

KOMPRESI DATA MENGGUNAKAN METODE HUFFMAN

Ari Wibowo

Jurusan Teknik Informatika, Politeknik Negeri Batam

wibowo@polibatam.ac.id

Abstrak

Kompresi data (pemampatan data) merupakan suatu teknik untuk memperkecil jumlah ukuran data (hasil kompresi) dari data aslinya. Pemampatan data umumnya diterapkan pada mesin komputer, hal ini dilakukan karena setiap simbol yang dimunculkan pada komputer memiliki nilai bit-bit yang berbeda. Pemampatan data digunakan untuk mengurangi jumlah bit-bit yang dihasilkan dari setiap simbol yang muncul. Dengan pemampatan ini diharapkan dapat mengurangi (memperkecil ukuran data) dalam ruang penyimpanan.

Pengkodean dengan metode Huffman dibangun dari panjang variabel kode-kode yang disusun dari bit-bit. Simbol dengan probabilitas yang tinggi akan memperoleh kode-kode paling pendek sedangkan simbol dengan probabilitas paling rendah akan memperoleh kode terpanjang. Kode Huffman mempunyai atribut unik yang sempurna, dalam arti kode-kode tersebut dapat mengembalikan (decoder) kode yang panjang dan decirable dalam arti kode Huffman tersebut tidak menjadi prefik kode Huffman yang lain.

Kata Kunci : Pemampatan, kode huffman, simbol

1. PENDAHULUAN

Secara umum pemampatan data merupakan merubah suatu simbol-simbol menjadi suatu kode-kode. Pemampatan dikatakan efektif jika ukuran perolehan kode-kode tersebut sangat kecil dibandingkan dengan ukuran kode simbol aslinya. Dari suatu kode-kode atau simbol-simbol dasar suatu model akan dinyatakan dalam kode khusus. Secara model sederhana suatu kumpulan data dan aturan-aturan untuk memproses masalah suatu simbol-simbol untuk menentukan suatu kode-kode sebagai hasil keluaran. Dalam pengkodean dengan metode Huffman, kode hasil dari keluaran ditentukan dari probabilitas (nilai kemungkinan yang muncul) paling tinggi akan memperoleh kode dengan nilai bit-bit yang paling sedikit. Simbol-simbol yang mempunyai nilai probabilitas kecil memperoleh kode dengan bit-bit terbanyak.

1.1 Latar Belakang Masalah

Semakin berkembangnya teknologi pada Ponsel maka semakin banyak pula data-data atau file-file yang ingin kita simpan, tetapi kadang kala kapasitas memory yang kita miliki tidak sebanding dengan data yang akan kita simpan. Oleh karena itu data-data yang akan disimpan perlu dikompres terlebih dahulu supaya ukurannya menjadi lebih kecil. Apabila ukuran data dapat dikompres menjadi lebih kecil dari ukuran aslinya, maka secara otomatis memory dapat menyimpan data lebih banyak lagi dan dari segi pengiriman pun akan semakin cepat. Pada saat ini banyak sekali perangkat lunak yang digunakan untuk menangani masalah kompresi data. Perangkat lunak tersebut digunakan untuk mengefisienkan kapasitas penyimpanan data, selain itu juga kompresi data dapat mencegah kerusakan data.

Dalam kompresi data, terdapat 4 (empat) faktor penting yang perlu diperhatikan, yaitu: *Time Process* (waktu yang dibutuhkan dalam menjalankan proses), *Completeness* (kelengkapan data setelah file-file tersebut dikompres), *Ratio Compress* (ukuran data setelah dilakukan kompresi), *Optimally* (perbandingan apakah ukuran file sebelum dikompres sama atau tidak sama dengan file yang telah dikompres).

1.2 Identifikasi Masalah

Masalah-masalah yang melatarbelakangi penulisan tugas akhir ini, diantaranya:

- a. Pada umumnya data yang ingin disimpan berukuran besar, sehingga membutuhkan *memory* yang besar pula. Sedangkan *memory* pada ponsel tidak terlalu besar, sehingga dibutuhkan kompresi data untuk memperkecil ukuran sebuah file.
- b. Data, bentuk atau formatnya bisa berupa teks, gambar, ataupun audio.

1.3 Maksud dan Tujuan

Maksud dari penelitian ini adalah untuk merancang program kompresi data menggunakan algoritma Huffman.

Tujuan dari penelitian ini adalah untuk memperkecil ukuran data yang disimpan pada memori, sehingga memori tersebut bisa menyimpan data dengan jumlah yang lebih banyak.

1.4 Batasan Masalah

Pada penulisan tugas akhir ini ruang lingkup kajian dibatasi hanya pada:

- a. Mensimulasikan dan menguji Aplikasi yang telah dibuat
- b. Algoritma yang digunakan untuk kompresi data yaitu algoritma *Huffman*
- c. Data-data yang akan dikompres berbentuk file teks, image, dan audio

1.5 Metodologi Penelitian

Metode penelitian merupakan cara-cara atau teknik dalam melakukan suatu penelitian. Dalam penulisan tugas akhir ini, penulis menggunakan metode penelitian eksperimental, yang berarti penulis mengadakan kegiatan percobaan untuk melihat hasil dari aplikasi kompresi data yaitu dengan cara melakukan perancangan aplikasi kompresi data, kemudian melakukan simulasi sistem dan terakhir adalah menguji aplikasi yang telah dibuat.

1.5.1 Metode Pengumpulan Data

Metode pengumpulan data yang digunakan penulis dalam penelitian ini ialah Studi Pustaka. Penulis melakukan kajian pustaka yaitu mempelajari buku – buku referensi dan hasil penelitian sejenis sebelumnya yang pernah dilakukan orang lain. Tujuannya adalah untuk mendapatkan landasan teori mengenai masalah yang akan diteliti. Teori merupakan pijakan bagi peneliti untuk memahami persoalan yang diteliti dengan benar dan sesuai dengan kerangka berfikir ilmiah. Bagian ini akan dikaji lebih dalam pada Bab IV.

1.5.2 Metode Rekayasa Perangkat Lunak

Metode rekayasa perangkat lunak yang dilakukan oleh penulis untuk menyusun tugas akhir ini adalah metode *Waterfall*. Model ini memerlukan pendekatan yang sistematis dan sekuensial di dalam pengembangan sistem perangkat lunaknya, pengembangannya dimulai dari analisis dan definisi kebutuhan, perancangan sistem dan perangkat lunak, implementasi dan pengujian unit, integrasi dan pengujian sistem, operasi dan perawatan. Dengan demikian, pada model ini terdapat aktifitas-aktifitas sebagai berikut :

1. Rekayasa Sistem
2. Analisis
3. Perancangan
4. Pemograman (*Coding*)
5. Pengujian
6. Pengoperasian dan Pemeliharaan

2. LANDASAN TEORI

Landasan teori berisi teori-teori yang mendukung pembuatan system.

2.1 Kompresi Data

Kompresi data (pemampatan data) merupakan suatu teknik untuk memperkecil jumlah ukuran data (hasil kompresi) dari data aslinya. Pemampatan data umumnya diterapkan pada mesin komputer, hal ini dilakukan karena setiap simbol yang dimunculkan pada komputer memiliki nilai bit-bit yang berbeda. Misal pada ASCII setiap simbol yang dimunculkan memiliki panjang bit 8 bit, misal kode A pada ASCII mempunyai nilai desimal = 65, jika dirubah dalam bilangan biner menjadi 01000001. Pemampatan data digunakan untuk mengurangi jumlah bit-bit yang dihasilkan dari setiap simbol yang muncul. Dengan pemampatan ini diharapkan dapat mengurangi (memperkecil ukuran data) dalam ruang penyimpanan.

2.2 Algoritma Huffman

Pengkodean dengan metode Huffman dibangun dari panjang variabel kode-kode yang disusun dari bit-bit. Simbol dengan probabilitas yang tinggi akan memperoleh kode-kode paling pendek sedangkan simbol dengan probabilitas paling rendah akan memperoleh kode terpanjang.

Contoh untuk string 'ABACCCA' mempunyai panjang bit sebanyak 56 bit karena 1 karakter dikodekan dengan 8 bit (ASCII) akan diperoleh jumlah bit untuk tiap simbolnya dengan jumlah yang lebih sedikit atau bitnya lebih pendek yaitu 13 bit, sehingga secara otomatis ukuran filenya berkurang.

2.3 Graf

2.3.1 Definisi Graf

Graf G merupakan representasi dari suatu masalah yang digambarkan sebagai sekumpulan noktah (simpul) yang dihubungkan dengan sekumpulan garis (sisi).

2.3.2 Jenis – jenis Graf

Graf dapat dikelompokkan menjadi beberapa jenis bergantung pada sudut pandang pengelompokannya. Pengelompokan graf dapat dipandang berdasarkan ada tidaknya sisi ganda atau sisi kalang. Berdasarkan jumlah simpul, atau berdasarkan orientasi arah pada sisi.

Berdasarkan ada tidaknya gelang atau sisi ganda pada suatu graf, maka secara umum graf digolongkan menjadi dua jenis:

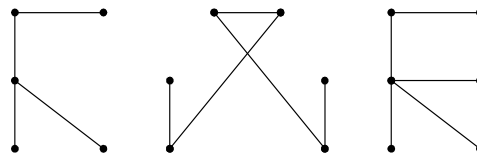
- Graf Sederhana (simple graph)
Graf yang tidak mengandung gelang maupun sisi ganda dinamakan graf sederhana.
- Graf Tak sederhana (unsimple graph)
Graf yang mengandung sisi ganda atau gelang dinamakan graf tak sederhana (unsimple graph). Ada dua macam graf tak sederhana, yaitu graf ganda (multi graph) dan graf semu (pseudograph).

2.4 Pohon (Tree)

Dari sekian banyak konsep graf, konsep pohon (tree) mungkin merupakan konsep paling penting, karena penerapannya sangat banyak pada kehidupan sehari-hari.

2.4.1 Definisi Pohon

Pohon adalah graf tak berarah terhubung yang tidak mengandung sirkuit. Gambar 2.8 dibawah ini menunjukkan contoh dari pohon.



Gambar 1 Tiga buah contoh dari pohon

3. ANALISIS DAN PERANCANGAN

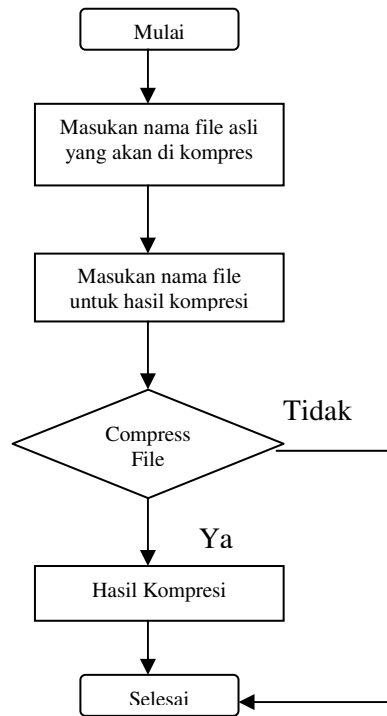
3.1 Deskripsi Masalah

Permasalahan yang dibahas dalam tugas akhir ini adalah membuat suatu simulasi program aplikasi kompresi data untuk digunakan pada ponsel berbasis system operasi symbian dengan menggunakan algoritma *Huffman*. Hasil uji coba algoritma *Huffman* yang telah terkompresi ini dianalisa dan selanjutnya dicari perbandingan kinerja metode kompresi tersebut. Data dari hasil penelitian metode *Huffman* dibandingkan dengan data aslinya. Perbandingan dari algoritma tersebut berupa hasil kompresi dan kode-kode bit yang dihasilkan oleh pohon-pohon biner.

3.2.1 Perancangan Aplikasi

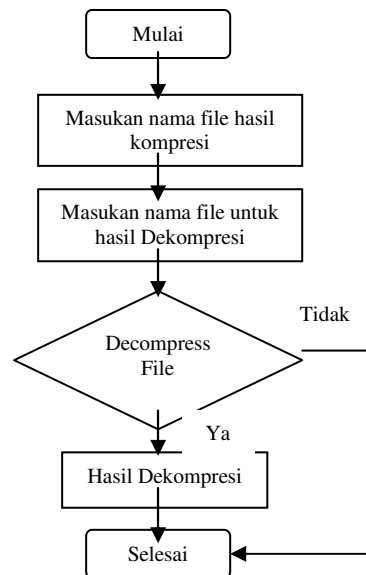
Aplikasi yang dibuat dalam penulisan tugas akhir ini dirancang khusus untuk ponsel berbasis sistem operasi symbian, terutama ponsel Nokia seri 9210i. Aplikasi ini dibuat dengan sederhana, sehingga diharapkan user dapat dengan mudah menggunakan program aplikasi ini.

3.2.2 Proses Kompresi



Gambar 2 – Proses Pemampatan Data

3.2.3 Proses Dekompresi



Gambar 3 – Proses Pemulihan Data

4. Pengujian

4.1 Hasil Pengujian

4.1.1 Pengujian kompresi data

Dalam pengujian ini, hasil kompresi data dengan file yang berjenis teks, image dan audio ditunjukkan pada tabel 4.1

Tabel 4.1 Kompresi data pada file teks, image, audio

Nama File	Ukuran File Asli	Ukuran File Kompres	Ratio	Waktu (s)
T1.txt	23 kb	13 kb	44%	0.1407
T2.pdf	449 kb	428 kb	5 %	0.6817
T3.doc	90 kb	50 kb	48 %	0.0687
G1.jpg	1298 kb	297 kb	1 %	0.2794
G2.gif	134 kb	128 kb	5 %	0.1309
G3.bmp	1.787 kb	322 kb	83%	0.7107
A1.mp3	2.570 kb	2.561 kb	1 %	2.8145
A2.wav	4.178 kb	3.473 kb	17%	4.0549
A3.wma	2.905 kb	2.871 kb	2 %	3.1551

4.1.2 Pengujian dekompresi data

Dalam pengujian ini, hasil dekompresi data dengan file yang berjenis teks ditunjukkan pada tabel 4.2

Tabel 4.2 Dekompresi data pada file teks, image, audio

Nama file	Ukuran File Kompres	Ukuran File Dekompres	Waktu (s)
T1.Text	13 kb	23 kb	0.0031
T2.pdf	428 kb	449 kb	0.3792
T3.doc	50 kb	90 kb	0.1189
G1.jpg	297 kb	298 kb	0.3011
G2.gif	128 kb	134 kb	0.1419
G3.bmp	322 kb	1.787 kb	0.8518
A1.mp3	2.561 kb	2570 kb	2.7239
A2.wav	3.473 kb	4178 kb	4.156
A3.wma	2.871 kb	2905 kb	3.1423

5. KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil pengujian algoritma *Huffman* untuk proses kompresi data dengan menggunakan VB, dapat ditarik kesimpulan sebagai berikut :

1. Rasio Kompresi yang dihasilkan cukup bervariasi tergantung jenis file yang dikompres. Untuk file audio dan image rasio kompresi yang dihasilkan lebih kecil sedangkan untuk file teks rasio kompresi yang dihasilkan lebih besar.
2. Tingkat keamanan data setelah dikompresi cukup terjaga, dengan kata lain file tidak berkurang atau mengalami kerusakan setelah proses kompresi data dilakukan
3. Kecepatan proses kompresi dan dekompresi data setara dengan ukuran dan jenis file

5.2 Saran

Untuk pengembangan selanjutnya diharapkan program aplikasi ini disarankan menggunakan algoritma kompresi data lainnya, agar didapatkan metode yang lebih efektif. Penulis juga menyarankan untuk pengembangan selanjutnya agar program aplikasi yang dibuat lebih variatif dan inovatif dari program aplikasi yang telah ada sebelumnya.

6. DAFTAR PUSTAKA

1. Halvsorson Michael, 2000. Microsoft Visual Basic 6.0 Step By Step. Jakarta; PT. Elex Media Komputindo
2. Bell, Timothy C., Cleary, John G., Witten, Ian H, (1990) "*Text Compression*", Prentice Hall, Englewood NJ
3. Torer, J.A., (1988) "*Data Compression*", Computer Science Press, Rockville, MD Witten, Ian H., Neal, Radford M.
4. Munir,Rinaldi. Matematika Diskrit. Informatika