

**PERBANDINGAN KOMPRESI DATA TEKS ANTARA SEQUITUR DAN
RUN-LENGTH ENCODING DENGAN ALGORITMA
BURROWS-WHEELER TRANSFORM**

Skripsi



Oleh
MICHAEL RODION TAMBUNAN
22104924

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNOLOGIINFORMASI
UNIVERSITAS KRISTEN DUTA WACANA
2016

**PERBANDINGAN KOMPRESI DATA TEKS ANTARA SEQUITUR DAN
RUN-LENGTH ENCODING DENGAN ALGORITMA
BURROWS-WHEELER TRANSFORM**

Skripsi



Diajukan kepada Program Studi Teknik Informatika Fakultas Teknologi Informasi

Universitas Kristen Duta Wacana
Sebagai Salah Satu Syarat dalam Memperoleh Gelar
Sarjana Komputer

Disusun oleh

MICHAEL RODION TAMBUNAN

22104924

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS KRISTEN DUTA WACANA
2016**

PERNYATAAN KEASLIAN SKRIPSI

Saya menyatakan dengan sesungguhnya bahwa skripsi dengan judul:

PERBANDINGAN KOMPRESI DATA TEKS ANTARA SEQUITUR DAN RUN-LENGTH ENCODING DENGAN ALGORITMA BURROWS-WHEELER TRANSFORM

yang saya kerjakan untuk melengkapi sebagian persyaratan menjadi Sarjana Komputer pada pendidikan Sarjana Program Studi Teknik Informatika Fakultas Teknologi Informasi Universitas Kristen Duta Wacana, bukan merupakan tiruan atau duplikasi dari skripsi kesarjanaan di lingkungan Universitas Kristen Duta Wacana maupun di Perguruan Tinggi atau instansi manapun, kecuali bagian yang sumber informasinya dicantumkan sebagaimana mestinya.

Jika dikemudian hari didapati bahwa hasil skripsi ini adalah hasil plagiasi atau tiruan dari skripsi lain, saya bersedia dikenai sanksi yakni pencabutan gelar kesarjanaan saya.

Yogyakarta, 12 Mei 2016



MICHAEL RODION TAMBUNAN
22104924

HALAMAN PERSETUJUAN

Judul Skripsi : PERBANDINGAN KOMPRESI DATA TEKS
ANTARA SEQUITUR DAN RUN-LENGTH
ENCODING DENGAN ALGORITMA
BURROWS-WHEELER TRANSFORM

Nama Mahasiswa : MICHAEL RODION TAMBUNAN

N I M : 22104924

Matakuliah : Skripsi (Tugas Akhir)

Kode : TIW276

Semester : Genap

Tahun Akademik : 2015/2016

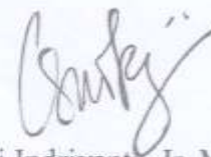
Telah diperiksa dan disetujui di
Yogyakarta,
Pada tanggal 12 Mei 2016

Dosen Pembimbing I



Lukas Chrisantyo, S.Kom., M.Eng.

Dosen Pembimbing II



Gani Indriyanta, Ir. M.T.

HALAMAN PENGESAHAN

PERBANDINGAN KOMPRESI DATA TEKS ANTARA SEQUITUR DAN RUN-LENGTH ENCODING DENGAN ALGORITMA BURROWS-WHEELER TRANSFORM

Oleh: MICHAEL RODION TAMBUNAN / 22104924

Dipertahankan di depan Dewan Penguji Skripsi
Program Studi Teknik Informatika Fakultas Teknologi Informasi
Universitas Kristen Duta Wacana - Yogyakarta
Dan dinyatakan diterima untuk memenuhi salah satu syarat memperoleh gelar
Sarjana Komputer
pada tanggal 26 Mei 2016

Yogyakarta, 27 Mei 2016
Mengesahkan,

Dewan Penguji:

1. Lukas Chrisantyo, S.Kom., M.Eng.
2. Gani Indriyanta, Ir. M.T.
3. Hendro Setiadi, M.Eng
4. Danny Sebastian, S.Kom., M.M., M.T.

DU TA WACANA

Dekan



(Signature)
(Budi Susanto, S.Kom., M.T.)

Ketua Program Studi

(Signature)

(Gloria Virginia, Ph.D.)

KATA PENGANTAR

Puji dan syukur penulis panjatkan pada Tuhan Yang Maha Esa, yang telah senantiasa melimpahkan rahmat dan berkat-Nya kepada penulis sehingga penulis dapat menyelesaikan skripsi dengan judul “Perbandingan Kompresi Data Teks Antara Sequitur Dan Run-Length Encoding Dengan Algoritma Burrows-Wheeler Transform” dengan baik dan tepat waktu.

Tugas akhir ini merupakan salah satu mata kuliah yang wajib ditempuh dalam studi fakultas Teknologi Informasi Universitas Kristen Duta Wacana. Selain itu, penulisan dan penyusunan skripsi ini disusun dalam rangka untuk memperoleh gelar Sarjana Komputer. Penulisan skripsi ini juga bertujuan melatih mahasiswa untuk menghasilkan suatu karya yang dapat dipertanggungjawabkan secara ilmiah, dan dapat bermanfaat bagi penggunanya.

Meskipun menghadapi kendala dan rintangan, penulis tidak lupa mengucapkan terima kasih kepada semua pihak yang telah membantu dalam menghadapi kendala dan rintangan dalam menyusun skripsi ini, antara lain :

1. Bapak Lukas Chrisantyo A.A., S.Kom., M.Eng. selaku dosen pembimbing I yang telah memberikan bimbingannya serta memberi arahan dan masukan yang sangat membantu sepanjang proses pembuatan skripsi.
2. Bapak Gani Indriyanta, Ir. M.T. selaku dosen pembimbing II yang memberikan bimbingannya serta memberi arahan dan masukan yang sangat membantu sepanjang proses pembuatan skripsi.
3. Keluarga tercinta yang selalu memberikan bantuan moral dan materi sehingga skripsi ini dapat diselesaikan.
4. Kepada Geo Fanny M. Rajaguguk yang telah memberikan dukungan dan motivasi sehingga skripsi ini dapat selesai.
5. Kepada seluruh dosen dan staf administrasi Fakultas Teknologi Informasi yang telah membantu baik secara langsung maupun tidak langsung dalam kelancaran penulisan skripsi ini.

6. Kepada semua pihak yang tidak dapat disebutkan satu-persatu yang telah memberikan dukungan baik secara langsung maupun tak langsung, sehingga skripsi ini dapat terselesaikan dengan baik.

Penulis menyadari bahwa dalam pembuatan skripsi ini masih jauh dari kata sempurna, baik dari segi penyusunan, pembahasan, ataupun penulisannya. Oleh karena itu penulis bersedia menerima kritik dan saran, semoga skripsi ini dapat bermanfaat bagi semua pihak.

Akhir kata, dengan segala kerendahan hati, penulis memohon maaf bila ada kesalahan dalam penulisan laporan skripsi ini.

Yogyakarta, 12 Mei 2016

Michael Rodion Tambunan

Abstrak

Data teks merupakan salah satu bentuk data yang sering digunakan pada saat ini. Data yang disimpan pada *file* teks semakin lama semakin bertambah sehingga menyebabkan ukurannya membesar. pengiriman *file* berukuran besar dapat mengganggu kelancaran pertukaran *file* karena pertukaran data dengan ukuran yang besar membutuhkan *bandwith* yang juga besar. Selain itu, secara tidak langsung proses kompresi data teks akan menyembunyikan bentuk asli dari teks sehingga dapat digunakan untuk keamanan data.

Dengan adanya permasalahan tersebut, penulis ingin melakukan kompresi data teks menggunakan metode Sequitur dan *Run-length encoding* (RLE) dengan algoritma *pre-processing* Burrows-Wheeler Transform (BWT) serta melakukan analisis untuk mengetahui kelebihan dan kekurangan dari kedua metode kompresi *file* teks tersebut.

Sistem mampu melakukan kompresi data teks, namun dalam beberapa kasus metode kompresi Sequitur tidak mampu melakukan kompresi data teks. Algoritma BWT mampu meningkatkan rasio kompresi metode kompresi RLE sekitar 13,5%, namun membutuhkan waktu yang lebih lama.

DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN SAMPUL DALAM	ii
PERNYATAAN KEASLIAN SKRIPSI.....	iii
HALAMAN PERSETUJUAN.....	iv
HALAMAN PENGESAHAN.....	v
KATA PENGANTAR	vi
ABSTRAK	viii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xi
DAFTAR GAMBAR	xii
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang Masalah.....	1
1.2 Pertanyaan Penelitian	2
1.3 Batasan Masalah.....	2
1.4 Tujuan Penelitian.....	2
1.5 Metodologi Penelitian	3
1.6 Sistematika Penulisan.....	3
BAB 2 TINJAUAN PUSTAKA	5
2.1 Tinjauan Pustaka	5
2.2 Landasan Teori	6
2.2.1 Kompresi Data	6
2.2.2 Jenis-jenis Kompresi Data.....	7
2.2.3 Sequitur.....	7
2.2.4 Burrows-Wheeler Transform.....	10

2.2.5	Run-length encoding.....	13
2.2.6	Parameter Pengukuran Performa Kompresi	15
BAB 3 ANALISIS DAN PERANCANGAN SISTEM		16
3.1	Analisis Kebutuhan	16
3.1.1	Analisis Data.....	16
3.1.2	Analisis Kebutuhan Perangkat Keras	16
3.2	Rancangan Sistem	17
3.2.1	Use Case	17
3.2.2	Diagram Alir (<i>Flowchart</i>).....	18
3.2.3	Perancangan Antarmuka Sistem	31
BAB 4 IMPLEMENTASI DAN ANALISIS SISTEM.....		33
4.1	Implementasi Sistem	34
4.1.1	Antarmuka Program.....	34
4.1.2	Implementasi <i>Input</i> dan <i>Output</i> Sistem	36
4.2	Analisis Sistem	38
4.2.1	Analisis metode kompresi sequitur.....	38
4.2.2	Analisis metode kompresi BWT+RLE.....	40
4.2.3	Pengaruh Algoritma BWT Terhadap Rasio Kompresi Metode Kompresi RLE.....	43
BAB 5 KESIMPULAN DAN SARAN		47
5.1	Kesimpulan.....	47
5.2	Saran	47
DAFTAR PUSTAKA		48
LAMPIRAN.....		49

DAFTAR TABEL

Tabel 4. 1. Hasil Pengujian metode Sequitur menggunakan 5 file teks yang berisi data berpola	38
Tabel 4. 2. Hasil Pengujian metode Sequitur menggunakan 20 file teks yang berisi data tidak berpola	39
Tabel 4. 3. Hasil Pengujian metode BWT+RLE menggunakan 5 file teks yang berisi data berpola	41
Tabel 4. 4. Hasil Pengujian metode BWT+RLE menggunakan 20 file teks yang berisi data tidak berpola	41
Tabel 4. 5. Hasil Pengujian rasio kompresi Sistem menggunakan 5 <i>file</i> teks yang berpola.....	43
Tabel 4. 6. Hasil Pengujian kecepatan kompresi Sistem menggunakan 5 <i>file</i> teks berpola.....	44
Tabel 4. 7. Hasil Pengujian rasio kompresi Sistem menggunakan 20 <i>file</i> teks tidak berpola.....	44
Tabel 4. 8. Hasil Pengujian kecepatan kompresi Sistem menggunakan 20 <i>file</i> teks tidak berpola.....	46

DAFTAR GAMBAR

Gambar 2. 1. Digram Uniqueness & Rule Utility	9
Gambar 2. 2. Hasil rotasi $n \times n$	11
Gambar 2. 3. Hasil sorting matrik M	11
Gambar 2. 4. Pembentukan string L.....	12
Gambar 2. 5. Pembentukan string L dan indeks I.....	13
Gambar 2. 6. Contoh dari penggunaan algoritma Run Length Eencoding	14
Gambar 2. 7. Rasio kompresi.....	15
Gambar 3. 1. usecase system.....	17
Gambar 3. 2. <i>Flowchart</i> Sistem	18
Gambar 3. 3. <i>Flowchart</i> metode sequitur	20
Gambar 3. 4. <i>Flowchart</i> transformasi BWT.....	22
Gambar 3. 5. <i>Flowchart</i> memindahkan karakter terakhir dari hasil sorting ke dalam array baru.....	23
Gambar 3. 6. <i>Flowchart</i> kompresi data metode RLE	25
Gambar 3. 7. <i>Flowchart</i> dekompresi data metode sequitur	27
Gambar 3. 8. <i>Flowchart</i> Dekompresi RLE	28
Gambar 3. 9. <i>Flowchart</i> Inverse BWT.....	29
Gambar 3. 10. <i>Form</i> halaman awal.....	31
Gambar 3. 11. <i>Form</i> kompresi	32
Gambar 3. 12. <i>Form</i> dekompresi	33
Gambar 4. 1. Tampilan Awal Sistem.....	34
Gambar 4. 2. Halaman Kompresi Teks	35
Gambar 4. 3. Halaman Dekompresi Teks	36
Gambar 4. 4. Halaman <i>input file</i> kompresi	37
Gambar 4. 5. Halaman <i>input file</i> dekompresi	37

Abstrak

Data teks merupakan salah satu bentuk data yang sering digunakan pada saat ini. Data yang disimpan pada *file* teks semakin lama semakin bertambah sehingga menyebabkan ukurannya membesar. pengiriman *file* berukuran besar dapat mengganggu kelancaran pertukaran *file* karena pertukaran data dengan ukuran yang besar membutuhkan *bandwith* yang juga besar. Selain itu, secara tidak langsung proses kompresi data teks akan menyembunyikan bentuk asli dari teks sehingga dapat digunakan untuk keamanan data.

Dengan adanya permasalahan tersebut, penulis ingin melakukan kompresi data teks menggunakan metode Sequitur dan *Run-length encoding* (RLE) dengan algoritma *pre-processing* Burrows-Wheeler Transform (BWT) serta melakukan analisis untuk mengetahui kelebihan dan kekurangan dari kedua metode kompresi *file* teks tersebut.

Sistem mampu melakukan kompresi data teks, namun dalam beberapa kasus metode kompresi Sequitur tidak mampu melakukan kompresi data teks. Algoritma BWT mampu meningkatkan rasio kompresi metode kompresi RLE sekitar 13,5%, namun membutuhkan waktu yang lebih lama.

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Perkembangan teknologi yang mulai meninggalkan penggunaan data fisik menuntut semua orang agar memiliki *device* yang dapat menyimpan dan mengakses data digital tersebut. Begitu juga dengan pengiriman data fisik kini beralih ke pengiriman data digital melalui internet. Namun dalam prakteknya, pengiriman *file* berukuran besar dapat mengganggu kelancaran pertukaran *file* karena pertukaran data dengan ukuran yang besar membutuhkan *bandwith* yang juga besar. Selain itu, secara tidak langsung proses kompresi data teks akan menyembunyikan bentuk asli dari teks sehingga dapat digunakan untuk keamanan data.

Salah satu cara untuk meminimalkan permasalahan kelancaran pertukaran data melalui internet tersebut adalah dengan memperkecil ukuran data tersebut atau biasa disebut dengan mengkompresi data. Kompresi data adalah proses yang bertujuan untuk memperkecil ukuran suatu data *digital*. Dengan menggunakan teknik kompresi data, data yang sebelumnya berukuran besar bisa diperkecil untuk menghemat penggunaan *bandwith* untuk mengirim data tersebut. Ada banyak macam-macam metode untuk mengkompresi data seperti yang dituliskan Salomon (2010) dalam buku Handbook of Data Compression diantaranya adalah metode Lempel-Ziv-Welch (LZW), Shannon-Fano, Run-length encoding (RLE), Sequitur, dan lain-lain.

Dengan adanya permasalahan tersebut, penulis ingin melakukan kompresi data teks menggunakan metode Sequitur dan Run-length encoding (RLE) dengan algoritma pre-processing Burrows-Wheeler Transform (BWT) serta melakukan analisis perbandingan rasio kompresi dan waktu (running time) kompresi dari kedua metode pemampatan *file* teks tersebut. Diharapkan dengan adanya studi ini selain menjadi solusi untuk meminimalkan penggunaan *bandwith* untuk mengirim

data juga memberikan wawasan tentang metode yang lebih baik antara metode Sequitur dan Run-length encoding (RLE).

1.2 Pertanyaan Penelitian

Pertanyaan penelitian yang dapat dijabarkan terkait dengan judul penelitian adalah sebagai berikut :

1. Bagaimana mengimplementasikan metode Sequitur dan Run-length encoding (RLE) dengan algoritma pre-processing Burrows-Wheeler Transform (BWT) pada kompresi *file* teks?
2. Apa kelebihan dan kekurangan dari metode Sequitur dan Run-length encoding (RLE) dengan algoritma pre-processing Burrows-Wheeler Transform (BWT) dalam mengkompresi data teks?

1.3 Batasan Masalah

Penelitian ini hanya terbatas pada *file* teks ASCII yang berekstensi (.txt) dengan ukuran *file* yang berbeda-beda yang diambil dari Calgary Corpus, Canterbury Corpus, dan beberapa file teks yang dibuat oleh penulis.

1.4 Tujuan Penelitian

Penelitian ini bertujuan untuk mengimplementasikan metode Sequitur dan Run-length encoding (RLE) dengan algoritma pre-processing Burrows-Wheeler Transform (BWT) dan melakukan analisis untuk mengetahui kelebihan dan kekurangan dari metode Sequitur dan Run-length encoding (RLE) dengan algoritma pre-processing Burrows-Wheeler Transform (BWT) dalam memampatkan *file* teks.

1.5 Metodologi Penelitian

Berikut ini merupakan metodologi penelitian yang akan dilakukan oleh penulis:

1. Melakukan studi literatur tentang konsep dan alur kerja metode kompresi Sequitur, Run Length Encoding dan Burrows-Wheeler Transform (BWT).
2. Mengumpulkan data untuk pengujian 25 data teks atau kutipan buku dengan format (.txt).
3. Menentukan langkah-langkah proses yang akan ditetapkan.
4. Menerapkan 2 metode kompresi data yaitu metode Sequitur dan Run-length encoding (RLE) dengan algoritma *pre-processing* Burrows-Wheeler Transform (BWT) dengan membuat program untuk pengujian.
5. Melakukan pengujian pada data teks melalui program kompresi data teks yang telah dibuat berdasarkan 2 metode diatas.
6. Mengolah analisis hasil pengujian data kompresi pada program, berdasarkan rasio kompresi dan waktu kompresi untuk menarik kesimpulan.

1.6 Sistematika Penulisan

Pada penulisan tugas akhir ini akan dibagi dalam 5 bab, yaitu :

BAB 1. PENDAHULUAN

Berisikan latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, metodologi penelitian dan sistematika penulisan.

BAB 2. TINJAUAN PUSTAKA

Berisikan tentang tinjauan pustaka dan landasan teori. Tinjauan pustaka menguraikan berbagai teori yang didapatkan dari berbagai sumber pustaka yang digunakan untuk penyusunan Tugas Akhir sedangkan landasan teori memuat penjelasan tentang konsep dan prinsip utama yang diperlukan untuk memecahkan masalah penelitian

BAB 3. PERANCANGAN SISTEM

Berisikan tentang Analisis teori-teori yang digunakan dan bagaimana menterjemahkannya ke dalam suatu sistem yang hendak dibuat

BAB 4. IMPLEMENTASI DAN ANALISIS SISTEM

Berisikan hasil riset / implementasi, dan pembahasan / analisis dari penelitian yang dilakukan yang sifatnya terpadu.

BAB 5. KESIMPULAN DAN SARAN

Berisikan tentang kesimpulan dari penelitian yang dilakukan oleh penulis dan saran-saran dari penulis untuk kegiatan penelitian ke depannya apabila ada mahasiswa yang nantinya akan meneliti topik yang berhubungan dengan penelitian yang penulis lakukan

©UKDW

BAB 5

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan pembahasan dan analisis terhadap hasil pengujian sistem yang dilakukan penulis pada bab 4, maka dapat ditarik beberapa kesimpulan sebagai berikut :

1. Sequitur memiliki kemampuan yang baik dalam melakukan pemampatan data berpola. Namun metode ini memiliki kelemahan dalam melakukan proses kompresi data yang tidak berpola. Dengan menggunakan *unicode 2 byte*, metode kompresi ini gagal melakukan kompresi. Selain kegagalan tersebut, waktu proses kompresi data tidak berpola juga sangat panjang.
2. BWT+RLE mampu melakukan kompresi data pada data uji berpola dan tidak berpola. Kekurangan dari metode ini adalah pada waktu proses kompresinya sangat bergantung pada ukuran data uji.
3. Waktu proses kedua metode kompresi sangat bergantung pada ukuran data uji. Semakin besar data uji maka proses kompresi akan lebih lama.
4. Algoritma BWT mampu meningkatkan rasio kompresi metode kompresi RLE sekitar 13,5%, namun membutuhkan waktu yang lebih lama.

5.2 Saran

Berikut ini adalah beberapa saran yang dapat diberikan oleh penulis untuk pengembangan lebih lanjut :

1. Menggunakan algoritma transformasi BWT untuk mengoptimalkan performa metode Sequitur.
2. Menggunakan *unicode 1,5 byte* sebagai *key* pada metode kompresi Sequitur.
3. Menggunakan *file* teks dengan format lain selain *.txt*. seperti *.doc* dan *odt*

Daftar Pustaka

- CALGARY CORPUS. (2016, Februari 11). *Calgary Corpus Compression Test*. Retrieved from Calgary Corpus Compression Test: <http://compression.ca/act/files/calgary.zip>
- Ervin, Proboyekti, U., & Krisnawati, L. D. (2007). Kompresi Data Teks Menggunakan Pendekatan Grammar Compression Dengan Algoritma Sequ Itur. *22 JURNAL INFORMATIKA, VOLUME 3 NOMOR 2*.
- Gozali, F., & Mervyn. (2004). Analisis Perbandingan Kompresi Data Dengan Teknik Arithmetic Coding Dan Run Length Encoding. *JETri, Volume 4, Nomor 1*.
- Langmead, B. (2013). Introduction to the Burrows-Wheeler Transform and FM Index. *Department of Computer Science, JHU*.
- Nevill-Manning, C. G., & Witten, I. H. (1997). Identifying Hierarchical Structure in Sequences: A linear-time algorithm. *Journal of Artificial Intelligence Research 7*.
- Salomon, D., & Motta, G. (2010). *Handbook of Data Compression*. London: Springer.
- THE CANTERBURY CORPUS. (2016, Februari 11). *CANTERBURY CORPUS*. Retrieved from CANTERBURY CORPUS: <http://compression.ca/act/files/canterbury.zip>
- Wibowo, H. A. (2009, April 29). *igilib.tes.telkomuniversity.ac.id*. Dipetik April 20, 2015, dari http://digilib.tes.telkomuniversity.ac.id/index.php?option=com_content&view=article&id=524:transformasi-burrows-wheeler-bwt&catid=15:pemrosesan-sinyal&Itemid=14
- Widyaningsih, E. N. (2015). Perbandingan Kompresi Data Teks Antara Metode Sequitur Dan Metode Lzw. 42.
- Wijaya, A. C. (2015). Perbandingan Kompresi Data Teks Antara Metode Sequitur, Static Huffman Dan Lzw. 43.