Visualization of Semantic Indexing Similarity over MeSH

Haixia Du, PhD and Terry S. Yoo, PhD
U.S. National Library of Medicine, National Institutes of Health, Bethesda, Maryland, USA

We present an interactive visualization system for the evaluation of indexing results of the MEDLINE database over the Medical Subject Headings (MeSH) structure in a graphical radial-tree layout. It displays indexing similarity measurements with 2D color coding and a 3D height field permitting the evaluation of the automatic Medical Text Indexer (MTI), compared with human indexers.

Radial Tree Layout A radial tree layout [1] is a set of concentric circles with the focusing node as the center and its children lying on different levels around it. For the MeSH hierarchy, the center is the root node linking the 16 MeSH categories together. Its immediate children, i.e., Category A, B, C, etc. form the surrounding smallest circle, and the immediate children of these category roots form the second smallest circle and so on. For terms with multiple appearances, we assign them to branches where they first appear and use additional edges linking them and their other parents. Users can further choose any MeSH term as the center node to view the hierarchical structure with a specified number of levels started from that term. Fig. 1 shows the entire mesh layout with additional edges for multiple-appearance terms.

Fig. 1 The MeSH Radial Tree Layout

Visualizing Indexing Similarity MeSH radial tree layout can be used to evaluate the performance of indexing systems such as the automatic Medical Text Indexer (MTI) compared with human indexers. The input includes indexing results for MEDLINE citations, comprising of a set of PMIDs and corresponding MeSH terms from different indexers, and semantic similarity values [2] by comparing these results. The similarity values can be displayed at either term level or document level.

Evaluation at the Term Level By copying the similarity value for each PMID to its index terms, each term has a similarity value list. Users can select to view statistical measurements on similarity values for each term and adjust thresholds of the high and low bounds of similarity values to further investigate the regions of strong or weak performance. The indexing similarity values can be viewed in either 2D or 3D. For the 2D view, we use color attributes of MeSH nodes to show different similarity values (Fig. 2 (a)). In 3D, the system draws a set of cylinders at MeSH nodes with heights proportional to their similarity values (Fig. 2 (b)).

Fig. 2 Indexing Similarity at the Term Level

Evaluation at the Document Level By highlighting index terms for a given document (Fig. 3 (a)), the system provides a way to directly compare different indexers by examining the term difference in the MeSH structure. Users can also evaluate whether the similarity value is measured appropriately with hierarchical and neighboring information displayed in the graphical layout. The computed similarity scores are displayed at the side panel for users’ reference (Fig. 3 (b)).

Fig. 3 Indexing Results at the Document Level

Conclusion The interactive visualization of indexing results helps users identify the strength and weakness of indexers quickly in a visually convincing way. We plan to incorporate other attributes of citations such as author and publication information for biomedical citation classification and clustering.

References