Abstract

Semantic similarity measurement using monolingual alignment has several constraints on the implementation of the system that is still becoming a study on how to modify the feature extraction used to align on word of the sentence. The existing system examine the feature extraction to be applied to the alignment. Semantic similarity is one of the measurements in the text mining subject that seeking a relation to represent between words in a sentence.

In this study, the method used is monolingual alignment by modifying feature extraction from several studies before. Feature extraction modification that is focused on the study was identical word, named entities recognition, paraphrase database, part-of-speech, synonyms, hypernym, hyponym and positional. The system built in this study using supervised and unsupervised techniques in data mining. Testing is done by extrinsic evalution by comparing the scores of the system with a score that has been made by experts. Techniques used for comparing scores of the system with gold standard score is using Pearson correlation. The test results of this study using unsupervised technique got correlation 0.7077 while supervised obtain correlation 0.7228.

Keyword : *semantic similarity, monolingual alignment, extrinsic evalution, identical word, named entities recognition, positional.*