

**RANCANG BANGUN APLIKASI BERBASIS WEB
DENGAN MENERAPKAN *GENERAL STATISTIC METHOD*
PADA PERINGKASAN ARTIKEL BERBAHASA INDONESIA**

TUGAS AKHIR



LAINATUSSIFA DALIMUNTHE

1122001003

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN ILMU KOMPUTER
UNIVERSITAS BAKRIE
JAKARTA
2017**

HALAMAN PERNYATAAN ORISINALITAS

Tugas Akhir ini adalah hasil karya saya sendiri,
dan semua sumber baik yang dikutip maupun dirujuk
telah saya nyatakan dengan benar.

Nama : Lainatussifa Dalimunthe

NIM : 1122001003

Tanda Tangan :



Tanggal : 14 September 2017

HALAMAN PENGESAHAN

Tugas Akhir ini diajukan oleh:

Nama : Lainatussifa Dalimunthe

NIM : 1122001003

Fakultas : Teknik dan Ilmu Komputer

Judul Skripsi : Rancang Bangun Aplikasi Berbasis Web dengan Menerapkan
General Statistic Method pada Peringkasan Artikel Berbahasa
Indonesia

Telah berhasil dipertahankan dihadapan Dewan Penguji dan diterima sebagai
bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana
Komputer pada Program Studi Informatika Fakultas Teknik dan Ilmu
Komputer, Universitas Bakrie

DEWAN PENGUJI

Pembimbing 1 : Yusuf Lestanto, S.T., M.Sc.

Penguji 1 : Berkah I. Santoso, S.T, M.T.I

Penguji 2 : Prof. Dr. Hoga Saragih, S.T, M.T

Ditetapkan di : Jakarta

Tanggal : 14 September 2017



KATA PENGANTAR

Puji dan syukur penulis panjatkan ke hadirat Allah SWT, karena atas izin-Nya lah skripsi ini dapat diselesaikan tepat pada waktunya. Skripsi ini berjudul "Rancang Bangun Aplikasi Berbasis Web dengan Menerapkan *General Statistic Method* pada Peringkasan Artikel Berbahasa Indonesia"

Terselesaikannya tugas akhir ini tidak luput dari bantuan serta partisipasi berbagai pihak, dengan segala kerendahan hati penulis menyampaikan terimakasih atas bimbingan dan bantuan dalam proses penyelesaian skripsi ini kepada:

1. Bapak Yusuf Lestanto, S.T., M.Sc. selaku Dosen Pembimbing Tugas Akhir penulis.
2. Bapak Berkah I. Santoso, S.T, M.T.I selaku Dosen Pembahas dan Dosen Penguji Tugas Akhir penulis.
3. Prof. Dr. Hoga Saragih, S.T, M.T selaku Ketua Program Studi Informatika dan Dosen Penguji Tugas Akhir penulis.
4. Keluarga besar khususnya Ibu, Ayah, dan adik-adik penulis atas dukungan, doa, motivasi, dan bantuan yang tak terhitung banyaknya.
5. Teman-teman mahasiswa Universitas Bakrie, khususnya mahasiswa Informatika angkatan 2012.
6. Seluruh keluarga besar Universitas Bakrie, pihak yang terlibat, saudara-saudara, dan teman-teman yang telah membantu dan memberikan doa dalam penyusunan Tugas Akhir ini.

Masih banyak kekurangan yang terdapat dalam penulisan Tugas Akhir ini. Maka dari itu penulis mengharapkan kritik yang membangun dari pembaca. Akhirnya, semoga laporan ini bermanfaat bagi pihak yang perkepentingan.

Jakarta, 14 September 2017



Lainatussifa Dalimunthe

HALAMAN PERSETUJUAN PUBLIKASI

Sebagai *civitas* akademik Universitas Bakrie, saya yang bertanda tangan di bawah ini:

Nama : Lainatussifa Dalimunthe
NIM : 1122001003
Program Studi : Informatika
Fakultas : Teknik dan Ilmu Komputer
Jenis Tugas Akhir : Rancang-Bangun Sistem

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Bakrie **Hak Bebas Royalti Noneksklusif** (*Non-exclusive Royalty-Free Right*) atas karya ilmiah yang berjudul:

Rancang Bangun Aplikasi Berbasis Web dengan Menerapkan *General Statistic Method* pada Peringkasan Artikel Berbahasa Indonesia

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneksklusif ini Universitas Bakrie berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta untuk kepentingan akademis.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Jakarta

Pada : 14 September 2017

Yang Menyatakan,



(Lainatussifa Dalimunthe)

**RANCANG BANGUN APLIKASI BERBASIS WEB DENGAN
MENERAPKAN *GENERAL STATISTIC METHOD* PADA PERINGKASAN
ARTIKEL BERBAHASA INDONESIA**

Lainatussifa Dalimunthe

ABSTRAK

Penerapan metode peringkasan teks sudah menjadi studi yang populer dalam beberapa dekade terakhir. Namun, belum banyak metode yang menerapkan peringkasan teks dalam berbahasa Indonesia. Penelitian ini berfokus pada peringkasan teks dokumen tunggal berbahasa Indonesia dengan menerapkan *General Statistic Method* (GSM). Terdapat delapan *sentence feature* yang digunakan pada metode GSM. Setiap *sentence feature* memberikan skor fitur kepada masing-masing kalimat. Skor total dari kalimat dihitung dengan menjumlahkan skor fitur dari delapan *sentence feature*. Kemudian, kalimat diurutkan berdasarkan total skor. Jumlah kalimat yang digunakan sebagai ringkasan adalah 20% dari jumlah kalimat dalam teks. Sehingga, 20% kalimat yang memiliki skor tertinggi dijadikan sebagai ringkasan. Hasil ringkasan kemudian diuji dengan menghitung nilai *precision*, *recall*, dan *f-measure* pada sepuluh data set yang berupa artikel opini berbahasa Indonesia. Nilai *precision*, *recall*, dan *f-measure* yang didapatkan adalah 83%, 68%, dan 74% secara berurut. Hasil dari penelitian ini adalah aplikasi berbasis web yang dapat melakukan peringkasan teks.

Kata Kunci:

peringkasan teks, dokumen tunggal, Bahasa Indonesia, *General Statistic Method*, *sentence feature*, *compression rate*, *f-measure*, *recall*, *precision*

DESIGN AND BUILDING WEB-BASED APPLICATION BY IMPLEMENTING GENERAL STATISTIC METHOD ON INDONESIAN LANGUAGE SUMMARY AUTOMIZATION

Lainatussifa Dalimunthe

ABSTRACT

The application of text summarization methods has become a popular study in recent decades. However, there are not many methods that apply text summarization in Indonesian language. This study focuses on the text summarization of a single Indonesian document by applying the General Statistic Method (GSM). There are eight sentence features used in the GSM method. Each sentence feature provides a feature score to each sentence. The total score of the sentence is calculated by summing the feature score of the eight sentence features. Then, sentences are sorted by the total score. The number of sentences used as the summary is 20% of the number of sentences in the text. Thus, the 20% of sentences with the highest score are the summary. The summary results are then tested by calculating the precision, recall, and f-measure values in ten data sets in the form of Indonesian-language articles of opinion. The precision, recall, and f-measure values were 83%, 68%, and 74% respectively. The results of this study is a web-based application that can perform text summarization.

Keyword:

summary automization, single document, Indonesian Language, General Statistic Method, sentence feature, compression rate, f-measure, recall, precision

DAFTAR ISI

HALAMAN PERNYATAAN ORISINALITAS.....	i
HALAMAN PENGESAHAN.....	iii
KATA PENGANTAR	iv
HALAMAN PERSETUJUAN PUBLIKASI.....	v
ABSTRAK	vi
ABSTRACT.....	vii
DAFTAR ISI.....	viii
DAFTAR GAMBAR	xii
DAFTAR TABEL	xiii
DAFTAR LAMPIRAN.....	xv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Pembatasan Masalah	3
1.4 Tujuan Penelitian.....	3
1.5 Manfaat Penelitian.....	3
1.6 Sistematika Penulisan.....	4
BAB II TINJAUAN PUSTAKA.....	5
2.1 Penelitian Terdahulu.....	5
2.2 Peringkasan Teks Otomatis.....	6
2.2.1 Definisi Peringkasan Teks Otomatis.....	6
2.2.2 Pendekatan Peringkasan Teks	7
2.2.3 Tahapan Membuat Ringkasan.....	7
2.2.4 Proses Peringkasan Teks	8
2.3 <i>Preprocessing</i>	9

2.3.1	<i>Sentence Tokenizing</i>	10
2.3.2	<i>Word Tokenizing</i>	10
2.3.3	<i>Stop Word Removal</i>	11
2.3.4	<i>Stemming</i>	12
2.4	<i>Sentence Feature</i>	15
2.4.1	<i>Title Feature</i>	15
2.4.2	<i>Sentence Length</i>	16
2.4.3	<i>Term Weight</i>	16
2.4.4	<i>Sentence Position</i>	17
2.4.5	<i>Sentence to Sentence Similarity</i>	18
2.4.6	<i>Proper Noun</i>	18
2.4.7	<i>Thematic Word</i>	19
2.4.8	<i>Numerical Data</i>	19
2.5	<i>Compression Rate</i>	20
2.6	<i>General Statistic Method</i>	20
2.7	<i>Metode Evaluasi</i>	22
2.7.1	<i>Recall dan Precision</i>	22
2.7.2	<i>F-Score</i>	23
2.8	<i>TextRank</i>	24
2.8	<i>Levensthein Distance</i>	26
BAB III METODE PENELITIAN		29
3.1	<i>Objek Penelitian</i>	29
3.2	<i>Alat Penelitian</i>	29
3.3	<i>Tahapan Penelitian</i>	30
3.4	<i>Teknik Pengumpulan Data</i>	30
3.5	<i>Analisis Penerapan Metode</i>	31

3.5.1	<i>Flowchart dan Pseudocode</i>	32
3.5.2	<i>Use Case</i>	33
3.5.3	<i>Sequence Diagram</i>	34
3.5.4	<i>Class Diagram</i>	36
3.5.5	Relasi Basis Data.....	37
3.6	Kerangka Penelitian	38
3.7	Rencana Kerja Penelitian	39
BAB IV IMPLEMENTASI		40
4.1	Proses Studi	40
4.2	Implementasi Sistem	41
4.2.1	Implementasi Perangkat Keras.....	41
4.2.2	Implementasi Perangkat Lunak.....	41
4.3	Implementasi Antar Muka.....	42
4.4	Implementasi Tahap <i>Preprocessing</i>	43
4.4.1	Implementasi Tahap <i>Sentence Tokenizing</i>	43
4.4.2	Implementasi Tahap <i>Word Tokenizing</i>	44
4.4.3	Implementasi Tahap <i>Stop Word Removal</i>	44
4.4.4	Implementasi Tahap <i>Stemming</i>	45
4.5	Implementasi Tahap <i>Sentence Feature</i>	46
4.5.1	Implementasi Tahap <i>Title Feature</i>	46
4.5.2	Implementasi Tahap <i>Sentence Length</i>	47
4.5.3	Implementasi Tahap <i>Term Weight</i>	48
4.5.4	Implementasi Tahap <i>Sentence Position</i>	49
4.5.5	Implementasi Tahap <i>Sentence to Sentence Similarity</i>	50
4.5.6	Implementasi Tahap <i>Proper Noun</i>	51
4.5.7	Implementasi Tahap <i>Thematic Word</i>	52

4.5.8	Implementasi Tahap <i>Numerical Data</i>	53
4.6	Implementasi Tahap <i>Sentence Scoring</i>	54
4.7	Implementasi Tahap <i>Sentence Selection</i>	55
4.8	<i>Web Application Testing</i>	56
4.8.1	<i>Functionality Testing</i>	56
4.8.2	<i>Performance Testing</i>	60
4.9	Pengujian	62
4.9.1	Skenario Pengujian.....	63
4.9.2	Ringkasan Ideal	64
4.9.2	Ringkasan Sistem	65
4.9.2	Ringkasan Metode Perbandingan.....	66
4.10	Evaluasi.....	67
BAB V_KESIMPULAN DAN SARAN.....		70
5.1	Kesimpulan.....	70
5.2	Saran	71
DAFTAR PUSTAKA		72
LAMPIRAN.....		75
LAMPIRAN I TEKS ARTIKEL DAN RINGKASAN		76
LAMPIRAN II PERBANDINGAN HASIL RINGKASAN		94
LAMPIRAN III PERBANDINGAN NILAI PRECISION, RECALL, DAN F-MEASURE.....		96
LAMPIRAN IV SKOR KALIMAT		98
LAMPIRAN V SRS.....		109

DAFTAR GAMBAR

Gambar 2.1 Tahap-tahap dalam <i>Preprocessing</i>	9
Gambar 2.2 Peringkasan Teks berdasarkan <i>General Statistic Method</i>	21
Gambar 2.3 Cara Menghitung <i>Recall</i>	22
Gambar 2.4 Cara Menghitung <i>Precision</i>	23
Gambar 3.1 <i>Flowchart</i> dari Proses Peringkasan Teks	32
Gambar 3.2 <i>Use Case</i> actor Pengguna	33
Gambar 3.3 <i>Sequence Diagram Preprocessing</i>	34
Gambar 3.4 <i>Sequence Diagram Sentence Scoring</i>	35
Gambar 3.5 <i>Class Diagram</i>	36
Gambar 3.6 Relasi <i>Database</i>	37
Gambar 3.7 Kerangka Penelitian	38

DAFTAR TABEL

Tabel 2.1 Perbandingan Penelitian Terdahulu.....	6
Tabel 2.2 Contoh <i>Sentence Tokenizing</i>	10
Tabel 2.3 Contoh <i>Word Tokenizing</i>	11
Tabel 2.4 Contoh <i>Stop Word Removal</i>	11
Tabel 2.5 <i>Stemmer</i> Nazief dan Andriani.....	13
Tabel 2.6 Algoritma <i>Levensthein Distance</i>	26
Tabel 3.1 Pseudocode Peringkasan Teks.....	33
Tabel 3.2 <i>Timeline</i> Kerja.....	39
Tabel 4.1 <i>Link Testing</i>	56
Tabel 4.2 <i>Form Testing</i>	57
Tabel 4.3 <i>Database Testing</i>	57
Tabel 4.4 Validasi HTML/CSS	59
Tabel 4.5 <i>Web Stress Testing</i>	60
Tabel 4.6 <i>Error Handling</i>	61
Tabel 4.7 Daftar Data Set Artikel.....	62
Tabel 4.8 Daftar Hasil Ringkasan Ideal.....	64
Tabel 4.9 Daftar Hasil Ringkasan Sistem.....	65
Tabel 4.10 Daftar Hasil Ringkasan <i>TextRank</i>	66
Tabel 4.11 Perbandingan Ringkasan Ideal dengan Ringkasan GSM.....	68
Tabel 4.12 Perbandingan Ringkasan Ideal dengan Ringkasan <i>TextRank</i>	69

DAFTAR RUMUS

Rumus 1 <i>Title Feature Score</i>	15
Rumus 2 <i>Sentence Length Feature Score</i>	16
Rumus 3 <i>Term Weight Score</i>	16
Rumus 4 <i>Term Weight Feature Score</i>	17
Rumus 5 <i>Sentence Position Feature Score</i>	17
Rumus 6 <i>Sentence to Sentence Similarity Feature Score</i>	18
Rumus 7 <i>Proper Noun Feature Score</i>	18
Rumus 8 <i>Thematic Word Feature Score</i>	19
Rumus 9 <i>Numerical Data Feature Score</i>	19
Rumus 10 <i>Total Feature Score</i>	21
Rumus 11 <i>F-measure</i>	23

DAFTAR LAMPIRAN

LAMPIRAN I Teks Artikel dan Ringkasan.....	76
LAMPIRAN II Perbandingan Hasil Ringkasan.....	94
LAMPIRAN III Perbandingan Nilai <i>Precision</i> , <i>Recall</i> , dan <i>F-Measure</i>	96
LAMPIRAN IV Skor Kalimat.....	98
LAMPIRAN V <i>Software Requirement Spesifications</i>	109