Comment Article Indexing for MEDLINE

Kristen B. Greenland, James G. Mork, Lou S. Knecht, Nancy K. Cox, & Rebecca L. Stanger
National Library of Medicine, Bethesda, MD

Investigation of Comment Indexing
In this study, we accessed the current state of comment indexing, examined the methods by which PubMed users access comment articles, and calculated the overlap of MeSH terms assigned to comment and referent articles. The purpose of this study was to evaluate the efficacy of the current comment indexing policy and to determine the feasibility of automatically indexing comment articles for MEDLINE.

Introduction
The number of published articles requiring indexing increases each year, without a proportionate increase in manpower or funding. This increase in indexed articles is due in part to selection of new titles for MEDLINE indexing. In order to keep up, the indexing process at the National Library of Medicine must become more and more efficient each year.

Use of Comments in PubMed
The number of direct searches for comment articles in March 2011 was determined, as well as the average number of clicks per day on comment links found in the abstract view of PubMed citations. Users rarely search directly for comments (of the ~76 Million total PubMed searches in March, only 52 were direct comment searches), but often click between referent and comment articles using comment links.

Number of MEDLINE Citations Per Year
- Approximately 30,000 comment articles are now indexed each year.
- The number of published articles requiring indexing increases each year.
- In order to keep up, the indexing process at the National Library of Medicine must become more and more efficient each year.

Opportunity for Increased Efficiency
NLM indexes provide high-quality hand-indexing of articles published in MEDLINE selected journals. This includes both original research articles (referent articles) and articles that comment on them (comment articles). Does hand indexing of both referent and comment articles represent a duplication of effort? In the example pair below, nine of the ten MeSH terms assigned to the comment article were also assigned to the referent article.

Example Referent Article/Comment Article Pair: Duplication of Effort?

Use of Comments in PubMed
The number of direct searches for comment articles in March 2011 was determined, as well as the average number of clicks per day on comment links found in the abstract view of PubMed citations. Users rarely search directly for comments (of the ~76 Million total PubMed searches in March, only 52 were direct comment searches), but often click between referent and comment articles using comment links.

Comparison of Assigned MeSH Terms
The percentage of MeSH terms assigned to comment articles that matched terms found on the referent article was determined for all comments published in 2009, 2009, 2010, and the complete MEDLINE comment set. Approximately 70% of terms were matched. For comparison, consistency between indices for Major Topic term assignment is 49-51%.

Non-Matching MeSH Terms
Approximately 30% of terms assigned to comment articles were not exact matches to terms assigned to referent articles (see above). Further analysis of non-matching terms revealed that two thirds were in the same MeSH tree as terms assigned to the referent articles. Of the terms found in the same tree, about half were within two levels of the MeSH hierarchy.

Automatic Indexing Options
Indexing terms applied to the ~30,000 comment articles indexed in 2009 were analyzed and compared to terms assigned to their referent articles, title terms suggested by the Medical Text Indexer (MTI), and a combination of the two sets of terms. Referent article terms had better overlap than MTI suggested terms, and the combination of the two sets had the best percent overlap with comment article terms.

Conclusions
Current comment indexing practices are costly, and based on the findings of this study, approximately 80% of terms assigned by indexers to comment articles are exact or close matches to terms assigned to their referent articles. This represents a duplication of work, and the library has therefore adopted an automatic indexing policy for comment articles. Comments are now indexed by automatically applying Major Topic terms from the referent article, along with select check tags. Implementation occurred on October 6, 2011.