

ABSTRAK

ABSTRAK

Pengembangan Algoritma Kompresi Text Berbasis Kata Untuk Dokumen Berbahasa Indonesia

Ardiles Sinaga

Supervisor : Hertog Nugroho, Ph.D

Co Supervisor : Dr. Adiwijaya S.Si., M.Si.

Teknologi informasi berkembang sangat pesat saat ini, khususnya untuk penanganan data. Data merupakan aset berharga bagi semua orang, terutama bagi perusahaan yang lebih besar. Khusus untuk perusahaan besar yang sudah memiliki cabang di beberapa tempat. Transmisi data dari kantor pusat ke kantor cabang membuat perusahaan harus menyediakan alat yang baik untuk melakukannya. Perusahaan-perusahaan ini juga membutuhkan alat yang dapat digunakan untuk kompres data sehingga dapat mengurangi ukuran data itu sendiri.

Ide utama dari Word-Based encoding adalah untuk mengidentifikasi setiap kata dari teks sumber, kemudian kata yang diidentifikasi itu akan dicek apakah terdapat huruf besar pada kata tersebut. Setelah itu, kata tersebut akan dicek kembali apakah terdapat simbol atau angka. Lalu setelah itu akan dilakukan stemming terhadap kata untuk memisahkan kata dasar dari imbuhan. Simbol, angka dan imbuhan akan diberi indeks sesuai dengan indeks yang sudah disimpan sebelumnya dalam kamus dasar. Sedangkan kata dasar yang diperoleh setelah proses stemming akan dicek kembali ke kamus dasar, apakah cocok atau tidak. Jika kata dasar tidak cocok pada kamus, maka kata ini akan menjadi entri baru ke dalam kamus.

Pada pengujian ini dilakukan pada data asli yang berukuran antara 10.000 Byte sampai 500.000 Byte dan menggunakan panjang bit kode 16 bit. Hasilnya menunjukkan bahwa rasio kompresi metode yang diusulkan sebanding dengan aplikasi populer RAR hingga 200 kbyte, sementara waktu pengolahannya jauh lebih baik daripada metode urutan dari karakter yang dibalik pada algoritma LZW

Kata kunci: Kompresi data, WB-LZW, Berbasis Kata, Stemming, Tree, Kamus dasar, Kamus utama