

Aplikasi Kode Huffman Sebagai Metode Kompresi Pada Mesin Faks

Juan Anton 13513013

Program Studi Teknik Informatika

Sekolah Teknik Elektro dan Informatika

Institut Teknologi Bandung, Jl. Ganesha 10 Bandung 40132, Indonesia

juan_anton_che@std.itb.ac.id

Abstrak—Mesin faks atau faksimile adalah suatu perangkat komunikasi yang digunakan untuk mengirim dokumen melalui jaringan telepon. Dalam mengirimkan dokumen, pertama-tama mesin faks memindai dokumen yang akan dikirim, hasil pemindaian tersebut merupakan suatu kode elektronik yang terdiri atas kombinasi angka 0 dan 1 (kode biner), hasil pemindaian ini kemudian dikirim melalui jaringan telepon kepada mesin faks tujuan. Saat mesin faks tujuan menerima kode elektronik tersebut, ia akan menerjemahkan kode tersebut ke dalam rangkaian titik-titik hitam dan putih yang kemudian digunakan untuk mencetak sebuah dokumen yang serupa dengan dokumen yang dikirim. Semakin kompleks dokumen yang akan dikirim, semakin panjang pula kode elektronik yang dibuat oleh mesin faks, oleh karena itu dibutuhkan sebuah metode kompresi agar kode elektronik yang dibuat oleh mesin faks tidak terlalu panjang. Salah satu metode kompresi yang dapat digunakan adalah Kode Huffman.

Keywords—Kode Huffman, Huffman, Mesin Faks, Metode Kompresi, Kode Biner, Faksimile.

I. PENDAHULUAN

Salah satu kendala yang ada dalam pengiriman dokumen adalah jarak. Untuk menempuh jarak yang jauh biasanya diperlukan waktu yang cukup lama. Sebelum ditemukannya mesin faks, pengiriman dokumen antarkota melalui layanan pos dapat memakan waktu sampai berhari-hari. Seiring dengan berkembangnya perekonomian dunia, kebutuhan akan kecepatan pengiriman dokumen juga meningkat. Dengan ditemukannya mesin faks pada abad ke-19, pengiriman dokumen yang sebelumnya dapat memakan waktu berhari-hari sampai berminggu-minggu dengan risiko dokumen yang dikirim hilang dalam perjalanan atau tiba ke tujuan yang tidak seharusnya berubah menjadi hanya memakan waktu beberapa menit sampai beberapa jam dan dokumen dipastikan tiba di tempat tujuan yang tepat. Namun pada kenyataannya, orang-orang tetap memerlukan suatu metode pengiriman dokumen yang lebih cepat khususnya untuk dokumen-dokumen yang bersifat mendesak.

Dalam mengirim dokumen yang sederhana (tidak terlalu banyak tulisan/gambar), kode elektronik yang dibuat oleh mesin faks tanpa metode kompresi apapun

akan relatif pendek, sehingga lama waktu yang diperlukan untuk mentransmisikan kode tersebut ke mesin faks lain melalui jaringan telepon tidak terlalu besar. Masalah sesungguhnya muncul saat mesin faks ingin mengirim dokumen yang cukup kompleks (mengandung banyak tulisan/gambar), untuk dokumen yang cukup kompleks tentu saja kode elektronik yang dibuat oleh mesin faks akan relatif panjang jika dibuat tanpa metode kompresi apapun. Hal ini menyebabkan mesin faks akan membutuhkan waktu yang lama untuk mengirimkan dokumen yang cukup kompleks.

Oleh karena itu, diperlukan suatu metode kompresi agar mesin faks tidak perlu lagi membuat kode yang panjang saat ingin mengirim dokumen yang kompleks dan mempersingkat waktu yang dibutuhkan untuk mengirim dokumen tersebut.

Salah satu metode kompresi yang cukup mudah untuk diimplementasikan dan dapat diandalkan adalah Kode Huffman. Dengan menggunakan Kode Huffman, kode elektronik yang panjang dapat dibuat menjadi lebih singkat.

II. DASAR TEORI

A. Pohon Biner

Pohon biner adalah suatu pohon dimana setiap simpul pada pohon hanya memiliki paling banyak 2 buah anak.

B. Kode Huffman

Kode Huffman adalah salah satu aplikasi dari pohon biner. Kode Huffman pada umumnya digunakan untuk mengompresi suatu rangkaian kode dengan panjang tertentu. Dalam membuat Kode Huffman, pertama-tama kita buat sebuah tabel yang berisi simbol-simbol kode yang digunakan beserta banyaknya kemunculan untuk setiap kode dan hitung probabilitas untuk setiap kode.

Misalkan kita ingin membuat Kode Huffman untuk rangkaian kode 'ABCABAACBDAD', maka kita buat tabel sebagai berikut:

Simbol	Banyak Kemunculan	Probabilitas
A	5	5/12
B	3	3/12
C	2	2/12
D	2	2/12

Tabel 1. Tabel Frekuensi Untuk Rangkaian Kode 'ABCABAACBDAD'

Kemudian dibentuklah sebuah pohon biner dari bawah ke atas dengan setiap simbol menjadi daun. Hal ini dilakukan dalam beberapa langkah sebagai berikut:

1. Pilih dua simbol dengan probabilitas terkecil, kombinasikan kedua simbol tersebut membentuk simpul orangtua dari kedua simbol tersebut yang berisi gabungan dari kedua simbol tersebut dengan probabilitas sejumlah probabilitas anaknya.
2. Kemudian, pilih dua simbol berikutnya, termasuk simbol yang baru dibentuk yang memiliki peluang terkecil.
3. Ulangi langkah satu dan dua sampai seluruh simbol habis.

Untuk rangkaian kode 'ABCABAACBDAD' diatas, maka pohon biner yang terbentuk adalah sebagai berikut:

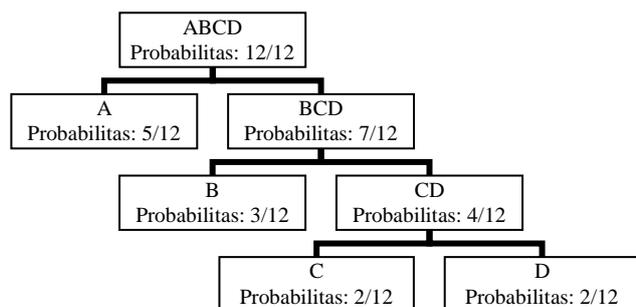


Diagram 1. Pohon Biner Untuk Rangkaian Kode 'ABCABAACBDAD'

Kemudian, setiap anak yang berada di sebelah kanan diberi kode biner 1, dan untuk setiap anak yang berada di sebelah kiri diberi kode biner 0. Kode Huffman untuk setiap simbol dapat diperoleh dengan cara menggabungkan kode 0 dan/atau 1 sampai tercapai daun yang sesuai. Untuk rangkaian kode 'ABCABAACBDAD' diatas, maka Kode Huffman yang terbentuk adalah sebagai berikut:

Simbol	Banyak Kemunculan	Probabilitas	Kode Huffman
A	5	5/12	0
B	3	3/12	10
C	2	2/12	110
D	2	2/12	111

Tabel 2. Tabel Kode Huffman Untuk Rangkaian Kode 'ABCABAACBDAD'

Setiap huruf memiliki 8 bit kode ASCII, jika kita ingin mengubah rangkaian kode 'ABCABAACBDAD' menjadi kode elektronik, maka akan dibutuhkan 8x12 bit yaitu sebanyak 96 bit. Tetapi jika kita menggunakan Kode Huffman, maka hanya dibutuhkan $(5 \times 1) + (3 \times 2) + (2 \times 3) + (2 \times 3)$ bit yaitu sebanyak 23 bit. Melalui contoh ini dapat dilihat bahwa dengan menggunakan Kode Huffman, kita dapat mengirimkan informasi yang sama dengan menggunakan jumlah bit yang lebih sedikit.

C. Run-Length Encoding (RLE)

Run-Length Encoding adalah metode kompresi data, dimana untuk sebuah nilai data yang sama dan berurut, disimpan sebagai sebuah nilai data yang merepresentasikan nilai dan banyaknya kemunculan nilai data tersebut yang terurut. Metode kompresi ini sangat berguna untuk mengompresi sebuah dokumen yang mengandung banyak data dengan nilai sama yang berurut.

Misalkan terdapat sebuah rangkaian kode 'AAAABBBBBCCCCDDD'. Jika kita ingin menyimpan rangkaian kode tersebut dalam bentuk kode biner, maka akan dibutuhkan banyak sekali bit kode. Dengan menggunakan RLE, kita dapat mengodekan rangkaian kode di atas sesuai dengan tabel berikut:

Simbol	Jumlah	Kode Biner
A	5	01011
B	4	01100
C	4	01011
D	3	00111

Tabel 3. Contoh Kode RLE Untuk Rangkaian Kode 'AAAABBBBBCCCCDDD'

Dengan mengikuti tabel diatas, maka rangkaian kode 'AAAABBBBBCCCCDDD' dapat disimpan cukup kode biner 01011 01100 01011 00111.

III. METODE KOMPRESI MESIN FAKS

Dalam perkembangannya, ada banyak macam metode kompresi yang digunakan oleh mesin faks. Beberapa metode kompresi yang digunakan dikembangkan oleh ITU-T (*International Telecommunications Union's Telecommunication Standardization Sector*) yang sebelumnya dikenal dengan nama CCITT (*Comite Consultatif International Telegraphique et Telephonique*). Pada umumnya, metode-metode yang dikembangkan oleh ITU-T diterima dan digunakan oleh orang-orang sebagai

standar dalam bidang telekomunikasi. Beberapa standar pengompresian data yang dianjurkan oleh ITU-T adalah T4 (dikenal juga sebagai *Group 3*) dan T6 (dikenal juga sebagai *Group 4*).

Metode kompresi *Group 3* didesain khusus untuk mengompresi dokumen hitam-putih, metode kompresi ini mengompresi dengan cukup baik dan dapat digunakan untuk berbagai macam dokumen hitam-putih. Biasanya metode kompresi *Group 3* dapat mengompresi dokumen dengan rasio 5:1 sampai 8:1. Metode kompresi *Group 3* dibagi menjadi 2, yaitu:

1. *Group 3 One-Dimensional (G31D)*, metode kompresi ini memanfaatkan Kode Huffman dan *Run-Length Encoding*, metode ini merupakan yang paling mudah untuk diimplementasikan.
2. *Group 3 Two-Dimensional (G32D)*, metode kompresi ini mengompres dokumen dengan cara menyimpan data yang merepresentasikan perbedaan yang terdapat di dalam dokumen tersebut.

Metode kompresi *Group 4* merupakan metode kompresi yang lebih efisien dibanding *Group 3*. Metode kompresi ini banyak digunakan sebagai metode kompresi untuk penyimpanan dokumen menggantikan metode kompresi *Group 3*. Meskipun *Group 4* dapat melakukan kompresi lebih baik dari *Group 3* (15:1 untuk *Group 4* dan 5:1 – 8:1 untuk *Group 3*), *Group 4* tidak memiliki kemampuan mengenali error seperti *Group 3*, sehingga *Group 4* tidak cocok untuk digunakan sebagai metode kompresi dalam pengiriman dokumen.

IV. APLIKASI KODE HUFFMAN DALAM METODE KOMPRESI MESIN FAKS

Dalam memindai sebuah dokumen, mesin faks membaca baris demi baris pada dokumen tersebut. Setiap baris yang dibaca oleh mesin faks terdiri atas titik-titik kecil hitam dan putih yang disebut *pels (Picture Elements)*. Resolusi pembacaan mesin faks secara horizontal adalah sebesar 8.05 *pels/mm*, sedangkan resolusi pembacaannya secara vertikal adalah 3.85 *baris/mm* atau 7.7 *baris/mm* (*baris* yang dimaksud adalah *baris* hasil pembacaan mesin faks bukan *baris* pada dokumen).

Untuk membuat kode yang digunakan dalam G31D, tim dari ITU-T menghitung semua kemunculan titik hitam dan putih yang berurut dalam 8 dokumen contoh yang dinilai cukup untuk merepresentasikan semua macam dokumen yang biasanya dikirim melalui mesin faks. 8 dokumen contoh yang digunakan adalah sebagai berikut:

1. Surat bisnis yang diketik (dalam bahasa Inggris).
2. Diagram sirkuit (digambar dengan tangan).
3. Faktur yang diketik (dalam bahasa Perancis).
4. Laporan penuh kata-kata yang diketik (dalam bahasa Perancis).
5. Artikel teknik yang ditulis dalam huruf cetak, dilengkapi dengan diagram dan rumus (dalam bahasa Perancis).
6. Graf dengan judul (dalam bahasa Perancis).
7. Dokumen yang penuh dengan huruf kanji.
8. Memo hasil tulisan tangan dengan huruf-huruf berwarna putih yang sangat besar (dalam bahasa Inggris).

Kemudian, mereka mencari Kode Huffman yang sesuai untuk setiap jumlah titik hitam dan putih yang berurut. Karena Kode Huffman dipakai untuk merepresentasikan jumlah titik hitam atau putih yang berurut, maka metode ini dikatakan memakai *Modified Huffman Coding* yaitu gabungan antara Kode Huffman dengan *Run Length Encoding*.

Tim dari ITU-T menemukan bahwa jumlah titik-titik berurut yang paling sering ditemukan adalah 2, 3, dan 4 titik-titik hitam, karena itu kode untuk jumlah tersebut merupakan kode yang paling pendek. Yang paling sering ditemukan selanjutnya adalah 2-7 titik-titik putih berurut yang kemudian diberi kode yang sedikit lebih panjang dibanding 2,3, dan 4 titik-titik hitam berurut.

Karena jumlah titik-titik berurut yang terdapat dalam sebuah dokumen dapat berjumlah sangat banyak, maka kode yang digunakan dibagi menjadi 2 bagian, yaitu kode untuk jumlah 1-63 dan kode untuk jumlah-jumlah kelipatan 64, selain itu terdapat juga kode khusus untuk menandakan akhir dari baris pembacaan (*End Of Line*).

Run length	White code-word	Black code-word	Run length	White code-word	Black code-word
0	00110101	0000110111	32	00011011	000001101010
1	000111	010	33	00010010	000001101011
2	0111	11	34	00010011	000011010010
3	1000	10	35	00010100	000011010011
4	1011	011	36	00010101	000011010100
5	1100	0011	37	00010110	000011010101
6	1110	0010	38	00010111	000011010110
7	1111	00011	39	000101000	000011010111
8	10011	000101	40	00101001	000001101100
9	10100	000100	41	00101010	000001101101
10	00111	0000100	42	00101011	0000011011010
11	01000	0000101	43	00101100	000011010111
12	001000	0000111	44	00101101	000001010100
13	000011	00000100	45	00000100	000001010101
14	110100	00000111	46	00000101	000001010110
15	110101	000011000	47	00001010	000001010111
16	101010	0000010111	48	00001011	000001100100
17	101011	0000011000	49	01010010	000001100101
18	0100111	0000001000	50	01010011	000001010010
19	0001100	00001100111	51	01010100	000001010011
20	0001000	00001101000	52	01010101	000000100100
21	0010111	00001101100	53	00100100	000000110111
22	0000011	00000110111	54	00100101	000000111000
23	0000100	00000101000	55	01011000	000000100111
24	0101000	00000010111	56	01011001	000000101000
25	0101011	00000011000	57	01011010	000001011000
26	0010011	000011001010	58	01011011	000001011001
27	0100100	000011001011	59	01001010	000000101011
28	0011000	000011001100	60	01001011	000000101100
29	00000010	000011001101	61	00110010	000001011010
30	00000011	000001101000	62	00110011	000001100110
31	00011010	000001101001	63	00110100	000001100111

Kode Untuk Jumlah Titik-Titik Berurut 1-63

Sumber : A Concise Introduction to Data Compression Chapter 2

Kode-kode yang digunakan untuk jumlah titik-titik berurut 1-63 disebut juga sebagai kode terminasi dan untuk jumlah kelipatan 64 disebut sebagai kode buatan. Kode hasil kompresi yang dihasilkan oleh mesin faks harus terdiri atas nol atau lebih kode buatan yang diakhiri oleh sebuah kode terminasi.

Run length	White code-word	Black code-word	Run length	White code-word	Black code-word
64	11011	000001111	1344	011011010	000001010011
128	10010	000011001000	1408	011011011	0000001010100
192	010111	000011001001	1472	010011000	0000001010101
256	0110111	000001011011	1536	010011001	0000001011010
320	00110110	000000110011	1600	010011010	0000001011011
384	00110111	000000110100	1664	011000	0000001100100
448	01100100	000000110101	1728	010011011	0000001100101
512	01100101	0000001101100	1792	00000001000	same as
576	01101000	0000001101101	1856	00000001100	white
640	01100111	0000001001010	1920	00000001101	from this
704	011001100	0000001001011	1984	000000010010	point
768	011001101	0000001001100	2048	000000010011	
832	011010010	0000001001101	2112	000000010100	
896	011010011	0000001110010	2176	000000010101	
960	011010100	0000001110011	2240	000000010110	
1024	011010101	0000001110100	2304	000000010111	
1088	011010110	0000001110101	2368	000000011100	
1152	011010111	0000001110110	2432	000000011101	
1216	011011000	0000001110111	2496	000000011110	
1280	011011001	0000001010010	2560	000000011111	

*Kode Untuk Jumlah Titik-Titik Berurut Kelipatan 64
Sumber : A Concise Introduction to Data
Compression Chapter 2*

Untuk setiap baris yang dibaca oleh mesin faks, mesin faks akan menambahkan sebuah titik putih di awal setiap pembacaan baris.

Misalkan terdapat sebuah baris yang akan dibaca oleh mesin faks sebagai berikut:



Pertama-tama, mesin faks akan mengenali titik-titik baris diatas sebagai 3 putih, 4 hitam, 1 putih, 3 hitam, 3 putih, dan EOL. Kemudian, mesin faks akan membuat kode sesuai dengan jumlah titik-titik yang berurutan. Kode yang dibuat untuk baris di atas adalah:

1000 011 000111 10 1000 000000000001

Kode 1000 adalah kode untuk 3 titik putih (1 titik putih yang ditambahkan oleh mesin faks dan 2 titik putih dari baris yang dibaca), kode 011 adalah kode untuk 4 titik hitam, kode 000111 adalah kode untuk 1 titik putih, kode 10 adalah kode untuk 3 titik hitam, kode 1000 adalah kode untuk 3 titik putih, dan kode 000000000001 adalah kode yang menandakan bahwa baris yang dibaca sudah berakhir (End Of Line).

Misalkan terdapat sebuah baris yang berisi 120 titik-titik hitam yang berurutan, maka mesin faks akan membuat kode dengan menggunakan sebuah kode buatan dan sebuah kode terminasi. Kode yang dibuat untuk 120 titik-titik hitam berurutan adalah:

000111 0000001111 000000101000 000000000001

Kode 0000001111 merupakan kode buatan untuk titik-titik hitam berurutan dengan jumlah 64, sedangkan kode

000000101000 adalah kode terminasi untuk titik-titik hitam berurutan dengan jumlah 56, sehingga menghasilkan jumlah total titik-titik hitam yang berurutan sebanyak 120.

Selain merepresentasikan suatu baris yang berisi titik-titik hitam dan putih ke dalam suatu rangkaian kode, mesin faks juga harus mampu untuk membaca suatu rangkaian kode dan mencetak baris yang berisi titik-titik hitam dan putih sesuai dengan rangkaian kode tersebut.

Karena kode yang digunakan memiliki panjang yang berbeda-beda, maka dalam membaca rangkaian kode mesin faks harus membaca satu bit demi satu bit kode sampai terbentuk suatu kombinasi kode yg dikenali.

Setiap kali mesin faks membaca sebuah rangkaian kode, maka mesin faks akan menerjemahkan dan mencetak kode-kode tersebut ke dalam titik-titik dengan warna dan jumlah yang sesuai. Pada setiap awal pencetakan, mesin faks akan secara otomatis mengurangi 1 titik putih di awal.

Dalam membaca rangkaian kode, mesin faks mencatat warna apa yang sebelumnya diterjemahkan. Melalui pencatatan warna ini, mesin faks dapat mengetahui kode yang saat ini sedang dibacanya merepresentasikan warna apa. Dengan adanya informasi warna ini, mesin faks cukup mencocokkan kode yang sedang dibacanya dengan kode-kode untuk warna yang sesuai.

Adanya penambahan satu titik putih di awal baris memudahkan mesin faks dalam menerjemahkan rangkaian kode yang ada. Pada awalnya mesin faks akan mencocokkan kode yang terbaca dengan kode-kode yang sesuai untuk titik-titik putih berurutan. Jika mesin faks membaca kode 000111 pada awal pembacaan, maka mesin faks akan mencocokkan kode-kode yang terbaca berikutnya dengan kode-kode yang sesuai untuk titik-titik hitam berurutan, karena kode 000111 menandakan bahwa hanya terdapat 1 buah titik putih dan yang berikutnya pastilah titik hitam. Tetapi jika mesin faks tidak membaca kode 000111 pada awal pembacaan, maka mesin faks akan tetap mencocokkan kode-kode yang terbaca dengan kode-kode yang sesuai untuk titik-titik putih berurutan.

Misalkan terdapat rangkaian kode 000111011011111000000000001, maka pertama-tama mesin faks akan membaca kode tersebut bit demi bit dan mencocokkan kombinasi kode yang terbentuk dengan kode-kode yang sesuai untuk titik-titik putih berurutan. Saat mesin faks membaca potongan kode 000111 maka mesin faks akan mengenali kode ini sebagai kode untuk satu buah titik putih dan tidak mencetak apa-apa, kemudian mesin faks akan melanjutkan pembacaan dan mencocokkan kombinasi kode yang terbentuk dengan kode-kode yang sesuai untuk titik-titik hitam berurutan. Saat mesin faks membaca potongan kode 011, maka mesin faks akan mengenali kode ini sebagai kode untuk 4 buah titik hitam dan mencetak 4 buah titik hitam. Begitu seterusnya sampai mesin faks membaca potongan kode 000000000001 yang mengakhiri pembacaan pada baris

tersebut. Untuk kode diatas, mesin faks akan menghasilkan baris:



Satu titik putih yang ada di awal kode tidak dicetak.

Sedangkan untuk kode 100001101111100000000001, mesin faks akan menghasilkan baris:



Satu dari tiga titik putih di awal kode tidak dicetak.

Karena kombinasi kode terpanjang yang dapat dikenali oleh mesin faks untuk suatu warna adalah 12 bit, maka mesin faks akan berhenti membaca dan mengirimkan pesan kesalahan jika mesin faks tidak berhasil mencocokkan kode hasil bacaannya dengan daftar kode-kode yang ada setelah membaca 12 bit kode.

V. KESIMPULAN

Untuk mempercepat pengiriman dokumen melalui mesin faks, maka perlu dilakukan kompresi terhadap data-data yang dikirim. Salah satu metode kompresi data yang banyak digunakan oleh mesin faks adalah *Group 3 One-Dimensional* yang merupakan metode kompresi data yang didasari oleh *Modified Huffman Coding* yaitu gabungan antara Kode Huffman dengan *Run-Length Encoding*.

REFERENSI

- [1] *CCITT (Huffman) Encoding*, (online), (http://www.fileformat.info/mirror/egff/ch09_05.htm, diakses 9 Desember 2014).
- [2] D. Salomon, *A Concise Introduction to Data Compression*. London: Springer, 2008, ch 2.
- [3] M. Rouse, *What is Fax?*, (online), (<http://searchnetworking.techtarget.com/definition/fax>, diakses 25 Juli 2012).
- [4] K. Adams, *Enforceability of Fax and Scanned Signature Pages*, (online), (<http://www.adamsdrafting.com/fax-and-scanned-signatures/>, diakses 25 juli 2012).
- [5] *T.4: Standardization of Group 3 facsimile terminals for document transmission*, ITU-T, 2003-07, (online), (<http://www.itu.int/rec/T-REC-T.4/en>, diakses 28 Desember 2013).
- [6] *International digital facsimile coding standards*, Hunter, R., and Robinson, A.H., *Proceedings of the IEEE* Volume 68 Issue 7, pp 854–867, July 1980.

PERNYATAAN

Dengan ini saya menyatakan bahwa makalah yang saya tulis ini adalah tulisan saya sendiri, bukan saduran, atau terjemahan dari makalah orang lain, dan bukan plagiasi.

Bandung, 9 Desember 2014

Juan Anton - 13513013