

Temu-Balik pertanyaan duplikat pada Forum Kesehatan Berbahasa Indonesia dengan Rank Fusion = Duplicate question retrieval in Indonesian Language Health Forum with Rank Fusion

Ilhan Firka Najia, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=9999920553221&lokasi=lokal>

Abstrak

Penelitian ini mengeksplorasi efektivitas penggunaan neural rerankers yang telah dilatih sebelumnya dalam meningkatkan kinerja model berbasis text matching seperti BM25 untuk digunakan dalam deteksi pertanyaan duplikat pada consumer health forum. Studi ini juga meneliti metode agregasi hasil reranking dari berbagai neural rerankers untuk menghasilkan performa yang lebih baik dibandingkan penggunaan reranker individual. Metode reranking pertama menggunakan BM25, diikuti oleh reranking kedua menggunakan model neural seperti cross-encoder/ms-marco-MiniLM-L-12-v2, paraphrase-MiniLM-L6-v2, dan lainnya. Tahap ketiga melibatkan teknik rank fusion seperti Borda Fuse, Condorcet, dan Weighted Combsum. Hasil menunjukkan bahwa kombinasi reranking dengan neural reranker secara signifikan meningkatkan efektivitas model BM25, terutama saat menggunakan teknik rank fusion yang lebih canggih seperti Weighted Combsum. Studi ini menyarankan bahwa agregasi hasil reranking dapat mengatasi kelemahan individual reranker dan memberikan hasil yang lebih konsisten dan efektif. Penelitian ini membuka jalan untuk eksplorasi lebih lanjut dalam optimisasi kombinasi model untuk pencarian informasi yang lebih akurat dan efisien.

.....This study explores the effectiveness of using pre-trained neural rerankers in improving the performance of text matching based models such as BM25 for use in duplicate question detection in textitconsumer health forum. This study also examines the method of aggregating reranking results from various neural rerankers to produce better performance than using individual rerankers. The first reranking method used BM25, followed by the second reranking using neural models such as cross-encoder/ms-marco-MiniLM-L-12-v2, paraphrase-MiniLM-L6-v2, and others. The third stage involves rank fusion techniques such as BordaFUSE, Condorcet, and Weighted COMBSUM. Results show that the combination of reranking with neural rerankers significantly improves the effectiveness of the BM25 model, especially when using more advanced rank fusion techniques such as Weighted COMBSUM. This study suggests that aggregation of reranking results can overcome the weaknesses of individual rerankers and provide more consistent and effective results. This research paves the way for further exploration in model combination optimization for more accurate and efficient information retrieval.