

INTISARI

Penelitian ini bertujuan untuk mengevaluasi efektivitas pendekatan representasi topikal, semantik, dan hybrid dalam sistem pencarian tafsir Al-Qur'an berbahasa Indonesia. Pendekatan yang dibandingkan meliputi TF-IDF + LSA, IndoBERT, serta kombinasi keduanya melalui penggabungan skor (score fusion). Korpus yang digunakan terdiri dari 13.934 entri tafsir yang diperoleh dari berbagai sumber terpercaya. Evaluasi dilakukan menggunakan desain Cranfield-style yang mencakup 50 query, penetapan relevansi otomatis, dan pengukuran kinerja menggunakan metrik precision@10, recall@10, dan Mean Average Precision (MAP). Hasil penelitian menunjukkan bahwa model berbasis TF-IDF + LSA menghasilkan performa paling rendah (precision@10 sebesar 0.1333 dan MAP sebesar 0.0784). Sementara itu, pendekatan IndoBERT memberikan peningkatan performa signifikan (precision@10 sebesar 0.5000 dan MAP sebesar 0.3758). Pendekatan hybrid dengan bobot $\alpha = 0.9$ menghasilkan kinerja terbaik dengan nilai precision@10 sebesar 0.8000 dan MAP sebesar 0.4706. Nilai MAP yang identik dengan recall@10 menunjukkan konsistensi hasil relevansi dalam peringkat dokumen teratas. Kesimpulan dari penelitian ini adalah bahwa pendekatan hybrid retrieval mampu mengintegrasikan keunggulan representasi semantik dan topikal secara efektif. Metode ini tidak hanya meningkatkan relevansi hasil pencarian, tetapi juga memberikan kontribusi terhadap pengembangan sistem IR dalam domain teks keagamaan berbahasa Indonesia.

Kata kunci: Information retrieval, LSA, IndoBERT, hybrid retrieval, tafsir Al-Qur'an

ABSTRACT

This study aims to evaluate the effectiveness of topical, semantic, and hybrid representation approaches in retrieving Indonesian-language Qur'anic exegesis (tafsir). The evaluated methods include TF-IDF with LSA, IndoBERT, and a hybrid approach using score fusion. The dataset comprises 13,934 tafsir entries collected from multiple reputable sources. The evaluation was conducted using a Cranfield-style experimental design involving 50 queries, automatic relevance judgment, and standard IR metrics including precision@10, recall@10, and Mean Average Precision (MAP). The results show that the TF-IDF + LSA approach yielded the lowest performance (precision@10: 0.1333, MAP: 0.0784). IndoBERT significantly outperformed this baseline (precision@10: 0.5000, MAP: 0.3758). The hybrid approach with a fusion weight $\alpha = 0.9$ achieved the best performance (precision@10: 0.8000, MAP: 0.4706). The identical values of recall@10 and MAP indicate consistent relevance within the top-ranked documents. It can be concluded that the hybrid retrieval approach effectively combines semantic and topical representations. This method not only improves search relevance but also contributes to the development of Information Retrieval systems for Indonesian religious texts.

Keywords: Information retrieval, LSA, IndoBERT, hybrid retrieval, Quranic exegesis