

Abstrak

Pada *information retrieval system*, *inverted index* digunakan untuk mengevaluasi suatu *query*. Semakin banyak dokumen yang harus disimpan, maka semakin besar pula *inverted index* yang terbentuk. Dan semakin banyak pula *query* yang harus diproses pada pencarian dokumen-dokumen tersebut. Maka, dibutuhkan suatu cara optimisasi performansi untuk memenuhi kebutuhan dalam penyimpanan *inverted index* yang semakin besar dan pemrosesan *query* yang semakin banyak, salah satunya adalah kompresi *inverted index*.

Kompresi *inverted index* diharapkan dapat mengurangi kebutuhan ruang penyimpanan *inverted index* dan meningkatkan penggunaan *cache* di memori. Salah satu metode kompresi *inverted index* adalah *Gamma code*, yang mengubah *integer* menjadi *binary codeword*. Data yang dikompresi berupa ID dokumen dan frekuensi term.

Pada tugas akhir ini, dilakukan pengujian penerapan kompresi *inverted index* pada *information retrieval system* dengan koleksi dokumen yang berukuran kecil dan koleksi dokumen yang berukuran besar. Dari analisis hasil pengujian, diperoleh kesimpulan bahwa *Gamma code* dapat menghasilkan performansi yang baik dalam hal ukuran *inverted index* pada koleksi dokumen yang besar, karena term-termnya tersebar di banyak dokumen, sehingga pengkodean *Gamma* lebih pendek pada setiap *posting*. Juga menghasilkan performansi yang baik dalam hal ukuran *inverted index* pada koleksi dokumen yang besar, karena rata-rata rasio waktu pemrosesan *query*-nya lebih kecil dibandingkan pada koleksi dokumen yang kecil .

Kata kunci: kompresi, *inverted index*, *Gamma code*, *integer*