

ABSTRAK

Nama : Rissa Qurotul Aini
NIM : 1217010061
Judul : Pengaruh Parameter *n-gram* pada Training *FastText* Terhadap *Similarity* Kata Menggunakan Data Al-Qur'an Terjemah Bahasa Inggris.

Analisis hubungan tematik dalam teks keagamaan seperti terjemahan Al-Qur'an merupakan tantangan karena kompleksitas semantic dan morfologinya. Meskipun model *word embedding* seperti *FastText* diakui unggul karena kemampuannya menganalisis sub-word, pengaruh spesifik dari parameter *n-gram* terhadap kualitas similaritas kata pada korpus ini belum banyak diteliti. Penelitian ini bertujuan untuk menganalisis secara sistematis pengaruh parameter *n-gram* pada model *FastText* untuk mengukur kemiripan kata dan mengungkap kluster-kluster tematik dalam data terjemahan Al-Qur'an bahasa Inggris. Metode penelitian menggunakan dataset terjemahan Al-Qur'an yang melalui tahap pra-pemrosesan teks, pelatihan model *FastText* dengan berbagai konfigurasi *n-gram*, dan evaluasi kuantitatif menggunakan *cosine similarity* serta visualisasi dengan PCA. Hasil penelitian menunjukkan bahwa parameter *n-gram* secara fundamental meningkatkan kemampuan model dalam menangani variasi struktur kata. Pengaturan *n-gram* terbukti krusial untuk membangun model yang tangguh dalam menangkap hubungan semantic yang mendalam pada teks keagamaan.

Kata Kunci : *FastText*, *n-gram*, *Word Embedding*, Kemiripan Kata, *Cosine Similarity*, Al-Qur'an, NLP

ABSTRACT

Name : Rissa Qurotul Aini

NIM : 1217010061

Thesis Title : Pengaruh Parameter *n*-gram pada Training FastText Terhadap Similarity Kata Menggunakan Data Al-Qur'an Terjemah Bahasa Inggris.

The analysis of thematic relationship in religious texts such as the Qur'an translation is challenging due to its semantic and morphological complexity. Although word embedding models such as FastText are recognized for their ability to analyze subwords, the specific influence of n-gram parameters on the quality of word similarity in this corpus has not been widely studied. This study aims to systematically analyze the influence of n-gram parameters on the FastText model to measure word similarity and reveal thematic clusters in English Quran translation data. The research method uses a Quran translation dataset that goes through text pre-processing stages, FastText model training with various n-gram configurations, and quantitative evaluation using cosine similarity and visualization with PCA. The results show that n-gram parameters fundamentally improve the model's ability to handle variations in word structure. The n-gram setting proves crucial for building a robust model in capturing deep semantic relationship in religious texts.

Keywords : FastText, n-gram, Word Embedding, Word Similarity, Cosine Similarity, Qur'an, NLP