Abstract

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. Application of PMI known to have a bias for the pair of words with low frequency, this led to the development of a variant form of the PMI measurements to avoid circumstances such bias. Positive pointwise Mutual Information Cosine (PPMIC) is one variant used in this thesis to calculate semantic similiarity. PPMIC value calculation performed on datasets obtained from brown corpus. PPMIC value obtained by the application of computed correlation with Word-Sim-353 which is an index of words based on the similiarity of human standpoint. The results of the research in this thesis is the correlation between the scores generated by the gold standard system SimLex-999, WordSim353 and Miller and Charles will resulting correlation value that would show how accurate the measurement method PPMIC.

Keywords*: Semantic Similiarity, Pointwise Mutual Information, Positive pointwise Mutual Information Cosine.*