Taxo Finder: A Graph-Based Approach for Taxonomy Learning

Taxonomy learning is an important task for knowledge acquisition, sharing, and classification as well as application development and utilization in various domains. To reduce human effort to build taxonomy from scratch and improve the quality of the learned taxonomy, we propose a new taxonomy learning approach, named Taxo Finder. Taxo Finder takes three steps to automatically build taxonomy. First, it identifies domain-specific concepts from a domain text corpus. Second, it builds a graph representing how such concepts are associated together based on their co-occurrences. As the key method in Taxo Finder, we propose a method for measuring associative strengths among the concepts, which quantify how strongly they are associated in the graph, using similarities between sentences and spatial distances between sentences. Lastly, Taxo Finder induces taxonomy from the graph using a graph analytic algorithm. Taxo Finder aims to build taxonomy in such a way that it maximizes the overall associative strengths among the concepts in the graph to build taxonomy. We evaluate Taxo Finder using gold-standard evaluation on three different domains: emergency management for mass gatherings, autism research, and disease domains. In our evaluation, we compare Taxo Finder with a state-of-the-art subsumption method and show that TaxoFinder is an effective approach significantly outperforming the subsumption method.