

**RANCANG BANGUN APLIKASI PERINGKAS TEKS OTOMATIS
BERBASIS WEBSITE DENGAN METODE *LATENT SEMANTIC
ANALYSIS (LSA)***

TUGAS AKHIR



MUHAMMAD FAUZAN ADHIM

1142001017

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN ILMU KOMPUTER**

UNIVERSITAS BAKRIE

JAKARTA

2020

**RANCANG BANGUN APLIKASI PERINGKAS TEKS OTOMATIS
BERBASIS WEBSITE DENGAN METODE LATENT SEMANTIC
ANALYSIS (LSA)**

TUGAS AKHIR

**Diajukan sebagai salah satu syarat untuk memperoleh gelar
Sarjana Komputer**



MUHAMMAD FAUZAN ADHIM

1142001017

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN ILMU KOMPUTER**

UNIVERSITAS BAKRIE

JAKARTA

2020

HALAMAN PERNYATAAN ORISINALITAS

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

Nama : Muhammad Fauzan Adhim

NIM : 1142001017

Tanda Tangan :



Tanggal : 8 April 2020

HALAMAN PENGESAHAN


Tugas Akhir ini diajukan oleh:

Nama : Muhammad Fauzan Adhim
NIM : 1142001017
Program Studi : Informatika
Fakultas : Teknik dan Ilmu Komputer
Judul Skripsi : Rancang Bangun Aplikasi Peringkat Teks Otomatis Berbasis Website Dengan Metode Lantent Semantic Analysis (LSA)

Telah berhasil dipertahankan di hadapan 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 : Prof. Dr. Hoga Saragih, S.T., M.T.  (.....)

Penguji : Yusuf Lestanto, S.T., M.Sc.  22.07.2020 (.....)

Penguji : Ihsan Ibrahim, S.T., M.T.  (.....)

Ditetapkan di : Jakarta

Tanggal : 13 Mei 2020

UNGKAPAN TERIMA KASIH

Puji dan syukur penulis panjatkan kepada Allah SWT, karena berkat rahmat dan karunia-Nya penulis dapat menyelesaikan tugas akhir ini, yang penulis beri judul “Rancang Bangun Aplikasi Peringkat Teks Otomatis Berbasis Website Dengan Metode Latent Semantic Analysis (LSA)”. Tujuan dari penyusunan tugas akhir ini ialah guna memenuhi salah satu syarat untuk bisa menempuh ujian sarjana pada Fakultas Teknik dan Ilmu Komputer Program Studi Informatika Universitas Bakrie.

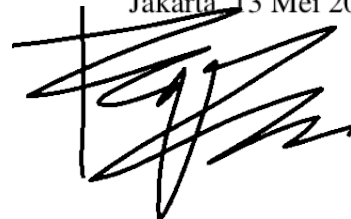
Dalam penyusunan tugas akhir ini, penulis mendapat banyak dukungan dari berbagai pihak hingga terselesaikannya skripsi ini. Penulis mengucapkan banyak terimakasih kepada pihak-pihak yang membantu diantaranya:

1. Tuhan Yang Maha Esa, atas segala rahmatnya penulis diberikan kesehatan dan kelancaran hingga bisa menyelesaikan tugas akhir ini;
2. Kedua orang tua, yang senantiasa memberi dukungan, motivasi dan doa
3. Bapak Prof. Dr. Hoga Saragih, S.T., M.T selaku Kepala Program Studi Informatika, Pembimbing Tugas Akhir, atas semua bimbingan dan motivasi yang tidak ada habisnya diberikan kepada penulis;
4. Bapak Refyul Rey Fatri, Drs., S.Si, M.Sc selaku dosen pembahas seminar proposal yang telah memberikan banyak gagasan serta kritik dan saran yang membangun dalam penyusunan skripsi ini;
5. Bapak Yusuf Lestanto, S.T., M.Sc. dan Bapak Ihsan Ibrahim, S.T., M.T. selaku dosen penguji yang telah memberikan arahan dan kritik serta saran yang membantu dalam penyusunan Tugas Akhir;
6. Seluruh Dosen, Staff dan Karyawan Program Studi Informatika Universitas Bakrie yang telah memberikan ilmu dan pengalaman selama masa perkuliahan hingga menyelesaikan Tugas Akhir;
7. Teman seperjuangan Informatika 2014 yang sudah menemani dan selalu menyemangati;
8. Keluarga Mahasiswa Pecinta Alam Wicaktala yang memberikan pengalaman dan banyak cerita menyenangkan setiap perjalanannya;

9. Keluarga besar Informatika yang mewarnai masa perkuliahan yang panjang, semoga sukses selalu dan dilancarkan dalam kerjaan ataupun pendidikannya;
10. Untuk semua pihak yang sudah banyak membantu yang tidak bisa disebutkan satu persatu.

Penulis tentunya menyadari bahwa dalam penyusunan Tugas Akhir ini masih banyak kekurangan, oleh sebab itu, penulis amat sangat mengharapkan kritik dan saran yang membangun untuk kedepannya. Semoga Tugas Akhir ini bisa berguna dan bermanfaat bagi pembaca.

Jakarta, 13 Mei 2020



Muhammad Fauzan Adhim

HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI

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

Nama : Muhammad Fauzan Adhim
NIM : 11402001017
Program Studi : Informatika
Fakultas : Teknik dan Ilmu Komputer

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

Rancang Bangun Aplikasi Peringkat Teks Otomatis Berbasis Website Dengan Metode Lantent Semantic Analysis (LSA)

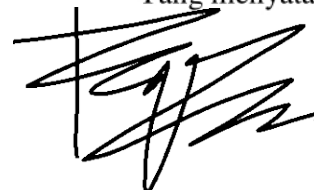
beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneksklusif ini Universitas Bakrie berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta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

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Jakarta

Pada tanggal : 13 Mei 2020

Yang menyatakan



Muhammad Fauzan Adhim

**RANCANG BANGUN APLIKASI PERINGKAS TEKS OTOMATIS
BERBASIS WEBSITE DENGAN METODE *LATENT SEMANTIC
ANALYSIS (LSA)***

Muhammad Fauzan Adhim

ABSTRAK

Penggunaan peringkasan teks secara otomatis adalah solusi agar pembaca atau pengguna dapat mengambil intisari dan informasi dari sebuah teks secara lebih cepat dari biasanya. Banyak algoritma dan pendekatan yang dapat dilakukan untuk membuat sebuah peringkasan teks otomatis, salah satu metode yang terbaru adalah dengan penggunaan *Latent Semantic Analysis (LSA)*. Sejak penggunaan LSA dalam bidang peringkasan pertama kali dikenalkan, sudah ada beberapa algoritma dalam LSA yang dikembangkan. Ringkasan yang merupakan hasil dari prosesnya dinilai sebagai ukuran seberapa efisien dan akurat kah algoritma yang digunakan. Banyak faktor yang memengaruhi penilaian tersebut, salah satunya adalah teks input yang dimasukkan. Pada penelitian ini akan dibangun sebuah aplikasi berbasis web yang dapat meringkas sebuah teks secara otomatis dengan algoritma berbasis LSA yang berbeda, sehingga pengguna nantinya dapat melihat, mengevaluasi dan memilih algoritma peringkasan yang dinilai lebih baik dan akurat hasil ringkasannya.

Kata Kunci: peringkasan teks otomatis, *Latent Semantic Analysis*, *WDLC*, *Natural Language Processing*

**DESIGN AND DEVELOPMENT WEB-BASED AUTOMATIC TEXT
SUMMARIZATION APPLICATION USING LATENT SEMANTIC
ANALYSIS (LSA)**

Muhammad Fauzan Adhim

ABSTRACT

The use of automatic text summarization is a solution that allows readers or users to extract the essence and information from a text faster than conventional method. Variety algorithms and approaches can be used to create an automatic text summary, one of the newest methods is the use of Latent Semantic Analysis (LSA). Since the use of LSA in the text summarization field was first introduced, there have been a number of algorithms in LSA developed. The summary which is the result of the process is assessed as a measure of how efficient and accurate the algorithm is. Many factors could potentially influence the assessment, one of which is the inputted text. In this research a web-based application will be built that can summarize a text automatically with several different LSA-based algorithms, so that users can later see, evaluate and choose a summary algorithm that is considered to be better and accurate.

Keywords: *Automatic Text Summarization, Latent Semantic Analysis, WDLC, Natural Language Processing*

DAFTAR ISI

HALAMAN PERNYATAAN ORISINALITAS	i
HALAMAN PENGESAHAN	ii
UNGKAPAN TERIMA KASIH	iii
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI	v
ABSTRAK	vi
ABSTRACT	vii
DAFTAR ISI	viii
DAFTAR GAMBAR	xi
DAFTAR TABEL	xiii
DAFTAR SINGKATAN	xiv
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	2
1.3. Batasan Masalah	2
1.4. Tujuan Penelitian	3
1.5. Manfaat Penelitian	3
1.6. Sistematika Penulisan	3
BAB II TINJAUAN PUSTAKA	5
2.1. Penelitian Terdahulu	5
2.2. Peringkasan Teks	9
2.2.1. <i>Extractive Summarization</i> dan <i>Abstractive Summarization</i>	10
2.2.2. <i>Single Document</i> dan <i>Multiple Documents</i>	10
2.2.3. <i>Generic</i> dan <i>Query-based</i>	10
2.2.4. <i>Supervised</i> dan <i>Unsupervised</i>	11
2.3. Latent Semantic Analysis	11
2.3.1. Pembuatan <i>Input Matrix</i>	12
2.3.2. Perhitungan SVD.....	13
2.3.3. <i>Sentence Selection</i>	14
2.4. Evaluasi Ringkasan	16
2.4.1. <i>Sentence Co-selection</i>	16
2.4.2. <i>Content-based methods</i>	17
2.4.3. <i>Relevance Correlation</i>	17

2.4.4.	<i>Task-based evaluations</i>	17
2.4.5.	Evaluasi berdasarkan LSA	18
2.5.	<i>Text Preprocessing</i>	19
2.5.1.	<i>Tokenizing</i>	19
2.5.2.	<i>Stemming</i>	20
2.5.3.	<i>Stopwords</i>	20
2.6.	<i>Web Development Life Cycle (WDLC)</i>	20
2.7.	Python	22
2.7.1.	Flask	23
2.8.	<i>Model View Controller (MVC)</i>	23
2.8.1.	<i>Model</i>	23
2.8.2.	<i>View</i>	23
2.8.3.	<i>Controller</i>	24
BAB III METODOLOGI PENELITIAN		25
3.1.	Gambaran Sistem Penelitian	25
3.2.	Analisis Kebutuhan	26
3.3.	Information Gathering (Graphical)	26
3.4.	Analysis (Graphical)	26
3.5.	Graphical Design	26
3.5.1.	Perancangan <i>User Interface</i>	26
3.6.	Information Gathering (Functional)	28
3.7.	Analysis (Functional)	29
3.8.	Functional Design	30
3.8.1.	<i>Use Case Diagram</i>	30
3.8.2.	<i>Use Case Scenario</i>	31
3.8.2.1.	Skenario Memilih Algoritma Peringkasan	31
3.8.2.2.	Skenario Memasukan Teks Dokumen	31
3.8.2.3.	Skenario Melihat Hasil Peringkasan	32
3.8.2.4.	Skenario Melihat Hasil Evaluasi Peringkasan	33
3.8.2.5.	Skenario Menghapus Data Peringkasan	33
3.8.3.	<i>Activity Diagram</i>	34
3.8.3.1.	Memilih Algoritma Peringkasan	34
3.8.3.2.	Memasukan Teks Dokumen	35
3.8.3.3.	Melihat Hasil Peringkasan	36

3.8.3.4.	Melihat Hasil Evaluasi Peringkasan.....	37
3.8.3.5.	Menghapus Data Peringkasan.....	38
3.8.4.	<i>Sequence Diagram</i>	38
3.8.4.1.	<i>Sequence Diagram</i> Memilih Algoritma Peringkasan	39
3.8.4.2.	<i>Sequence Diagram</i> Memasukkan Teks Dokumen	40
3.8.4.3.	<i>Sequence Diagram</i> Melihat Hasil Peringkasan.....	41
3.8.4.4.	<i>Sequence Diagram</i> Melihat Evaluasi Peringkasan	42
3.8.4.5.	<i>Sequence Diagram</i> Menghapus Data Peringkasan	42
3.8.5.	<i>Model View Control Diagram</i>	42
3.8.5.1.	<i>MVC Class Structure</i>	43
3.8.5.2.	<i>Model Class Diagram</i>	43
BAB IV	HASIL DAN PEMBAHASAN	44
4.1.	Implementation	44
4.1.1.	Spesifikasi <i>Software</i> dan <i>Hardware</i> yang digunakan.	44
4.1.2.	Implementasi Rancangan <i>User Interface</i>	45
4.1.3.	Implementasi Fungsi dan Algoritma.....	48
4.1.3.1.	<i>LSA-based Summarization</i>	48
4.1.3.2.	<i>Summary Evaluation</i>	51
4.1.3.3.	<i>Routes</i>	52
4.2.	Testing	54
BAB V	PENUTUP	62
5.1.	Simpulan	62
5.2.	Saran	62
DAFTAR PUSTAKA	63

DAFTAR GAMBAR

Gambar 2.1 First Singular Vector	18
Gambar 2.2 Web Development Life Cycle.....	21
Gambar 3.1 Gambaran Sistem Keseluruhan	25
Gambar 3.2 Rancangan UI Halaman Utama.....	27
Gambar 3.3 Rancangan Pengaturan Peringkasan.....	27
Gambar 3.4 Rancangan Halaman Evaluation Data.....	28
Gambar 3.5 Use Case Diagram.....	30
Gambar 3.6 Activity Diagram Memilih Algoritma Peringkasan	35
Gambar 3.7 Activity Diagram Memasukan Teks Dokumen.....	36
Gambar 3.8 Activity Diagram Melihat Hasil Peringkasan	37
Gambar 3.9 Activity Diagram Melihat Hasil Evaluasi Peringkasan.....	37
Gambar 3.10 Activity Diagram Menghapus Data Peringkasan	38
Gambar 3.11 Sequence Diagram Memilih Algoritma Peringkasan	39
Gambar 3.12 Sequence Diagram Memasukkan Teks Dokumen	40
Gambar 3.13 Sequence Diagram Melihat Hasil Peringkasan	41
Gambar 3.14 Sequence Diagram Melihat Evaluasi Peringkasan	42
Gambar 3.15 Sequence Diagram Menghapus Data Peringkasan	42
Gambar 3.16 Struktur Kelas MVC	43
Gambar 3.17 Model Class Diagram	43
Gambar 4.1 User Interface Halaman Utama	45
Gambar 4.2 User Interface Menu Summarization Setting	46
Gambar 4.3 User Interface Halaman Evaluation-Result	47
Gambar 4.4 User Interface Halaman Evaluation-Data	47
Gambar 4.5 Implementasi Algoritma Preprocessing–Tokenizer	48
Gambar 4.6 Implementasi Algoritma Preprocessing–Stemming dan Stopword ..	49
Gambar 4.7 Implementasi Algoritma Pembuatan Document-term Matrix	49
Gambar 4.8 Implementasi Algoritma Pembuatan SVD	50
Gambar 4.9 Implementasi Algoritma LSA Gong&Liu	50
Gambar 4.10 Implementasi Algoritma LSA Steinberger&Jezek	51
Gambar 4.11 Implementasi Algoritma LSA Cross Method	51
Gambar 4.12 Implementasi LSA-based Evaluation	52

Gambar 4.13 Routes index.....	52
Gambar 4.14 Routes summarization.....	53
Gambar 4.15 Routes save_evaluation.....	53
Gambar 4.16 Routes evaluasi.....	54
Gambar 4.17 Routes detail evaluasi.....	54
Gambar 4.18 Uji Coba Teks Input.....	55
Gambar 4.19 <i>Default Setting</i> Kode.....	55
Gambar 4.20 Peringatan Minimal Input.....	56
Gambar 4.21 Syarat Kondisi Minimal Input Kode.....	56
Gambar 4.22 Syarat Kondisi Evaluasi.....	56
Gambar 4.23 Syarat Kondisi Evaluasi.....	57
Gambar 4.24 JSON Response Hasil Peringkasan.....	57
Gambar 4.25 Query Evaluation Summary.....	58
Gambar 4.26 Halaman Evaluation Summary.....	58
Gambar 4.27 Query Evaluation Details.....	59

DAFTAR TABEL

Tabel 2.1 Rangkuman Penelitian Terkait.....	7
Tabel 3.1 Use Case Scenario Memilih Algoritma Peringkasan.....	31
Tabel 3.2 Use Case Scenario Memasukan Teks Dokumen.....	31
Tabel 3.3 Use Case Scenario Melihat Hasil Peringkasan	32
Tabel 3.4 Use Case Scenario Melihat Hasil Evaluasi Peringkasan	33
Tabel 3.5 Use Case Scenario Menghapus Data Peringkasan	33
Tabel 4.1 Spesifikasi Hardware	44
Tabel 4.2 Spesifikasi Software.....	44
Tabel 4.3 Blackbox Testing	60

DAFTAR SINGKATAN

LSA	<i>Latent Semantic Analysis</i>
NLP	<i>Natural Language Processing</i>
IR	<i>Information Retrieval</i>
Tf-IDF	<i>Term Frequency – Inverse Document Frequency</i>
MMR	<i>Maximum Marginal Relevance</i>
SVD	<i>Singular Value Decomposition</i>
DTM	<i>Document Term Matrix</i>
WDLC	<i>Web Development Life-Cycle</i>
MVC	<i>Model View Controller</i>
URL	<i>Uniform Resource Locator</i>