Reviewer's report

Title: A MEDLINE Categorization Algorithm

Version: 1 Date: 13 July 2005

Reviewer: Stuart Nelson

Reviewer's report:

Categorization: Refining Retrieval of Indexed Materials

In a large bibliographic database such as MEDLINE, finding materials of particular interest to a specialty group, or relevant to a particular audience, can be difficult. Indexing of citations is aimed at indicating topics discussed, not at indicating the persons or specialty to whom the citation might be of interest. This information model has the advantage of not requiring that all possible views of why a citation might be of interest be recognized before indexing. The disadvantage is that a search for information of interest in broad categories is not well supported. The use in MeSH of a specialty name as an index term implies that the article cited is about the specialty, not that the article would be of interest to that specialty. For MEDLINE, MeSH is used to index the most specific aspects of citations. In contrast, some other indexing systems intend to capture the broadest aspects of a citation, or possibly even the disciplines involved.

In this paper, Darmoni and associates describe a method for categorizing citations as potentially of interest to particular specialty groups. The method depends on selecting the group, generating semantic links between MeSH descriptors and qualifiers and that particular group, and using the resulting strategy in searches of PubMed. The first parts of the method depend heavily on the use of a senior reference librarian. Subsequent search strategies and categorizing the indexed materials can then be done algorithmically.

The paper would be helped by clarifying the discussion in the Results section. Some of the material in the first paragraph seems more appropriate to the Materials and Methods section. Furthermore, a description of how to use the CISMef MEDLINE Categorization system should be separated from the result of creating one.

In a similar vein, the second bullet in the Discussions section would be improved by stating clearly that what was under discussion as an example was a categorization of the areas of coverage of the BMC Medical Informatics and Decision Making Journal. The reader only discovers this when reviewing Figure 3.

As noted in the discussion, similar categorization attempts are made with the PubMed subsets. I am also personally aware of other similar efforts. What distinguishes this effort is the degree of precision achieved, and that it has covered a number of important areas of specialty within medicine.

I would recommend that this paper be accepted after discretionary revisions. The article describes techniques which may be important to others with closely related interests.

I have no competing interests.

Stuart J. Nelson, MD
Head, Medical Subject Headings
What next?: Accept after discretionary revisions

Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Acceptable

Statistical review: No

Declaration of competing interests:

I have not competing interests