literature.[2] Based on the 234,338 records from the first collection, we retrieved the corresponding PMC IDs (if available) from the PMID to the PMCID converter.[3] With the PMCIDs, we were able to collect full-text articles for a subset of 14,660 records from the first collection. In this second collection, we created two indexes for the 14,660 records: one with article titles, abstracts, and full text, and the other with all of the former fields plus tags. The comparison between these two indexes reveals the impact of tags in the full-text environment.

Collection 3: Tags and author keywords with bibliographic records. Of the 234,338 records in the first collection, 17,264 have author-assigned keywords available from PMC. Using the same method to connect PMID with PMCID, we collected the author keywords for a subset of the first collection and added them to the metadata field. Therefore, in Collection 3, every record consists of the fields from Collection 1 plus author keywords. Similarly, we created different indexes to study the influence of tags and author keywords. The first index includes only titles and abstracts, which serves as a benchmark. The second index includes every field from the first one plus tags. The third index includes everything from the first one plus author keywords. The fourth index includes titles, abstracts, tags, and author keywords. We compared the impact of tags and author keywords in the abstract environment in this collection.

Test Topics

As in the Ninth Text REtrieval Conference (TREC-9) (Voorhees & Harman, 2000), we used the MeSH headings as test topics. Every article in PM has a number of MeSH headings assigned by professional indexers to describe its content. Accordingly, each article is considered to be relevant to its MeSH headings. That is, if we were to search for "Hypertension," a MeSH heading, articles with this MeSH heading are intended to be topically relevant results. Based on this, we randomly selected 50 headings as test topics and collected additional information from their Scope Note, Entry Term, and Previous Indexing fields through the MeSH browser.[4] In the query-construction procedure, each of the authors independently generated her or his own queries according to the information from Scope Note, Entry Term, and Previous indexing. The overlapping query terms (i.e., where the authors agreed) were kept for the experiments. An expert in physicians' information seeking and information management was consulted for any controversial terms (i.e., where the authors disagreed). One query was built for each topic, with phrases preferred when applicable. A phrase is defined as a sequence of words that constitutes one single concept. For example, "Tissue Restoration," "Chromosome Banding," and "Cardiovascular Diseases" are considered to be phrases in the experiments. We did not limit ourselves to any particular retrieval system or query length when constructing the queries.

Phrase queries versus single-word queries. In the process of query construction, phrases were used when the authors considered them applicable. Therefore, the queries consist of phrases (e.g., "Blood vessels") and single words (e.g., "heart"). The phrase search is performed through an ordered window that matches the words in a phrase in the same order as they appear in the document. However, note that most users do not use advanced features such as phrase search in real search (Spink, Wolfram, Jansen, & Saracevic, 2001). It would be of interest to know how the retrieval effectiveness of tags and author keywords varies if only single-word queries are submitted. Therefore, we processed the phrase queries from the manually constructed queries into single-word queries. This process includes breaking down the phrases into their constituent words and removing the duplicate words in a query. For example, "Blood vessels" will become "Blood" and "vessels" in the single-word queries. The difference between the phrase queries and the single-word queries is that the words will be matched separately in single-word queries, whereas the phrase will be matched as a whole in phrase queries. Both phrase queries and single-word queries will be tested in our experiments to provide a comprehensive view. Note that being next to each other does not necessarily make the single words a phrase. In our study, we manually constructed phrase queries to ensure that our phrase searches were appropriate. Figure 2 summarizes the query-construction process.

Relevance Judgments

As mentioned in the previous section, articles are considered to be relevant to a topic if they are assigned the corresponding MeSH heading, so we obtained our relevance judgments through the MeSH headings. The use of MeSH headings in relevance assessments also has been employed in other benchmark environments. The TREC-9 filtering track (Robertson & Hull, 2000) used MeSH headings for relevance judgments. In the TREC Genomics track (Hersh, Cohen, Roberts, & Rekapalli, 2006), expert judges were employed to identify relevant passages, and MeSH headings were assigned to designate their aspects of relevance. However, note that using MeSH headings for relevance judgments is indeed a simplification of realistic tasks and relevance judgments. It has been shown that users' perspectives on relevance are multidimensional constructs that go beyond the content relevance or topicality (Barry, 1994; Greisdorf, 2003; Maglaughlin & Sonnenwald, 2002); however, our first interest in this study is to conduct a controlled experiment that focuses on the impact of tags and author keywords. This simplification helps to control many confounding variables that would be introduced in realistic tasks and relevance judgments (Hjørland, 2010; Saracevic, 1975, 2007).

[2]http://www.ncbi.nlm.nih.gov/pmc/

[3]http://www.ncbi.nlm.nih.gov/sites/pmctopmid

[4]http://www.nlm.nih.gov/mesh/MBrowser.html

DOI: 10.1002/asi