Using a Similarity Measurement to Partition a Vocabulary of Medical Concepts

Huanying (Helen) Gu[1], James Geller[1], Li-min Liu[1], Michael Halper[2]

[1]CIS Dept., New Jersey Institute of Technology, Newark, NJ 07102 USA
{helen,geller,limin}@homer.njit.edu
[2]Math & Computer Science Dept., Kean University, Union, NJ 07083 USA
mhalper@turbo.kean.edu

Abstract

Controlled medical vocabularies have become increasingly important in a range of medical informatics applications. However, the extensive size of most vocabularies often makes it difficult for users to gain an understanding of their contents. In previous work, we have investigated the partitioning of a large semantic-network based medical vocabulary into smaller units, for the purpose of easier graphical display and comprehension. The partitioning process relied heavily on a domain expert. In this paper, we propose a structural method for automating the partitioning of a vocabulary. The structural method is based on a definition of the similarity of a pair consisting of a child concept and its parent concept in the semantic network. A distribution over these similarities for all pairs in the semantic network is then computed. Based on this distribution, the semantic network can be partitioned into more manageable pieces. The approach has been applied to the InterMED and a complex portion of the MED, two large medical vocabularies.

H. Gu, J. Geller, M. Halper, L. Liu. Using a similarity measurement to partition a vocabulary of medical concepts, in T. Bench-Capon, G. Soda, and A. M. Tjoa, editors, *Proceedings of the 10th International Conference on Database and Expert Systems Applications: DEXA'99*, Florence, Italy, August 1999. Published as: Lecture Notes in Computer Science, 1677, Springer-Verlag, New York, pages 712-723. (SCI)