

Abstract

Natural Language Processing (NLP) is one of science disciplines which focusing to generate knowledge from human written text which is not structured. The measuring of Semantic Similarity to word pair is one of the task in Natural Language Processing that the main idea is to find a semantic similarity score to word pair. This score is representing how similar the word pair is. One of methods for measuring semantic similarity is Pointwise Mutual Information (PMI).

Semantic similiarity is one type of measurement in the text mining to describe how the relationship between words. The purpose of this semantic association measurement is to obtain a value that represents how much the association. Pointwise Mutual Information (PMI) is a statistical measurement of the semantic relationship that has been widely used. Pointwise Mutual Information (PMI) is one variant used in this thesis to calculate semantic similiarity. PMI value calculation performed on datasets obtained from wikipedia Indonesia language. PMI value obtained by the application of computed correlation. The results of the research in this thesis is the correlation between the scores generated by the gold standard system SimLex-999, WordSim-353 and Miller and Charles will resulting correlation value that would show how accurate the measurement method PMI. The highest correlation value is 0,256 by using window size = 25.

Keywords: *Semantic, Text Mining, Pointwise Mutual Information.*