

Perolehan informasi lintas bahasa indonesia-inggris berdasarkan korpus paralel dengan menggunakan metoda mutual information dan metoda similarity thesaurus

Syandra Sari, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=126210&lokasi=lokal>

Abstrak

Perolehan informasi lintas bahasa adalah bidang perolehan informasi yang semakin berkembang seiring dengan pesatnya perkembangan Internet di seluruh dunia. Perolehan informasi lintas bahasa adalah sistem perolehan informasi yang memungkinkan pengguna memberikan query dalam bahasa yang berbeda dengan bahasa dokumen. Dalam perolehan informasi lintas bahasa, masalah perbedaan bahasa diatasi dengan proses penerjemahan dokumen atau penerjemahan query. Sumber daya bahasa yang digunakan untuk proses penerjemahan pada perolehan informasi lintas bahasa dapat berupa kamus dwi bahasa, mesin penerjemah atau korpus paralel. Penelitian ini mengembangkan sistem perolehan informasi lintas bahasa Indonesia-Inggris berdasarkan korpus paralel. Padanan kata dalam bahasa Indonesia-Inggris diperoleh dengan menggunakan dua metoda yaitu informasi mutual dan similarity thesaurus. Selanjutnya untuk meningkatkan hasil perolehan dokumen berdasarkan korpus paralel, maka diterapkan teknik perluasan query yaitu pseudo-relevance feedback. Kinerja terbaik dari metoda informasi mutual dan penerapan teknik perluasan query mencapai 41,86 % dari kinerja perolehan informasi lintas bahasa monolingual; 80,5 % dari kinerja perolehan informasi lintas bahasa Indonesia-Inggris dengan menggunakan kamus; serta mencapai 54% dari kinerja sistem perolehan informasi lintas bahasa Indonesia-Inggris menggunakan mesin penerjemah Transtool. Sedangkan kinerja terbaik dari metoda similarity thesaurus dan penerapan teknik perluasan query mencapai 55,61 % dari kinerja perolehan informasi lintas bahasa monolingual; mencapai 107% dari kinerja perolehan informasi lintas bahasa Indonesia-Inggris dengan menggunakan kamus; serta mencapai 71% dari kinerja perolehan informasi lintas bahasa Indonesia-Inggris dengan menggunakan mesin penerjemah Transtool. Berdasarkan hasil tersebut diatas, terlihat bahwa perluasan query dengan teknik pseudo-relevance feedback yang diterapkan pada sistem perolehan informasi lintas bahasa Indonesia-Inggris berdasarkan korpus paralel dengan metoda informasi mutual dan similarity thesaurus dapat meningkatkan kinerja sampai dengan 25 % dibanding tanpa proses perluasan query. Kemudian, dari hasil tersebut diatas juga dapat diambil kesimpulan umum bahwa sistem perolehan informasi lintas bahasa Indonesia-Inggris berdasarkan korpus paralel dapat menjadi alternatif bagi pengembangan sistem perolehan informasi lintas bahasa Indonesia-Inggris. Dan penelitian lebih lanjut mengenai metoda berdasarkan korpus paralel yang lebih baik serta usaha perbaikan kualitas korpus paralel perlu terus dilakukan bagi peningkatan kinerja sistem perolehan informasi lintas bahasa Indonesia-Inggris di masa yang akan datang.