

ANALISIS DAN IMPLEMENTASI KOMPRESI DATA MENGGUNAKAN METODE PPMC

Bagus Rachmad Bintoro¹, Mahmud Imrona², M.t³

¹Teknik Informatika, Fakultas Teknik Informatika, Universitas Telkom

Abstrak

Kompresi data sudah menjadi hal yang penting sekarang ini, dimana kebutuhan akan data dan informasinya akan semakin meningkat namun sebisa mungkin akan menggunakan kapasitas media transfer ataupun media penyimpanan yang kecil. Dengan menggunakan kompresi, pengiriman data akan menjadi lebih cepat dan hemat waktu. Salah satu algoritma kompresi data yaitu "Prediction by Partial Matching" (PPM) yang dikembangkan oleh Cleary dan Witten yang mampu menghasilkan tingkat kompresi yang sangat tinggi.

Dalam tugas akhir ini, penulis menggunakan metode PPMC, yang kemudian dibandingkan hasil kompresinya dengan metode lossless lain yaitu dengan metode Huffman. Dari hasil pengujian, maka didapat hasil rasio kompresi rata-rata untuk metode Huffman adalah 42,26%. Sedangkan pada metode PPMC didapatkan hasil rasio kompresi rata-ratanya 68,72 %. Dengan demikian, rasio kompresi metode PPMC lebih baik dibanding metode Huffman.

Kata Kunci : kompresi, lossless , PPMC, Huffman

Abstract

Data compression has become an important right now, where requirement for data and information will be increase as much as possible, but will use any media capacity storage media transfer or small. By using compression, data transmission will be faster and save time. One of the data compression algorithms are "Prediction by Partial Matching" (PPM) developed by Cleary and Witten were able to produce a very high compression rate.

In this final project, the author uses the method PPMC, which is then compared to the results of the compression with other lossless method is by Huffman method. From the test results, the obtained results an average compression ratio for Huffman method was 42.26%. While the PPMC method obtained results the average compression ratio of 68.72%. Thus, the compression ratio PPMC method is better than the Huffman method.

Keywords : Compression, Lossless, PPMC, Huffman

1. PENDAHULUAN

1.1 Latar Belakang

Seiring dengan kemajuan teknologi informasi, berkembang pula berbagai teknik dan teknologi untuk kompresi data. Melalui teknologi kompresi, ukuran data akan menjadi lebih kecil dari ukuran aslinya sehingga proses pengiriman dan penyebaran data dapat menjadi lebih cepat dan juga dapat mengurangi kebutuhan kapasitas media penyimpanan. Kompresi merupakan proses pengubahan sekumpulan data menjadi bentuk kode dengan tujuan untuk menghemat kebutuhan tempat penyimpanan dan waktu untuk transmisi data.

Kompresi data dilakukan dengan mengkodekan setiap karakter di dalam pesan atau data dengan mengkodekan setiap karakter di dalam pesan atau data dengan kode yang lebih pendek. Saat ini banyak sekali algoritma yang digunakan untuk mengkompresi data, diantaranya seperti algoritma Huffman, RLE, *Arithmetic Coding*, *Prediction by Partial Matching* (PPM), dan lain-lain.

Terdapat 2 jenis teknik kompresi, yang pertama kompresi *Lossy*. Dimana, jenis kompresi ini menimbulkan perubahan data output setelah terjadinya proses kompresi. Kompresi jenis ini biasanya digunakan didalam file multimedia seperti *file image*, *audio* dan *video*. Yang kedua, Teknik kompresi ini bersifat *lossless* dimana tidak menghilangkan informasi sedikitpun. Salah satu Algoritma yang menggunakan teknik adalah algoritma PPMC. *Prediction by Partial Matching* (PPM) merupakan metode untuk memprediksi simbol berikutnya dari pengetahuan tentang simbol-simbol yang sudah ada/dikenali sebelumnya. Metode PPM dapat diimplementasikan dengan cara digabungkan dengan teknik kompresi lain seperti *Arithmetic Coding*, *Huffman*, dan teknik-teknik kompresi lain yang berbasis *dictionary*. Proses kompresi terhadap *English text* dengan *text* yang biasanya tentu berbeda. Hal ini dikarenakan pada *English text* banyak memiliki frase/kombinasi yang sering muncul dalam kalimat pada *English text* tersebut, seperti kata *the*, *is*, *not*, dll. Hal inilah kemudian menjadi ide dimana algoritma PPM ini akan menampung semua kata kata tersebut dalam sebuah table yang berbasis *dictionary*. Dimana ketika sebuah kata tersebut muncul

kembali, maka akan diperiksa pada tabel/kamus kata yang ada, dan akan hanya akan menambah isi tabel jika ketika kata yang ditemui tidak terdapat dalam tabel/kamus data yang telah dibuat tersebut. Oleh karena itulah, implementasi algoritma ini perlu digabungkan dengan algoritma kompresi lain yang berbasis *dictionary* sehingga akan menghasilkan tingkat kompresi yang tinggi.

1.1 Perumusan Masalah

Rumusan masalah yang akan dikaji dalam Tugas Akhir ini adalah bagaimana cara mengimplementasikan algoritma kompresi *Prediction by Partial Matching type C* (PPMC) pada *file* tipe teks (dokumen). Kemudian meneliti pengaruh algoritma kompresi *Prediction by Partial Matching* (PPM) pada hasil kompresi. Dan mengukur performansi dari penerapan algoritma *Prediction by Partial Matching* (PPMC) ditinjau dari rasio kompresi.

Hipotesis awal dari Tugas Akhir ini adalah metode algoritma kompresi *Prediction by Partial Matching type C* (PPMC) adalah lebih baik dalam kompresi jenis teks dibanding metode lain, seperti Huffman.

1.2 Tujuan

Tujuan yang ingin dicapai dalam pelaksanaan Tugas Akhir ini adalah studi literatur mengenai teknik algoritma *Prediction by Partial Matching* (PPM) untuk kompresi teks, dengan menggunakan varian PPM tipe C. Setelah itu mengimplementasikan kedua algoritma tersebut dalam sebuah aplikasi perangkat lunak. Kemudian menganalisis dan membandingkan performansi dilihat dari rasio kompresi dari teknik algoritma *Prediction by Partial Matching type C* (PPM) dengan metode kompresi *lossless* lain seperti metode Huffman.

1.3 Batasan Masalah

Batasan masalah untuk Tugas Akhir ini adalah :

1. Data uji yang akan dikompresi adalah jenis file teks, misalnya kumpulan file text didalam *Canterbury Corpus* dan *file* uji tambahan seperti: .doc, .xls, .pdf, .rtf.
2. Hal yang akan diukur pada kompresi adalah size kompresi (rasio kompresi).

3. *Text formatting* yang akan di uji menggunakan kode ASCII 8 bit (256 simbol).
4. Pada aplikasi ini maksimal order tidak lebih dari dua.

1.4 Metodologi Penyelesaian Masalah

a. Studi Literatur

Mengumpulkan, mempelajari berbagai bahan literatur terkait dengan kompresi data, macam-macam algoritma kompresi, serta melakukan berbagai pencarian referensi dan sumber-sumber relevan lain sebagai acuan dalam pembuatan Tugas Akhir ini.

b. Analisis dan Desain

Melakukan analisis serta perancangan desain dan sistem yang terstruktur dalam implementasi Tugas Akhir.

c. Implementasi sistem

Melakukan proses pembangunan perangkat lunak/*Coding* sebagai realisasi implementasi algoritma kompresi *Prediction by Partial Matching type C*.

d. Analisis hasil dan Testing

Pada tahap ini dilakukan kegiatan analisis terhadap hasil pengerjaan dan proses penyusunan perangkat lunak. Lalu dilakukan uji terhadap perangkat yang telah dibangun apakah telah berjalan sesuai prosedur dan algoritma. Pengujian juga dilakukan dengan membandingkan ukuran *file* kompresi dengan *file* original secara keseluruhan.

e. Penyusunan laporan

Pada tahap terakhir ini akan dilakukan penyusunan laporan hasil penelitian yang telah dilakukan sejak awal dan rincian langkah-langkah pengerjaan Tugas Akhir.

1.5 Sistematika Penulisan

Sistematika Penulisan dalam Tugas Akhir ini adalah sebagai berikut :

Bab 1. Pendahuluan

Pengenalan dan teori singkat mengenai dunia kompresi.

Bab 2. Landasan Teori

Pemahaman serta penggalian bahan mengenai kompresi dan algoritma PPMC, dan algoritma Huffman.

Bab 3. Analisis Kebutuhan Sistem dan Perancangan Sistem

Gambaran spesifikasi sistem aplikasi perangkat lunak yang akan dibangun ber-dasarkan algoritma yang diajukan.

Bab 4. Implementasi dan Analisis Hasil Pengujian

Proses pembuatan aplikasi perangkat lunak sesuai dua algoritma yang diajukan. Selanjutnya, dilakukan analisis dan perbandingan performansi atas aplikasi perangkat lunak yang dibangun.

Bab 5. Kesimpulan dan Saran

Tahap akhir sebagai penyajian kesimpulan serta saran dari pengerjaan Tugas Akhir.



5. KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil analisis dan implementasi atas algoritma *Prediction by Partial Matching type C* dan algoritma Huffman, maka didapat kesimpulan sebagai berikut:

1. Dari rasio hasil kompresi metode PPMC, yaitu: nilai 75,63% + 66,60% + 63,94% yang dikalkulasikan dan mendapat hasil rata-rata 68,72 % dari semua rata-rata rasio kompresi *file* uji secara umum.
2. Panjang konteks sedikit banyak mempengaruhi hasil kompresi, pada aplikasi diatas kompresi rata-rata tertinggi dicapai pada order $N=1$.
3. Algoritma Huffman memiliki rata-rata rasio kompresi 41,93%.
4. Dari nilai hasil rata-rata rasio kompresi PPM dan rata-rata rasio kompresi Huffman adalah kurang lebih 7 berbanding 4.
5. Jika *file* input yang akan dikompresi mempunyai banyak perulangan dalam informasinya maka semakin baik pula nilai ratio kompresinya.
6. Untuk N lebih besar dari 2 ditemukan banyak *error flushing memory* sehingga pada waktu di dekompresi ke semula menjadi error. Penyebabnya adalah banyaknya symbol untuk *dictionary* yang terdapat didalam file uji selain kode kode ASCII.

5.2 Saran

Saran untuk penelitian mengenai algoritma ini adalah:

1. Perlu untuk menguji teknik kompresi dengan menggunakan algoritma PPM yang lain, kemudian diujikan dengan teknik kompresi lossles lain, seperti misalnya algoritma yang berbasis *dictionary*.
2. Perlu penambahan parameter kecepatan dalam melakukan decompresi maupun kompresi.
3. Dalam aplikasi ini dapat diperbaiki sehingga dapat menghitung konteks nilai N lebih besar dari 2.

REFERENSI

[1] David, Solomon. (2010) "Hand book of data Compression," Springer: Californiav.



DAFTAR PUSTAKA

- [1] Bloom, Charles. "Solving the Problems Context Modeling". California Institute of Technology.
- [2] Cleary, J.G., Witten, I.H. 1984. Data compression using adaptive Coding and partial string matching. IEEE Transactions on Communications, 32(4), 396-402
- [3] David, Solomon. (2010) "Hand book of data Compression," Springer : Californiav.
- [4] <http://corpus.canterbury.ac.nz/>
- [5] Moffat, A. (1990) "Implementing the PPM data compression scheme," IEEE Transactions on Communications, 38(11), 1917-1921.
- [6] Sayood, Khalid. (2006) "Introduction to Data Compression Third Edition", Elsevier: San Francisco, 143-152.
- [7] Uribe, C.(2003) "High performance PPMC Compression Algorithm". Proceedings of the Fourth Mexican International Conference. 135-142.

Telkom
University