

Abstract

To measure the semantic correlation between words there are many methods that can be used, one of which is word alignment. Word alignment is a method that aligns words that have a letter correlation or meaningful correlation between two sentences. This study focuses on using word alignment in the translation of the Al-Quran verse. This method was developed to align the sentence pair data but can be used to measure the semantic correlation between verses. By using the algorithm back to basic word alignment developed by Sultan et al.[5] the researcher re-develops to research the alignment between verses in the Al-Quran, to find out the effect if used in the translation of the Al-Quran as the dataset. The Al-Quran dataset used will be converted into the MSR-RTE[1] dataset format by researchers, with the aim of providing new research results in the context of the Al-Quran word alignment. In Back to Basic Word Alignment there is a pipeline alignment that contains the use of a tour map sequence, the tour used in this research, align identical word, align PPDB, align word sequences, align named entities, align content words (dependency), align content words using surrounding words (text neighbor), align stop words, align PPDB Extended[5]. These features will be combined to determine the correlation value between two Al-Quran verses (F1 score). The best correlation value between verses that can be produced in this study was 51.02 % compared to the baseline research by Sultan et al. that is 91.7%. The correlation value between verses in this study can be concluded as a sufficient value, and can still be improved by adding features, knowledge base, or using a combination of different translators of the Al-Quran .

Keywords: Back to Basic Word Alignment, Monolingual Alignment, MSR-RTE, Al-Quran, Text mining.
