Abstract

Semantic similarity is one of the text mining's measurement to find the value of the similarity between word's meaning. This semantics similarity can be applied in various applications. The measurement's background is caused where the computer not able yet to equate human's perspective related to measurement of the similarity between words. Therefore, this thesis will discuss about semantics similarity between words in Bahasa Indonesia by using GloVe method. GloVe method is a model for unsupervised learning on words representation that surpass another models in word analogy, word similarity and named entity recognition. With the input of a Wikipedia corpus of Bahasa Indonesia and the correlation value from resulted score is calculated with correlation pearson by comparing it with gold standard score from WordSim-353, SimLex-999, and Miller Charles. The final result from this thesis produce a correlation value in gold standard with the obtained correlation value is 0.1165 for Miller Charles, 0.2280 for SimLex-999 and 0.2849 for WordSim-353.

Keywords: Text mining, Semantics Similarity, GloVe