

Pengembangan sistem perolehan informasi untuk dokumen legal Indonesia

Yans Sukma Pratama, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=122890&lokasi=lokal>

Abstrak

Dokumen legal Indonesia memiliki karakteristik yang membedakannya dengan dokumen umum.

Karakteristik tersebut ada yang kasat mata seperti struktur penulisan dan ada pula yang tidak kasat mata seperti frekuensi kata dan sebaran kata dalam dokumen. Adanya karakteristik tersebut membuat perolehan informasi pada dokumen legal berbeda dengan dokumen teks biasa, misalnya isu pemanfaatan struktur dokumen legal sebagai unit perolehan informasi dan penggunaan stopwords khusus dokumen legal. Dalam penelitian tugas akhir ini, dikembangkan sistem perolehan informasi untuk dokumen legal Indonesia. Sistem yang dikembangkan mencakup tiga komponen utama dalam perolehan informasi, yakni pengindeksan, pencarian, dan pemeringkatan hasil. Pengindeksan dilakukan dengan menggunakan dua buah pendekatan: pengindeksan elemen terbesar dan pengindeksan elemen yang dianggap berharga sebagai unit perolehan informasi, yaitu elemen bab dan elemen pasal. Pendekatan ini diambil dengan tujuan memanfaatkan struktur penulisan pada dokumen legal. Untuk itu, dalam penelitian ini digunakan koleksi dokumen legal yang sudah ditandai dengan tag XML. Isu penggunaan stopwords khusus dokumen legal tidak ditangani dalam penelitian ini, mengingat kata-kata yang umum muncul dalam dokumen legal direpresentasikan dengan tag-tag XML. Sejumlah ujicoba dilakukan untuk mengevaluasi sejauh mana kinerja setiap jenis pendekatan. Data yang digunakan sebagai ujicoba adalah undang-undang Republik Indonesia yang diundangkan mulai tahun 1983 hingga 2009. Hasil ujicoba menunjukkan bahwa elemen terbesar, yakni elemen undang-undang merupakan unit perolehan informasi terbaik dibanding elemen bab dan elemen pasal. Sementara itu, elemen pasal merupakan elemen yang memiliki dokumen relevan terbanyak dalam hasil pencarian.

<hr>Indonesian legal documents have some characteristics that differs it from general documents. The characteristics can be classified into two types: characteristic that can be seen clearly and characteristic that can be detected only by using statistical linguistic methods. The structural writing of legal document is a subset of the first type characteristic while words frequency and words distribution among collection can be included into the second type. Existence of those characteristics made information retrieval in legal document has differences compare to retrieval in general text collection, such as using document structure as a retrieval unit and using special stopwords for legal document. In this undergraduate thesis, we developed information retrieval system for Indonesian legal document. The system employed three information retrieval main components: indexing, searching, and ranking. We used two approaches in indexing step: made biggest element and valuable element (chapter and article) as an indexing unit. The approaches taken have a purpose to exploit the structural writing of legal document. We used legal document which has been tagged with XML syntax to make it easier to afford the purpose. The issue of using special stopwords for document legal were omitted in this research because it has been replaced by XML syntax. Finally, we conducted some experiments to evaluate performance of each indexing types. We used Indonesian law documents which released from 1983 until 2009 as experiments data. We conclude that the biggest element is the best indexing unit among others. Beside that, we found that article element is the most frequent

element which occur in search result lists.