Larry Mongin, Javed Mostafa, and John Fieber (2001) Using a Spring Embedding Algorithm to Display Term Relationships from a Medical Concept Discovery System. Poster at the 2001 Annual Conference of the American Medical Informatics Association, pp. 661-662.

Using a Spring Embedding Algorithm to Display Term Relationships from a Medical Concept Discovery System

Larry Mongin, MS, Javed Mostafa, Ph.D., John Fieber, MS
School of Library and Information Science, Indiana University, Bloomington, Indiana

Abstract. Visualization can make explicit the subtle and interdependent relationships that exist among terms in a broad topical area. Such visualizations may help in refining retrieval or enhance the user's understanding a complex area. We will demonstrate an application consisting of a term generation service, UMLS categorization of the terms, and a spring embedded visualization of the term set relationships.

Background. The SIFTER (Smart Information Filtering Technology for Electronic Resources) group at Indiana University is experimenting with software methodologies for information extraction based on automated document analysis techniques. Research has been conducted in the area of term and association discovery at IU for several years[1]. Recently, this work was expanded to include external categorization of terms into standard medical concepts by using the UMLS Metathesaurus services of the National Institute of Health[2]. The goal is to apply a visualization technique to display complex networks of relationships that may exist among different medical concepts.

System. Terms are discovered using an algorithm based on an automated thesaurus generation procedure, and subsequently associations among terms are established using a method known as Maxi-min distance clustering[3]. After terms are discovered, and before cluster analysis, a Java servlet queries the Knowledge Source Server of UMLS to identify an appropriate broad category (a single concept) for each term in the term set. The user is presented with a table consisting of the discovered terms and the broad UMLS category if one exists for each term. Users then can select individual terms for cluster analysis and visualization.

The clustering algorithm computes distance between weighted vectors using a cosine similarity measure. The output of the cluster analysis includes term centroids of clusters, related terms for each cluster, and distances between terms. The Java applet, shown in figure 1, displays the results of the cluster analysis by using a spring embedding algorithm. The spring embedding algorithm[4] uses a heuristic whereby edges are springs and nodes are rings. The system oscillates and then settles at a place of minimum force. Based on the similarity data from the cluster analysis, the stable configuration roughly represents term similarity in a 2-D visual representation.

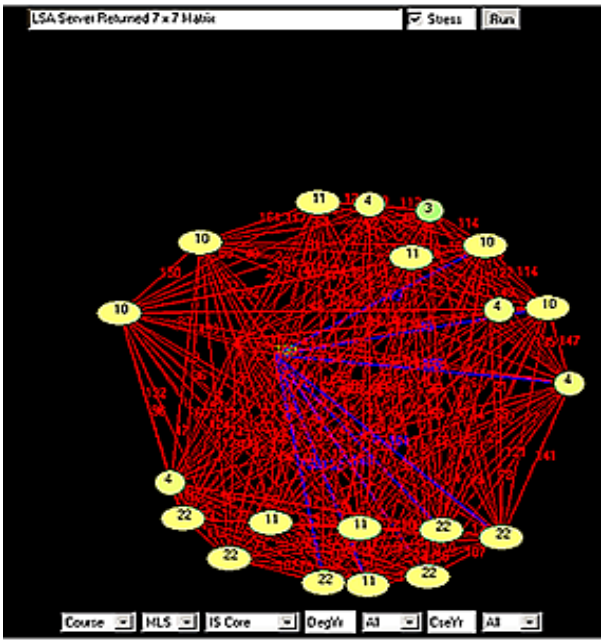

Fig. 1. A display produced by a Java Applet implementation of the spring embedding algorithm.

Acknowledgement

This project was partially supported through the NSF ITR Grant 0081944. We are also grateful to Katy Borner and Yuezheng Zhou for the enhancement they made to the spring embedding applet originally developed at Sun Microsystems.

References

1. SIFTER. http:// sifter.indiana.edu
2. NIH. Unified Medical Language System. D.C. 2000.
3. Mostafa, J. Quiroga, LM, Palakal, M. Filtering medical documents using automated and human classification methods. J American Soc of Info Sci 1998; 49:1304-1318.
4. Sun Microsystems. Graph.java. California.1996.