

ABSTRAK

GEMPAR PERKASA TAHIR. Implementasi Metode *Hybrid Fuzzy Jaro-Winkler* Dan *Cosine Similarity* Pada Sistem Pencarian Ayat Al-Qur'an Berbasis Transliterasi Latin (Dibimbing Oleh Emil Agusalim Habi Talib, S.Pd., M.Eng., dan Fahrim Irhamna Rachman S.Kom., M.T.).

Penelitian ini membahas tantangan dalam pencarian ayat Al-Qur'an berbasis transliterasi Latin, yang terhambat oleh ketiadaan ortografi baku sehingga menimbulkan berbagai variasi ejaan. Penelitian ini bertujuan untuk merancang dan mengimplementasikan sistem pencarian informasi hibrida yang mengintegrasikan *Fuzzy Jaro-Winkler* untuk kesamaan leksikal serta *Cosine Similarity* pada *embedding DistilBERT* yang telah di *fine-tune* untuk relevansi semantik. Alur kerja sistem dimulai dengan tahap praproses dan normalisasi dataset, dilanjutkan dengan pemilihan kandidat awal menggunakan *Jaro-Winkler*, serta *reranking* akhir melalui penilaian kesamaan semantik. Evaluasi dilakukan dengan *black-box testing* pada berbagai skenario, termasuk kueri ideal, variasi ejaan, kueri tidak lengkap, dan variasi panjang kueri. Hasil penelitian menunjukkan akurasi tinggi pada kueri ideal (96%) dan variasi ejaan (92%), dengan kinerja yang meningkat seiring bertambahnya panjang kueri, mencapai 96% pada kueri empat kata. Pendekatan hibrida ini secara efektif menjembatani kesenjangan leksikal dan semantik, melampaui performa metode tunggal, serta menunjukkan ketangguhan dalam menangani transliterasi non standar pada pencarian teks Al-Qur'an.

Kata Kunci: *Jaro-Winkler*, *Cosine Similarity*, pencarian Al-Qur'an, transliterasi, pencarian semantik.

ABSTRACT

GEMPAR PERKASA TAHIR. *Implementation of Hybrid Fuzzy Jaro-Winkler and Cosine Similarity Method in Qur'anic Verse Retrieval System Based on Latin Transliteration (Supervised by Emil Agusalim Habi Talib, S.Pd., M.Eng., and Fahrir Irhamna Rachman S.Kom., M.T.).*

This research addresses the challenge of retrieving Qur'anic verses in Latin transliteration, which is hindered by the absence of a standardized orthography, leading to diverse spelling variations. The study aims to design and implement a hybrid information retrieval system that integrates Fuzzy Jaro-Winkler for lexical similarity and Cosine Similarity on fine-tuned DistilBERT embeddings for semantic relevance. The system workflow begins with preprocessing and normalization of the dataset, followed by initial candidate selection using Jaro-Winkler, and final reranking through semantic similarity scoring. Evaluation was conducted using black-box testing across scenarios including ideal queries, spelling variations, incomplete queries, and varying query lengths. Results show high accuracy for ideal (96%) and varied spelling queries (92%), with performance improving as query length increases, reaching 96% for four-word queries. The hybrid approach effectively bridges lexical and semantic gaps, outperforming single-method baselines, and demonstrates robustness in handling non-standard transliteration in Qur'anic text retrieval.

Keywords : *Jaro Winkler; Cosine Similarity, Qur'an retrieval, transliteration, semantic search.*