

BAB 1 PENDAHULUAN

Pada bab ini dipaparkan latar belakang penelitian, rumusan masalah, tujuan penelitian, ruang lingkup penelitian, tahapan penelitian, dan sistematika penulisan laporan.

1.1 Latar Belakang

Undang-undang sebagai sebuah aturan dan landasan hukum terdiri atas kata-kata, kalimat-kalimat yang memiliki makna dan direpresentasikan dalam bentuk tulisan atau teks tertulis. Sebagai sebuah dokumen teks, undang-undang memiliki ciri khas yang membedakannya dari dokumen teks pada umumnya seperti artikel koran, email, surat, atau buku. Ciri tersebut bersifat kasat mata sehingga seorang pembaca dapat dengan segera mengetahui walau hanya dengan membaca sekilas. Ciri tersebut adalah struktur penulisan undang-undang, yakni bagaimana ketentuan-ketentuan yang ada dirumuskan ke dalam bagian-bagian seperti bab-bab dan pasal-pasal.

Struktur penulisan undang-undang Republik Indonesia mengikuti suatu aturan baku sebagaimana diatur oleh Undang-undang nomor 10 tahun 2004 tentang Pembentukan Peraturan Perundang-undangan. Undang-undang tersebut membagi isi undang-undang menjadi empat bagian pokok, yakni judul, pembukaan, batang tubuh, dan penutup. Bagian judul berisi nama aturan, nomor dan tahun undang-undang. Pada bagian pembukaan dibahas alasan dan landasan hukum pembentukan undang-undang. Sementara itu, pada bagian batang tubuh, inti dari undang-undang dikelompokkan dalam bab-bab, bagian, subbagian, paragraf, pasal, dan ayat-ayat. Bagian penutup memuat pengesahan undang-undang serta pejabat yang mengesahkan undang-undang.

Struktur penulisan dan pengelompokkan isi tersebut merupakan ciri yang kasat mata. Adapula ciri dari dokumen undang-undang yang tidak kasat mata seperti frekuensi kata, panjang kata, dan distribusi kata dalam koleksi sebagaimana

dipaparkan oleh Noortwijk. Noortwijk menggunakan metode statistik linguistik untuk mengukur ciri-ciri tersebut pada dua jenis koleksi, yakni dokumen umum berbahasa Belanda dan dokumen legal Belanda (undang-undang dan kasus hukum) dengan total 50 juta kata [NOO96]. Hasil penelitiannya sebagai berikut.

- a. Jumlah kata unik pada dokumen umum lebih banyak ketimbang dokumen legal.
- b. Frekuensi pengulangan kata unik pada dokumen legal lebih tinggi daripada dokumen umum.
- c. Pada peringkat teratas daftar frekuensi kata, kata-kata pada dokumen legal mempunyai frekuensi lebih tinggi ketimbang dokumen umum.
- d. Distribusi kata pada dokumen legal cenderung lebih rendah, sering ditemukan kata yang tidak universal, yakni kata-kata yang digunakan hanya pada sedikit dokumen saja.

Selain melakukan penelitian terhadap ciri-ciri dokumen legal, Noortwijk juga memaparkan pemanfaatan hasil penelitiannya, yakni pembuatan sistem perolehan informasi dengan korpus dokumen legal tersebut. Noortwijk tidak memaparkan bagaimana keterkaitan antara karakteristik dokumen legal dengan pemanfaatannya dalam pembuatan sistem perolehan informasi.

Penggunaan sistem perolehan informasi dokumen legal pada dasarnya bertujuan memberikan akses yang cepat ke dokumen undang-undang, peraturan-peraturan hukum, dan putusan perkara pengadilan berformat digital yang jumlahnya terus bertambah [BAR06]. Misalnya, jumlah dokumen undang-undang yang ada di Indonesia saat ini hampir mencapai 1200 undang-undang. Jumlah tersebut baru dari satu jenis dokumen legal saja, belum termasuk peraturan-peraturan hukum atau putusan pengadilan. Adanya sistem perolehan informasi tersebut tentu dapat memberikan manfaat kepada mereka yang berkecimpung di dunia hukum terutama saat menghadapi suatu perkara.

Sistem perolehan informasi dengan korpus dokumen legal perlu perhatian tersendiri, terutama isu pemanfaatan karakteristik dokumen legal dalam perancangan sistem perolehan informasi. Karakteristik yang tidak kasat mata, dapat dimanfaatkan dengan membuat sebuah daftar *stopwords* khusus untuk dokumen legal. Kata-kata yang khas dan sering muncul dalam penulisan dokumen legal, seperti “menimbang”, “undang-undang”, “republik”, “indonesia”, “tentang”, “memutuskan”, dapat diabaikan saat proses pengindeksan. Dampaknya indeks yang dibuat akan lebih ramping. Sementara itu, untuk karakteristik yang kasat mata, yaitu struktur penulisan undang-undang dapat dimanfaatkan dengan fitur yang dapat mengeksploitasi struktur dokumen sehingga pengguna memperoleh bagian dokumen yang spesifik dan relevan dengan kueri.

Terkait dengan eksploitasi struktur dokumen, pada umumnya sistem perolehan informasi memberikan hasil berupa dokumen undang-undang secara keseluruhan. Padahal bisa saja yang dibutuhkan pengguna bukan undang-undang secara keseluruhan, melainkan bagian tertentu saja seperti dasar hukum, bab atau pasal. Pada kasus tersebut, mengembalikan hasil pencarian berupa dokumen undang-undang secara utuh bukan pilihan bijak. Apabila yang diinginkan pengguna adalah informasi seputar pasal, maka sistem diharapkan dapat memberikan jawaban kepada pengguna berupa daftar pasal yang terurut berdasarkan kecocokan terhadap kueri.

Salah satu solusi permasalahan tersebut adalah melakukan *pre-processing* terhadap korpus, yakni dengan *tagging* (memberi penanda) pada struktur dokumen legal. Dokumen legal yang strukturnya telah diberi penanda dapat dieksploitasi sehingga memberikan jawaban yang lebih spesifik [MOE05]. Salah satu standar yang telah dipergunakan secara luas dalam *tagging* atau strukturisasi dokumen adalah *eXtensible Markup Language* (XML).

Pemanfaatan XML untuk penandaan dokumen legal Indonesia telah dimulai sejak tahun 2006 oleh Mawaddah. Mawaddah membuat sistem ekstraksi informasi yang mampu memberi penanda XML pada dokumen undang-undang Republik

Indonesia secara otomatis. Penelitian serupa dengan pendekatan berbeda dilakukan oleh Nuraminah [NUR07]. Setahun kemudian, Violina menyempurnakan penelitian yang dilakukan Mawaddah agar sistem dapat mengenali undang-undang jenis lain, yakni undang-undang perubahan [VIO08]. Violina juga menambahkan fitur untuk menggambarkan keterkaitan antar undang-undang.

Hingga saat ini, penelitian tentang dokumen legal Indonesia baru terbatas pada dua hal tersebut, yakni strukturisasi dokumen legal dengan format XML dan pembuatan fitur pengenalan keterkaitan antar undang-undang. Belum ada penelitian lanjutan tentang pemanfaatan undang-undang XML yang telah dibuat oleh Mawaddah, Nuraminah, dan Violina, misalnya penerapannya dalam sistem perolehan informasi. Padahal di beberapa negara lain, penelitian tentang perolehan informasi dengan topik dokumen legal berformat XML sudah lama dimulai.

Perolehan informasi dengan korpus dokumen berformat XML lazim disebut dengan *XML retrieval* atau sistem perolehan informasi XML. Adapula yang menyebutnya dengan *structured retrieval* karena korpusnya merupakan dokumen yang memiliki struktur. Pada sistem perolehan informasi XML, pengguna berharap yang dikembalikan saat proses perolehan informasi adalah bagian dokumen yang paling spesifik [MAN08]. Untuk itu, diperlukan pengetahuan tentang unit mana dari struktur dokumen XML yang akan diindeks. Unit pengindeksan merupakan unit yang akan dikembalikan saat perolehan informasi. Menentukan unit pengindeksan pada dokumen XML bukanlah hal yang mudah. Apabila unit yang diindeks terlalu besar, muncul kekhawatiran akan lebih banyak informasi (elemen-elemen) yang tidak relevan diperoleh.

Pencarian tentang “keselamatan penerbangan” bisa jadi yang dikembalikan adalah satu unit undang-undang dimana kasus kecelakaan hanya dibahas di salah satu bab dari belasan bab dalam undang-undang tersebut. Sebaliknya apabila unit pengindeksan terlalu kecil muncul kekhawatiran akan ada informasi penting yang

terlewatkan. Manning menyebutkan beberapa strategi pengindeksan elemen XML sebagai berikut [MAN08].

- Membuang semua elemen kecil
- Membuang elemen yang diperkirakan tidak akan dicari oleh pengguna
- Pertahankan elemen yang dianggap berharga sebagai hasil pencarian untuk dijadikan unit pengindeksan.

Untuk mengetahui elemen mana yang berharga, Kamps melakukan perbandingan 2 jenis pendekatan, yakni dengan mengindeks elemen terbesar dan mengindeks seluruh elemen [KAM03]. Hasilnya, elemen terbesar dengan elemen yang lebih kecil memiliki hasil yang kompetitif, tetapi baik pengguna maupun penilai bersepakat bahwa unit XML terbesar merupakan elemen yang berharga.

Penelitian dengan metode serupa dilakukan oleh Sigurbjornsson, yakni membandingkan elemen terbesar dan seluruh elemen. Ujicoba perolehan informasi dilakukan untuk setiap jenis indeks. Hasil penelitian menunjukkan bahwa proses perolehan informasi dengan target elemen terbesar memberikan hasil lebih baik dibanding elemen lainnya [SIG03].

Laporan tugas akhir ini membahas penelitian tentang penerapan sistem perolehan informasi pada dokumen legal Indonesia, lebih khusus dokumen undang-undang. Dokumen undang-undang yang digunakan merupakan dokumen yang sudah diberi penanda elemen XML menggunakan sistem yang dikembangkan oleh Violina. Pendekatan yang digunakan adalah pengindeksan elemen terbesar, yakni elemen dokumen (selanjutnya disebut undang-undang), pengindeksan elemen bab dan elemen pasal.

Pemilihan elemen undang-undang didasarkan pada penelitian Sigurbjornsson dan Kamps yang menunjukkan bahwa elemen terbesar merupakan elemen yang informatif dan berharga sebagai hasil pencarian. Elemen pasal dipilih karena seluruh substansi peraturan perundang-undangan dimuat dalam elemen tersebut [IND04]. Sementara itu, pemilihan elemen bab disebabkan pada elemen ini terjadi pengelompokkan pasal-pasal yang memiliki kesamaan materi.

1.2 Rumusan Permasalahan

Berdasarkan latar belakang yang telah disampaikan, masalah yang terkait penelitian tugas akhir ini sebagai berikut:

1. Bagaimana penerapan sistem perolehan informasi pada dokumen legal Indonesia, yakni undang-undang Republik Indonesia berformat XML?
2. Sejauh mana kinerja sistem perolehan informasi pada unit pengindeksan undang-undang, bab, dan pasal?

1.3 Tujuan Penelitian

Tujuan penelitian tugas akhir ini adalah mengembangkan sistem perolehan informasi untuk koleksi undang-undang Republik Indonesia.

1.4 Ruang Lingkup Penelitian

Penelitian tugas akhir ini dibatasi dalam ruang lingkup sebagai berikut.

1. Korpus berupa dokumen undang-undang berformat XML yang dihasilkan oleh Sistem Ekstraksi Informasi yang dibuat oleh Violina [VIO08].
2. Proses *indexing* dan *retrieval* menggunakan *open source search engine* Indri versi 2.8 [LEM09].
3. Sistem memberikan pilihan kepada pengguna untuk memilih elemen yang ingin diperoleh sehingga hasil pencarian berupa daftar berkas sesuai dengan pilihan pengguna.
4. Sistem yang dikembangkan merupakan sistem berbasis web.

1.5 Tahapan Penelitian

Pelaksanaan penelitian tugas akhir ini meliputi tahapan-tahapan sebagai berikut:

- Studi pustaka tentang penelitian di bidang *xml retrieval* dan *legal retrieval*, serta penelitian tentang perbandingan *tools* perolehan informasi.
- Mempelajari sistem yang dikembangkan oleh Violina serta mempelajari dokumen undang-undang XML yang dihasilkan sistem tersebut.
- Perancangan sistem perolehan informasi yang meliputi perancangan pembuat unit pengindeksan dan perancangan antarmuka.

- Implementasi fitur-fitur yang telah dirancang.
- Pengujian kinerja sistem dan evaluasi hasil pengujian.

1.6 Sistematika Penulisan Laporan

Laporan tugas akhir ini terdiri atas enam bab, dengan sistematika penulisan sebagai berikut.

- Bab 1 Pendahuluan. Bab ini terdiri dari latar belakang, rumusan masalah, tujuan penelitian, ruang lingkup penelitian, tahapan penelitian, dan sistematika penulisan laporan.
- Bab 2 Teori Penunjang. Bab ini berisi landasan teori tentang sistem perolehan informasi, sistem perolehan informasi XML, undang-undang Republik Indonesia berformat XML, *open source search engine*, *Indri retrieval model*, metode evaluasi, dan pembahasan.
- Bab 3 Perancangan Sistem. Bab ini terdiri dari perancangan pembuat unit pengindeksan dokumen XML dan perancangan antarmuka.
- Bab 4 Implementasi. Bab ini berisi implementasi pembuat unit pengindeksan, konfigurasi pengindeks Indri, konfigurasi pemroses kueri Indri, dan implementasi antarmuka.
- Bab 5 Ujicoba dan Analisis Hasil. Bab ini menjelaskan koleksi data ujicoba, skenario ujicoba, lingkungan ujicoba, hasil ujicoba, dan analisis hasil ujicoba.
- Bab 6 Penutup. Bab ini berisi kesimpulan dan saran.