

Abstract

The Quran is the way of life of Muslims. The Quran has several unique characteristics including the repetition of the same verses in other chapter and discussion of a topic on a certain verse which also discussed in other verses in different chapters. Therefore, by the necessity to understanding the Quran, it is necessary to study the text of the Quran. One of the basic research in text mining is about STS research, the results of this research can be used for another research in the text mining and NLP fields, such as clustering, word sense disambiguation, question answering and other research. This research measured the semantic similarity between two segmented verses of the Quran in Indonesian translation. Semantic similarity measurements in this study conducted by implementing word alignment approach. In addition, the gold standard and data collection paraphrases the Quran are built, the data were became a contribution from writer and the gold standard can be used as a benchmark to determine the correlation between the results of the system measurement with the result of human intuition. In this study, we want to know how the correlation value is generated by the system, and how the influence of alignment features that implemented on the system. In the test results showed that the highest correlation value reaches 75% and known alignment feature that the most influential in this implementation of STS in the Quranic verse is identical word alignment.

Keywords: The Quran, The Quranic verse, semantic similarity, Word Alignment, correlation, gold standard, feature alignment