

ABSTRAK

SEQRAN: COMPOSITE SCORING AND RERANKING TECHNIQUES FOR REFINING QURAN SEARCH ENGINE RESULT

Ray Ramadita – 1217050117

Jurusan Teknik Informatika

Universitas Islam Negeri Sunan Gunung Djati Bandung

Sistem *Information Retrieval* (IR) sangat penting dalam pengembangan *search engine* yang akurat dan relevan, terutama pada aplikasi domain spesifik seperti Sequran. Meskipun *fine-tuning* merupakan pendekatan optimasi yang umum, metode ini seringkali memerlukan sumber daya komputasi yang besar, *dataset* yang beragam, dan *hyperparameter tuning* yang memakan waktu, dengan risiko penurunan performa. Untuk mengatasi tantangan tersebut, penelitian ini memperkenalkan dan mengevaluasi arsitektur baru yang meningkatkan relevansi pencarian dengan mengintegrasikan sinyal leksikal dan semantik, menawarkan alternatif praktis untuk *fine-tuning* yang membutuhkan sumber daya besar. Arsitektur yang diusulkan melibatkan proses dua tahap. Tahap pertama, *composite scoring*, meningkatkan frekuensi kata tradisional (BM25/S) dengan *semantic intent booster* untuk menghasilkan peringkat awal yang lebih bervariasi. Tahap kedua menggunakan arsitektur *Cross-Encoder* (jina-reranker-v2-base-multilingual) untuk refine peringkat hasil awal berdasarkan relevansi kontekstual yang lebih dalam. Evaluasi pada *dataset* spesifik pertanyaan-jawaban tentang Islam menunjukkan bahwa kombinasi ini (tanpa *keyword extraction*) menghasilkan hasil yang paling seimbang. Temuan menunjukkan peningkatan moderat namun konsisten pada berbagai metrik, seperti Precision@10 dan Recall@10, dengan peningkatan masing-masing 6.3% dan 14.1% dibandingkan dengan model dasar. Peningkatan ini melibatkan *trade-off* yang sangat jelas, karena waktu evaluasi total pada *dataset* meningkat menjadi 18.5 detik. Implikasi utama dari penelitian ini adalah validasi arsitektur praktis untuk meningkatkan sistem IR, menawarkan alternatif yang layak untuk konteks domain spesifik seperti Sequran.

Kata Kunci: Composite Scoring, Information Retrieval, Reranking, Mesin Pencari, Quran

ABSTRACT

SEQURAN: COMPOSITE SCORING AND RERANKING TECHNIQUES FOR REFINING QURAN SEARCH ENGINE RESULT

Ray Ramadita – 1217050117

Department of Informatics Engineering

Sunan Gunung Djati State Islamic University Bandung

Information Retrieval (IR) systems are crucial in the development of accurate and relevant search engines, especially in domain-specific applications such as Sequran. While fine-tuning is a common optimization approach, this method often requires significant computational resources, diverse datasets, and time-consuming hyperparameter tuning, with the risk of performance degradation. To address these challenges, this research introduces and evaluates a novel architecture that enhances search relevance by integrating lexical and semantic signals, offering a practical alternative to resource-intensive fine-tuning. The proposed architecture involves a two-stages process. The first stage, composite scoring, enhances traditional term frequency (BM25/S) with a semantic intent booster to generate a more varied initial ranking. The second stage utilizes the Cross-Encoder architecture (jina-reranker-v2-base-multilingual) to further refine the ranking of the initial results based on deeper contextual relevance. Evaluation on a specific Islamic query-answer dataset showed that this combination (without keyword extraction) produced the most balanced results. The findings showed a moderate but consistent improvement in various metrics, such as Precision@10 and Recall@10, with increases of 6.3% and 14.1%, respectively, compared to the baseline model. This improvement involves a very clear trade-off, as the total evaluation time on the dataset increased to 18.5 seconds. The main implication of this research is the validation of a practical architecture for improving IR systems, offering a viable alternative for domain-specific contexts such as Sequran.

Keywords: *Composite Scoring, Information Retrieval, Reranking, Search Engine, Quran*