

***AUTOMATIC DOCUMENT
SUMMARIZATION MENGGUNAKAN METODE MAXIMUM
MARGINAL RELEVANCE DENGAN TOP-N TF-IDF
TOKEN QUERY EXPANSION PADA DOKUMEN PENDEK***

TUGAS AKHIR

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

Oleh

SELLA NOVY ANGGRENI

11051201645



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

2014

LEMBAR PENGESAHAN

AUTOMATIC DOCUMENT SUMMARIZATION MENGUNAKAN METODE *MAXIMUM MARGINAL RELEVANCE* DENGAN *TOP-N TF-IDF TOKEN QUERY EXPANSION* PADA DOKUMEN PENDEK

TUGAS AKHIR

Oleh :

SELLA NOVY ANGGRENI
11051201645

Telah dipertahankan di depan sidang dewan penguji
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 2 Oktober 2014.

Pekanbaru, 2 Oktober 2014
Mengesahkan

Ketua Jurusan



Dra. Hj. Venita Morena, M.Si
NIP. 19601125 198503 2 002

Muhammad Irsyad ST., MT
NIP. 19780508 200710 1 007

DEWAN PENGUJI

Ketua : Drs. Martius M.Hum
Sekretaris : Muhammad Safrizal, ST., M.Cs
Anggota I : Dra. Okfalisa, ST., M.Sc
Anggota II : Rahmad Abdillah, M.T

AUTOMATIC DOCUMENT SUMMARIZATION MENGUNAKAN METODE MAXIMUM MARGINAL RELEVANCE DENGAN TOP-N TF-IDF TOKEN QUERY EXPANSION PADA DOKUMEN PENDEK

SELLA NOVY ANGGRENI
11051201645

Tanggal Sidang : 02 Oktober 2014

Periode Wisuda : November 2014

Jurusan Teknik Informatika

Fakultas Sains dan Teknologi

Universitas Islam Negeri Sultan Syarif Kasim Riau

ABSTRAK

Peringkasan teks otomatis (*Automatic Text Summarization*) adalah pembuatan versi yang lebih singkat dari sebuah teks dengan memanfaatkan aplikasi yang dijalankan pada komputer. Peringkasan teks otomatis berguna untuk membantu manusia dalam mendapatkan ringkasan dari suatu bacaan tanpa harus membaca semua isi dari bacaan. Metode yang digunakan dalam *Automatic document summarization* ini adalah metode TF-IDF dan MMR. Penelitian ini diawali dengan lima tahap *text preprocessing*: pemecahan kalimat, *case folding*, *tokenizing*, *filtering*, dan *stemming*. Proses selanjutnya menghitung bobot tf-idf, bobot *query relevance* dan bobot *similarity*. Ringkasan dihasilkan dari ekstraksi kalimat dengan menggunakan *metode maximum marginal relevance*. Metode ekstraksi *maximum marginal relevance* merupakan metode yang digunakan untuk mengurangi redundansi dalam perangkangan kalimat. Data uji coba diambil dari berita berbahasa Indonesia online sejumlah 3 berita. Hasil pengujian dibandingkan dengan ringkasan manual menghasilkan akurasi pada *compression* 25% adalah 64 % sedangkan pada *compression* 50% sebesar 76%.

Kata Kunci : *Automatic document summarization, Maximum Marginal Relevance, Query Expansion, Text Processing, TF-IDF.*

KATA PENGANTAR

Assalammu'alaikum wa rahmatullahi wa barakatuh.

Alhamdulillah, Puji dan Syukur penulis panjatkan kehadiran Allah SWT, karena berkat rahmat dan karunia-Nya penulis dapat menyelesaikan Skripsi ini yang disusun untuk memenuhi salah satu syarat kelulusan pada Program Studi Strata 1 pada jurusan Teknik Informatika Universitas Islam Negeri Sultan Syarif Kasim Riau dengan judul "*Automatic document summarization menggunakan metode Maximum Marginal Relevance dengan Top-N TF-IDF token Query Expansion Pada Dokumen Pendek*". Banyak sekali pihak yang telah membantu penulis dalam penyusunan laporan ini, baik berupa bantuan materi ataupun berupa motivasi dan dukungan kepada penulis. Semua itu tentu terlalu banyak bagi penulis untuk membalasnya, namun pada kesempatan ini penulis hanya dapat mengucapkan terima kasih kepada :

1. Kedua Orang Tua Penulis, serta kakak dan adik-adik yang tiada hentinya memanjatkan doa, memberikan dukungan dan semangat untuk kesuksesan penulis.
2. Bapak Prof. Dr. H. M. Nazir, selaku Rektor Universitas Islam Negeri Sultan Syarif Kasim Riau.
3. Ibu Dra. Hj. Yenita Morena, M.Si, selaku Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau.
4. Bapak Muhammad Irsyad, ST., MT selaku Ketua Jurusan Teknik Informatika Fakultas Sains dan Teknologi UIN SUSKA RIAU.
5. Bapak Surya Agustian, S.T, M. Kom dan Bapak Muhammad Safrizal, ST., M.Cs Selaku dosen pembimbing tugas akhir. Terimakasih untuk waktu yang selalu bapak luangkan untuk penulis, ilmu, semangat, dan motivasi yang luar biasa.
6. Ibu Dra. Okfalisa, ST., M.Sc selaku dosen penguji 1 yang banyak membantu dan memberi masukan penulis dalam penyempurnaan Laporan Tugas Akhir ini.
7. Bapak Rahmad Abdillah, M.T, selaku dosen penguji 2, terimakasih pak untuk ilmu, saran, perbaikan, dan masukannya.

8. Muhammad Affandes, M.T, sebagai koordinator tugas akhir yang telah memberi masukan-masukan untuk penyelesaian tugas akhir ini, dan sangat sabar membantu penulis dalam mempersiapkan semua kebutuhan penulis dalam penyelesaian Tugas Akhir ini.
9. Ibu Rizqa Raiqa Bintana, S.T terimakasih buk sudah memberikan ilmu, saran, dan masukkan kepada penulis dalam menyelesaikan laporan Tugas Akhir ini.
10. Bapak dan Ibu Dosen Jurusan Teknik Informatika Fakultas Sains dan Teknologi UIN SUSKA RIAU yang telah banyak memberikan ilmunya kepada penulis.
11. Yhozi Aulia Rahman yang selalu menyemangati, memberikan support, dan doanya selama ini untuk penulis. Thanks for all dear.
12. Buat teman-teman kesayangan “TIF C” Mak Ipi, Mak Ayu, Mak Aliyah, Didit Master, Berry Blogger, Yudhi Galau, Rizky Guntang, Wahyudi Ucok, Bang Risman, Idham, Pak Gub Ari, Ihsan Pendiam, Mas oko, Dega Cino, Ust Ahyar, Dede Everglade, Ocu Rian, Wendi, Iqbal, Galiyong, Kelik Abi, Aldy, Rovi, Mbak Neneng, Lihong, Fya, Ziah poenya, Yayan dan teman-teman sekelas. Tetap semangat dan tetap Berjaya kawan.
13. Abe selaku kakak yang selalu memberi motivasi dan menemani dalam pengerjaan tugas akhir ini.

Penulis menyadari bahwa dalam penulisan laporan ini masih banyak kesalahan dan kekurangan, oleh karena itu kritik dan saran yang sifatnya membangun sangat penulis harapkan untuk kesempurnaan laporan ini. Akhirnya penulis berharap semoga laporan ini dapat memberikan sesuatu yang bermanfaat bagi siapa saja yang membacanya. Amin.

Wassalamu'alaikum wa rahmatullahi wa barakatuh

Pekanbaru, 02 Oktober 2014

Penulis

DAFTAR ISI

LEMBAR PERSETUJUAN.....	ii
LEMBAR PENGESAHAN	iii
LEMBAR HAK ATAS KEKAYAAN INTELEKTUAL.....	iv
LEMBAR PERNYATAAN.....	v
LEMBAR PERSEMBAHAN	vi
ABSTRAK	vii
ABSTRACT.....	viii
KATA PENGANTAR	ix
DAFTAR ISI.....	xi
DAFTAR GAMBAR	xiv
DAFTAR TABEL.....	xv
DAFTAR LAMPIRAN.....	xvi
DAFTAR RUMUS	xvii
BAB I. PENDAHULUAN	I-1
1.1 Latar Belakang.....	I-1
1.2 Rumusan Masalah.....	I-3
1.3 Batasan Masalah	I-3
1.4 Tujuan Penelitian.....	I-3
1.5 Sistematika Penulisan.....	I-3
BAB II. LANDASAN TEORI.....	II-1
2.1 Kata	II-1
2.2 Kalimat	II-2
2.3 Paragraf.....	II-2
2.4. Tipe Evaluasi	II-3
2.5. Pemrosesan Teks	II-3
2.4.1. Pemacahan Kalimat.....	II-4
2.4.2. <i>Case Folding</i>	II-5
2.4.3. <i>Filtering</i> Kalimat.....	II-5
2.4.3. <i>Tokenizing</i> Kata.....	II-5

2.4.5 <i>Stemming</i> dengan Algoritma Nazief dan Andriani.....	II-6
2.6. <i>Automatic Text Summarization</i>	II-8
2.6.1 Riset Tentang Peringkat teks Sebelumnya.....	II-9
2.6.2 Hipotesa	II-10
2.7. Algoritma TF-IDF.....	II-10
2.8. <i>Cosine Similarity</i>	II-11
2.9. Teknik <i>text summarization</i> dengan MMR	II-12
2.10. Tahapan Algoritma pada <i>Automatic Document Summarization</i>	II-13
BAB III. METODOLOGI PENELITIAN	III-1
3.1. Studi Pustaka.....	III-1
3.2. Perencanaan dan Analisa.....	III-2
3.3. Perancangan Sistem	III-3
3.4. Implementasi	III-3
3.5. Pengujian.....	III-3
BAB IV. ANALISA DAN PERANCANGAN.....	IV-1
4.1. Analisa Sistem Lama.....	IV-1
4.2. Analisa Sistem Baru	IV-1
4.3. Deskripsi Fungsional.....	IV-3
4.3.1. <i>Data Flow Diagram</i> (DFD).....	IV-3
4.3.2. <i>Entity Relation Diagram</i> (ERD)	IV-4
4.3.1. <i>Flowchart</i>	IV-4
4.3.3.1. Analisa <i>Text Preprocessing</i>	IV-4
4.3.3.2. Analisa Pembobotan TF-IDF	IV-12
4.3.3.3. Analisa Pembobotan <i>Relevance Query</i>	IV-13
4.3.3.4. Analisa Pembobotan <i>Similarity</i> Kalimat.....	IV-13
4.3.3.5. Analisa <i>Maximum Marginal Relevance</i>	IV-14
4.3.3.6. Analisa <i>Query Expansion</i>	IV-27
4.4. Perancangan <i>Interface</i>	IV-28
4.4.1. <i>Interface</i> Menu Ringkasan Dokumen Ideal.....	IV-29
4.4.2. <i>Interface</i> Menu Ringkasan Dokumen Website.....	IV-30
BAB V. IMPLEMENTASI DAN PENGUJIAN	V-1
5.1. Tahap Implementasi	V-1

5.1.1. Batasan Implementasi	V-1
5.1.2. Lingkungan Implementasi.....	V-1
5.1.3. Implementasi <i>Interface</i> Aplikasi	V-2
5.2. Pengujian Aplikasi	V-5
5.2.1 Pengujian <i>Black Box</i>	V-5
5.3.2 Pengujian Hasil Ringkasan	V-6
5.3. Kesimpulan Pengujian	V-9
BAB VI. PENUTUP	VI-1
6.1. Kesimpulan	VI-1
6.2. Saran.....	VI-1
DAFTAR PUSTAKA	
LAMPIRAN	
DAFTAR RIWAYAT HIDUP	