And, so far, the inverse relationship between recall and precision seems to true for the full-text databases.

ACKNOWLEDGEMENT

This publication was supported in part by grant number LM04605 from the National Library of Medicine. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the National Library of Medicine.

REFERENCES

1. McKinin EJ, Sievert ME, Johnson ED, Mitchell JA. The MEDLINE/Full-Text Research Project. Journal of the American Society for Information Science. 1991;42(4):297-307.
2. Directory of Online Databases. Cuadra/Elsevier; 1991:viii.
3. Cleverdon CW. Report on Testing and Analysis of an Investigation into the Comparative Efficiency of Indexing Systems. Cranfield, England: College of Aeronautics; 1962.
4. Cleverdon CW. Factors Determining the Performance of Index Languages. Cranfield, England: College of Aeronautics; 1966.
5. Lancaster FW. Information Retrieval Systems: Characteristics, Testing and Evaluation. New York: John Wiley & Sons; 1979:129-130.
6. Sievert ME, Boyce BR. Hedge Trimming and the Resurrection of the Controlled Vocabulary. Online Review 1983;7:489-494.
7. Saracevic T, Kantor P. A study of information seeking and retrieving. III. Searches, searchers, and overlap. Journal of the American Society for Information Science 1988;39:197-216.
8. Blair DC, Maron ME. An evaluation of retrieval effectiveness for a full-text document-retrieval system. Communications of the ACM 1985;28:289-299.
9. Blair DC, Maron ME. Full-text information retrieval: Further analysis and clarification. Information Processing & Management 1990;26:437-447.
10. Tenopir C. Full text database retrieval performance. Online Review 1985;9:149-164.
11. Abbott JP, Smith CR. Full-text and bibliographic ACS databases: Rivals or companions? In: Sixth National Online Meeting (pp. 5-9) Medford, NJ: Learned Information, Inc; 1985.
12. Gey F, Dabney DP. Letter to the Editor. Journal of the American Society for Information Science December, 1990; 41(8):pp. 613-614.
13. Ro JS. An evaluation of the applicability of ranking algorithms to improving the effectiveness of full-text retrieval. I: On the effectiveness of full-text retrieval. Journal of the American Society for Information Science 1988;39:73-78.
14. Tenopir C. Full text database retrieval performance. Online Review 1985;9:149-164.
15. McKinin, et al., 1991.
16. Tenopir C, Ro JS. Full-Text Databases. New York: Greenwood Press; 1990.
17. Sievert ME, McKinin EJ. Why full-text misses some relevant documents: An analysis of documents not retrieved by CCML or MEDIS. ASIS '89: Proceedings of the 52nd Annual Meeting of the American Society for Information Science 1989;26:34-39.
18. McKinin EJ, Sievert ME. A comparison of full-text and abstracts for information retrieval in clinical medicine. In: Tenth National Online Meeting (pp. 295-301) Medford, NJ: Learned Information; 1989.
19. Hersh WR, Greenes RA. SAPHIRE-An Information Retrieval System Featuring Concept Matching, Automatic Indexing, Probabilistic Retrieval, and Hierarchical Relationships. Computers and Biomedical Research 1990;23:410-425.
20. Hersh WR, Greenes RA. Information Retrieval in Medicine: State of the Art. M.C. Computing 1990;7:302-311.