

BAB 1 PENDAHULUAN

Bab ini membahas latar belakang penelitian, perumusan masalah, tujuan penelitian, rumusan masalah, ruang lingkup penelitian, metodologi penelitian, dan sistematika penulisan dalam riset yang dilakukan oleh penulis.

1.1 Latar Belakang

Perkembangan teknologi informasi saat ini telah mengubah bentuk repositori dokumen yang ada dari media penyimpanan berupa berkas-berkas yang berbentuk fisik (buku, jurnal, dsb.) menjadi media penyimpanan elektronik. Hal ini juga mendorong perkembangan sistem repositori dokumen dari bentuk perpustakaan ataupun gudang arsip menjadi sebuah perpustakaan digital, yaitu perpustakaan yang memiliki struktur dan kapasitas untuk penyimpanan seluruh koleksinya dalam bentuk digital [Baeza-Yates,1999].

Turut berkembang pesatnya internet, maka jumlah data yang disimpan dan diakses di dalam dunia maya akan semakin banyak dan beragam. Selain itu, internet juga meningkatkan kebutuhan akan informasi yang sangat beragam serta dinamis. Oleh karena itu, mesin pencari seperti menjadi alat yang penting dalam dunia Internet untuk dapat mengakses informasi yang diinginkan.

Kebutuhan informasi di Internet juga tidak terbatas dalam satu bahasa. Menurut penelitian oleh Global Reach , pemakai internet yang berbahasa Inggris berkisar 35.8%, sisanya, adalah orang-orang yang berkomunikasi dengan bahasa selain Inggris. Dengan begitu terlihat bahwa kebutuhan informasi serta ketersediaan informasi di internet terdiri dari berbagai macam bahasa. Selain itu, untuk mendapatkan informasi yang sesuai, setiap orang menginginkan informasi tanpa harus terbatas bahasa. Dengan begitu muncul kenaikan pesat untuk dapat mengakses informasi tanpa terhalang budaya dan bahasa.

Seringkali informasi yang lengkap tidak tersedia hanya dalam satu bahasa saja. Dengan begitu, muncul permintaan yang kuat untuk dapat mengakses informasi dalam bahasa asing serta dapat membaca, menginterpretasikan dan menggabungkan informasi-informasi tersebut. Akan tetapi, sistem perolehan

informasi tidak dapat menginterpretasikan berbagai macam bahasa yang ada di dunia. Hal ini menghambat akses ke informasi yang diinginkan.

Untuk dapat mengatasi masalah tersebut, maka perolehan informasi lintas bahasa (*Cross Language Information Retrieval*) berkembang. Perolehan informasi lintas bahasa diteliti dan dikembangkan untuk mengatasi masalah utama yang muncul untuk pengaksesan informasi lintas bahasa yaitu penerjemahan. Bila suatu kata kunci yang digunakan untuk memperoleh informasi diterjemahkan tanpa memperhatikan konteks atau topik dari informasi yang diinginkan maka akan menyebabkan hasil perolehan informasi yang tidak tepat karena pada umumnya suatu kata diterjemahkan lebih dari satu kata. Untuk dapat meningkatkan hasil perolehan informasi yang tidak terbatas dalam satu bahasa maka sistem perolehan informasi lintas bahasa terus dikembangkan.

Pada penelitian tentang perolehan informasi lintas bahasa sudah banyak teknik yang digunakan untuk melakukan penerjemahan antara lain menggunakan kamus dwi-bahasa, penerjemahan dengan mesin penerjemah, penerjemahan dengan menggunakan koleksi dokumen dalam dua bahasa, ataupun dengan memanfaatkan pencatatan mengenai hubungan kata kunci dengan dokumen yang diinginkan [Wei Gao, et. al 2007]. Akan tetapi penelitian dalam perolehan informasi lintas bahasa Inggris – Indonesia masih sangat terbatas meskipun saat ini telah tersedia kamus dwi-bahasa Inggris-Indonesia yang cukup banyak dan gratis. Selain itu, Indonesia juga telah mempunyai sumber-sumber informasi yang cukup banyak dan dapat dimanfaatkan oleh orang-orang asing jika dibutuhkan. Hal inilah yang memotivasi penulis untuk melakukan penelitian dalam bidang perolehan informasi dengan topik perolehan informasi lintas bahasa untuk kueri bahasa Inggris untuk memperoleh dokumen berbahasa Indonesia.

1.2 Rumusan Masalah

Penelitian ini akan terfokus pada masalah-masalah sebagai berikut:

1. Teknik apa yang sebaiknya digunakan untuk meningkatkan perolehan dokumen yang relevan pada sistem perolehan informasi lintas bahasa Inggris-Indonesia?

2. Bagaimana pengaruh pengelompokan dokumen bahasa Inggris maupun dokumen bahasa Indonesia terhadap kualitas perolehan informasi lintas bahasa Inggris-Indonesia?
3. Bagaimana pengaruh teknik perbaikan kueri bahasa Inggris sebelum penerjemahan terhadap hasil perolehan informasi?
4. Bagaimana pengaruh teknik perbaikan hasil penerjemahan bahasa Inggris ke bahasa Indonesia oleh kamus dwi-bahasa Inggris-Indonesia terhadap hasil perolehan informasi?
5. Bagaimana pengaruh teknik perluasan kueri pada hasil perolehan informasi?

1.3 Tujuan

Tujuan dari penelitian ini adalah:

1. Mengimplementasikan teknik-teknik penerjemahan pada perolehan informasi lintas bahasa Inggris-Indonesia
2. Melakukan analisis hasil yang diperoleh oleh teknik-teknik yang digunakan dalam proses penerjemahan yaitu teknik untuk perbaikan hasil penerjemahan kamus, teknik perluasan kueri, teknik pengelompokan dokumen, dan teknik perbaikan kueri sebelum penerjemahan.
3. Mengevaluasi dan membandingkan teknik-teknik penerjemahan yaitu teknik perbaikan hasil penerjemahan kamus (*Post-Translation*), teknik perbaikan kueri sebelum penerjemahan (*Pre-Translation*), teknik pengelompokan dokumen (*Documen Cluster*) yang digunakan untuk mengurangi keambiguan hasil terjemahan serta menjalankan Perluasan Kueri (*Query Expansion*) untuk hasil optimal yang diperoleh dari teknik-teknik diatas.

1.4 Ruang Lingkup

Penelitian dilakukan pada perolehan informasi lintas bahasa Inggris-Indonesia. Koleksi dokumen yang digunakan yaitu koleksi dokumen berbahasa Indonesia dan koleksi dokumen berbahasa Inggris. Koleksi dokumen berbahasa Indonesia merupakan kumpulan berita Kompas dan Tempo yang diperoleh melalui ILPS¹ (*Information and Language Processing System*) yang merupakan bagian dari

¹ <http://ilps.science.uva.nl>

University of Amsterdarm, sedangkan koleksi dokumen bahasa Inggris merupakan hasil pengumpulan secara manual menggunakan mesin pencari Google, sehingga koleksi berasal dari berbagai macam website, antara lain People's Daily², Jakarta Post³ dan Antara⁴.

Kamus dwi-bahasa akan digunakan penulis untuk melakukan proses penerjemahan pada perolehan informasi lintas bahasa Inggris Indonesia. Kamus dwi-bahasa Inggris-Indonesia diperoleh dengan menggabungkan tiga kamus yaitu Hantarto Widjaja⁵, Kamus.net⁶, kamus BPPT (Badan Penerapan dan Pengkajian Teknologi) serta dilakukan penambahan perbendaharaan kata-kata berdasarkan kueri yang dilakukan secara manual oleh penulis.

Teknik-teknik yang akan diterapkan dalam penelitian perolehan informasi lintas bahasa berbasiskan penerjemahan dengan kamus dwi-bahasa Inggris Indonesia. Teknik yang digunakan yaitu teknik perbaikan hasil penerjemahan kamus (*Post-Translation*), teknik perbaikan kueri sebelum penerjemahan (*Pre-Translation*), teknik pengelompokan dokumen (*Documen Cluster*) yang digunakan untuk mengurangi keambiguan hasil terjemahan serta menjalankan Perluasan Kueri (*Query Expansion*) untuk hasil optimal yang diperoleh dari teknik-teknik diatas. Pada akhirnya teknik diatas dievaluasi terhadap hasil perolehan informasi untuk satu bahasa.

1.5 Metodologi Penelitian

Penelitian ini dilakukan dengan menggunakan metodologi penelitian eksperimental. Dalam melakukan penelitian ini, penulis melakukan tiga tahapan besar yaitu:

1. Studi literatur. Penulis mempelajari berbagai pustaka seperti jurnal, makalah penelitian, dan literatur lainnya yang membahas penelitian mengenai perolehan informasi khususnya topik perolehan informasi lintas bahasa. Pustaka yang dipelajari tersebut berasal dari berbagai penelitian mengenai

² <http://english.people.com.cn>

³ <http://www.thejakartapost.com>

⁴ <http://www.antara.co.id>

⁵ <http://hantarto.definitionroadsafety.org>

⁶ <http://www.kamus.net>

perolehan informasi lintas bahasa yang pernah dilakukan oleh orang lain sebelumnya.

2. Eksperimen. Penulis menyiapkan skenario serta masukan untuk melakukan eksperimen terhadap perolehan informasi lintas bahasa Inggris-Indonesia. Penulis membuat data masukan yang berupa kumpulan dokumen yang diperoleh dari artikel surat kabar *online* dan situs-situs berita di internet, pembuatan kumpulan kamus dwi-bahasa Inggris-Indonesia, serta pembuatan penilaian relevansi dokumen terhadap kueri. Eksperimen dilakukan menggunakan teknik perbaikan hasil penerjemahan kamus (*Post-Translation*), teknik perbaikan kueri sebelum penerjemahan (*Pre-Translation*), teknik pengelompokan dokumen (*Documen Cluster*) yang digunakan untuk mengurangi keambiguan hasil terjemahan serta menjalankan Perluasan Kueri (*Query Expansion*) untuk hasil optimal yang diperoleh dari teknik-teknik diatas. Evaluasi perolehan informasi dilakukan dengan membandingkan hasil-hasil yang diperoleh dari penerapan teknik tersebut serta melihat hasil perolehan informasi dengan menggunakan penilaian relevansi dan daftar perolehan informasi sebagai masukan.
3. Analisis. Penulis menganalisis hasil eksperimen terhadap penerapan metode yang digunakan, melakukan perbandingan antara teknik-teknik perolehan informasi lintas bahasa kemudian menarik kesimpulan.

1.6 Sistematika Penulisan

Susunan Laporan penelitian tugas akhir ini terdiri dari 5 bab dengan sistematika sebagai berikut. Bab 1 yaitu Pendahuluan, membahas mengenai latar belakang, rumusan masalah, ruang lingkup penelitian, tujuan penelitian, metode penelitian yang dilakukan dan sistematika penulisan. Pada bab 2 dibahas landasan teori yang menjadi dasar dari penelitian ini. Bab 3 membahas eksperimen yang dilakukan oleh penulis untuk proses perolehan informasi lintas bahasa menggunakan kamus dwi-bahasa. Hasil dari eksperimen serta analisis akan dibahas pada bab 4. Bab 5 menutup laporan dengan memberikan kesimpulan dari penelitian serta saran untuk penelitian berikutnya.