

Sistem Temu-Kembali Dokumen Hukum dengan Menggunakan Ekspansi Istilah = Legal Document Retrieval System by Using Term Expansion

Arya Bintang Pratama Kumaladjati, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=9999920551861&lokasi=lokal>

Abstrak

Penelitian ini bertujuan untuk meningkatkan efektivitas model Temu-kembali Informasi di bidang hukum dengan memanfaatkan metode Term Expansion. Latar belakang penelitian ini didasarkan pada kebutuhan praktisi hukum akan sistem yang dapat melakukan pencarian informasi secara otomatis dan efisien dalam database dokumen hukum yang besar dan kompleks. Pendekatan yang diusulkan dalam penelitian ini menggunakan arsitektur Cascade-Ranking Model, yang terdiri dari dua tahapan utama: Sparse Retrieval Model untuk pencarian awal dan algoritma Re-ranking untuk penyusunan ulang hasil pencarian. Penelitian ini berfokus pada eksplorasi kinerja Sparse Retrieval Model melalui penerapan teknik Term Expansion, yang mencakup Query Expansion dan Document Expansion. Query Expansion dilakukan dengan berbagai pendekatan, termasuk non-contextual word embedding, contextual word embedding dengan BERT, dan Local Analysis. Document Expansion melibatkan penambahan prediksi query yang relevan ke dalam dokumen untuk meningkatkan kecocokan dengan query pengguna. Hasil dari eksperimen menunjukkan bahwa tidak ada skenario yang secara mutlak unggul dalam semua metrik evaluasi yang digunakan. Beberapa skenario menunjukkan keunggulan dalam satu metrik tertentu namun mengalami penurunan pada metrik lainnya. Pada salah satu skenario berhasil meningkatkan $R@20$ dari 0,896 menjadi 0,906, selain itu pada skenario lainnya berhasil meningkatkan metrik $P@5$ dari 0,186 menjadi 0,192. Selain itu, pada skenario yang menggunakan stopword dengan konteks hukum untuk melakukan filter terhadap kata yang ingin diekspansi dengan word embedding, berhasil meningkatkan $R@20$ dari 0,896 menjadi 0,916 dan mendapatkan p-value sebesar 0,045. P-value yang dibawah 0,05 ini berarti terjadi peningkatan yang signifikan pada skenario tersebut. Meskipun dapat meningkatkan beberapa metrik, Query Expansion tidak selalu meningkatkan efektivitas Sparse Retrieval Model secara signifikan, terutama karena kurang presisinya kata yang diekspansi dan keterbatasan vocabulary model embedding yang digunakan. Penggunaan BERT untuk kontekstual checking juga belum memberikan hasil yang optimal.

.....This research aims to improve the effectiveness of Information Retrieval models in the legal field by utilizing Term Expansion methods. The background of this research is based on the need of legal practitioners for a system that can automatically and efficiently search for information in large and complex legal document databases. The proposed approach in this research uses a Cascade-Ranking Model architecture, which consists of two main stages: a Sparse Retrieval Model for initial searches and a Re-ranking algorithm for reordering search results. This study focuses on exploring the performance of the Sparse Retrieval Model through the application of Term Expansion techniques, which include Query Expansion and Document Expansion. Query Expansion is conducted using various approaches, including non-contextual word embedding, contextual word embedding with BERT, and Local Analysis. Document Expansion involves adding predicted relevant queries into the documents to enhance their match with user queries. The results of the experiments show that no scenario is absolutely superior in all evaluation metrics used. Some scenarios show advantages in one particular metric but experience a decline in others. In one

scenario, the R@20 metric was successfully increased from 0.896 to 0.906, while in another scenario, the P@5 metric was increased from 0.186 to 0.192. Additionally, in a scenario that used legal context stopwords to filter the words to be expanded with word embedding, R@20 was increased from 0.896 to 0.916, achieving a p-value of 0.045. This p-value below 0.05 indicates a significant improvement in that scenario. Although some metrics improved, Query Expansion did not always significantly enhance the effectiveness of the Sparse Retrieval Model, primarily due to the imprecision of the expanded words and the limitations of the embedding model's vocabulary used. The use of BERT for contextual checking also did not yield optimal results.