

ABSTRACT

Abstract - *The Qur'an is the Muslim holy book as the primary source of knowledge and guidance, consisting of 114 surahs, 30 juz, and has approximately 6200 verses in it. Searching for connections or similarities between words in the Qur'an takes a long time to find and summarize them. There is a need for a dictionary, encyclopedia, or thesaurus of the Al-Qur'an vocabulary, which contains each word entry related to other words. This study discusses the interrelations and semantic similarities between words in the Qur'an, which aims to help in searching between related words in them. The approach taken is a distributional similarity which is an important part of word embedding. Measurement of word relevance is measured by semantic similarity which is one of the lessons learned in Natural Language Processing (NLP). Semantic similarity measures the closeness of word vectors using cosine similarity. The process of changing words in vector form uses the FastText algorithm which is a development of the Word2vec algorithm. The dataset used is the translation of the word Al-Qur'an in English and Indonesian. The word becomes an input in the system and then produces a score that represents the interrelationship between words. Evaluation of system output results using the Pearson correlation method involving the gold standard. Evaluation of the use of the FastText algorithm produces a correlation value of 0.3398 for Indonesian translation corpus and 0.2326 for English translation corpus.*

Keywords— Qur'an, semantic similarity, Word embedding, FastText, Pearson correlation