

ABSTRAK

Abstrak - Alquran adalah kitab suci umat Islam sebagai sumber utama pengetahuan dan petunjuk, yang terdiri dari 114 surah, 30 juz, dan memiliki sekitar 6200 ayat di dalamnya. Pencarian keterkaitan ataupun kesamaan antara kata-kata dalam Al Qur'an membutuhkan waktu lama untuk menemukan dan merangkumnya. Ada kebutuhan untuk kamus, ensiklopedia, atau tesaurus dari kosa kata Al-Qur'an, yang berisi setiap entri kata yang terkait dengan kata-kata lain. Studi ini membahas tentang keterkaitan dan kesamaan semantik antara kata-kata dalam Alquran, yang bertujuan untuk membantu dalam mencari antara kata-kata terkait di dalamnya. Pendekatan yang diambil adalah kesamaan distribusional yang merupakan bagian penting dari *word embedding*. Pengukuran keterkaitan kata diukur dengan kesamaan semantik yang merupakan salah satu pelajaran dalam *Natural Language Processing* (NLP). Kesamaan semantik mengukur kedekatan vektor kata menggunakan *cosine similarity*. Proses mengubah kata dalam bentuk vektor menggunakan algoritma FastText yang merupakan pengembangan dari algoritma Word2vec. Dataset yang digunakan adalah terjemahan dari kata Al-Qur'an dalam bahasa Inggris dan Indonesia. Kata menjadi input dalam sistem dan kemudian menghasilkan skor yang mewakili keterkaitan antar kata. Evaluasi hasil keluaran sistem menggunakan metode *Pearson correlation* yang melibatkan *Gold Standard*. Evaluasi penggunaan algoritma FastText menghasilkan nilai korelasi 0,3398 untuk corpus terjemahan Indonesia dan 0,2326 untuk corpus terjemahan bahasa Inggris.

Keywords— *Quran, semantic similarity, Word embedding, FastText, Pearson correlation*