Abstract

Searching for information is part of people's needs, specially in using colloquial. For example Bahasa. In searchingforinformationeffectively,humanandcomputersneedtohavethesameknowledgeinprocessing the information. People can easily get the information, because people know how to process the word they need. They have knowledge about how one word relates to another words, but computers can't do that because computers don't know any sense of the words. Therefore, computers need to find similarity value for each words. Based on the idea, similarity's method that is choosen for calculating semantic similarity value between two words in Bahasa is PMImax that is a derivative from PMI method. This method was chosen because this method can give similarity value based on the words co-occurrence in a corpus. This method also gave a good result in English words. This study examines if this method can be implemented in Bahasa for calculating similarity value, and also examines how good this method in the implemantation. Using pearson correlation, the result of this study is PMImax gave good results when it is implemented in Bahasa compared to PMI and Word2Vec method. The correlation's scores are 0,26 in Miller and Charles, 0,33inSimLex-999,0,52inWordSim-353Similarity.