

LETTER TO THE EDITOR

SPIRS, WinSPIRS, and OVID: a question of free-text versus thesaurus retrieval?

Background

Dirk Schoonbaert is to be commended for his interesting analysis of the SPIRS, WinSPIRS, and OVID CD-ROM interfaces for MEDLINE [1]. The author makes a very detailed comparison of the three systems, but he also declares that the aim of his study is not to determine which of the systems is superior. To us the author's choice of Silver-Platter products is somewhat confusing. Today the DOS-oriented SPIRS interface is being rapidly replaced by the Windows products WinSPIRS (for PC), and MacSPIRS (for Macintosh). It is hard to see the point of reviewing a product which soon will be used by just a few "mouse-haters." We are also surprised that MacSPIRS is not included in Schoonbaert's investigation. This interface is certainly more widely used than SPIRS. The displays of MacSPIRS and WinSPIRS are quite different. Schoonbaert is of the opinion that the WinSPIRS screen is "rather heavy." We think this product has a very helpful and user-friendly screen for the majority of the users, but this is of course a matter of taste. It should also be added that the WinSPIRS version 1.0 reviewed in the article was replaced by the 2.0 version in 1995.

Schoonbaert's article raises a very important question: the merits or disadvantages of free-text versus controlled vocabulary searching. The MEDLINE thesaurus, MeSH, is an excellent tool for retrieving relevant articles [2]. However, we believe that the reliability of using controlled vocabularies is somewhat exaggerated. Schoonbaert is of the opinion that the use of thesaurus terms (i.e., the MeSH vocabulary) "presents no illusions as free-text searching often does." We

believe that thesaurus terms can present illusions too. The best way to accomplish good results is to combine MeSH and free-text searching [3]. Here we find one advantage of the WinSPIRS system. When a free-text word or phrase is used, WinSPIRS also automatically searches for the term—or part of the term—in the MeSH field. In our opinion, Schoonbaert doesn't explain this MeSH field retrieval clearly. This approach often gives a satisfactory result, but naturally synonyms, suffixes, and different spellings must also be considered.

Discussion

If we search for the concept "myocardial infarction" and "quality of life" with MeSH (by exploding and including all subheadings), seventy-five hits are retrieved for the years 1993–1996:(May). A simple free-text search for the same concepts (myocardial and infarction and quality of life) retrieves 131 postings for the same time period. We also find that all the postings from the MeSH search are included among the 131 free-text hits. Furthermore, several relevant articles are to be found among the free-text postings which were not indexed under the headings "myocardial infarction" and "quality of life."

The OVID search system defaults to controlled vocabularies (i.e., MeSH terms) for MEDLINE. We agree with Schoonbaert, who points out that "the disadvantage of this approach is a lack of flexibility," especially for experienced professionals. But is automatic thesaurus mapping (as in OVID) really the best solution for the inexperienced user? Does only free-text searching present illusions? When a word or phrase is mentioned in the title of the article and when there is a corresponding MeSH term for this concept, the searcher naturally expects to find the most relevant articles when using the MeSH term. Too often this is not

true. Sometimes the lines of thought of the indexers can be difficult to follow for the novice searcher, and MEDLINE—as well as other databases—suffers to a considerable degree from incomplete indexing [4]. When searching for articles dealing with "screening for breast cancer," a free-text search for these words (i.e., breast and cancer and screening) in the title field retrieves 531 postings for the period 1991–1996:(May). If the searcher uses MeSH, the keywords "breast neoplasms" and "mass screening" seem to be correct. However, this is a rather unsuccessful strategy, as 84 of the 531 free-text hits were not retrieved here. The MeSH term "mammography" must also be taken into consideration. When using the keywords "breast neoplasms" and ("mass screening" or "mammography"), the result will be better, but still there are several relevant articles which are not found when using all these keywords. There are numerous examples where the indexing is ambiguous. Choosing the correct MeSH term is often difficult, especially when searching for articles in the "soft" disciplines. When searching for "alcohol problems in the family" there are several seemingly relevant MeSH terms to choose from, i.e., Alcoholism, alcoholic beverages, alcohol drinking, alcohols, family, family therapy, domestic violence, social environment, etc. Many important references here are indexed under the MeSH heading "child of impaired parents." How is the searcher supposed to find or even think of this heading? It must also be difficult for the indexers to choose the correct keywords in such cases.

Sometimes it is very important to search specifically with MeSH terms in order to get satisfactory retrievals. MeSH is no doubt the best alternative when there are adequate "explode" terms with several subordinate terms and when

the MeSH heading does not appear in the subordinate terms, or only in some of them. Exploding "antipsychotic agents" yields more relevant results than text-wording for this concept. Generic drug names are subordinated under the "exploded" MeSH heading "antipsychotic agents." All these generic terms are easily retrieved when using the thesaurus approach. Another danger with using free-text is that synonyms very often must be taken into consideration, especially when exhaustive retrieval is wanted. MeSH uses "neoplasms" for all diseases connected with cancer. When using a free-text approach, the searcher must think of many synonyms, e.g., cancer, carcinoma, tumor, tumour, etc. However, the term "neoplasms" is not frequently used by the authors themselves.

Some librarians and information specialists believe that using MeSH to search MEDLINE is the only reliable way to get satisfactory results. We don't share that opinion. The specific indexing in MEDLINE is an advantage from the specialist's point of view, but the novice searcher may find it difficult to find the correct keyword. Even the indexers themselves may sometimes have problems assigning subject headings [5, 6]. We agree with Gun Soremark's advice to the database suppliers: "Be generous when allocating broader terms to an assigned descriptor" [7].

New terms are also a problem because it sometimes takes a long time for new concepts to be included in the MeSH vocabulary. In the meantime, the searcher must rely

solely on free-text searching. Concepts like "Protein C" and "Low molecular weight heparin" can be found in MEDLINE as far back as 1976 and 1972 respectively, but only if a free-text search is done. The term "protein c" was not included in the MeSH until 1987 and "low molecular weight heparin" was included sixteen years after it had been introduced, i.e., in 1988.

Conclusion

In our experience MEDLINE—as well as most databases—suffers from insufficient and ambiguous descriptor indexing. We believe that successful MEDLINE searching is not a question of using thesaurus or free-text terms. Both approaches must be applied. For optimal results, the searcher should first consult the thesaurus and then look at free-text words in the titles or abstracts. Surprisingly, the searcher will find relevant articles which were not retrieved by using thesaurus terms alone. With this in mind the WinSPIRS default free-text model for simultaneous searching in several fields (primarily the title, abstract, and MeSH fields) offers a simple and comfortable technique for the inexperienced user, as well as for the experienced professionals. As we pointed out before, there are many pitfalls to be taken into consideration. Many of these can be avoided if the users are instructed to consult MeSH to look up terms or concepts, check spelling and synonyms, and explode subject headings as needed. The printed MeSH is perhaps easier to use than the online version. It con-

tains valuable information that should be on every user's desk. If a searcher is aware of the advantages and weaknesses of thesaurus versus free-text searching and has basic knowledge of features like truncation and Boolean operators, the result of the search is likely to be very satisfactory.

Edgar Fremer
Librarian

Birgitta Larsson
Librarian

Medical Library
Malmö University Hospital
S-205 02 Malmö
Sweden

References

1. SCHOONBAERT D. SPIRS, WinSPIRS, and OVID: a comparison of three MEDLINE-on-CD-ROM interfaces. *Bull Med Libr Assoc* 1996 Jan;84(1):63-70.
2. LOWE HJ, BARNETT GO. Understanding and using the Medical Subjects Headings (MeSH) vocabulary to perform literature searches. *JAMA* 1994 Apr 13;271(14):1103-8.
3. FREMER E. Understanding MeSH for literature searches. *JAMA* 1995 Jan 18;273(3):184-5.
4. SOREMARK G. MEDLINE versus EMBASE: comparing search quality. *Database* 1990 Dec;13:66-7.
5. LARSSON B. The indexing systems for MEDLINE and EMBASE: a comparison. Report; Swedish School of Library and Information Science. University College of Borås, 1994. In Swedish.
6. GRETZ M, THOMAS M. Indexierung in biomedizinischen Literaturdatenbanken: eine vergleichende Analyse (Teil 1). *Nachrichten für Dokumentation* 1991;42(5):337-42.
7. SOREMARK, op. cit.