

Sistem Temu-Balik Pengetahuan untuk Helpdesk PDDikti dengan Reranking dan Ekspansi Dokumen = Knowledge Retrieval System for The PDDikti Helpdesk with Reranking and Document Expansion

Abdul Rahman, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=9999920567384&lokasi=lokal>

Abstrak

Website Helpdesk PDDikti memiliki sekumpulan knowledge yang belum didukung oleh sistem pencarian yang efektif. Penelitian ini bertujuan untuk membangun system temu-balik informasi yang efektif dengan skor metrik yang tinggi terhadap knowledge Helpdesk PDDikti. Metode yang digunakan dalam penelitian ini adalah penerapan model temu-balik informasi berbasis text matching pada berbagai model, yang kemudian disempurnakan dengan dua metode perbaikan ranking dokumen: (1) metode ekspansi istilah pada dokumen menggunakan prediksi kueri dari model doc2query, dan (2) metode reranking dokumen menggunakan model LambdaMART. Hasil penelitian menunjukkan bahwa di antara beberapa model temu-balik informasi berbasis text matching, model BM25 memberikan kinerja terbaik dengan skor MRR 0,781. Selanjutnya, hasil ranking dokumen dari model BM25 dapat ditingkatkan akurasi melalui metode ekspansi istilah dengan lima kueri menggunakan pemilihan kandidat random sampling, yang meningkatkan skor MRR menjadi 0,799. Namun, penggunaan metode reranking dengan model LambdaMART untuk meningkatkan akurasi hasil ranking dokumen belum memberikan hasil yang lebih baik dibandingkan metode ekspansi istilah. Meskipun demikian, terdapat varian model reranking LambdaMART yang menggunakan fitur semantic similarity dan fitur skor agregat, yang mampu mengalahkan model BM25 tanpa ekspansi istilah, dengan skor MRR terbaik masing-masing 0,782 dan 0,787. Meskipun begitu, peningkatan ini masih belum cukup signifikan.

.....The PDDikti Helpdesk website has a collection of knowledge that is not yet supported by an effective search system. This study aims to develop an effective information retrieval system with high metric scores for the PDDikti Helpdesk knowledge base. The method used in this research involves applying text matching-based information retrieval models across various models, which are then refined using two document ranking improvement methods: (1) term expansion in documents using query predictions from the doc2query model, and (2) document reranking using the LambdaMART model. The research results indicate that among several text matching-based information retrieval models, the BM25 model provides the best performance with an MRR score of 0.781. Furthermore, the document ranking results from the BM25 model can be improved in accuracy through the term expansion method with five queries using random sampling for candidate selection, which increases the MRR score to 0.799. However, the use of the reranking method with the LambdaMART model to improve document ranking accuracy did not yield better results compared to the term expansion method. Nevertheless, there are variants of the LambdaMART reranking model that use semantic similarity features and aggregate score features, which managed to outperform the BM25 model without term expansion, with the best MRR scores of 0.782 and 0.787, respectively. However, these improvements are still not significant enough.