

ABSTRAK

Nama : Rahmaniyah A'laa Pohan
NIM : 1217010056
Judul : Penerapan *Word2Vec* Untuk Analisis Hubungan Semantik Kata pada Dataset Terjemahan Bahasa Inggris Al-Qur'an

Al-Qur'an memiliki kedalaman makna yang sering sulit ditangkap secara utuh dalam terjemahan, khususnya dari bahasa Arab ke bahasa Inggris. Perbedaan struktur gramatikal dan semantik antar bahasa ini dapat menimbulkan ambiguitas makna, sehingga dibutuhkan pendekatan analisis yang mampu merepresentasikan hubungan semantik antar kata secara lebih akurat. Penelitian ini menggunakan model *Word2Vec* untuk menganalisis kemiripan semantik kata dalam terjemahan Al-Qur'an berbahasa Inggris. Dataset yang digunakan mencakup 6.236 ayat terjemahan dari *Tanzil.net*, dengan total 146.163 kata dan 11.214 kosakata unik, yang diproses melalui *data cleaning*, *case folding*, tokenisasi, *stopword removal*, dan lemmatisasi. Model dilatih dengan dua arsitektur *Word2Vec*, yaitu CBOW dan *Skip-Gram*, menggunakan variasi dimensi vektor, *window size*, dan *epoch*. Evaluasi dilakukan dengan *cosine similarity* pada kata-kata religius yang dipilih secara tematik. Hasil penelitian menunjukkan bahwa model *CBOW* dengan konfigurasi dimensi 100 dan *window size* 3 memberikan hasil terbaik dengan nilai *similarity* 0.9905 pada kata target *prophet*. Nilai *similarity* yang tinggi tersebut juga tercermin dari kata-kata hasil perhitungan yang memiliki kedekatan makna dengan kata *prophet*, sehingga semakin menguatkan bahwa representasi vektor yang dihasilkan sudah sesuai secara semantik. Dengan demikian, *Word2Vec* terbukti efektif dalam memetakan makna kata pada teks terjemahan Al-Qur'an dan berpotensi mendukung pengembangan analisis semantik dalam studi keislaman.

Kata kunci: *Word2Vec*, *semantic similarity*, Al-Quran, *cosine similarity*, CBOW, *Skip-Gram*

ABSTRACT

Name : Rahmaniyah A'laa Pohan

NIM : 1217010056

Thesis Title : Penerapan *Word2Vec* Untuk Analisis Hubungan Semantik Kata pada Dataset Terjemahan Bahasa Inggris Al-Qur'an

The Qur'an contains profound meanings that are often difficult to fully capture in translation, particularly from Arabic into English. Differences in grammatical structures and semantic nuances between the two languages may lead to ambiguity of meaning, thereby requiring an analytical approach capable of representing semantic relationships between words more accurately. This study employs the Word2Vec model to analyze semantic similarity of words in the English translation of the Qur'an. The dataset consists of 6,236 translated verses from Tanzil.net, comprising a total of 146,163 words and 11,214 unique vocabularies, which were preprocessed through data cleaning, case folding, tokenization, stopwords removal, and lemmatization. The model was trained using two Word2Vec architectures, namely CBOW and Skip-Gram, with variations in vector dimensions, window sizes, and epochs. Evaluation was conducted using cosine similarity on thematically selected religious terms. The results show that the CBOW model with a vector dimension of 100 and a window size of 3 achieved the best performance, yielding a similarity score of 0.9905 for the target word prophet. This high similarity score is further reflected in the fact that the computed similar words indeed share close semantic meanings with prophet, thereby reinforcing that the generated vector representations are semantically appropriate. Consequently, Word2Vec is proven to be effective in mapping word meanings in the Qur'an translation text and holds potential to support semantic analysis in Islamic studies.

Keywords: *Word2Vec, semantic similarity, Qur'an, cosine similarity, CBOW, Skip-Gram*