

**Pembandingan Aplikasi Peringkasan Multi Dokumen
menggunakan Sentence Scoring dan Maximum
Marginal Relevance dengan K- Means**

TUGAS AKHIR

**Diajukan Untuk Memenuhi Sebagai Persyaratan Mencapai
Derajat Sarjana Teknik Informatika**



Oleh:

VALENTINA BERLIAN PROVITASARI

12 07 06893

PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS TEKNOLOGI INDUSTRI

UNIVERSITAS ATMA JAYA YOGYAKARTA

2016

Halaman Pengesahan

Tugas Akhir Berjudul

PEMBANDINGAN APLIKASI PERINGKASAN MULTI DOKUMEN MENGUNAKAN SENTENCE SCORING DAN MAXIMUM MARGINAL RELEVANCE DENGAN K- MEANS

Disusun oleh:

Valentina Berlian Provitasaki (NIM: 12 07 06893)

Dinyatakan telah memenuhi syarat
Pada tanggal : Mei 2016

Oleh:

Pembimbing I,

(Y. Sigit P.W.P., S.T., M.Kom)

Pembimbing II,

(Kusworo Anindito, S.T., M.T)

Tim Penguji:

Penguji I

(Y. Sigit P.W.P., S.T., M.Kom)

Penguji II,

(Th. Devi Indriasari, S.T., M.Sc.)

Penguji III,

(Thomas Suselo, S.T., M.T.)

Yogyakarta, Mei 2016

Universitas Mahadewa Jaya Yogyakarta
Fakultas Teknologi Industri



FAKULTAS
TEKNOLOGI INDUSTRI

(Dr. A. Teguh Siswanto, M.Sc.)

Pernyataan Originalitas

Saya yang bertanda tangan di bawah ini:

Nama : Valentina Berlian Provitasari

NPM : 12 07 06893

Dengan ini menyatakan bahwa Tugas Akhir saya dengan judul "Pembandingan Aplikasi Peringkasan Multi Dokumen menggunakan Sentence Scoring dan Maximum Marginal Relevance dengan K- Means" merupakan hasil penelitian saya pada Tahun Akademik 2015/2016 yang bersifat original dan tidak mengandung plagiasi dari karya manapun.

Bilamana di kemudian hari ditemukan ketidaksesuaian dengan pernyataan ini, maka saya bersedia dituntut dan diproses sesuai dengan ketentuan yang berlaku termasuk untuk dicabut gelar Sarjana yang telah diberikan Universitas Atma Jaya Yogyakarta kepada saya.

Demikian pernyataan ini dibuat dengan sesungguhnya dan dengan sebenar-benarnya.

Yogyakarta, Mei 2016

Yang Menyatakan,



Valentina Berlian Provitasari

KONTRAK KERJA PENELITIAN

Kami yang bertanda tangan di bawah ini,

Nama : Yohanes Sigit Purnomo W.P., S.T., M.Kom.
NPP : 04.01.710
Jabatan : Dosen

yang selanjutnya disebut sebagai Pihak Pertama, dan

Nama : Valentina Berlian Provitasari
NPM : 12 07 06893
Jabatan : Mahasiswa

yang selanjutnya disebut sebagai Pihak Kedua.

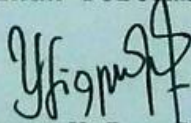
Pihak pertama dan pihak kedua sepakat untuk melakukan kontrak kerja yaitu sebagai mahasiswa peserta Skripsi/Tugas Akhir dengan ketentuan sebagai berikut:

1. Pihak kedua menandatangani kontrak kerja dan melakukan Skripsi/Tugas Akhir dengan topik yang berkaitan dengan penelitian yang dilakukan oleh pihak pertama.
2. Topik dari Skripsi/Tugas Akhir adalah "Pembandingan Aplikasi Peringkasan Multi Dokumen Menggunakan Sentence Scoring dan Maximum Marginal Relevance dengan K- Means".
3. Masa kerja mahasiswa dalam menyelesaikan Skripsi/Tugas Akhir selama 6 bulan sejak ditandatanganinya kontrak kerja ini dan dapat diperpanjang maksimal 1 kali untuk periode 3 bulan.
4. Pihak kedua wajib memberikan laporan hasil perkembangan kerja kepada pihak pertama sekurang-kurangnya 2 minggu sekali.

Demikian isi kontrak kerja ini, semoga dapat dipergunakan sebagaimana mestinya, dan masing-masing pihak yang terlibat akan menunaikan kewajibannya hingga selesai.

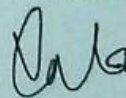
Disetujui,

Pihak Pertama



Y. Sigit Purnomo W.P., S.T., M.Kom.

Pihak Kedua



Valentina Berlian Provitasari

Halaman Persembahan

Tugas Akhir ini dipersembahkan untuk:
Tuhan Yesus,
Teman yang selalu mendampingi dikala duka dan suka.

“

Besides, the world isn't split into good people and Death Eaters. We've all got both light and dark inside us. What matters is the part we choose to act on. That's who we really are.

”

-Sirius Black

Kata Pengantar

Puji dan syukur penulis panjatkan kepada Tuhan karena atas berkat dan rahmat-Nya, tugas akhir ini dapat terselesaikan. Penulisan laporan tugas akhir ini bertujuan untuk memenuhi salah satu syarat memperoleh gelar kesarjanaan di Jurusan Teknik Informatika, Fakultas Teknologi Industri, Universitas Atma Jaya Yogyakarta.

Penulis menyadari bahwa dalam pembuatan tugas akhir ini telah melibatkan banyak pihak yang mendukung dan menolong dalam berbagai hal baik secara langsung maupun tidak langsung. Oleh karena itu, pada kesempatan ini penulis ingin mempersembahkan rasa terima kasih kepada:

1. Tuhan Yesus yang atas berkat-Nya, penulis bisa menyelesaikan tugas akhir yang penulis sempat ragu bisa selesai.
2. Bapak Y. Sigit Purnomo W.P., S.T., M.Kom., selaku dosen pembimbing yang mengadakan pertemuan skripsi mewah tiap minggunya, terima kasih atas topik skripsi yang sudah dipercayakan kepada penulis.
3. Bapak Kusworo Anindito, S.T., M.T., selaku dosen pembimbing yang sabar menghadapi penulis yang bimbingan hampir setiap hari.
4. Ferdinandus Satriyo Putro Pamungkas Supit, selaku pacar yang setia mendampingi penulis setiap hari saat mengerjakan tugas akhir.
5. Seluruh dosen dan staff Fakultas Teknologi Industri Universitas Atma Jaya Yogyakarta, atas bimbingan dan bantuannya selama ini.
6. Deta dan Vena, selaku teman seperjuangan dari semester 1, atas pendampingannya.

7. Tante Kokok dan Bukti, atas tumpangan tempat mengerjakan skripsi.
8. Teman-teman Kelompok 43, Pengos B, Egik, Vania, Widi, Rizky, Arik, yang menambah cerita hidup penulis selama terdampar di Samigaluh.
9. Putu dan Mak May, selaku teman yang berbagi cerita dan pengalaman sampai sekarang.
10. Siska, Santa, Retta, Priska, Desti yang selalu memberi semangat kepada penulis.
11. Teman-teman dan pihak lain yang tidak mungkin disebutkan satu per satu yang telah membantu selama pengerjaan skripsi.
12. Semua orang yang secara tidak langsung memberikan dukungan dan semangat.

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna karena keterbatasan waktu dan pengetahuan yang dimiliki penulis. Oleh karena itu segala kritik dan saran yang bersifat membangun sangat diharapkan. Akhir kata, semoga skripsi ini dapat berguna dan bermanfaat bagi semua pihak.

Yogyakarta, Mei 2016

Penulis

Daftar Isi

Halaman Pengesahan.....	ii
Pernyataan Originalitas.....	iii
Halaman Persembahan.....	v
Kata Pengantar.....	vi
Daftar Isi.....	viii
Daftar Tabel.....	x
Daftar Gambar.....	xii
Daftar Lampiran.....	xiii
Intisari.....	xiv
BAB 1 PENDAHULUAN.....	1
1.1 Latar belakang.....	1
1.2 Rumusan Masalah.....	3
1.3 Batasan Masalah.....	4
1.4 Tujuan Penelitian.....	4
1.5 Metodologi Penelitian.....	4
1.6 Sistematika Penulisan.....	6
BAB 2 TINJAUAN PUSTAKA.....	7
BAB 3 LANDASAN TEORI.....	11
3.1 <i>Text Mining</i>	11
3.2 <i>Automatic Text Summarization</i>	12
3.3 <i>K-Means</i>	13
3.4 <i>Metode Maximum Marginal Relevance (MMR)</i>	14
BAB 4 ANALISIS DAN PERANCANGAN PERANGKAT LUNAK.....	19
4.1 Analisis Latar Belakang Sistem.....	19
4.2 Analisis Sistem yang Akan Dibangun.....	20
4.3 Perancangan Sistem.....	24
BAB 5 IMPLEMENTASI DAN PENGUJIAN SISTEM.....	26
5.1 Definisi CLUSUM.....	26

5.2	Alur Pengelompokan Dokumen.....	27
5.3	Alur Peringkasan Dokumen Tunggal.....	44
5.4	Alur Peringkasan Multi Dokumen.....	54
5.5	<i>Web Service</i>	97
5.6	Implementasi Sistem.....	98
5.7	Perbandingan dengan Summary pada KUBERIN.....	106
5.8	Hasil Pengujian Perangkat Lunak.....	114
5.9	Perbandingan Kelebihan Kekurangan CLUSUM dengan KUBERIN.....	116
BAB 6	PENUTUP.....	117
6.1	Kesimpulan.....	117
6.2	Saran.....	117
	Daftar Pustaka.....	118

Daftar Tabel

Tabel 2.1	Tabel Perbandingan.....	10
Tabel 5.1	Dokumen 1.....	30
Tabel 5.2	Dokumen 2.....	31
Tabel 5.3	Filter Kata.....	33
Tabel 5.4	Penghitungan Frekuensi.....	34
Tabel 5.5	Penghitungan IDF.....	36
Tabel 5.6	Penghitungan Bobot.....	38
Tabel 5.7	Penghitungan Skalar Centroid dengan Dokumen.....	40
Tabel 5.8	Penghitungan Panjang Vektor.....	42
Tabel 5.9	Pemecahan Kalimat Dokumen Tunggal.....	47
Tabel 5.10	Filter Kata Dokumen.....	48
Tabel 5.11	Frekuensi Kata pada Dokumen.....	49
Tabel 5.12	Perhitungan Term Frequency.....	51
Tabel 5.13	Sentence Scoring Dokumen.....	53
Tabel 5.14	Pemecahan Kalimat pada Dokumen Gabungan.....	58
Tabel 5.15	Pencacahan Kata dalam Kalimat.....	59
Tabel 5.16	Penghitungan IDF.....	62
Tabel 5.17	Penghitungan Bobot.....	64
Tabel 5.18	Penghitungan Bobot Query dikali Bobot Kalimat.....	67
Tabel 5.19	Jumlah Bobot Tiap Kalimat.....	70
Tabel 5.20	Penghitungan Kuadrat Bobot.....	71
Tabel 5.21	Penjumlahan Kuadrat Bobot.....	74
Tabel 5.22	Akar Kuadrat Penjumlahan Bobot.....	74
Tabel 5.23	Cosine similarity Query dengan Kalimat.....	75
Tabel 5.24	Penghitungan Kata tiap Kalimat.....	76
Tabel 5.25	Penghitungan IDF.....	79
Tabel 5.26	Penghitungan Bobot tiap Kata.....	80
Tabel 5.27	Penghitungan Bobot Kalimat 1 dengan Bobot Lain.....	83
Tabel 5.28	Penjumlahan Perkalian Bobot Kalimat 1 dengan Kalimat lain.....	86
Tabel 5.29	Penghitungan Kuadrat Bobot.....	87
Tabel 5.30	Penghitungan Penjumlahan Kuadrat Bobot Kalimat.....	90
Tabel 5.31	Penghitungan Akar Kuadrat Penjumlahan Bobot Kalimat.....	91
Tabel 5.32	Cosine Similarity antar Kalimat 1 dengan Kalimat Lain.....	91
Tabel 5.33	Cosine Similarity antar Kalimat.....	92
Tabel 5.34	Hasil Penghitungan MMR.....	96

Tabel 5.35 Hasil Penyusunan Ringkasan Multidokumen.....	96
Tabel 5.36 Contoh Hasil Pengelompokan dan Peringkasan	98
Tabel 5.37 Tabel Perbandingan KUBERIN dengan CLUSUM..	106
Tabel 5.38 Hasil Pengujian Perangkat Lunak.....	114
Tabel 5.39 Perbandingan CLUSUM dengan KUBERIN.....	116



Daftar Gambar

Gambar 3.1 Text Mining Process.....	12
Gambar 3.2 Skema Peringkasan.....	15
Gambar 4.1 Arsitektur Perangkat Lunak CLUSUM.....	23
Gambar 4.2 Use Case Diagram.....	23
Gambar 4.3 Entity Relationship Diagram.....	24
Gambar 4.4 Perancangan Arsitektur Web CLUSUM.....	25
Gambar 5.1 Flow Chart K-Means.....	27
Gambar 5.2 Flowchart Sentence Scoring.....	45
Gambar 5.3 Flowchart Peringkasan Multi Dokumen.....	55
Gambar 5.4 Contoh JSON.....	97

Daftar Lampiran

- I. Spesifikasi Kebutuhan Perangkat Lunak (SKPL)
CLUSUM
- II. Deskripsi Perancangan Perangkat Lunak (DPPL)
CLUSUM



**Pembandingan Aplikasi Peringkasan Multi Dokumen
menggunakan Sentence Scoring dan Maximum Marginal
Relevance dengan K- Means**

Disusun oleh:

Valentina Berlian Provitasari

NIM: 12 07 06893

Intisari

Informasi dalam bentuk teks berita telah menjadi salah satu komoditas yang paling penting dalam era informasi ini. Perkembangan teknologi internet berdampak bertambahnya jumlah situs berita berbahasa Indonesia dan menciptakan ledakan informasi. Banyak berita-berita yang dihasilkan sebenarnya memiliki inti yang sama tapi narasi yang berbeda. Banyaknya informasi yang beredar tentu membuat kebutuhan akan informasi yang relevan semakin meningkat.

Guna mempermudah pengguna dalam mendapatkan informasi yang sesuai, penulis membangun aplikasi pengelompokan dokumen dan peringkasan multidokumen. Aplikasi ini dilengkapi dengan pengelompokan berita sehingga berita yang diringkas saling bersesuaian. Peringkasan berita mampu mengefisienkan pengguna dalam membaca berita.

Penulis membangun aplikasi pengelompokan dokumen dengan metode K-Means dan peringkasan dokumen menggunakan Maximum Marginal Relative. Aplikasi ini mampu memudahkan pengguna sehingga tidak perlu membaca topik yang sama dari berbagai macam sumber, tetapi cukup membaca dari 1 sumber saja.

Keyword: berita, pengelompokan, peringkasan, K-Means,
Maximum Marginal Relative

Pembimbing I : Y. Sigit Purnomo W.P., S.T., M.Kom.

Pembimbing II : Kusworo Anindito, S.T., M.T.

Jadwal Pendadaran : 10 Mei 2016