

**RANCANG BANGUN APLIKASI PERINGKAS TEKS
OTOMATIS ARTIKEL BERBAHASA INDONESIA
MENGGUNAKAN METODE *TERM FREQUENCY INVERSE
DOCUMENT FREQUENCY (TF-IDF)* DAN *K-MEAN
CLUSTERING***

TUGAS AKHIR

Diajukan Sebagai Salah Satu Syarat
Untuk Memperoleh Gelar Sarjana Teknik
Pada Jurusan Teknik Informatika

Oleh

FENDRA PRATAMA

11051100003



UIN SUSKA RIAU

**FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SULTAN SYARIF KASIM RIAU
PEKANBARU
2014**

LEMBAR PERSETUJUAN

RANCANG BANGUN APLIKASI PERINGKAS TEKS
OTOMATIS ARTIKEL BERBAHASA INDONESIA
MENGGUNAKAN METODE *TERM FREQUENCY INVERSE*
DOCUMENT FREQUENCY (TF-IDF) DAN *K-MEAN*
CLUSTERING

TUGAS AKHIR

Oleh

FENDRA PRATAMA
11051100003

Telah diperiksa dan disetujui sebagai laporan tugas akhir
di Pekanbaru, pada tanggal 23 April 2014

Pembimbing



Lestari Handayani, S.T., M.Kom
NIP. 19811113 200710 2 003

LEMBAR PENGESAHAN

RANCANG BANGUN APLIKASI PERINGKAS TEKS
OTOMATIS ARTIKEL BERBAHASA INDONESIA
MENGGUNAKAN METODE *TERM FREQUENCY INVERSE
DOCUMENT FREQUENCY (TF-IDF) DAN K-MEAN
CLUSTERING*

TUGAS AKHIR

Oleh

FENDRA PRATAMA
11051100003

Telah dipertahankan di depan sidang pengaji
Sebagai salah satu syarat untuk memperoleh gelar sarjana Teknik Informatika
Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau
Di Pekanbaru, pada tanggal 22 April 2014

Pekanbaru, 23 April 2014

Mengesahkan,

Ketua Jurusan

Dekan



Dra. Hj. Yenita Morena, M.Si
NIK. 19601125 198503 2 002


Elin Haerani, S.T., M.Kom
NIP. 19810523 200710 2 003

DEWAN PENGUJI

Ketua	:	Dra. Hj. Yenita Morena, M.Si
Sekretaris	:	Lestari Handayani, S.T., M.Kom
Penguji I	:	Dr. Okfalisa, S.T, M.Sc
Penguji II	:	Rahmad Abdillah, M.T



**RANCANG BANGUN APLIKASI PERINGKAS TEKS
OTOMATIS ARTIKEL BERBAHASA INDONESIA
MENGGUNAKAN METODE *TERM FREQUENCY INVERSE*
DOCUMENT FREQUENCY (TF-IDF) DAN K-MEAN
*CLUSTERING***

FENDRA PRATAMA

11051100003

Tanggal Sidang: 22 April 2014

Periode Wisuda: Juni 2014

Jurusan Teknik Informatika

Fakultas Sains dan Teknologi

Universitas Islam Negeri Sultan Syarif Kasim Riau

ABSTRAK

Membaca teks artikel secara keseluruhan membutuhkan waktu yang lama untuk mendapatkan informasi penting. Peringkas teks otomatis menawarkan solusi untuk membantu memperoleh informasi penting dengan membaca isi artikel dalam bentuk deskripsi singkat berupa ringkasan (*summary*). Dengan adanya ringkasan, informasi bisa diakses dengan cepat dan tidak perlu membutuhkan waktu yang lama dengan cara mengambil isi yang paling penting dari teks bacaan tersebut bagi penggunanya. Penerapan metode TF-IDF bertujuan untuk menghitung bobot tiap kalimat dan pada metode *K-Mean Clustering* bertujuan untuk meminimalisasikan variasi kalimat di dalam suatu *cluster* dan memaksimalisasikan variasi kalimat antar *cluster*. Pemilihan kalimat penting berdasarkan bobot kalimat paling besar pada tiap *cluster* dengan mengambil *compression* sebesar 25% dan 40%. Pengujian hasil ringkasan sistem pada *compression* 40% menghasilkan rata-rata *recall* 60%, *precision* 62%, dan *f-measure* 61%. Sedangkan untuk *compression* 25% menghasilkan rata-rata *recall* 37%, *precision* 44%, dan *f-measure* 40%. Penilaian hasil ringkasan sistem oleh *expert judgement* pada *compression* 40% menghasilkan nilai rata-rata 3,15 atau pada *range* diterima, sedangkan pada *compression* 25% menghasilkan nilai rata-rata 1,83 atau berada pada *range* cukup diterima. Dari hasil pengujian dan penilaian terhadap hasil ringkasan sistem tersebut membuktikan bahwa hasil ringkasan teks artikel dengan *compression* 40% lebih baik daripada hasil ringkasan dengan *compression* 25%.

Kata kunci: *F-Measure, K-Mean Clustering, Precision, Recall, TF-IDF*

***DESIGN AND APPLICATION OF AUTOMATIC TEXT
SUMMARIZATION ARTICLES IN INDONESIAN LANGUAGE
USING TERM INVERSE DOCUMENT FREQUENCY
FREQUENCY (TF-IDF) AND K-MEAN CLUSTERING***

FENDRA PRATAMA

11051100003

Final Exam Date: April 22th, 2014

Graduation Ceremony Period: June 2014

Informatics Engineering Department

Faculty of Science and Technology

State Islamic University of Sultan Syarif Kasim Riau

ABSTRACT

Reading the text of the article as a whole takes a long time to get the important information. Automatic text summarization offers a solution to help get important information by reading the contents of the article in the form of a short description in the form of a summary. With the summary, the information can be accessed quickly and do not need to take a long time by taking the most important content of the reading text for its users. Application of TF-IDF method aims to calculate the weight of each sentence and the K-Mean Clustering methods aim to minimize variations of a sentence in cluster and maximize variation of a sentence inter-cluster. Selection of important sentences based on the weight of the greatest sentences in each cluster by taking a compression of 25% and 40 %. The test results summary on the compression 40 % generates an average of 60 % recall, 62 % precision, and f-measure of 61 %. As for the compression of 25 % produces an average of 37 % recall, 44 % precision, and f -measure of 40%. Assessment results summary of the system by expert judgement at 40 % compression produces an average value of 3.15 or acceptable range, whereas at 25 % compression produces an average value of 1.83 or be in the range is quite acceptable. From the results of the testing and assessment of the system summary results prove that the results summary text articles with compression of 40 % better than the summary results with compression of 25 %.

Key Words: *F-Measure, K-Mean Clustering, Precision, Recall, TF-IDF*

KATA PENGANTAR

Puji syukur kehadirat Allah SWT, karena atas segala limpahan rahmat dan karunia-Nya yang diberikan sehingga penulis dapat menyelesaikan penelitian sekaligus penulisan laporan tugas akhir ini. Shalawat serta salam Penulis kirimkan untuk junjungan Rasulullah Muhammad SAW.

Laporan tugas akhir ini merupakan salah satu prasyarat untuk memenuhi persyaratan akademis dalam rangka meraih gelar kesarjanaan di Jurusan Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Sultan Syarif Kasim Riau (UIN SUSKA RIAU). Selama menyelesaikan tugas akhir ini, penulis telah banyak mendapatkan bantuan, bimbingan, dan petunjuk dari banyak pihak baik secara langsung maupun tidak langsung. Untuk itu dalam kesempatan ini penulis ingin mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Prof. Dr. H. M. Nazir, selaku Rektor Universitas Islam Negeri Sultan Syarif Kasim Riau.
2. Dra. Yenita Morena, M.Si, selaku Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau.
3. Elin Haerani, S.T, M.Kom selaku Ketua Jurusan Teknik Informatika, Fakultas Sains dan Teknologi.
4. Lestari Handayani, S.T, M.Kom selaku dosen pembimbing.
5. Dr. Okfalisa, S.T, M.Sc selaku dosen penguji 1.
6. Rahmad Abdillah, M.T selaku dosen penguji 2.
7. M. Affandes, S.T., M.T selaku koordinator TA.
8. Arini AR, S.Pd selaku *Expert Judgement* 1.
9. H. Rushan, M.Pd selaku *Expert Judgement* 2.
10. Chairul. Y (Ayah), Yurmaidas (Mama), Fitria Alvionika dan Fadhil Chairul yang selalu memberikan doa, dorongan dan semangat untuk keberhasilan penulis.
11. Isra Wati Fazri yang selalu memberikan doa dan motivasi untuk keberhasilan penulis.

12. Teman-teman seperjuangan TIF B '10 dan teman-teman sesama Mahasiswa Teknik Informatika yang juga turut memberi semangat luar biasa.
13. Dan terakhir, terima kasih pula penulis ucapkan untuk Almamater Teknik Informatika, Fakultas Sains dan Teknologi, UIN Sultan Syarif Kasim Riau serta pihak-pihak lain yang tidak dapat penulis sebutkan satu persatu. Terima kasih banyak atas bantuan dan dukungannya yang sangat berharga.

Penulis menyadari dalam penulisan laporan ini masih terdapat kekurangan. Oleh karena itu, saran dan kritik sangat penulis harapkan untuk kemajuan penulis secara pribadi. Terima kasih.

Pekanbaru, 22 April 2014

Penulis

DAFTAR ISI

HALAMAN JUDUL	i
LEMBAR PERSETUJUAN	ii
LEMBAR PENGESAHAN	iii
LEMBAR HAK ATAS KELAYAKAN INTELEKTUAL	iv
LEMBAR PERNYATAAN	v
LEMBAR PERSEMBAHAN	vi
ABSTRAK	vii
<i>ABSTRACT</i>	viii
KATA PENGANTAR	ix
DAFTAR ISI	xi
DAFTAR GAMBAR	xv
DAFTAR TABEL	xvi
DAFTAR RUMUS	xviii
DAFTAR LAMPIRAN	xix
BAB I PENDAHULUAN	I-1
1.1. Latar Belakang	I-1
1.2. Rumusan Masalah	I-3
1.3. Batasan Masalah	I-4
1.4. Tujuan Penelitian	I-4
1.5. Sistematika Penulisan	I-4
BAB II LANDASAN TEORI	II-1
2.1. Peringkas Teks Otomatis.....	II-1
2.1.1. Definisi Peringkas Teks Otomatis	II-1
2.1.2. Tahapan Membuat Ringkasan.....	II-1

2.1.3. Rasio Kompresi (<i>Compression Rate</i>).....	II-2
2.1.4. Pendekatan Peringkas Teks.....	II-2
2.1.5. Tujuan Ringkasan.....	II-3
2.1.6. Artikel	II-3
2.2. Pra Proses (<i>Preprocessing</i>)	II-4
2.3. Proses Peringkasan Teks	II-8
2.4. Algoritma TF-IDF.....	II-9
2.5. Algoritma <i>K-Means Clustering</i>	II-9
2.6. Teknik Evaluasi Peringkasan Teks	II-11
BAB III METODOLOGI PENELITIAN	III-1
3.1. Studi Pendahuluan	III-2
3.2. Perumusan Masalah	III-2
3.3. Pengumpulan Data	III-2
3.4. Analisa Sistem.....	III-3
3.4.1. Analisa Kebutuhan Data	III-3
3.4.2. Analisa Peringkasan Teks	III-3
3.4.3. Analisa Fungsional Sistem.....	III-4
3.4.4. Analisa Data Sistem	III-4
3.5. Perancangan Sistem	III-4
3.5.1. Perancangan Basis Data	III-4
3.5.2. Perancangan Antar Muka (<i>Interface</i>).....	III-5
3.5.3. Perancangan <i>Procedural</i>	III-5
3.6. Implementasi	III-5
3.7. Pengujian	III-5
3.7.1. Pengujian <i>Blackbox</i>	III-5
3.7.2. Pengujian <i>Expert Judgement</i>	III-5
3.8. Kesimpulan dan Saran.....	III-6
BAB IV ANALISA DAN PERANCANGAN	IV-1
4.1. Analisa.....	IV-1
4.1.1. Analisa Kebutuhan Data	IV-1
4.1.2. Analisa Peringkasan Teks	IV-1
a. <i>Text Preprocessing</i>	IV-1

a.1.	Pemecahan Kalimat.....	IV-3
a.2.	<i>Case Folding</i>	IV-6
a.3.	Penghapusan <i>Stopword</i>	IV-8
a.4.	<i>Stemming</i>	IV-10
a.5.	<i>Tokenizing</i>	IV-11
b.	<i>Topic Identification</i>	IV-12
c.	<i>Interpretation</i>	IV-13
c.1.	TF-IDF	IV-13
c.2.	<i>K-Mean Clustering</i>	IV-15
d.	<i>Generating</i>	IV-27
4.1.3.	Analisa Fungsional Sistem.....	IV-31
a.	<i>Contex Diagram</i>	IV-32
b.	<i>Data Flow Diagram</i>	IV-32
4.1.4.	Analisa Data Sistem	IV-35
4.2.	Perancangan	IV-35
4.2.1.	Perancangan Basis Data	IV-35
4.2.2.	Perancangan <i>Interface</i> Sistem	IV-36
a.	Rancangan <i>Interface Front-End</i>	IV-36
4.2.3.	Perancangan <i>Procedural</i>	IV-38
BAB V	IMPLEMENTASI DAN PENGUJIAN	V-1
5.1.	Implementasi	V-1
5.1.1.	Batasan Implementasi	V-1
5.1.2.	Lingkungan Implementasi.....	V-2
a.	Kebutuhan Perangkat Keras (<i>Hardware</i>)	V-2
b.	Kebutuhan Perangkat Lunak (<i>Software</i>)	V-2
5.1.3.	Implementasi <i>Interface</i> Sistem.....	V-3
a.	Implementasi <i>Interface Front-End</i>	V-3
5.2.	Pengujian Sistem	V-5
5.2.1.	Pengujian <i>Blackbox</i>	V-5
5.2.2.	Pengujian <i>Expert Judgement</i>	V-14
a.	Prosedur Pengujian <i>Expert Judgement</i>	V-14
b.	Hasil Pengujian <i>Expert Judgement</i>	V-14

c. Kesimpulan Hasil Pengujian <i>Expert Judgement</i>	V-36
5.3. Penilaian Hasil Ringkasan Sistem	V-37
BAB VI PENUTUP	VI-1
6.1. Kesimpulan	VI-1
6.2. Saran.....	VI-2
DAFTAR PUSTAKA	
DAFTAR RIWAYAT HIDUP	