

Automated Taxonomy Generation for Summarizing Multi-type Relational Datasets

Tao Li and Sarabjot S. Anand

May 27, 2008

Taxonomy construction provides an efficient navigating and browsing mechanism to people by organizing large amounts of information into a small number of hierarchical clusters. Compared with manually editing taxonomies, Automated Taxonomy Generation has numerous advantages and has therefore been applied to categorize document collections. However, the utility of this technique to organize and represent relational datasets has not been investigated, because of its unaffordable computational complexity. In this paper we propose a new ATG method based on the relational clustering framework *DIVA*. By incorporating the idea of Representative Objects, the computational complexity can be greatly reduced. Moreover, we analyze the divergence of the data attributes and label the taxonomic nodes accordingly. The quality of the derived taxonomy is quantitatively evaluated by a synthesized criterion that considers both the intra-node homogeneity and inter-node heterogeneity. Theoretical analysis and experimental results prove that our approach is comparably effective and more efficient than other ATG algorithms.