

Abstract

The research about measuring semantic score on Al-Quran verse pairs still very few. The research is still limited on searching verses that divided into level 0, level 1, and level 2 in Qursim research. Therefore, it needs other methods, one example is monolingual alignment method.

On this final project research, it used monolingual alignment and addition of supervised learning system in measuring semantic similarity score on Al-Quran verse pairs. The results of alignment features system on 2016 dataset got its highest correlation score with 0.728 and on 2017 dataset with 0.698. While on the alignment features system that added supervised learning system, the maximum correlation score for 2016 dataset is 0.728 and for 2017 dataset is 0.7. Based on the results, we can conclude that identical words as the basic feature of alignment still become the best feature compared with other feature which are PPDB and WordNet. Correlation score from addition of supervised learning has gotten closer from the correlation score of alignment system.

Keywords: semantic textual similarity, monolingual alignment, supervised learning, gold standard, correlation.