A Combined Weighting for the Feature-Based Method on Topological Parameters in Semantic Taxonomy Using Social Media

Ali Muttaleb Hasan 1, Noorhuzaimi Mohd Noor 2, Taha Hussein Rassem 3*, Ahmed Muttaleb Hasan 4, Waleed A.Hammood 5

Faculty of Computing University Malaysia Pahang,
26300 Gambang, Kuantan, Pahang DarulMakmur, Malaysia
alimatlab65@yahoo.com nhuzaimi@ump.edu.my tahahussein@ump.edu.my
ahmed.matlab11@gmail.com engwaleed54@yahoo.com

Abstract:

The textual analysis has become most important task due to the rapid increase of the number of texts that have been continuously generated in several forms such as posts and chats in social media, emails, articles, and news. The management of these texts requires efficient and effective methods, which can handle the linguistic issues that come from the complexity of natural languages. In recent years, the exploitation of semantic features from the lexical sources has been widely investigated by researchers to deal with the issues of "synonymy and ambiguity" in the tasks involved in the Social Media like document clustering. The main challenges of exploiting the lexical knowledge sources such as 1WordNet 3.1 in these tasks are how to integrate the various types of semantic relations for capturing additional semantic evidence, and how to settle the high dimensionality of current semantic representing approaches. In this paper, the proposed weighting of features for a new semantic feature-based method as which combined four things as which is "Synonymy, Hypernym, non-taxonomy, and Glosses". Therefore, this research proposes a new knowledge-based semantic representation approach for text mining, which can handle the linguistic issues as well as the high dimensionality issue. Thus, the proposed approach consists of two main components: a feature-based method for incorporating the relations in the lexical sources, and a topic-based reduction method to overcome the high dimensionality issue. The proposed method approach will evaluated using WordNet 3.1 in the text clustering and text classification.

Keywords: Component; Text Mining; Text Classification; Sentiment Analysis; Semantic Representation; Weighting; Topological Parameter; Non-Taxonomy

Acknowledgement