

ABSTRAK

Nama : Galih Fikri Pratama

NIM : 1207010025

Judul : *Clustering* Ayat Al-Qur'an Bahasa Indonesia Menggunakan Algoritma *Spectral Clustering* Untuk Mengungkap Relasi Semantik Kata

Al-Qur'an sebagai kitab suci umat Islam memiliki pesan-pesan yang dalam dan kompleks, yang sering kali diungkapkan melalui susunan kata-kata yang kaya akan makna. Oleh karena itu, berbagai versi terjemahan dan tafsir Al-Qur'an dalam Bahasa Indonesia telah dikembangkan untuk memperluas pemahaman terhadap makna ayat-ayat suci. Namun, proses penerjemahan tidak selalu mampu menangkap seluruh makna semantik dari kata-kata asli dalam bahasa Arab, terutama ketika satu kata dapat memiliki beragam makna tergantung pada konteksnya. Misalnya di antara istilah penting yang digunakan untuk menyebut Tuhan adalah *Rabb* dan *Ilah*. Meskipun keduanya sering kali diterjemahkan serupa dalam bahasa Indonesia, keduanya digunakan dalam konteks yang berbeda dan memiliki nuansa makna yang khas. Penelitian ini menggunakan *word2vec* untuk menghasilkan representasi numerik dari kata-kata dalam teks terjemahan Al-Qur'an Bahasa Indonesia. Metode *clustering* digunakan untuk mengelompokkan representasi numerik dari kata-kata ini. Menggunakan algoritma *Spectral Clustering*, analisis hasil *clustering* menunjukkan bahwa *cluster* terbaik berdasarkan logika linguistik memiliki nilai *silhouette coefficient* yang bervariasi dan nilai *cosine similarity* yang tinggi, menunjukkan keterikatan semantik yang kuat antar kata dalam *cluster*.

Kata kunci : *Natural Language processing, Word2vec, Data Mining, Spectral clustering, Silhouette Coefficient, Conductance Score*

ABSTRACT

Name : Galih Fikri Pratama

NIM : 1207010025

Title : *Clustering of Indonesian Quran Verses Using Spectral
Clustering Algorithm to Reveal Semantic Relations of Words*

The Quran, as the holy book of Muslims, contains profound and complex messages, often expressed through richly meaningful words. Therefore, various translations and interpretations of the Quran in Indonesian have been developed to broaden understanding of the meaning of the holy verses. However, the translation process is not always able to capture the full semantic meaning of the original Arabic words, especially when a single word can have multiple meanings depending on the context. For example, among the important terms used to refer to God are Rabb and Ilah. Although both are often translated similarly in Indonesian, they are used in different contexts and have distinctive nuances of meaning. This study uses word2vec to generate numerical representations of words in the Indonesian translation of the Quran. The clustering method is used to group these numerical representations. Using the Spectral Clustering algorithm, analysis of the clustering results shows that the best clusters based on linguistic logic have varying silhouette coefficient values and high cosine similarity values, indicating strong semantic ties between words in the cluster.

Keyword : *Natural Language Processing, Word2vec, Data Mining, Spectral Clustering, Silhouette Coefficient, conductance score*