

ABSTRAK

Kemajuan teknologi internet telah memberikan dampak pada ketersediaan informasi melalui web yang begitu pesat. Demikian pula, jumlah informasi yang tersedia pada situs web bertambah secara cepat. Di sisi lain informasi yang dapat diperoleh dari web search engine merupakan semua hal yang berkaitan dengan kata kunci yang dicari. Hal ini menyebabkan pengguna terpaksa menyaring untuk mendapatkan dokumen yang relevan. Oleh karena itu diperlukan cara untuk mengelompokkan (*clustering*) banyaknya informasi yang tersedia, yang dibutuhkan pengguna sehingga memudahkan pengguna untuk mendapatkan dokumen yang diinginkan.

Dalam tugas akhir ini dibuat perangkat lunak untuk mengelompokkan dokumen web dengan menggunakan metode *Suffix Tree Clustering*. Metode ini menggunakan algoritma suffix tree dalam mengidentifikasi secara efisien sekumpulan dokumen yang berbagi frase, dan menggunakan informasi tersebut untuk membuat kelompok dan meringkas kesimpulan yang berupa kelompok gabungan yang akan diberikan ke pengguna. Metode ini dilakukan dengan cara persiapan dokumen, pembersihan dokumen, pembuatan suffix tree, pengidentifikasian kelompok dasar, penggabungan kelompok dasar menjadi kelompok gabungan, pemberian nilai terhadap kelompok gabungan. Perangkat lunak yang dibuat menerima masukan berupa dokumen web hasil pencarian dari web search engine tertentu berdasarkan kata kunci yang diberikan pengguna. Dari hasil yang diperoleh tadi, kemudian dilakukan proses pengelompokan dengan menggunakan algoritma suffix tree. Setelah dilakukan pemrosesan, hasil pengelompokan tersebut ditampilkan pada pengguna.

Perangkat lunak tersebut telah dilakukan ujicoba verifikasi untuk mengetahui kesalahan yang mungkin terjadi. Uji coba ini dilakukan dengan menggunakan berbagai kata kunci dan jumlah dokumen maksimal yang bervariasi. Hasil uji coba menunjukkan bahwa jumlah maksimal pencarian mempengaruhi waktu yang diperlukan untuk mengelompokkan dokumen. Selain itu, adanya parameter (α) untuk mencari kemiripan dokumen, dari hasil uji coba menunjukkan bahwa nilai 0.6 merupakan parameter yang optimal karena untuk nilai yang lebih dan kurang dari 0.6 kenaikan jumlah kelompok relatif kecil.