

**IMPLEMENTASI ALGORITMA *MAXIMAL MARGINAL
RELEVANCE* UNTUK MERINGKAS TEKS BERITA
BERBAHASA INDONESIA**

SKRIPSI



Diajukan kepada Program Studi Teknik Informatika Fakultas Teknologi Informasi
Universitas Kristen Duta Wacana

Sebagai Salah Satu Syarat dalam Memperoleh Gelar Sarjana Komputer

Disusun Oleh :

STEFANI DESSY RAHMAWATI

22104992

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS KRISTEN DUTA WACANA
YOGYAKARTA

2014

PERNYATAAN KEASLIAN SKRIPSI

Saya menyatakan dengan sesungguhnya bahwa skripsi dengan judul:

IMPLEMENTASI ALGORITMA MAXIMAL MARGINAL RELEVANCE UNTUK MERINGKAS TEKS BERITA BERBAHASA INDONESIA

yang saya kerjakan untuk melengkapi sebagian persyaratan menjadi Sarjana Komputer pada pendidikan Sarjana Program Studi Teknik Informatika Fakultas Teknologi Informasi Universitas Kristen Duta Wacana, bukan merupakan tiruan atau duplikasi dari skripsi kesarjanaan di lingkungan Universitas Kristen Duta Wacana maupun di Perguruan Tinggi atau instansi manapun, kecuali bagian yang sumber informasinya dicantumkan sebagaimana mestinya.

Jika dikemudian hari didapati bahwa hasil skripsi ini adalah hasil plagiat atau tiruan dari skripsi lain, saya bersedia dikenai sanksi, yakni pencabutan gelar kesarjanaan saya.

Surabaya, 25 September 2014

METERAI

AD97DACF325012B

3000

STEFANI DESSY RAHMAWATI

22104992

HALAMAN PERSETUJUAN

Judul Skripsi : IMPLEMENTASI ALGORITMA MAXIMAL
MARGINAL RELEVANCE UNTUK MERINGKAS
TEKS BERITA BERBAHASA INDONESIA

Nama Mahasiswa : STEFANI DESSY RAHMAWATI

N I M : 22104992

Matakuliah : Skripsi (Tugas Akhir)

Kode : TIW276

Semester : Gasal

Tahun Akademik : 2014/2015

Tesis diperiksa dan disetujui di
Yogyakarta,
Pada tanggal 25 September 2014

Dosen Pembimbing I


Budi Susanto, SKom., M.T.

Dosen Pembimbing II


Gloria Virginia, S.Kom., MAI, Ph.D.

HALAMAN PENGESAHAN

IMPLEMENTASI ALGORITMA MAXIMAL MARGINAL RELEVANCE UNTUK MERINGKAS TEKS BERITA BERBAHASA INDONESIA

Oleh: STEFANI DESSY RAHMAWATI / 22104992

Dipertahankan di depan Dewan Penguji Skripsi
Program Studi Teknik Informatika Fakultas Teknologi Informasi
Universitas Kristen Duta Wacana - Yogyakarta
Dan dinyatakan diterima untuk memenuhi salah satu syarat memperoleh gelar
Sarjana Komputer
pada tanggal 25 September 2014

Yogyakarta, 25 September 2014
Mengesahkan,

Dewan Penguji:

1. Budi Susanto, SKom., M.T.
2. Gloria Virginia, S.Kom., M.M. Ph.D.
3. Aditya Wikan Mahatama, S.Kom.
4. Lucia Duta Krisnawati, M.T.



Dekan

(Drs. Wimmie Handiwidjono, M.T.)

Ketua Program Studi

(Nugroho Agus Haryono, M.Si)

UCAPAN TERIMAKASIH

Segala puji syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa yang telah melimpahkan rahmat dan anugerahNya, sehingga penulis dapat menyelesaikan Tugas Akhir yang berjudul *Implementasi Algoritma Maximal Marginal Relevance untuk Meringkas Teks Berita berbahasa Indonesia* ini dapat diselesaikan dengan baik.

Penulis mengucapkan terimakasih yang sebesar-besarnya kepada semua pihak yang telah banyak membantu dan memberikan dukungan kepada penulis selama penyusunan Tugas Akhir ini, diantaranya :

1. Bapak Budi Susanto, S.Kom.,M.T. selaku dosen pembimbing I dan Ibu Gloria Virginia, S.Kom., MAI., Ph.D selaku dosen pembimbing II yang telah memberikan ide, masukan, kritik, saran dan nasehat dalam pembuatan tugas akhir dan penulisan laporan.
2. Bapak Marsudi, Ibu Astuti Yuli Setyani, Sandra Galuh, Valentino Guntur sebagai keluarga yang selalu memberikan kasih sayang, semangat dan doa kepada penulis.
3. Devry Pramesti, Ika Nur Kurnia, Dewi Ratna, Khudzaifah selaku sahabat yang selalu menjadi teman diskusi, bercerita dan berbagi.
4. Mas Christian Puji Nugraha yang selalu memberikan pengertian, saran, dukungan, pengorangan, nasehat dan kasih sayang kepada penulis.
5. Aan Setawan, Dharip Pramono, Gusti Putu Fajar, Roby Chandra, Rico Revianto, Esmannia Preticia, Yohanes Dedeo, Cornelia Aventina, Norbertus Yunendra, Deasy Nathalia, semua teman-teman Kelas Gokil 2010 selaku sahabat yang selalu siap memberikan semangat, dukungan, bantuan dan penghiburan.
6. Keluarga Besar Biro 3 UKDW yang telah memberikan banyak semangat.
7. Pihak-pihak lain yang telah mendukung baik secara langsung maupun tidak langsung yang tidak dapat penulis tuliskan satu-persatu.

Yogyakarta, September 2014

Penulis

INTISARI

Berita merupakan informasi yang dibutuhkan oleh masyarakat. Semakin berkembangnya teknologi, jumlah berita yang dapat diakses melalui internet berkembang pesat. Banyaknya berita mengakibatkan pembaca membutuhkan waktu yang cukup untuk membaca. Maka dari itu dapat dikembangkan peringkasan teks otomatis berita yang membantu pembaca untuk mempersingkat waktu membaca. Algoritma *Maximum Marginal Relevance (MMR)* merupakan salah satu metode *extractive summary* yang dapat digunakan untuk meringkas dokumen.

Tahap menghasilkan ringkasan adalah *Text preprocessing*, penghitungan *TF-ISF*, *Cosine Similarity* dan penghitungan *MMR*. *Text Preprocessing* terdiri dari pembentukan kalimat, *case folding*, tokenisasi, *filtering* dan *stemming*. Data yang diuji berupa 40 berita dengan topik ekonomi. Pengujian sistem dilakukan menggunakan proses *stemming* dan tanpa proses *stemming*.

Hasil ringkasan sistem menggunakan proses *stemming* dibandingkan dengan ringkasan manual menghasilkan rata-rata *Precision* 56,13%, *Recall* 66,36%, dan *F-measure* 59,56%. Sedangkan untuk sistem tanpa menggunakan proses *stemming* menghasilkan rata-rata *Precision* 56,10%, *Recall* 64,83%, *F-measure* 58,81%. Berdasarkan hasil tersebut, dapat diindikasikan bahwa *MMR* dapat membantu dalam menghasilkan ringkasan sebuah teks. Selain itu proses *stemming* dapat membantu meningkatkan nilai *F-measure*.

Kata kunci : Peringkasan Teks, *Maximal Marginal Relevance*, *Stemming*, *F-measure*.

DAFTAR ISI

HALAMAN JUDUL	i
PERNYATAAN KEASLIAN SKRIPSI	ii
HALAMAN PERSETUJUAN	iii
HALAMAN PENGESAHAN	Error! Bookmark not defined.
UCAPAN TERIMA KASIH	iv
INTISARI.....	vi
DAFTAR ISI.....	vii
DAFTAR TABEL	ix
DAFTAR GAMBAR.....	xi
DAFTAR LAMPIRAN	xiv
BAB I PENDAHULUAN	1
1.1.Latar Belakang Masalah	1
1.2.Perumusan Masalah.....	2
1.3.Batasan Masalah	2
1.4.Tujuan Penelitian	3
1.5.Metode Penelitian	3
1.6 Sistem dan Penulisan	4
BAB II LANDASAN TEORI	6
2.1. Tinjauan Pustaka	6
2.2. Landasan Teori	7
2.2.1.Konsep Ringkasan	7
2.2.2.Berita	9
2.2.3.Text Mining	10
2.2.4.Text Preprocessing	10
2.2.5.TF- ISF	15
2.2.6.Metode Cosine Similarity	19
2.2.7.Algoritma MMR (Maximal Marginal Relevance)	21
BAB III ANALISIS DAN PERANCANGAN SISTEM	25

3.1. Spesifikasi Kebutuhan	25
3.1.1. Kebutuhan Fungsional	25
3.1.2. Usecase Diagram	26
3.1.3. Spesifikasi Perangkat	27
3.2. Blok Diagram Sistem	28
3.3. Rancangan Proses	29
3.4. Rancangan <i>Database</i>	40
3.4.1. Kamus Data	40
3.4.2. Rancangan Model Relasi	41
3.5. Rancangan Antar Muka	42
3.5.1. Struktur Web	42
3.5.2. Rancangan Halaman	43
3.6. Rancangan Pengujian Sistem	51
BAB IV IMPLEMENTASI DAN ANALISIS SISTEM	54
4.1. Implementasi Sistem	54
4.1.1. Blok Diagram Sistem	54
4.1.2. Implementasi Antarmuka Sistem	55
4.2. Evaluasi dan Analisis Sistem	80
4.2.1. Evaluasi Sistem	80
4.2.2. Analisis Sistem	105
BAB V KESIMPULAN DAN SARAN	114
5.1. Kesimpulan	114
5.2. Saran	115
DAFTAR PUSTAKA	Error! Bookmark not defined.
LAMPIRAN	119

DAFTAR TABEL

Nomor	Nama Tabel	Halaman
2.1	Daftar Partikel Akhiran Infeksional	12
2.2	Daftar Partikel Kata Ganti	12
2.3	Daftar Awalan Pertama	13
2.4	Daftar Awalan Kedua	14
2.5	Daftar Partikel Akhiran Derivasional	14
2.6	Bobot <i>TF-ISF</i>	18
2.7	Matriks Kemiripan antara Kalimat dengan <i>Query</i>	20
2.8	Matriks Kemiripan antara Kalimat dengan Kalimat	21
2.9	Matriks Hasil Penghitungan MMR	24
2.10	Hasil Bobot MMR Maksimum dari iterasi MMR	24
3.1	Kamus Data	40
3.2	Rancangan Tabel Hasil Ringkasan Ideal (<i>Evaluator</i>)	51
3.3	Rancangan Tabel Hasil Ringkasan Sistem	52
3.4	Tabel <i>Precision</i> dan <i>Recall</i>	52
3.5	Rancangan Tabel Evaluasi	53
4.1	Daftar Jumlah Kalimat Hasil Ringkasan	81
4.2	Hasil Evaluasi Sistem Tanpa Proses Stemming dengan Ringkasan Manusia	87
4.3	Hasil Ringkasan Berita Nomor 24	90
4.4	Hasil Ringkasan Berita Nomor 5	91
4.5	Hasil Ringkasan Berita Nomor 8	93
4.6	Hasil Ringkasan Berita Nomor 9	94
4.7	Hasil Ringkasan Berita Nomor 11	95
4.8	Hasil Ringkasan Berita Nomor 20	97
4.9	Hasil Ringkasan Berita Nomor 32	98

Nomor	Nama Tabel	Halaman
4.10	Hasil Ringkasan Berita Nomor 35	100
4.11	Hasil Ringkasan Berita Nomor 29	101
4.12	Hasil Ringkasan Berita Nomor 37	103
4.13	Perbedaan Nilai <i>Precision</i> , <i>Recall</i> dan <i>F-measure</i>	107
4.14	Evaluasi Jumlah Ringkasan Dibanding Jumlah Kalimat Berita	107
4.15	Data Jumlah Berita Berdasarkan Nilai Precision, Recall dan F-measure	109

@UKDW

DAFTAR GAMBAR

Nomor	Nama Gambar	Halaman
2.1	Gambar Tabel Kontingensi Proses <i>Precision</i> dan <i>Recall</i>	9
3.1	<i>Usecase Diagram</i>	26
3.2	Blok Diagram Sistem	29
3.3	Alur Kerja Sistem	30
3.4	Flowchart Text Preprocessing	31
3.5	<i>Flowchart</i> Pembentukan Kalimat	31
3.6	<i>Flowchart</i> proses <i>case folding</i>	33
3.7	<i>Flowchart</i> Proses Tokenisasi	34
3.8	Flowchart Filtering	35
3.9	<i>Flowchart</i> Proses <i>stemming</i>	35
3.10	<i>Flowchart</i> Pembobotan <i>TF-ISF</i>	36
3.11	<i>Flowchart</i> Normalisasi <i>TF-ISF</i>	36
3.12	<i>Flowchart</i> Cosine Similiarity	37
3.13	<i>Flowchart</i> MMR	38
3.14	Rancangan Relasi Tabel	41
3.15	Struktur Web	42
3.16	Rancangan Halaman upload file	43
3.17	Rancangan Halaman <i>input file</i>	44
3.18	Rancangan Halaman Hasil Ringkasan	45
3.19	Rancangan Halaman <i>Text Preprocessing</i>	45
3.20	Rancangan Halaman <i>TF-ISF</i>	46

3.21	Rancangan Halaman Penghitungan Cosine Similiarity	47
3.22	Rancangan Halaman Penghitungan <i>MMR</i>	48
3.23	Rancangan Halaman <i>Kamusku</i>	49
3.24	Rancangan Halaman Tentang	50
4.1	Blok Diagram Sistem	54
4.2	Halaman Awal Sistem	55
4.3	<i>Pseudo-code Text Preprocessing</i>	56
4.4	Pseudo-code Stemming	59
4.5	Tampilan Daftar Kalimat hasil <i>Text Preprocessing</i>	60
4.6	Tampilan Daftar Token hasil Teks Preprocessing	60
4.7	<i>Pseudo-code Menghitung SF dan ISF</i>	62
4.8	<i>Pseudo-code Menghitung TF</i>	62
4.9	<i>Pseudo-code Menghitung TF-ISF</i>	63
4.10	<i>Pseudo-code Menghitung Normalisasi TF-ISF</i>	63
4.11	Tampilan Halaman Nilai TF-ISF	65
4.12	Pseudo-code Menghitung Cosine Similarity	65
4.13	Tampilan Halaman <i>Cosine Similarity</i> Kalimat dengan Query	67
4.14	Tampilan Halaman <i>Cosine Similarity</i> Kalimat dengan Kalimat	68
4.15	<i>Pseudo-code Menghitung MMR</i>	69
4.16	Tampilan Halaman <i>MMR</i>	73
4.17	Tampilan Halaman Hasil Ringkasan	74
4.18	Halaman Evaluasi	75
4.19	<i>Pseudo-code Hapus Partikel</i>	75
4.20	<i>Pseudo-code Hapus Kata Ganti Kepemilikan</i>	76

4.21	<i>Pseudo-code</i> Hapus Awalan	77
4.22	<i>Pseudo-code</i> Hapus Awalan Kedua	79
4.23	<i>Pseudo-code</i> Hapus Akhiran.	79
4.24	Grafik Evaluasi Precision, Recall dan F-measure Sistem Menggunakan Proses Stemming	112
4.25	Grafik Evaluasi Precision, Recall dan F-measure Sistem Tanpa Menggunakan Proses Stemming	113

@UKDW

DAFTAR LAMPIRAN

Daftar Lampiran	Keterangan	Halaman
LA.1	Hasil Iterasi <i>MMR</i> Menggunakan Proses <i>Stemming</i>	Lampiran A-1
LA.2	Hasil Iterasi <i>MMR</i> Tanpa Menggunakan Proses <i>Stemming</i>	Lampiran A-2
LB.1	Rekap Kalimat Ringkasan Ideal	Lampiran B-1
LB.2	Rekap Kalimat Ringkasan Hasil Pengujian Sistem	Lampiran B-2

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Kebutuhan masyarakat akan informasi sangatlah besar, terutama informasi dalam bentuk berita, jurnal dan artikel. Bahkan terkadang orang mampu untuk membaca berita dengan jumlah paragraf yang banyak demi mendapat sebuah informasi. Dengan berkembangnya teknologi, jumlah berita yang dapat diakses melalui internet secara umum meningkat pesat. Hal ini didukung dengan menjamurnya situs berita misalnya Kompas.com, Detik.com, vivanews.com dan yang lainnya. Berita yang ada pada situs internet pada umumnya bersifat *real time* dan *up to date* dimana berita tersebut selalu diperbaharui setiap saat. Hal tersebut mengakibatkan pembaca harus meluangkan waktu yang cukup untuk membaca berita.

Melihat dari permasalahan tersebut, dapat dikembangkan sebuah sistem peringkas teks otomatis yang mampu menghasilkan sebuah ringkasan teks sebuah berita. Sistem peringkas teks berita dapat membantu mengurangi waktu dalam membaca dan memberikan sebuah ringkasan (*summary*) berupa informasi yang lebih ringkas. Untuk membuat sistem peringkas teks otomatis dapat dilakukan dengan menghitung bobot TF-IDF (*Term Weight - Inverse Document Frequency*) dimana metode ini mampu menghitung bobot kalimat dan menghasilkan sebuah ringkasan. Namun dalam metode ini tidak ada solusi untuk mengatasi resiko adanya redundansi (kemiripan) kalimat yang dihasilkan dalam ringkasan.

Algoritma *Maximum Marginal Relevance (MMR)* merupakan salah satu metode *extractive summary* yang dapat digunakan untuk meringkas dokumen tunggal maupun multidokumen. Metode MMR bermanfaat untuk mengurangi

reduksi kalimat yang akan diringkas. Hasil ringkasan didapatkan dari kombinasi dua matrix *cosine similarity* yaitu antara kalimat dengan *query* dan antara kalimat dengan kalimat. Kalimat yang memiliki bobot MMR tinggi akan ditampilkan sebagai ringkasan.

1.2. Perumusan Masalah

Berdasarkan latar belakang masalah diatas, rumusan masalah dalam penelitian ini adalah mengukur seberapa besar akurasi yang dihasilkan system menggunakan algoritma MMR dalam menghasilkan ringkasan berita yang dibangun dalam sistem peringkas berita berbahasa Indonesia.

1.3. Batasan Masalah

Adapun batasan sistem yang akan dibuat dalam penelitian ini adalah sebagai berikut :

- a. Dokumen yang digunakan adalah berita dengan topik ekonomi yang diambil dari website Kompas antara bulan Februari 2014 hingga Maret 2014 (<http://bisniskeuangan.kompas.com/>) yang dipilih secara acak.
- b. Bentuk masukan ke sistem berupa teks dengan format *.txt* dan memiliki ukuran maksimal 3 KB.
- c. Tidak ada perbaikan untuk adanya kesalahan penulisan kata.
- d. Dokumen berita yang diringkas hanya 1 dokumen atau dokumen tunggal.
- e. Dokumen yang akan diringkas harus memiliki judul.
- f. Peringkas teks menggunakan metode ekstraktif sehingga hasil keluaran sistem berupa teks yang diambil dari teks utama tanpa mengubah isi kalimat.

1.4. Tujuan Penelitian

Melalui penelitian ini, maka tujuan yang ingin dicapai oleh penulis adalah :

- a. Menghasilkan sistem yang dapat melakukan peringkasan teks berita ekonomi berbahasa Indonesia.
- b. Mengetahui seberapa efektif algoritma MMR dalam meringkas teks berita berbahasa Indonesia.

Penelitian ini akan memberikan kontribusi yaitu membantu pengguna dalam mendapatkan informasi berita dan mempersingkat waktu pengguna dalam membaca berita.

1.5. Metode Penelitian

Metode yang digunakan dalam penelitian ini adalah :

- a. Studi Pustaka

Studi pustaka dilakukan dengan mempelajari teori-teori melalui buku, artikel, jurnal dan bahan lain yang berkaitan dengan metode tahap yang digunakan untuk meringkas teks dan algoritma *Maximal Marginal Relevance*.

- b. Pengumpulan Data

Pengumpulan data dilakukan secara manual. Dimana data yang diambil merupakan berita dengan sumber *www.kompas.com*. Data dipilih secara acak, disalin, kemudian disimpan dalam format file .txt. Perancangan sistem

Tahap ini berisi perancangan basis data dan perancangan antarmuka untuk sistem peringkasan teks berita yang akan dibangun.

- c. Pembangunan Sistem

Tahap ini merupakan tahap pembuatan program sesuai dengan rancangan yang telah dibuat. Pertama yaitu pembuatan *database* yang akan digunakan untuk menyimpan kamus kata dan *stopword list* yang telah diunduh sebelumnya. Kedua pembuatan program yang mampu meng *upload* berkas yang akan diringkas dan menghasilkan sebuah ringkasan. Dalam menghasilkan ringkasan, terdapat proses yaitu ⁽¹⁾*teks preprocessing* (berupa tokenisasi, *case folding*, *stopword filtering*, dan *stemming*), ⁽²⁾*pembobotan TF – ISF*, ⁽³⁾*penghitungan cosinesimilarity*, dan ⁽⁴⁾*pembobotan MMR*

d. Pengujian dan Analisis Sistem

Pengujian terhadap program dilakukan dengan mengevaluasi hasil ringkasan yang dihasilkan sistem. Kemudian akan dilakukan analisis hasil ringkasan dari sistem dan mengukur keefektifan dari program yang telah dibangun.

1.6 Sistematika Penulisan

Sistematika penulisan laporan Tugas Akhir ini dikelompokkan menjadi 5 bab, yaitu :

Bab I, Pendahuluan. Bab ini berisi gambaran umum mengenai penelitian yang akan dilakukan yang mencakup antara lain latar belakang, perumusan masalah, batasan masalah, tujuan penelitian, metode penelitian, dan sistematika penulisan penelitian.

Bab II, Landasan teori. Bab ini berisi mengenai tinjauan pustaka dan landasan teori yang digunakan dalam penelitian ini. Tinjauan pustaka menguraikan berbagai teori yang didapat dari berbagai sumber terkait dengan penelitian ini.

Bab III, Perancangan Sistem. Bab ini berisi mengenai rancangan sistem yang akan dibangun dalam penelitian ini. Rancangan sistem yang akan dibuat

berupa spesifikasi dari sistem, rancangan diagram sistem, rancangan antarmuka sistem berupa input dan output.

Bab IV, Implementasi Sistem. Bab ini berisi tentang implementasi sistem yang telah dirancang sebagaimana sudah dirancang ada bab III.

Bab V, Kesimpulan dan Saran. Bab ini berisi tentang kesimpulan hasil penelitian yang telah dilakukan dan saran yang mungkin dapat dilakukan untuk pengembangan penelitian selanjutnya.

@UKDW

BAB V

KESIMPULAN DAN SARAN

5.1. Kesimpulan

Dari proses perancangan hingga implementasi sistem peringkasan berita berbahasa Indonesia yang telah dibangun, diperoleh kesimpulan sebagai berikut :

- Penggunaan algoritma *MMR* dapat membantu dalam menghasilkan ringkasan sebuah teks. Hasil ringkasan yang dihasilkan mampu mewakili *query*.
- Di dalam penelitian ini, efektifitas algoritma *MMR* dihitung berdasarkan performa sistem. Alat bantu yang digunakan adalah *Precision*, *Recall* dan *F-measure*. Berdasarkan hasil evaluasi terhadap sistem, nilai rata-rata *Precision*, *Recall* dan *F-measure* untuk pengujian menggunakan proses *stemming* adalah 56,13%, 66,36%, dan 59,76%. Sedangkan nilai rata-rata *Precision*, *Recall* dan *F-measure* untuk pengujian sistem tanpa menggunakan proses *stemming* adalah 56,10%, 64,83%, dan 58,81%.
- Berdasarkan perbandingan nilai rata-rata *Precision*, *Recall* dan *F-measure* di atas, dapat dilihat bahwa penggunaan proses *stemming* dapat membantu meningkatkan nilai *Precision*, *Recall* dan *F-measure* walaupun tidak terlalu signifikan.
- Ringkasan yang berukuran setengah atau kurang dari setengah jumlah kalimat teks asli mampu mewakili *query* dan cukup efektif dijadikan sebagai ringkasan.
- Penelitian ini juga memberikan kontribusi berupa data yang dapat digunakan untuk penelitian selanjutnya. Data tersebut adalah :

- 1) Dokumen penelitian berupa berita dengan topik ekonomi yang berjumlah 40 berita. Berita diambil dari website bisniskeuangan.kompas.com yang di-posting pada bulan Februari hingga Maret 2014.
- 2) *Relevant document* yang dapat digunakan dalam proses evaluasi. Dokumen dihasilkan dengan mengikutsertakan *evaluator* sebanyak 10 orang mahasiswa prodi teknik informatika.
- 3) Modul *stemming* menggunakan algoritma Porter *Stemming* untuk Bahasa Indonesia.

5.2. Saran

Saran untuk pengembangan dan perbaikan sistem ini adalah :

- Hasil ringkasan yang dihasilkan oleh sistem dalam penelitian ini berupa kalimat berdasarkan penghitungan dan kemampuan *query* saja. Untuk penelitian berikutnya, diharapkan ringkasan yang dihasilkan memiliki sistematika dan alur yang lebih baik terutama memperhatikan penggunaan tanda baca untuk kalimat langsung. Hasil kalimat yang dihasilkan mungkin dapat diurutkan sesuai dengan nomor kalimatnya agar ringkasan yang dihasilkan dapat memiliki sistematika yang lebih baik.
- Algoritma *stemming* yang digunakan pada penelitian ini masih belum mencakup semua aturan untuk *stemming* Bahasa Indonesia. Misalnya kata “mengusik” seharusnya memiliki katadasar “usik” dan kata “mengupas” menjadi “kupas” namun pada penelitian ini penggunaan awalan “meng-” yang apabila diikuti huruf “u” dan “e” akan diubah menjadi huruf “k” sehingga kata “mengusik” menjadi “kusik”.

Jadi dapat ditambahkan aturan-aturan lain agar hasil *stemming* menjadi lebih baik misalnya dapat dikembangkan adanya pencocokan terhadap daftar katadasar untuk setiap prosesnya. Pengembangan juga dapat dilakukan dengan penggunaan algoritma *stemming* yang lain yang dapat meningkatkan kualitas hasil *stemming*.

- Pada penelitian ini, *evaluator* yang dilibatkan adalah mahasiswa program studi teknik informatika. Muncul indikasi bahwa *evaluator* kurang teliti dalam menghasilkan ringkasan ideal. Untuk penelitian selanjutnya, dapat dilibatkan orang yang memiliki latar belakang pendidikan Bahasa Indonesia.

@UKDW

Daftar Pustaka

- Andriani, M., Asian, J., Nazief, B., Taghoghi, S., & Williams, H. (2007). Stemming Indonesian: A Confix-stripping approach. *ACM Transactions on Asian Language Information Processing*, Vol. 6, No. 4. Article 13, p. 1-33.
- Assegaf, D. (1991). *Jurnalistik Masa Kini*. Jakarta: Ghalia Indonesia.
- Feldman, R., & Sanger, J. (2007). *The Text Mining Handbook*. New York: Cambridge University Press.
- Fu'adi, M., Abidin, Z., & Nurhayati, H. (2014, Maret 12). *Pembuatan Ringkasan Otomatis Dokumen Berita Berbahasa Indonesia Menggunakan Algoritma Dijkstra*. Dipetik September 10, 2014 dari <http://saintek.uin-malang.ac.id/pembuatan-ringkasan-otomatis-dokumen-berita-berbahasa-indonesia-menggunakan-algoritma-dijkstra/>
- Goldstein, J., & Carbonel, J. (1998). Summarization: (1) Using MMR for Diversity - Based Reranking and (2) Evaluating Summaries. *TIPSTER Proceedings of a workshop*, (pp. 181-195). Baltimore, Maryland.
- Grossman, D., & Ophir, F. (1998). *Information Retrieval :Algorithm and Heuristics*. Kluwer Academic Publisher.
- Hidayah, O., Maharani, W., & Leksitawening, A. (2013). *Opinion Summarization Review Produk dalam Bahasa Indonesia dengan Metode Maximum Marginal Relevance (MMR)*. Dipetik September 09, 2014, dari cdndata.telkom.ac.id/tapa/jurnal/113090272.pdf
- Hovy, E., & Lin, C. (1998). Automated text summarization in summarist. *Proceeding TIPSTER Proceedings of a workshop*, (pp. 197-214). Baltimore, Maryland.
- Intan, R., & Defeng, A. (2006). HARD: Subject-based Search Engine menggunakan TF-IDF dan Jaccard's Coefficient. *Jurnal Teknik Industri*, Vol. 8 (1), p. 61-72.
- Mani, I., & Maybury, T. (. (1999). *Advance in Automatic Text Summarization*. Cambridge: The MIT Press.
- Manning, C. D., Raghavan, P., & Schutz, H. (2009). *Introduction to Information Retrieval*. England: Cambridge University Press.
- Mustaqfiri, M., Abidin, Z., & Kusumawati, R. (2011). Peringkasan Teks Otomatis Berita Berbahasa Indonesia Menggunakan Metode Maximum Marginal Relevance. *MATICS*, Vol. 4, No. 9.

- Smith, A. (1776). *An Inquiry into the Nature and Cause of the Wealth of Nation*.
- Sparck, K., & Galliers, J. (1996). *Evaluating Natural Language Processing System*. Cambridge: Springer.
- Tala, Z. (2003, Februari). *A Study Of Stemming Effects On Information Retrieval In Bahasa Indonesia*. Dipetik Maret 03, 2014, dari illc: www.illc.uva.nl/Research/Publications/Reports/MoL-2003-02.text.pdf
- Vega, V. (2001). *Information Retrieval for The Indonesian Language. Master Thesis National Universty of Singapore (Unpublished)* .
- Xie, S., & Liu, Y. (2010). *Using Corpus and Knowledge-based Similiarity Measure in Maximum Marginal Relevance for Meeting Summarization*. Dipetik Maret 24, 2014, dari <http://www.icsi.berkeley.edu/pubs/speech/usingcorpus08.pdf>

@UKDWN