

Evaluasi teknik perolehan informasi pada dokumen web multibahasa

Ananda Budi Prasetya, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=123722&lokasi=lokal>

Abstrak

Perolehan informasi Web adalah cabang perolehan informasi yang mengkhususkan pencarian informasi terhadap dokumen Web. Dokumen yang terdapat pada lingkungan Web memiliki struktur yang berbeda dengan dokumen teks biasa, sehingga menyebabkan pendekatan perolehan informasi pada dokumen Web berbeda dengan dokumen biasa. Pada penelitian ini digunakan koleksi dokumen EuroGOV yang merupakan koleksi dokumen Web multibahasa yang digunakan pada WebCLEF, dan kueri atau topik yang digunakan adalah kueri yang diberikan pihak WebCLEF. Koleksi dokumen dan kueri yang digunakan terdiri dari 15 bahasa Eropa. Pada penelitian ini dipelajari pengaruh berbagai teknik perolehan informasi, yaitu operator kedekatan kata, umpan balik relevan semu, pendeteksian bahasa, analisis pranala, PageRank, dan pengelompokan dokumen, terhadap hasil perolehan informasi Web multi bahasa yang dihasilkan teknik-teknik tersebut. Selain itu pada penelitian ini juga dipelajari kemampuan dari sistem perolehan informasi Indri Search Engine dalam mencari dan memperoleh dokumen Web multibahasa.

Penelitian ini menemukan bahwa dengan melakukan evaluasi secara umum tanpa melihat bahasa, kenaikan nilai perolehan hanya didapatkan dengan menggunakan teknik pendeteksian bahasa sebesar 0.72% dan teknik umpan balik relevan semu sebesar 1.02% - 1.61%. Sedangkan pada teknik analisis pranala, PageRank, dan pengelompokan dokumen, nilai perolehan informasi yang didapatkan menurun sebesar -46.3431% - -0.60%. Dengan melakukan evaluasi berdasarkan bahasa, didapatkan bahwa setiap teknik yang diteliti dan dipelajari memiliki keunggulan masing-masing berdasarkan bahasa. Untuk teknik pengurutan ulang pendeteksian bahasa, peningkatan nilai perolehan informasi terdapat pada 6 bahasa sebesar 0.1478% - 27.1431%, dan penurunan terdapat pada 10 bahasa sebesar -0.7066% - -0.0156%. Untuk teknik analisis pranala dalam, peningkatan nilai perolehan informasi terdapat pada 4 bahasa sebesar 6.5585% - 77.1804%, dan penurunan terdapat pada 4 bahasa sebesar -60.9202% - -0.3339%, dan untuk analisis pranala luar terdapat kenaikan pada 9 bahasa sebesar 0.7151% - 134.9881%, dan penurunan terdapat pada 6 bahasa sebesar -23.2941% - -1.6330%, sedangkan untuk kombinasi analisis pranala dalam dan luar terdapat pada 5 bahasa sebesar 0.2809% - 100%, dan penurunan terdapat pada 10 bahasa sebesar -92.0830% - -3.1030%. Untuk teknik PageRank, peningkatan nilai informasi terdapat pada 4 bahasa sebesar 5.1083% - 77.1804%, dan penurunan terdapat pada 10 bahasa sebesar -96.4844% - -2.6968%. Dan untuk pengelompokan berdasarkan bahasa, peningkatan nilai informasi terdapat pada 3 bahasa sebesar 3.6440% - 38.7337%, dan penurunan terdapat pada 9 bahasa sebesar -53.6322% - -1.6715%. Untuk pengelompokan berdasarkan topik, peningkatan tertinggi terdapat pada satu bahasa, yaitu Spanyol sebesar 12.4754% dan penurunan terdapat pada 5 bahasa sebesar -97.5600% - -35.6133%. Untuk teknik umpan balik relevan semu, peningkatan nilai perolehan informasi terdapat pada 7 bahasa sebesar 0.0472% - 5.1523%, penurunan hanya terdapat pada dua bahasa sebesar -4.2597% - -0.2760%. Berdasarkan penelitian ini disimpulkan bahwa sistem perolehan informasi Indri Search Engine dapat mencari dan memperoleh dokumen Web multibahasa dengan baik, dan teknik perolehan Web standar dapat diterapkan juga pada koleksi dokumen

Web multibahasa.