Abstract

Ontologies currently are a hot topic in the areas of Semantic Web. The current clustering research emphasizes the development of a more efficient clustering method and mainly focuses on term weight calculation without considering the domain knowledge. This paper investigates how ontologies can also be applied to the clustering process. To complement the traditional
clustering method, more informative features including concept weight are important based on recent developments in the area of the Semantic technologies. The proposed system presents the concept weight for text clustering system developed based on a k-means algorithm in accordance with the principles of ontology so that the important of words of a cluster can be identified by the weighted values. To a certain extent, it has resolved the semantic progeny in specific areas. The experimental results performed using dissertations papers from Google Search Engine and the proposed method demonstrated its effectiveness and practical value.

Reference

- A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities by TIM BERNERS-LEE, JAMES HENDLER and ORA LASSILA
- A. Hotho and S. Staab, "Ontology based Text clustering".
- http://www.textfixer.com/resources/common-english-words.txt
- Stefan Brueggemann,Using Domain Knowledge Provided by Ontologies for Improving Data Quality Management.

Index Terms

Computer Science Semantic Web

Key words

Clustering Concept Weight Document

clustering

Feature Selection

Ontology