

ABSTRACT

In natural language processing (NLP), a semantic similarity measure has an important role. The measurement results are often used as the basis to perform advance natural language processing tasks. The methods used to measure the semantic similarity one including: vector-based method, path-based method, information content-based methods, methods gloss-based, and a hybrid method. In this thesis, the information content-based method is implemented in the semantic similarity measure to determine the accuracy of the similarity between the words. Content of information used in implementation is come from WordNet. The implementation results are compared to the dataset Gold Standard for measuring the correlation. The dataset used for testing the dataset 353 WordSim Similarity, WordSim 353 Relatedness, and SimLex-999. Based on the testing performed, Pearson and Spearman Correlation Lin's method to noun and verb SimLex-999 gold standard respectively for 0.532 and 0.543. Intrinsic information content method represented by Lingling Meng's method produces result Pearson and Spearman Correlation respectively for 0.475 and 0.551. Basic information content method that uses a combination WordNet corpus with the structure and method of the intrinsic information content produce a performance strong positive.

Keywords: Natural Language Processing, Semantic Similarity, Content-Based Method of Information, WordNet, Gold Standard