Complementary use of the SciSearch database for improved biomedical information searching*†

By Cecelia M. Brown, M.L.S., Ph.D.
Assistant Professor
Chemistry-Mathematics Librarian

University of Oklahoma
Norman, Oklahoma 73019

The use of at least two complementary online biomedical databases is generally considered critical for biomedical scientists seeking to keep fully abreast of recent research developments as well as to retrieve the highest number of relevant citations possible. Although the National Library of Medicine's MEDLINE is usually the database of choice, this paper illustrates the benefits of using another database, the Institute for Scientific Information's SciSearch, when conducting a biomedical information search. When a simple query about red wine consumption and coronary artery disease was posed simultaneously in both MEDLINE and SciSearch, a greater number of relevant citations were retrieved through SciSearch. This paper also provides suggestions for carrying out a comprehensive biomedical literature search in a rapid and efficient manner by using SciSearch in conjunction with MEDLINE.

INTRODUCTION

Choosing a database with which to search the vast and expanding body of biomedical literature is not a simple task. Many databases are available, and vary in size, comprehensiveness, and subject coverage, as well as cost. The database of choice for searching the biomedical literature is usually the National Library of Medicine's (NLM) MEDLINE. However, it is generally agreed that it is necessary to search two or more bibliographic databases to achieve a comprehensive online search for biomedical information [1–4].

This paper examines the use of the Institute for Scientific Information's (ISI) SciSearch as a database that complements NLM's MEDLINE in an online literature search for biomedical information. To facilitate the searching and comparison of the databases, all searches were performed with the same platform, the Knight-Ridder Information Service's DIALOG.

SUBJECT COVERAGE IN SCISEARCH VERSUS MEDLINE

SciSearch is an international database that indexes not only the literature of medicine, but also that of science and technology. The print counterpart to SciSearch is *Science Citation Index*. Online searchers of SciSearch also have access to ISI's table of contents index, Current Contents. What sets SciSearch apart from MEDLINE is that SciSearch offers citation indexing and therefore allows searching by cited references.

SciSearch indexes articles, review papers, meeting abstracts, letters, editorials, book reviews, and correction notices from approximately 4,500 major scientific and technical journals. Approximately 3,800 of these journals are further indexed by the references cited within each article, enabling the unique citation-searching feature of SciSearch. Seven hundred additional journals drawn from ISI's Current Contents are also indexed in SciSearch. Approximately 75% of the citations in SciSearch include author-written abstracts. SciSearch indexes publications dealing with medicine and pharmacology, agriculture and food technology, astronomy and astrophysics, engineering, nuclear science, and zoology.

MEDLINE indexes articles from approximately 3,700 journals and includes author-written abstracts for about 69% of these items. MEDLINE is equivalent to a combination of three print indexes: *Index Medicus, Index to Dental Literature,* and *International Nursing Index*. A broad range of biomedical literature is covered, including dentistry, nursing, paramedical professions, communication disorders, population and reproductive biology, and clinical and experimental medicine. Thus, SciSearch and MEDLINE differ in terms of sub-

* This program was supported by the Research Council of the University of Oklahoma.

† A preliminary version of this paper appeared as an abstract in *Tracings* (Spring;10, 1996) after the author received the University of Oklahoma School of Library and Information Studies Alumni Association's 1996 Student Paper Award.

ject coverage. It follows that the use of SciSearch to tap the literature beyond the biomedical disciplines indexed by MEDLINE could prove helpful for many patrons seeking to enhance information retrieval and achieve a comprehensive literature search.

USES FOR SCISEARCH

Citation searching

The unique aspect of SciSearch, which sets it apart from MEDLINE and all other databases except other citation indexes produced by ISI, is the feature that allows retrieval of articles through citation indexing. Citation searching allows the retrieval of articles through subject relationships established by the author's references to prior articles. Citation indexing is particularly valuable when relevant material on a subject is scattered through more than one discipline or buried in the text of a seemingly irrelevant paper [5]. The indexing approach used in typical bibliographic databases may overlook a topic of particular relevance to a patron. Citation indexing enables direct searching by the name of an article's first author; title, year and volume number of the journal; and the first page number of the article. The information is immediately available from the SciSearch full record, which provides the citing article's reference list.

Faculty review

The SciSearch citation retrieval capability can be used not only for research purposes, but also for making personnel decisions. When faculty members are evaluated for tenure or promotion it is common practice to assess the impact of the candidate's work in their particular area of research. SciSearch can be used to ascertain how many times the candidate's work has been cited in the journal literature. Courtois and Matthews [6] suggest that a similar practice could be used during the selection of new faculty members.

When searching for a cited author, the user searches the "author" field rather than the basic index. The name of the researcher is entered by using the "ca=cited author" command without any reference to a specific publication. To make a search more comprehensive, truncation or expansion can be used to retrieve the many possible variations on an author's name. However, the expand command may produce a retrieval too large to provide meaningful results. To avoid this possibility, the search can be refined by using the "cr=cited reference" command rather than the "ca" command. The "cr" command retrieves information about specific cited articles, not just the citations to the author. For example, the search can be refined to pinpoint an article written in a particular year by using the command "s cr=cited author, 1996?".

Collection development

Librarians and information specialists can use SciSearch's citation searching feature to help make decisions about journal subscriptions. The more often a journal is cited, the more prominence it has in a discipline and the greater the demand for it; less-cited journals may be viewed as less important. However, citation analysis is sometimes criticized for bias, errors, and poor coverage of the current year [7], possible drawbacks that need to be kept in mind. Nonetheless, a combination of citation information and a journal's authority, circulation data, demand, and cost effectiveness can be a very useful indicator for collection development. *Science Citation Index Journal Citation Reports* [8] can help in such decisions because it provides the impact factor of each journal as ascertained by citation enumeration.

To learn the number of citations to a journal with SciSearch users can enter the command "s cw=nature", where "cw" means cited work. The use of this command reveals, for example, that the journal *Nature* was cited more than 730,000 times from 1988 to 1995. After taking into consideration other aspects such as cost and authority, librarians might mark less-frequently cited journals for cancellation. It would be unlikely, of course, that titles cited as often as *Nature* would be cancelled.

Care should be taken when searching for journal titles to make collection development decisions. One potential problem is that the "cw" field is phrase-indexed in SciSearch. Therefore, journals such as *Nutrition and Cancer* need to be entered with single quotation marks ('nutrition and cancer') or one of the letters in "and" must be masked. Another problem is a lack of standard abbreviations for use in the "cw" field. The problem can be illustrated by executing a search for *Nutrition and Cancer* with "s cw='nutrition and cancer' ". This search results in no citations. Similarly, the use of the expand command yields no citations. However, a review of the expanded list shows that *Nutrition and Cancer* has indeed been cited from 1988 to the present under various abbreviations: nutri cancer, nutr can, nut cancer, nutr canc, and nutr cancer. Although cancellation of this journal would be highly unlikely because of the authority of *Nutrition and Cancer,* it might have been considered unless the resulting list was carefully perused. In contrast, MEDLINE uses specific journal title abbreviations published by NLM in the *List of Journals Indexed in Index Medicus* [9].

Finding secondary authors through citation searching

Another drawback to citation searching with SciSearch is the failure to retrieve articles when the author used for the search is not the first author of the article. This disadvantage may be illustrated with an article written

by nine authors—J.M. Gaziano, E.J. Johnson, R.M. Russell, J.E. Manson, M.J. Stampher, P.M. Ridker, B. Frei, C.H. Hennekens, and N.I. Krinsky—that appeared in 1995 in the *American Journal of Clinical Nutrition*. When a search is conducted with the "ss cr=russell rm, 1995?" command, no hits are retrieved even though R.M. Russell contributed to the paper. Similarly, no citations are found to M.I. Krinsky in 1995 or to J.E. Manson when restricting the search to *American Journal of Clinical Nutrition*, 1995, as the cited work. Consequently, when seeking to determine the impact of a particular author on a given field of research, the exclusive use of SciSearch could underestimate the author's importance. It would be necessary to use another database such as MEDLINE to retrieve articles by authors who are not listed as the primary author.

Discerning the impact of recent articles

Even though SciSearch is updated weekly, there is a nine-to-twelve-month time lag between the publication of an article and its citation by other authors. Consequently, a searcher can become frustrated and confused when seeking to discern the impact of the most recent work of an author. Aguillo [10] argues that the poor coverage of the current calendar year by citation indexing combined with lost, misspelled, or incorrect citations are serious shortcomings of citation indexing. Although it is not possible to learn the impact of a very recent work, once an article begins to be cited, the weekly additions to SciSearch ensure that the information will be available in a timely manner.

Inaccuracies in citation analysis

Other problems with citation searching are suggested by Moed and Vriens [11], who examined twenty-nine articles cited in SciSearch. Although their sample was small, Moed and Vriens suggested that errors found in SciSearch were largely due to errors in the original text rather than the data entry and processing by ISI.

To test this hypothesis, this author conducted a search in SciSearch to find citations to an article by C.M. Brown published in 1988 in *The Journal of Nutrition*. The following command was entered: "e cr=brown cm, 1988". The following was retrieved: "3 cr=brown cm, 1988, v118, p1294, jnutr". Next, citations to 1989 articles were searched for, even though no articles were published by C.M. Brown in 1989. The command, "cr=brown cm, 1989?" retrieved fifty-three citations. By scrolling through the list of articles that cited C.M. Brown from 1989, this author found an article by Boozer et al. in 1993 on a topic similar to that of the 1988 *The Journal of Nutrition* article. An examination of the list of citations in the article by Boozer et al. provided by the SciSearch full record revealed that C.M. Brown's 1988 article was cited as having been published in 1989. This discrepancy was confirmed by a review of the citations in Boozer's original paper. Clearly this error is not attributable to SciSearch itself but rather to the incorrect citation in the original text. These types of errors can lead to an underestimation of author's influence in a research field and reduce the retrieval of important articles. The complementary use of MEDLINE could help increase the retrieval by finding the correct articles.

Errors can occur not only in the year of publication in SciSearch, but also in the volume and page numbers. For example, an article published by L.R. Krumholz et al. in the *Archives of Microbiology*, 1985, volume 143, page 313, was listed in SciSearch as being cited two times, once correctly ("cr=krumholz lr, 1985, v143, p313, arch microbiol") and once incorrectly ("cr=krumholz lr, 1986, v142, p313, arch microbiol"). Consequently, searchers should not only make parallel use of another database such as MEDLINE, but also use the expand and truncation commands and thoroughly examine the records retrieved, to ensure that all of the citations to a given query are found.

Subject searching

SciSearch uses a natural language index, meaning that keywords are assigned by the authors rather than professional indexers. Therefore, natural language is used when searching by subject. Words in the article title and the abstract are also used to create the basic index. This approach is a striking contrast to the use in MEDLINE of the sophisticated controlled vocabulary Medical Subject Headings (MeSH). The preface of the print version of SciSearch, *Science Citation Index*, states that a natural language system offers many advantages in that if a user recalls the title of a paper or even one or two words of that title, then the entire article may be located [12]. The editors of *Science Citation Index* also believe that the search terms chosen by a scientist are more likely to coincide with those chosen by other researchers in their field than are indexer-assigned, and perhaps outdated, subject headings. Keywords assigned by the author for articles added to SciSearch since January of 1991 are in the basic index field ("Author Keywords"), and can be searched with the suffix "/DE". The basic index in SciSearch also includes the "Keywords Plus" ("/ID") field, which consists of computer-generated, frequently occurring words from the titles of cited references. This feature is also available for records added after January 1991.

Searchers who prefer the use of a controlled vocabulary such as MeSH can take advantage of the "Research Fronts" ("/RF") field available in the basic index of SciSearch. This field brings together clusters of often-cited, closely related papers under one code number. Table 1 illustrates the research fronts displayed when the command "e rf=" is entered. From the list shown in Table 1 the searcher learns that 86-

Table 1
Research fronts displayed in response to the command "e rf="

rf = 86-0001 (pars-plana vitrectomy; proliferative diabeti. . . .)
rf = 86-0001 001 (pars-plana vitrectomy; prolifer)
rf = 86-0001 003 (pars-plana vitrectomy; prolifer)
rf = 86-0001 004 (pars-plana vitrectomy; prolifer)
rf = 86-0001 005 (pars-plana vitrectomy; prolifer)
rf = 86-0002 (cryptosporidium oocysts; outbreak of cryptos. . .)
rf = 86-0001 001 (cryptosporidium oocysts; outbre)
rf = 86-0001 002 (cryptosporidium oocysts; outbre)
rf = 86-0001 003 (cryptosporidium oocysts; outbre)

0001 represents articles about pars-plana vitrectomy in proliferative diabetes whereas the research front 86-0002 encompasses articles about cryptosporidium oocysts. In the future the searcher can search SciSearch by using the code "rf=86-0002" to find articles about cryptosporidium oocysts. This is a convenient time- and money-saving device. Furthermore, the searcher gains confidence that the correct terminology is being used to obtain the high retrieval and precision provided when MeSH terms are used for searching.

COMPLEMENTARY USE OF SCISEARCH AND MEDLINE

To demonstrate the benefits of using both SciSearch and MEDLINE for searching, information was sought about a contemporary topic—red wine consumption and coronary artery disease. Two approaches were taken. First, the basic index of each database was searched individually to determine how many citations were retrieved from each database. Second, the basic indexes of both databases were searched simultaneously. Because DIALOG does not identify duplicates when two or more files are searched simultaneously, the duplicates were removed manually. This process resulted in a set of unique records and eliminated the need to compare the citations resulting from individual database searches.

To search the basic indexes of both SciSearch (1988–present) and MEDLINE (1985–present) the following search string was employed; "coronary(w)artery(w) disease and red(w)wine", where the proximity operator "w" was used to request that the terms be adjacent to each other in the order specified. The search in SciSearch resulted in thirty-five citations, whereas the search in MEDLINE yielded only two citations.

The same search string was used again to search both databases simultaneously. The result was a total of thirty-seven citations. When the duplicates were removed, thirty-five citations remained. The number of duplicates (two) indicated that the search in MEDLINE did not yield any unique citations; close examination of the citations showed this to be the case. Therefore, a MEDLINE search using this specific string of terms would fail to discover thirty-three of the possible thirty-five articles about red wine consumption and coronary artery disease. There may have been some reduction in the retrieval due to the limited flexibility in the index fields searched. To overcome this limitation, fields can be added to the search statement with the "OR" Boolean operator. Although the use of the operator "OR" will result in no hits from the database that does not use the index specified, it will find relevant citations from the other database.

For example, to discover articles that cite an author of one of the articles found about red wine and coronary artery disease, S. Gorinstein, the search string "or cr=gorinstein s" was added to the original query and then searched in both databases simultaneously. This approach resulted in an additional forty-five citations that a search in MEDLINE alone would not have discovered. Similarly, the unique MEDLINE index fields such as the Chemical Abstracts Service (CAS) Registry Number field, "rn", can be tapped by using the "OR" operator. For example, specific information about one of the active ingredients in red wine, quercetin, can be found in MEDLINE by using the "OR" operator followed by quercetin's CAS Registry number, 117-39-5.

Table 2 further illustrates the utility of complementary searching in SciSearch and MEDLINE. It presents the number of citations retrieved from nine different searches made in SciSearch and MEDLINE, then performed in both databases simultaneously and followed by removal of any duplicates. In eight of nine searches SciSearch added to the retrieval from MEDLINE.

Cost-effectiveness of complementary database searching

A hindrance to complementary database searching is the additional cost incurred. MEDLINE is much less costly to search and has been made available free of charge as the PubMed database on the World Wide Web by the National Library of Medicine. When DIALOG is used as the search platform, searches in MEDLINE cost $30.00 per hour, whereas those in SciSearch cost $60.00 per hour for subscribers to the print *Science Citation Index* and $120.00 per hour for nonsubscribers. Display costs also differ significantly. MEDLINE displays cost $0.17 per citation, whereas a display in SciSearch costs $1.70 per citation. However, as shown in Table 2, the cost of searching both databases and then removing the duplicates is not excessive. Yet, when displaying several citations from SciSearch, high costs may be incurred. The searcher should therefore first determine whether there is enhanced retrieval with SciSearch and then decide whether the need for enhanced retrieval justifies the added cost of display from SciSearch.

SUMMARY

This paper demonstrates the utility of SciSearch in searches for biomedical information. When searching

Table 2
Number of citations retrieved and cost of complementary searching with SciSearch and MEDLINE*

Search string	SciSearch		MEDLINE		SciSearch and MEDLINE Simultaneously/duplicates removed	
	No.	Cost	No.	Cost	No.	Cost
1. wom?n(w)physician? or nurse(w)midwi???	4	4.21	84	2.09	87	6.71
2. lipoprotein(w)lipase and weight (w)loss	65	3.17	46	1.80	91	2.45
3. vanishing(w)bile(w)duct(w)syndrome	210	1.54	49	1.19	215	1.21
4. nintendo(n)enuresis	1	1.54	1	1.19	1	1.00
5. californium(w)252	12	1.54	40	0.58	51	0.88
6. garlic(n)longevity	4	1.54	2	1.19	4	1.18
7. fifth?(2)disease	14	1.54	54	1.19	55	1.88
8. asthma and exercise and inhaler(n)use	4	1.54	4	0.58	5	1.00
9. blood(w)stem(2)cell and transplant?	69	1.54	29	1.19	71	1.06

* Costs are estimated costs in U.S. dollars. All searches were limited to articles in English (la = english). Search 1 was limited to articles published in 1997 (py = 1997).

for articles dealing with a simple concept, red wine consumption and coronary artery disease, SciSearch achieved retrieval that was superior to that of the same search in MEDLINE. These results support the findings of Pao, who discovered that citation searching added approximately 24% to the retrieval of a search performed in MEDLINE [13]. Several other searches also exhibit the utility of searching both databases in a complementary fashion to enhance citation retrieval.

Unfortunately, the usefulness of SciSearch is undermined somewhat by the many errors that appear in the records as a result of inaccuracies in the original text. There is also the drawback of the high added cost of displaying SciSearch records, compared to the cost of displaying citations from MEDLINE. SciSearch and MEDLINE are both authoritative, international indexes to the literature of the biomedical sciences. Because MEDLINE focuses on medical literature while SciSearch extends to other scientific disciplines (e.g., agriculture and food technology, astronomy and astrophysics, engineering, nuclear science, and zoology), the complementary use of both databases is recommended when a comprehensive search for biomedical information is required. Further research is needed to identify other databases that provide a useful complement to MEDLINE to further enhance the results of a biomedical information search query.

ACKNOWLEDGMENTS

The author wishes to acknowledge the consultation provided by Michael Havener, M.L.S., Ph.D., associate professor, School of Library and Information Studies, University of Oklahoma, in preparation of this manuscript.

REFERENCES

1. KLEIJNEN J, KNIPSCHILD P. The comprehensiveness of MEDLINE and Embase computer searches. Pharm Weebl Sci 1992 Oct;14(5):316–20.
2. PAO ML. Term and citation retrieval: a field study. Info Proc Man 1993 Jan– Feb;29(1):95–112.
3. VANCAMP AJ. Online sources for alternative medicine information. Database 1993 Oct;16(5):100–3.
4. RAMOS-REMUS C, SUAREZ-ALMAZOR M, DORGAN M, GOMEZ-VARGAS A, RUSSELL AS. Performance of online biomedical databases in rheumatology. J Rheumatol 1995 Oct;21(10): 1912–21.
5. BOORKMAN JA. Electronic bibliographic databases. In: Roper FW, Boorkman JA, eds. Introduction to reference sources in the health sciences. Metuchen NJ: Scarecrow Press 1994:75–95.
6. COURTOIS MP, MATTHEWS KJA. Tips for searching the ISI citation indexes for personnel decisions. Database 1993 June;16(3):60–7.
7. AGUILLO IF. Increasing the between-year stability of the impact factor in the science citation index. Scientometrics 1996 Feb;35(2):279–82.
8. GARFIELD E, ED. Science citation index journal citation reports. Philadelphia: ISI, 1996.
9. NATIONAL LIBRARY OF MEDICINE. List of journals indexed in Index Medicus. Washington, DC: U.S. Department of Health and Human Services, 1991.
10. AGUILLO, op. cit.
11. MOED HF, VRIENS M. Possible inaccuracies occurring in citation and analysis (using SciSearch). J Info Sci 1989;15(2): 95–107.
12. GARFIELD E. How to use the Science Citation Index. In: Garfield E. Ed. Science Citation Index. Philadelphia:ISI, 1996: 24–9.
13. PAO, op. cit.

Received March, 1997; accepted June 1997