

Hierarchical Text Categorization Using Fuzzy Relational Thesaurus.

Domonkos Tikk; Jae Dong Yang; Sun Lee Bang

Abstract: Text categorization is the classification to assign a text document to an appropriate category in a predefined set of categories. We present a new approach for the text categorization by means of Fuzzy Relational Thesaurus (FRT). FRT is a multilevel category system that stores and maintains adaptive local dictionary for each category. The goal of our approach is twofold; to develop a reliable text categorization method on a certain subject domain, and to expand the initial FRT by automatically added terms, thereby obtaining an incrementally defined knowledge base of the domain. We implemented the categorization algorithm and compared it with some other hierarchical classifiers. Experimental results have been shown that our algorithm outperforms its rivals on all document corpora investigated.

Keywords: text mining; knowledge base management; multi-level categorization; hierarchical text categorization;

AMS Subject Classification: 68W99; 62P30;