Abstract

The implementation of semantic equality measurement has a very important role in some areas of NLP, where the results are often used as the basis for performing further NLP tasks. One of its application is by doing the measurement of multilingual semantic similarities between words. This measurement is motivated by a problem where many information search systems now have to deal with text or multilingual documents. For this time, as far as writer know, there is still a lack of a computer system that can measure the value of semantic similarity between words in two languages. A pair of words are said to have a semantic similarity if the word that pair has a similarity meaning or concept. There are many methods that can be used to measure semantic similarity between words, one of then is statistical measurement using Pointwise Mutual Information (PMI).In this research, the calculation of semantic similarity is implemented between words in different languages namely English and Spanish.TThe corpus used in this study is Europarl Parallel Corpus in English and Spanish. The word context is sourced from the Swadesh list, as well as the results of its semantic similarities compared to the Gold Standard SemEval 2017 Crosslingual Semantic Similarity dataset for measured the correlation values. BThe result shows that the measurement of PMI method yields a correlation 0.577 for harmonic mean between Pearson correlation and Spearman correlation.The correlation value of the system is higher than the correlation value by the Rufino team which also uses the PMI method for cross-lingual semantic similarity measurement on the dataset semEval 2017. The correlation value by the Rufino team is 0.340.

Keywords: *Semantic Similarity,Crosslingual Semantic Similarity, Pointwise Mutual Information*