

INTISARI

Model CLIR (*Cross-Language Information Retrieval*) untuk Pencarian Teks Fikih Berbahasa Arab Menggunakan Kueri Bahasa Indonesia

Oleh

MUHAMMAD SYIFAURROHMAN

22/505198/PA/21748

Dalam studi fikih, sumber rujukan utama mayoritas tersedia dalam Bahasa Arab, sementara banyak pembelajar di Indonesia menghadapi keterbatasan dalam merumuskan kueri pencarian menggunakan bahasa tersebut. Kesenjangan akses ini mendorong perlunya evaluasi terhadap sistem *Cross-Language Information Retrieval* (CLIR) yang mampu menjembatani masukan kueri Bahasa Indonesia dengan dokumen target Bahasa Arab. Penelitian ini mengevaluasi arsitektur pencarian lintas bahasa pada korpus digital kitab *Fathul Mu'in* yang telah direstorasi menggunakan dataset uji berisi 153 kueri tervalidasi pakar. Eksperimen dilakukan pada pendekatan pencarian leksikal (*sparse retrieval* BM25) berbasis terjemahan, pencarian semantik (*dense retrieval* multibahasa yang di-*fine-tune* menggunakan data sintetis JH-POLO (mBERT dan DistilBERT), model *dense retrieval* ekabahasa (AraDPR), serta pendekatan (*hybrid*) menggunakan metode *Reciprocal Rank Fusion* (RRF). Performa sistem dievaluasi secara kuantitatif menggunakan metrik *Mean Reciprocal Rank* (MRR) dan *Success@k*. Hasil pengujian menunjukkan bahwa pendekatan leksikal murni mencapai MRR sebesar 0,360 dan terbukti superior pada kueri kata kunci. Meskipun model *dense retrieval* multibahasa memiliki skor keseluruhan yang lebih rendah (MRR 0,297), model ini terbukti paling tangguh dalam memahami konteks kueri naratif berupa studi kasus. Penggabungan kedua pendekatan tersebut melalui metode *hybrid* RRF berhasil memberikan performa optimal dengan capaian MRR tertinggi sebesar 0,394. Penelitian ini memberikan landasan empiris mengenai efektivitas berbagai arsitektur pencarian lintas bahasa pada domain teks hukum Islam.

Kata-kata kunci : *Cross-Language Information Retrieval*, Fikih, *Sparse Retrieval*, *Dense Retrieval*, *Reciprocal Rank Fusion*, JH-POLO.

ABSTRACT

Cross-Language Information Retrieval Models for Indonesian–Arabic Fiqh Text Retrieval

By

MUHAMMAD SYIFAURROHMAN

22/505198/PA/21748

In the study of Islamic jurisprudence (*fiqh*), primary reference sources are predominantly available in Arabic, while many learners in Indonesia face limitations in formulating search queries in that language. This access gap necessitates the evaluation of Cross-Language Information Retrieval (CLIR) systems capable of bridging Indonesian queries with target Arabic documents. This study evaluates cross-lingual retrieval architectures on a restored digital corpus of *Fathul Mu'in* using a test dataset of 153 expert-validated queries. Experiments were conducted on translation-based lexical retrieval (BM25 sparse retrieval), multilingual semantic retrieval fine-tuned using the JH-POLO synthetic dataset (mmBERT and DistilBERT dense retrieval), monolingual dense retrieval (AraDPR), and a hybrid approach using Reciprocal Rank Fusion (RRF). System performance was quantitatively evaluated using Mean Reciprocal Rank (MRR) and Success@k metrics. The test results indicate that the pure lexical approach achieved an MRR of 0.360 and was superior for keyword queries. Although the multilingual dense retrieval model had a lower overall score (MRR 0.297), it proved to be the most robust in understanding the context of narrative case-study queries. Combining these approaches via the RRF hybrid method successfully yielded optimal performance, achieving the highest MRR of 0.394. This study provides an empirical foundation regarding the effectiveness of various cross-lingual retrieval architectures within the Islamic law text domain.

Keywords : Cross-Language Information Retrieval, Fiqh, Sparse Retrieval, Dense Retrieval, Reciprocal Rank Fusion, JH-POLO.