



**IMPLEMENTASI METODE *MAXIMUM MARGINAL  
RELEVANCE* PADA PERINGKASAN TEKS  
OTOMATIS ARTIKEL BERITA**

SKRIPSI

diajukan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana  
Pendidikan Program Studi Pendidikan Teknik Informatika dan Komputer

Oleh

Winda Yulita NIM.5302411139

**JURUSAN TEKNIK ELEKTRO  
FAKULTAS TEKNIK  
UNIVERSITAS NEGERI SEMARANG  
2015**

## PERSETUJUAN PEMBIMBING

Nama : Winda Yulita  
NIM : 5302411139  
Program Studi : S-1 Pendidikan Teknik Informatika dan Komputer  
Judul Skripsi : Implementasi Metode *Maximum Marginal Relevance* pada Peringkasan Teks Otomatis Artikel Beria

Skripsi ini telah disetujui oleh pembimbing untuk diajukan ke sidang Panitia Ujian Skripsi Program Studi S-1 Pendidikan Teknik Informatika dan Komputer FT. UNNES.

Semarang, Juni 2015

Pembimbing



Feddy Setio Pribadi S.Pd., M.T.

NIP.197808222003121002

## PENGESAHAN

Skripsi dengan judul “Implementasi Metode *Maximum Marginal Relevance* pada Peringkasan Teks Otomatis Artikel Berita” telah dipertahankan di depan sidang Panitia Ujian Skripsi Fakultas Teknik UNNES pada tanggal 25 bulan Juni tahun 2015.

Oleh

Nama : Winda Yulita  
NIM : 5302411139  
Program Studi : S-1 Pendidikan Teknik Informatika dan Komputer

Panitia :

Ketua Panitia

Drs. Suryono, MT

NIP.195503161985031001

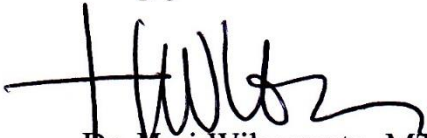
Sekretaris



Feddy Setio Pribadi, S.Pd., MT

NIP.196803121989031001

Penguji I



Dr. Mari Wibawanto, MT

NIP. 196501071991021001

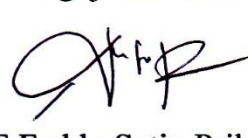
Penguji II



Aryo Baskoto Utomo, ST,MT

NIP.198409092012121002

Penguji III/Pembimbing



Feddy Setio Pribadi S.Pd.,MT

NIP.197808222003121002

Mengetahui :

Dekan Fakultas Teknik UNNES



Drs. Muhammad Harlanu, M.Pd.

NIP.196602151991021001

## **PERNYATAAN KEASLIAN**

Dengan ini saya menyatakan bahwa :

1. Skripsi ini adalah asli dan belum pernah diajukan untuk mendapatkan gelar akademik (sarjana, magister, dan/atau doktor), baik di Universitas Negeri Semarang (UNNES) maupun di perguruan tinggi lain.
2. Karya tulis ini adalah murni gagasan, rumusan, dan penelitian saya sendiri, tanpa bantuan pihak lain, kecuali arahan Pembimbing dan masukan Tim Penguji.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat yang telah ditulis atau dipublikasikan orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan dicantumkan dalam daftar pustaka.
4. Pernyataan ini saya buat dengan sesungguhnya dan apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka saya bersedia menerima sanksi akademik berupa pencabutan gelar yang telah diperoleh karena karya ini, serta sanksi lainnya sesuai dengan norma yang berlaku di perguruan tinggi ini.

Semarang, Juni 2015

yang membuat pernyataan,

Winda Yulita

NIM. 5302411139

## MOTTO DAN PERSEMBAHAN

### MOTTO

*Setiap tangisan akan berujung dengan senyuman, ketakutan akan berakhir dengan rasa aman, dan kegelisahan akan sirna oleh kedamaian (La Tahzan)*

### PERSEMBAHAN

*Skripsi ini saya persembahkan kepada :*

- ✚ Bapak dan Ibu yang selalu mendoakan dan menyayangi saya*
- ✚ Abangku, David dan adikku Dicky yang aku sayangi*
- ✚ Abah Yai Masyrohan beserta keluarga besar PPDAW*
- ✚ Kak Siddiq Supriyanto yang memberikan saran*
- ✚ Teman-teman Al-Kholik dan eks-AlKholik*

## ABSTRAK

**Yulita, Winda. 2015.** *Implementasi Metode Maximum Marginal Relevance pada Peringkasan Teks Otomatis Artikel Berita.* Dibimbing oleh Feddy Setio Pribadi, S.Pd., M.T. Program studi Pendidikan Teknik Informatika dan Komputer.

Sejak tahun 2000, di Indonesia media *online* berkembang dengan pesat, terlihat dari jumlah portal berita *online* yang meningkat. Berdasarkan konten berita, portal berita *online* ada yang berskala nasional dan lokal. Struktur kalimat pada portal berskala nasional lebih dominan menggunakan kalimat yang panjang, sedangkan berita pada portal berskala lokal lebih banyak menggunakan kalimat yang lebih ringkas. Kalimat yang panjang memiliki informasi yang lebih lengkap jika dibandingkan dengan kalimat yang ringkas. Meningkatnya jumlah portal berita *online* menimbulkan tingginya jumlah berita yang disuguhkan kepada masyarakat, sehingga kebutuhan peringkasan teks semakin diperlukan untuk kemudahan dan kehematan waktu manusia mengetahui inti berita. Oleh karena itu dibuatlah sistem peringkasan teks otomatis dengan metode *Maximum Marginal Relevance* dan pembobotan kata dengan algoritma TF-IDF-DF untuk melihat keefektifan dalam menghasilkan ringkasan. Dokumen uji coba diambil dari antaranews.com untuk portal *online* berskala nasional, sedangkan untuk portal *online* berskala lokal dari tribunjateng.com dan radarmagelang.com. Proses peringkasan melalui tahap *text preprocessing*, yang terdiri dari segmentasi kalimat, *case folding*, *tokenizing filtering* dan *stemming*. Tahap selanjutnya perhitungan TF-IDF-DF untuk menentukan bobot kata dan menentukan ringkasan dengan MMR. Hasil pengujian dibandingkan dengan ringkasan manual yang menghasil nilai *precision* 76,39 %, *recall* 65,28 %, dan *f-measure* 70,4 % untuk dokumen berita dari portal *online* berskala nasional, sedangkan peringkasan artikel berita dari portal berita *online* lokal menghasilkan nilai *precision* 45,83 %, *recall* 45,83 %, dan *f-measure* 45,83%.

**Kata kunci** : peringkasan teks, artikel berita, portal *online* berskala nasional dan lokal, *Maximum Marginal Relevance*, TF-IDF-DF.

## KATA PENGANTAR

Segala puji dan syukur penulis ucapkan kehadirat Allah SWT dan mengharapkan ridho yang telah melimpahkan rahmat-Nya, sehingga penulis dapat menyelesaikan skripsi yang berjudul Implementasi Metode *Maximum Marginal Relevance* pada Peringkasan Teks Otomatis Artikel Berita. Skripsi ini disusun sebagai salah satu persyaratan meraih gelar Sarjana Pendidikan pada Program Studi S-1 Pendidikan Teknik Informatika dan Komputer Universitas Negeri Semarang. Shalawat dan salam disampaikan kepada junjungan alam Nabi Muhammad SAW, mudah-mudahan kita semua mendapatkan safaat Nya di yaumul akhir nanti, Amin.

Penyelesaian karya tulis ini tidak lepas dari bantuan berbagai pihak, oleh karena itu pada kesempatan ini penulis menyampaikan ucapan terima kasih kepada:

1. Prof. Dr. Fathur Rokhman, M.Hum, Rektor Universitas Negeri Semarang yang memberikan kesempatan untuk menempuh studi di Universitas Negeri Semarang.
2. Drs. M. Harlanu, M.Pd., Dekan Fakultas Teknik, Drs. Suryono, M.T., Ketua Jurusan Teknik Elektro, dan Feddy Setio Pribadi, S.Pd., M.T., Ketua Program Studi Pendidikan Teknik Informatika dan Komputer, yang memberikan kesempatan dan izin dalam penyusunan skripsi ini.
3. Feddy Setio Pribadi, S.Pd., M.T., selaku dosen pembimbing yang telah memberikan masukan dan arahan sehingga skripsi ini dapat diselesaikan dengan baik.
4. Dr. Hari Wibawanto, MT dan Aryo Baskoro, ST, MT, sebagai Penguji I dan Penguji II yang telah memberikan pertanyaan, saran dan perbaikan sehingga menambah bobot dan kualitas skripsi ini.
5. Para dosen Pendidikan Teknik Informatika dan Komputer FT. UNNES yang telah memberikan bekal dan pengetahuan yang berharga.
6. Orang tua dan keluarga saya yang telah memberikan dukungan dan doa.
7. Semua pihak yang telah membantu dalam penelitian ini.

Penulis berharap semoga skripsi ini dapat bermanfaat untuk perkembangan metode peringkasan otomatis.

Semarang, Juni 2015

Penulis



## DAFTAR ISI

HALAMAN JUDUL.....	i
PERSETUJUAN PEMBIMBING.....	ii
PERNYATAAN KEASLIAN.....	iii
PENGESAHAN .....	iv
MOTTO DAN PERSEMBAHAN .....	v
ABSTRAK .....	vi
KATA PENGANTAR .....	vii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xii
DAFTAR GAMBAR .....	xiv
DAFTAR LAMPIRAN.....	xv
BAB I PENDAHULUAN	
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	3
1.3 Tujuan Penelitian .....	3
1.4 Manfaat Penelitian .....	4
1.5 Pembatasan Masalah .....	4
1.6 Penegasan Istilah.....	4
BAB II PUSTAKA DAN LANDASAN TEORI	
2.1 Landasan Teori.....	7
2.1.1 Konsep Dasar Peringkasan Teks Otomatis.....	7
2.1.2 Morfologi Bahasa Indonesia.....	9
2.1.3 <i>Text Preprocessing</i> .....	11
2.1.4 Konsep Dasar <i>Term Frequency-Inverse Document Frequency- Document Frequency (TF-IDF-DF)</i> .....	21
2.1.5 Konsep Dasar <i>Cosine Similarity</i> .....	23
2.1.6 Konsep Dasar <i>Maximum Marginal Relevance (MMR)</i> .....	24

2.1.7	Evaluasi Peringkasan Teks .....	27
2.1.8	Peringkasan Teks Otomatis Menggunakan Metode TF-IDF-DF dan Metode MMR .....	28
2.2	Penelitian Terdahulu .....	30
2.3	Perangkat Pembangunan Aplikasi .....	32
2.4	Kerangka Berfikir .....	34
<b>BAB III METODE PENELITIAN</b>		
3.1	Metode Pengumpulan Data .....	36
3.1.1	Observasi .....	36
3.1.2	Studi Pustaka .....	36
3.2	Pengembangan Sistem .....	37
3.2.1	Arsitektur Sistem .....	37
3.2.2	Desain .....	38
3.2.3	Pengkodean.....	41
3.3	Perancangan Pengujian .....	47
3.3.1	Bahan Pengujian .....	47
3.3.2	Tujuan Pengujian .....	47
3.3.3	Skenario dan Kriteria Pengujian.....	47
<b>BAB IV HASIL DAN PEMBAHASAN</b>		
4.1	Spesifikasi <i>Hardware</i> dan <i>Software</i> .....	49
4.2	Hasil Pengujian dan Analisis .....	50
4.2.1	Hasil Responden.....	50
4.2.2	Hasil Sistem.....	73
4.2.2.1	<i>Text preprocessing</i> .....	86
4.2.2.2	Algoritma TF-IDF-DF .....	92
4.2.2.3	Algoritma <i>cosine similarity</i> .....	95
4.2.2.4	Metode <i>MMR</i> .....	96
4.2.2.5	Evaluasi sistem .....	101
4.3	Perbandingan Hasil Ringkasan Portal Berita <i>Online</i> Berskala Nasional dan	

Lokal .....	104
BAB V PENUTUP	
5.1 Kesimpulan .....	109
5.2 Saran .....	110
DAFTAR PUSTAKA .....	111

## DAFTAR TABEL

Tabel 2.1. Proses segmentasi .....	12
Tabel 2.2. Proses <i>case folding</i> .....	13
Tabel 2.3. Proses <i>tokenizing</i> kata.....	13
Tabel 2.4. Proses <i>filtering</i> .....	15
Tabel 2.5. Hasil <i>stemming</i> .....	15
Tabel 2.6. Kombinasi awalan akhiran yang tidak diizinkan .....	18
Tabel 2.7. Aturan Pemenggalan Awalan <i>Stemmer</i> Nazief & Adriani .....	18
Tabel 2.8. Jenis awalan berdasarkan tipe awalannya.....	19
Tabel 2.9. Hasil penghitungan bobot masing-masing kata .....	22
Tabel 4.1. Ringkasan manual portal berita <i>online</i> berskala nasional.....	50
Tabel 4.2. Ringkasan manual portal berita <i>online</i> berskala lokal .....	62
Tabel 4.3. Ringkasan sistem portal berita <i>online</i> berskala nasional .....	74
Tabel 4.4. Hasil segmentasi kalimat .....	87
Tabel 4.5. Hasil <i>case folding</i> .....	88
Tabel 4.6. Hasil <i>tokenizing</i> .....	89
Tabel 4.7. Hasil <i>filtering</i> .....	90
Tabel 4.8. Hasil <i>stemming</i> .....	91
Tabel 4.9. Hasil penghitungan bobot masing-masing kata .....	92
Tabel 4.10. <i>Relevansi</i> antara judul dan semua kalimat .....	95
Tabel 4.11. <i>Similarity</i> antara kalimat .....	95
Tabel 4.12. Perhitungan MMR iterasi ke-1.....	97
Tabel 4.13. Nilai $S_i$ untuk perhitungan MMR iterasi ke-2.....	98
Tabel 4.14. Nilai MMR iterasi ke-2 .....	99
Tabel 4.15. Nilai <i>similarity</i> yang digunakan.....	99
Tabel 4.16. Hasil perhitungan MMR .....	100
Tabel 4.17. Hasil perangkaian kalimat .....	101

Tabel 4.18. Nilai akurasi pada portal berita <i>online</i> nasional.....	104
Tabel 4.19. Nilai akurasi pada portal berita <i>online</i> lokal.....	104
Tabel 4.20. Pengaruh panjang kalimat.....	106
Tabel 4.21. Pengaruh banyak <i>query</i> .....	106

## DAFTAR GAMBAR

Gambar 2.1. Proses <i>text preprocessing</i> .....	12
Gambar 2.2. Kerangka berfikir .....	35
Gambar 3.1. Arsitektur sistem peringkasan teks otomatis .....	38
Gambar 3.2. Alur proses pengembangan sistem .....	39
Gambar 3.3. Gambar interface proses peringkasan .....	40
Gambar 3.4. Gambar interface menyimpan ringkasan .....	40
Gambar 3.5. <i>Pseudocode</i> pemecahan paragraf dan <i>tokenizing</i> .....	41
Gambar 3.6. <i>Pseudocode</i> penyaringan simbol dan <i>stopword</i> .....	42
Gambar 3.7. <i>Pseudocode</i> penghitungan nilai TF dan DF .....	43
Gambar 3.8. <i>Pseudocode</i> penghitungan nilai TF-IDF-IDF .....	44
Gambar 3.9. <i>Pseudocode</i> penghitungan nilai <i>cosine similarity</i> .....	45
Gambar 3.10. <i>Pseudocode</i> penghitungan nilai MMR .....	46

## DAFTAR LAMPIRAN

Lampiran 1. Perhitungan <i>cosine similarity</i> .....	114
Lampiran 2. Surat Tugas Panitia Ujian Sarjana .....	117
Lampiran 3. Surat Penetapan Dosen Pembimbing.....	118

## BAB 1

### PENDAHULUAN

#### 1.1. Latar Belakang

Sejak tahun 2000, di Indonesia media *online* berkembang dengan pesat, terlihat dari jumlah portal berita *online* yang meningkat (Nugroho, *et al.*, 2012:83). Portal berita *online* ada berskala nasional dan lokal. Antaranews.com merupakan portal berita *online* berskala nasional yang beritanya mencakup peristiwa atau kejadian di lokal, nasional, bahkan internasional dan distribusi media cetaknya hampir meliputi semua wilayah (pulau, provinsi, kota/kabupaten). Berbeda dengan portal berita *online* lokal, sebagian besar beritanya hanya seputar peristiwa yang terjadi di suatu daerah saja dan distribusi media cetaknya mencakup satu provinsi dan kota/kabupaten dalam provinsi tersebut, seperti tribunjateng.com dan radarmagelang.com. Meningkatnya jumlah portal berita *online* menimbulkan tingginya jumlah berita yang disuguhkan kepada masyarakat. Oleh karena itu, kebutuhan peringkasan teks semakin diperlukan untuk kemudahan dan kehematan waktu manusia.

Dalam dunia komputer peringkasan teks dikenal dengan peringkasan teks otomatis (*automatic text summarization*). Peringkasan teks otomatis adalah secara otomatis komputer mengekstrak ringkasan dari artikel asli. Dalam keadaan ideal, ringkasan dapat mendeskripsikan isi



utama dari artikel secara akurat dan memahami serta bahasa ringkasan koheren (Lahari, *et al.*, 2014:1519).

Peringkasan teks otomatis telah banyak diteliti oleh beberapa peneliti terdahulu dengan menggunakan berbagai metode dan diterapkan dalam berbagai bahasa. Di antaranya *word frequency* (Luhn, 1958), *sentence position* (Hovy dan Lin, 1997), *lexical chain* (Barzilay, 1997) dan *maximum marginal relevance* (Cabonell dan Goldstein, 1998). Adapun metode dan algoritma yang diterapkan pada teks bahasa Indonesia, seperti algoritma genetika (Aristoteles, 2013), Metode *Relevance Measure* dan *Latent Semantic Analysis* (Triwibowo, *et al.*, 2010) serta *Maximum Marginal Relevance* (Mustaqhfiri, 2011).

Metode *Maximum Marginal Relevance* memiliki beberapa kelebihan yaitu, sederhana, efisien dan dapat mengurangi redundansi (Xie dan Liu, 2008:4985). Kelebihan tersebut menyebabkan banyak peneliti yang tertarik terhadap metode tersebut, seperti Mustaqhfiri, *et al.*, (2011) yang meneliti mengenai peringkasan teks otomatis berbahasa Indonesia pada berita Tempo Interaktif menggunakan metode *Maximum Marginal Relevance* (MMR). Kesimpulan dari penelitiannya yaitu metode MMR dapat meringkas *single* dokumen secara otomatis dengan akurasi 66 %.

Pesatnya perkembangan portal berita *online* yang tanpa kendali menyebabkan banyaknya portal berita *online* yang tampil dengan gaya dan struktur yang berbeda. Pada portal berita *online* berskala nasional, dominan

menggunakan kalimat yang panjang, sedangkan pada portal berita *online* berskala lokal, lebih banyak menggunakan kalimat yang pendek. Kalimat yang panjang memiliki informasi yang lebih lengkap jika dibandingkan dengan kalimat yang pendek.

Berdasarkan permasalahan tersebut, maka peneliti tertarik untuk melakukan penelitian terhadap peringkasan teks otomatis pada portal berita *online* berskala nasional dan lokal dengan judul “Implementasi Metode *Maximum Marginal Relevance* pada Peringkasan Teks Otomatis Artikel Berita”.

## **1.2. Rumusan Masalah**

Berdasarkan latar belakang yang telah dikemukakan, maka terdapat masalah yang timbul, yaitu : bagaimana keefektifan metode *summarization* dengan menggunakan pembobotan kata TF-IDF-DF yang diterapkan dalam metode *Maximum Marginal Relevance* (MMR) sebagai alat bantu dalam proses peringkasan teks pada artikel berita?

## **1.3. Tujuan Penelitian**

Tujuan penulisan skripsi ini adalah mengetahui keefektifan metode *summarization* dengan menggunakan pembobotan kata dengan metode TF-IDF-DF yang diterapkan dalam metode *Maximum Marginal Relevance* (MMR) sebagai alat bantu dalam proses peringkasan teks pada artikel berita.

#### 1.4. Manfaat Penelitian

Manfaat dari penelitian ini adalah menghemat waktu bagi para pembaca berita dalam memahami berita melalui ringkasan berita, mengetahui kemampuan metode MMR (*Maximum Marginal Relevance*) dalam memperoleh ringkasan berita secara otomatis, serta menambah ilmu pengetahuan, menjadi bahan referensi dan perbandingan untuk penelitian yang berkaitan dengan penggunaan metode MMR.

#### 1.5. Pembatasan Masalah

Permasalahan dibatasi pada proses peringkasan teks otomatis yang telah dikembangkan yaitu menggunakan metode *Maximum Marginal Relevance* (MMR). Inputan teks yang digunakan berupa *single document*, artinya ringkasan dihasilkan dari satu dokumen sumber. Teks yang dapat diringkas hanya sebatas karakter a-z.

#### 1.6. Penegasan Istilah

Berikut dijelaskan beberapa istilah yang berkaitan dengan judul penelitian. Beberapa istilah tersebut adalah :

a) Peringkasan Teks Otomatis (*Automatic Summarization Text*)

*Automatic Summarization Text* adalah komputer secara otomatis mengekstrak ringkasan dari artikel asli. Dalam kasus ideal, ringkasan bisa menggambarkan isi utama dari artikel (Lahari, *et al.*, 2014 : 1519). Peringkasan teks otomatis yang dimaksud dalam penelitian ini adalah suatu aplikasi yang dijalankan oleh komputer untuk memproses artikel

berita menjadi ringkasan yang lebih padat yang dapat mewakili isi dari suatu artikel berita.

b) Artikel

Artikel adalah sebuah susunan tulisan yang lengkap, mengenai suatu masalah tertentu, dalam surat kabar, majalah, dan sebagainya (Shahab, 2008:94).

c) Berita

Berdasarkan pendapat Harviyanto (2013:34), berita adalah laporan yang berisi suatu peristiwa atau kejadian penting yang menarik perhatian orang banyak dan berita itu berisi tentang fakta atau suatu yang baru yang dapat dipublikasikan melalui media cetak atau media elektronik.

d) TF-IDF-DF

Metode TF-IDF-DF (*Term Frequency-Inverse Document Frequency-Document Frequency*) merupakan modifikasi dari metode TF-IDF untuk mendapatkan bobot perwakilan dari kata-kata yang diekstrak dari data informasi dengan mempertimbangan penyebaran kata di dokumen lain (Pramono, *et al.*, 2013).

e) *Maximum Marginal Relevance* (MMR)

*Maximum Marginal Relevance* (MMR) adalah algoritma untuk menentukan relevansi hasil ringkasan dengan dokumen dan *query* (judul artikel berita) yang diberikan oleh *user* berdasarkan bobot

kesamaannya serta dapat mengurangi redundansi dalam peringkasan (Mustaqfiri, *et al.*, 2011 : 135).

## **BAB II**

### **TINJAUAN PUSTAKA**

#### **2.1 Landasan Teori**

##### **2.1.1 Konsep Dasar Peringkasan Teks Otomatis**

Teks yang terdapat pada buku, koran, majalah dan artikel yang terdiri dari banyak kalimat terkadang membutuhkan ringkasan untuk mengetahui inti bacaan. Ringkasan merupakan miniatur karangan aslinya sehingga struktur dan kelengkapan unsur ringkasan harus sama dengan karangan asli. Konsep sederhana dari meringkas adalah mengambil bagian penting dari suatu teks yang menggambarkan keseluruhan dokumen.

Dalam dunia komputer peringkasan teks dikenal dengan Peringkasan Teks Otomatis. Peringkasan Teks Otomatis (*Automatic Summarization Text*) adalah komputer secara otomatis mengekstrak ringkasan dari artikel asli. Dalam kasus yang ideal, ringkasan bisa menggambarkan isi utama dari artikel secara akurat dan komprehensif, dan bahasa ringkasan koheren dan halus (Lahari, *et al.*, 2014 : 1519). Menurut Muchammad Mustaqhfiri peringkasan teks otomatis (*automatic text summarization*) adalah pembuatan bentuk yang lebih singkat dari suatu teks dengan memanfaatkan aplikasi yang dijalankan dan dioperasikan pada komputer (Mustaqhfiri, *et al.*, 2011 : 134).

Peringkasan teks otomatis telah banyak diteliti oleh beberapa peneliti terdahulu dengan menggunakan berbagai metode dan pendekatan. Teknik untuk peringkasan dengan teknik statistika diantaranya yaitu teknik *word*

*frequency* (Luhn, 1958) dan *sentence position* (Hoovy dan Lin, 1997). Teknik pendekatan dengan natural language analysis yaitu *maximum marginal relevance* (Cabonell dan Goldstein, 1998).

Peringkasan teks otomatis diklasifikasikan menjadi 2 kategori, yaitu ekstraksi dan abstraksi. Peringkasan ekstraksi adalah pilihan kalimat atau frase dari teks asli dengan skor paling besar tanpa mengubah teks sumber. Metode peringkasan abstraksi menggunakan metode linguistik untuk memeriksa dan menafsirkan teks. Sebagian besar metode yang digunakan untuk peringkasan teks adalah metode peringkasan ekstraksi.

Berdasarkan jumlah sumbernya peringkasan teks otomatis dapat dibagi menjadi dua yaitu *single document* dan *multi document* (Mulyana *et al.*, : 2). *Single document* merupakan sumber dokumen yang akan diekstraksi berupa dokumen tunggal sedangkan, *multi document* adalah dokumen yang akan diekstrak terdiri dari beberapa dokumen. Berdasarkan metodenya, peringkasan dokumen dapat dibedakan menjadi peringkasan generik (*generic summarization*) dan peringkasan berdasarkan *query* (*query-based summarization*) (Jesek dan Steinberger, 2008:3).

Metode peringkasan dilakukan dengan menggunakan metode perangkingan untuk memilih kalimat dengan informasi yang sesuai dengan kebutuhan *user*. Hal ini mengakibatkan kurangnya keefektifan suatu kalimat yang dipilih, karena tidak memperhitungkan adanya duplikasi informasi terhadap kalimat yang menjadi ringkasan. Oleh karena itu, dibutuhkan suatu metode untuk menangani duplikasi informasi (Yusintan *et al.*, 2010:1-2).

Penulis memilih menggunakan metode *Maximum Marginal Relevance* (MMR) untuk meringkas artikel berita.

### 2.1.2 Morfologi Bahasa Indonesia

Secara etimologi kata morfologi berasal dari kata *morf* yang berarti ‘bentuk’ dan kata *logi* yang berarti ‘ilmu’. Secara harfiah kata morfologi berarti ilmu mengenal bentuk. Jadi, ujung dari proses morfologi adalah terbentuknya kata dalam bentuk dan makna sesuai dengan keperluan dalam satu tindak tutur (Chaer, 2008:3). Morfologi terbagi menjadi 2 bentuk, yaitu :

#### a. Bentuk inflektif

Dalam pembentukan kata inflektif, identitas leksikal kata yang dihasilkan sama dengan identitas leksikal bentuk dasarnya. Pembentukan kata membeli dari dasar beli adalah sebuah kasus inflektif. Hal ini karena beli dan kata membeli sama-sama berkategori verba.

Kasus inflektif dalam bahasa Indonesia hanya terdapat dalam pembentukan verba transitif, yaitu dengan prefik me-, prefik di-, prefik ter-, dan prefik zero. Bentuk dasarnya dapat berupa :

- 1) Pangkal verba akar yang memiliki komponen makna, seperti baca, beli dan tulis.
- 2) Pangkal bersufiks –kan, seperti selipkan dan daratkan.
- 3) Pangkal bersufiks –i, seperti tangisi dan nasihati
- 4) Pangkal bersufiks per-, seperti perpanjang dan perluas.
- 5) Pangkal berkonfiks per-kan seperti persembahkan.
- 6) Pangkal berkonfiks per-i, seperti perbaiki.



Keenam tipe pangkal tersebut dapat diberi afiks me-, di-, ter-, dan zero. Contoh:

- 1) me-  
di- + baca  
ter-  
Ø
- 2) me-  
di- + selipkan  
ter-  
Ø
- 3) me-  
di- + nasihati  
ter-  
Ø
- 4) me-  
di- + perpanjang  
ter-  
Ø
- 5) me-  
di- + persembahkan  
ter-  
Ø
- 6) me-  
di- + perbaiki  
ter-  
Ø

b. Bentuk derivatif

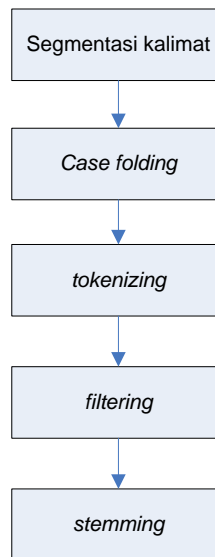
Proses pembentukan derivatif identitas bentuk yang dihasilkan tidak sama dengan identitas leksikal bentuk dasarnya. Pembentukan kata pembeli dari dasar beli adalah kasus derivatif. Dasar beli dan pembeli tidak sama kategorinya. Beli adalah verba dan pembeli adalah nomina. Pada derivatif juga terdapat prefiks me-, prefiks di-, dan prefiks ter-, seperti pada bentuk inflektif.

Misal :

- 1) Prefiks me-, seperti melompat dan membengkak.
- 2) Prefiks di-, seperti dimaksud.
- 3) Prefiks ter-, seperti terlena dan tertidur.

### **2.1.3 *Text Preprocessing***

*Text preprocessing* adalah suatu tahap untuk mengolah teks berita yang merupakan bahan mentah menjadi kata-kata yang telah siap dihitung bobot katanya. Beberapa proses dari *text preprocessing*, yaitu segmentasi kalimat, *case folding*, *tokenizing*, *filtering*, dan *stemming*. Berikut adalah diagram alirnya terlihat pada gambar 2.1.



Gambar 2.1 Proses *text preprocessing*

a. Segmentasi kalimat

Segmentasi kalimat merupakan langkah awal dari proses *text preprocessing*. Dalam proses ini, teks berita yang terdiri dari paragraf yang dipecah menjadi beberapa kalimat. Pemisahan setiap kalimat berdasarkan tanda baca, seperti tanda titik (.), tanda seru (!) dan tanda tanya (?). Contoh pemecahan dokumen menjadi kalimat pada tabel 2.1.

Tabel 2.1 Proses segmentasi

<b>Dokumen</b>	<b>Hasil segmentasi kalimat</b>
Membuang sampah sembarangan dapat menyebabkan banjir.	- Membuang sampah sembarangan dapat menyebabkan banjir
Aliran air akan tersumbat karena sampah. Banjir merugikan manusia.	- Aliran air akan tersumbat karena sampah - Banjir merugikan manusia

b. *Case Folding*

Paragraf berita yang telah dipotong menjadi kalimat akan menjalankan proses *case folding*. *Case folding* adalah proses mengubah semua teks menjadi karakter dengan huruf kecil dan membuang semua karakter selain a-z. Jika terdapat tanda baca, bilangan numerik dan simbol akan dihilangkan. Contohnya terlihat pada tabel 2.2.

Tabel 2.2 Proses *case folding*

<b>Kalimat</b>	<b>Hasil <i>case folding</i></b>
- Membuang sampah sembarangan dapat menyebabkan banjir	- membuang sampah sembarangan dapat menyebabkan banjir
- Aliran air akan tersumbat karena sampah	- aliran air akan tersumbat karena sampah
- Banjir merugikan manusia	- banjir merugikan manusia

c. *Tokenizing*

Suatu proses untuk mengubah bentuk kalimat menjadi kata-kata tunggal. Pemotongan kalimat berdasarkan delimiter yang menyusunnya, yaitu spasi (" "). Proses ini bertujuan agar nantinya dapat melakukan proses *stemming*. Contoh hasil *tokenizing* kata terlihat pada tabel 2.3.

Tabel 2.3 Proses *tokenizing* kata

<b>Hasil filtering</b>	<b>Hasil <i>tokenizing</i> kata</b>
- membuang sampah sembarangan dapat menyebabkan banjir	- membuang - sampah
- aliran air akan tersumbat karena sampah	- sembarangan - dapat

- banjir merugikan manusia	- menyebabkan
	- banjir
	- aliran
	- air
	- akan
	- tersumbat
	- karena
	- sampah
	- banjir
	- merugikan
	- manusia

#### d. *Filtering*

Pada proses *filtering* dilakukan pembuangan *stopword*. *Stopword* adalah kata-kata yang tidak memiliki makna atau kata yang kurang berarti dan sering muncul dalam kumpulan kata-kata. Proses pembuangan kata-kata yang tidak penting dengan mengecek pada kamus *stopword*. Jika kata ada yang sama dengan kata pada kamus *stopword*, maka kata tersebut akan dibuang atau dihapus. Seandainya *stopword* tidak dihilangkan maka *stopword* akan memiliki bobot yang besar karena sering muncul dalam teks, sehingga akan mempengaruhi hasil ringkasan karena kalimat yang terpilih adalah kalimat yang memiliki bobot paling besar. *Stopword* dapat berupa kata penghubung, kata ganti, preposisi, dll, seperti : dia, antara, akan, demi, karena, atau, bahwa, bila, juga, kalau, hingga, bagi, kecuali, oleh, dan lain-lain. Contoh *filtering* terlihat pada tabel 2.4.

Tabel 2.4 Proses *filtering*

<b>Kalimat</b>	<b>Hasil <i>filtering</i></b>
- membuang sampah sembarangan dapat menyebabkan banjir	- membuang sampah sembarangan menyebabkan banjir
- aliran air akan tersumbat karena sampah	- aliran air tersumbat sampah
- banjir merugikan manusia	- banjir merugikan manusia

e. *Stemming*

Proses selanjutnya *stemming*, yaitu mengembalikan suatu kata ke bentuk akarnya (*root word*) dengan aturan-aturan tertentu, sehingga setiap kata memiliki representasi yang sama. *Stemming* dalam penelitian ini menggunakan algoritma Nazief & Adriani (1996). Contoh *stemming* terlihat pada tabel 2.5.

Tabel 2.5 Hasil *stemming*

<b>Hasil <i>tokenizing</i> kata</b>		<b>Hasil <i>stemming</i></b>	
Membuang	air	buang	air
sampah	tersumbat	sampah	sumbat
sembarangan	sampah	sembarang	sampah
menyebabkan	banjir	sebab	banjir
banjir	merugikan	banjir	rugi
aliran	manusia	alir	manusia

Algoritma *stemming* Nazief dan Adriani (1996), dikembangkan berdasarkan morfologi bahasa Indonesia dengan mengelompokkan imbuhan menjadi awalan (*prefix*), akhiran (*suffix*), dan gabungan awalan-akhiran (*confixes*). Algoritma yang dibuat oleh Bobby Nazief dan Mirna Adriani ini memiliki tahap-tahap sebagai berikut:

1. Cari kata yang akan distem dalam kamus. Jika ditemukan maka diasumsikan bahwa kata tersebut adalah *root word*. Maka algoritma berhenti.
2. *Inflection Suffixes* (“-lah”, “-kah”, “-ku”, “-mu”, atau “-nya”) dibuang. Jika berupa *particles* (“-lah”, “-kah”, “-tah” atau “-pun”) maka langkah ini diulangi lagi untuk menghapus *Possesive Pronouns* (“-ku”, “-mu”, atau “-nya”), jika ada.
3. Hapus *Derivation Suffixes* (“-i”, “-an” atau “-kan”). Jika kata ditemukan di kamus, maka algoritma berhenti. Jika tidak maka ke langkah 3a.
  - a. Jika “-an” telah dihapus dan huruf terakhir dari kata tersebut adalah “k”, maka “-k” juga ikut dihapus. Jika kata tersebut ditemukan dalam kamus maka algoritma berhenti. Jika tidak ditemukan maka lakukan langkah 3b.
  - b. Akhiran yang dihapus (“-i”, “-an” atau “-kan”) dikembalikan, lanjut ke langkah 4.
4. Hapus *Derivation Prefix*. Jika pada langkah 3 ada sufiks yang dihapus maka pergi ke langkah 4a, jika tidak pergi ke langkah 4b.

- a. Periksa tabel kombinasi awalan-akhiran yang tidak diijinkan pada tabel 2.6. Jika ditemukan maka algoritma berhenti, jika tidak pergi ke langkah 4b.
  - b. For  $i = 1$  to 3, tentukan tipe awalan kemudian hapus awalan. Jika *root word* belum juga ditemukan lakukan langkah 5, jika sudah maka algoritma berhenti. Catatan: jika awalan kedua sama dengan awalan pertama algoritma berhenti.
5. Melakukan *Recoding*.
  6. Jika semua langkah telah selesai tetapi tidak juga berhasil maka kata awal diasumsikan sebagai *root word*. Proses selesai.

Tipe awalan ditentukan melalui langkah-langkah berikut:

1. Jika awalannya adalah: “di-”, “ke-”, atau “se-” maka tipe awalannya secara berturut-turut adalah “di-”, “ke-”, atau “se-”.
2. Jika awalannya adalah “te-”, “me-”, “be-”, atau “pe-” maka dibutuhkan sebuah proses tambahan untuk menentukan tipe awalannya.
3. Jika dua karakter pertama bukan “di-”, “ke-”, “se-”, “te-”, “be-”, “me-”, atau “pe-” maka berhenti.
4. Jika tipe awalan adalah “none” maka berhenti. Jika tipe awalan adalah bukan “none” maka awalan dapat dilihat pada Tabel 2.8. Hapus awalan jika ditemukan.



Tabel 2.6 Kombinasi awalan akhiran yang tidak diizinkan

Awalan	Akhiran yang tidak diizinkan
be-	-i
di-	-an
ke-	-i, -kan
me-	-an
se-	-i, -kan

Tabel 2.7 Aturan Pemenggalan Awalan *Stemmer* Nazief & Adriani

Aturan	Format Kata	Pemenggalan
1	berV...	ber-V...   be-rV
2	berCAP...	ber-CAP... dimana C!=’r’ & P!=’er’
3	berCAerV...	ber-CaerV... dimana C!’r’
4	Belajar...	bel-ajar
5	beC l erC2...	be-C lerC2... dimana C1!= {’r’ ’I’}
6	terV...	Ter-V...   te-rV...
7	terCerV...	Ter-CerV... dimana C!=’r’
8	terCP...	Ter-CP... dimana C!=’r’ dan P!=’er’
9	teC l erC2...	te-C l erC2... dimana C1!=’r’
10	me{l r w y}V...	me-{l r w y}V...
11	mem{b f v}...	mem-{b f v}...
12	mempe{r l}...	mem-pe...
13	mem{rV V}...	me-m{rV V}...   me-p{rV V}...
14	men{c d j z}...	men-{c d j z}...
15	menV...	me-nV...   me-tV
16	meng{g h q}...	meng-{g h q}...
17	mengV...	meng-V...   meng-kV...
18	menyV...	meny-sV...
19	mempV...	mem-pV... dimana V!=’,e’
20	pe{w y}V...	pe-{w y}V...
21	perV...	per-V...   pe-rV...
23	perCAP	per-CAP... dimana C!=’r’ dan P!=’er’
24	perCAerV...	per-CAerV... dimana C!=’r’
25	pem{b f V}...	pem-{b f V}...

Aturan	Format Kata	Pemenggalan
26	pem{rV V}...	pe-m{rV V}...   pe-p{rV V}...
27	pen{c d j z}...	pen-{c d j z}...
28	penV...	pe-nV...   pe-tV...
29	peng{g h q}...	peng-{g h q}...
30	pengV...	peng-V...   peng-kV...
31	penyV...	peny-sV...
32	peIV...	pe-IV... kecuali “pelajar” yang menghasilkan “ajar”
33	peCerV...	per-erV... dimana C!={r w y l m n}
34	peCP...	pe-CP... dimana C!={r w y l m n} dan P!="er"

Keterangan simbol huruf :

C : huruf konsonan

V : huruf vokal

A : huruf vokal atau konsonan

P : partikel atau fragmen dari suatu kata, misalnya “er”

Tabel 2.8 Jenis awalan berdasarkan tipe awalannya

Tipe awalan	Awalan yang harus dihapus
di-	-di
ke-	ke-
se-	se-
te-	te-
ter-	ter-
ter-luluh	ter

Untuk mengatasi keterbatasan pada algoritma di atas, maka ditambahkan aturan-aturan dibawah ini:

1. Aturan untuk reduplikasi.
  - a. Jika kedua kata yang dihubungkan oleh kata penghubung adalah kata yang sama maka *root word* adalah bentuk tunggalnya, contoh: “buku-buku” *root word*-nya adalah “buku”.
  - b. Kata lain, misalnya “bolak-balik”, “berbalas-balasan, dan ”seolah-olah”. Untuk mendapatkan *root word*-nya, kedua kata diartikan secara terpisah. Jika keduanya memiliki *root word* yang sama maka diubah menjadi bentuk tunggal, contoh: kata “berbalas-balasan”, “berbalas” dan “balasan” memiliki *root word* yang sama yaitu “balas”, maka *root word* “berbalas-balasan” adalah “balas”. Sebaliknya, pada kata “bolak-balik”, “bolak” dan “balik” memiliki *root word* yang berbeda, maka *root word*-nya adalah “bolak-balik”
2. Tambahan bentuk awalan dan akhiran serta aturannya.
  - a. Untuk tipe awalan “mem-“, kata yang diawali dengan awalan “memp-” memiliki tipe awalan “mem-”.
  - b. Tipe awalan “meng-“, kata yang diawali dengan awalan “mengk-” memiliki tipe awalan “meng-”.

#### 2.1.4 Konsep Dasar *Term Frequency-Inverse Document Frequency-Document Frequency* (TF-IDF-DF)

Metode *Term Frequency-Inverse Document Frequency-Document Frequency* (TF-IDF-DF) merupakan modifikasi dari metode TF-IDF, karena metode TF-IDF memiliki kekurangan dalam pembobotan kata. Kekurangannya yaitu adanya anggapan bahwa kata yang tersebar dalam dokumen lain tidak penting, sehingga dianggap tidak ada. Padahal kata yang sering muncul dalam kalimat lain bisa jadi merupakan kata yang penting. Akibatnya, nilai bobot yang tinggi diperoleh pada kata yang memiliki frekuensi tinggi dalam dokumen, sedangkan kata yang tersebar di dokumen lain memiliki perhitungan bobot yang kecil. Oleh karena itu, metode TF-IDF ini dikembangkan lebih lanjut untuk mendapatkan bobot perwakilan dari kata-kata yang diekstrak dengan mempertimbangan penyebaran kata di dokumen lain. *Document Frequency* (DF) yang mengandung kata ke-i berpengaruh pada topik keseluruhan dokumen, sehingga nilai dalam pembobotan kata akan dikalikan dengan DF kata ke-i (Pramono, 2013). Rumus pembobotan kata dari modifikasi TF-IDF adalah :

$$w_{i,j} = \left( tf_{i,j} \times \log \left( \frac{N}{df_i} \right) \right) \times df_i$$

Dimana  $w_{i,j}$  adalah bobot dari kata ke-i dari dokumen ke-j.  $tf_{i,j}$  adalah *term frequency*, adalah jumlah kata dari kata ke-i dalam dokumen ke-j.  $\log \left( \frac{N}{df_i} \right)$  merupakan rumus *Invers Document Frequency* (IDF),  $N$  merupakan jumlah seluruh dokumen atau kalimat.  $df_i$  yaitu *document frequency*

merupakan banyaknya kalimat yang berisi kata ke-i dalam kumpulan dokumen. Contoh perhitungan algoritma TF-IDF-DF dengan *term* “sampah” sesuai dengan tabel 2.5 sebagai berikut :

$$w_{i,j} = \left( tf_{i,j} \times \log \left( \frac{N}{df_i} \right) \right) \times df_i$$

$$w_{2,1} = \left( 1 \times \log \left( \frac{3}{1} \right) \right) \times 2 \longrightarrow \text{bobot term "sampah" pada dokumen 1 (D1)}$$

$$= 0,81093$$

$$w_{2,2} = \left( 1 \times \log \left( \frac{3}{1} \right) \right) \times 2 \longrightarrow \text{bobot term "sampah" pada dokumen 2 (D2)}$$

$$= 0,81093$$

$$w_{2,3} = \left( 0 \times \log \left( \frac{3}{1} \right) \right) \times 2 \longrightarrow \text{bobot term "sampah" pada dokumen 3 (D3)}$$

$$= 0$$

Pada tabel 2.1 diketahui bahwa jumlah dokumen (N) sebanyak 3 kalimat. Hasil perhitungan bobot kata dengan metode TF-IDF-DF terlihat pada tabel 2.9.

Tabel 2.9 Hasil penghitungan bobot masing-masing kata

Term	tf			Df	Idf	W		
	D1	D2	D3			D1	D2	D3
buang	1	0	0	1	1,098612	1,098612	0	0
sampah	1	1	0	2	0,405465	0,810930	0,810930	0
sembarang	1	0	0	1	1,098612	1,098612	0	0
sebab	1	0	0	1	1,098612	1,098612	0	0
banjir	1	0	1	2	0,405465	0,810930	0	0,810930
alir	0	1	0	1	1,098612	0	1,098612	0
air	0	1	0	1	1,098612	0	1,098612	0
sumbat	0	1	0	1	1,098612	0	1,098612	0
rugi	0	0	1	1	1,098612	0	0	1,098612
manusia	0	0	1	1	1,098612	0	0	1,098612

Keterangan :

Term : kata yang akan dihitung bobotnya

Tf : jumlah kata muncul dalam satu kalimat/dokumen

Df : jumlah dokumen yang berisi kata tersebut

Idf : kemunculan kata pada kumpulan kalimat/dokumen

W : bobot kata

Di : dokumen atau kalimat ( $i=1,2,3$ )

### 2.1.5 Konsep Dasar *Cosine Similarity*

*Cosine similarity* digunakan untuk menghitung relevansi *query* dengan dokumen. Nilai relevansi didapat dengan mengukur kesamaan antara 2 vektor, yaitu vektor *query* dan vektor dokumen. Semakin besar nilai relevansi, maka *query* dan dokumen akan semakin relevan.

Menurut Patidar *et al.* (2013: 90), ukuran kesamaan (*similarity measure*) adalah jarak antara berbagai titik data. *Similarity measure* juga digunakan dalam mengukur kesamaan antara set berdasarkan persimpangan dua set. *Similarity measure* juga dikenal sebagai fungsi yang menghitung tingkat kesamaan antara sepasang objek teks. Singkatnya, *similarity* adalah jumlah yang mencerminkan kekuatan hubungan antara dua data.

Ada beberapa jenis *similarity measure* seperti *Dice coefficient*, *Jaccard Similarity*, *Cosine Similarity*, *Euclidean Distance* dan lain-lain. *Similarity measure* dapat mewakili kesamaan antara dua kalimat dan memungkinkan untuk merangking informasi yang diambil dalam urutan dengan anggapan itu penting.

Salah satu ukuran kesamaan yang paling umum digunakan adalah *cosine similarity*. *Cosine similarity* merupakan dasar perhitungan untuk mendapatkan nilai relevansi antara *query* dengan dokumen dan relevansi antara dokumen. *Cosine similarity* adalah cosine sudut  $\theta$  antara vektor. *Cosine similarity* memiliki rumus seperti dibawah ini :

$$\text{sim}(S_1, S_2) = \frac{\sum_i t_{1i} t_{2i}}{\sqrt{\sum_i t_{1i}^2} \times \sqrt{\sum_i t_{2i}^2}}$$

Keterangan :

$S_1$  = vektor bobot kata yang menjadi kandidat

$S_2$  = vektor bobot kata selain kandidat

Dimana  $t_i$  merupakan bobot kata dari kata  $w_i$ . Berikut adalah contoh perhitungan *cosine similarity* antara kalimat 1 (D1) dengan kalimat 2 (D2) :

$$\begin{aligned} \text{sim}(D_1, D_2) &= \frac{\sum_i t_{1i} t_{2i}}{\sqrt{\sum_i t_{1i}^2} \times \sqrt{\sum_i t_{2i}^2}} \\ &= \frac{(1,098612 * 0) + (0,81093 * 0,81093) + (1,098612 * 0) + (1,098612 * 0) + (0,81093 * 0) + (0 * 1,098612) + (0 * 1,098612) + (0 * 1,098612) + (0 * 0) + (0 * 0)}{\sqrt{1,098612^2 + 0,81093^2 + 1,098612^2 + 1,098612^2 + 0,81093^2 + 0^2 + 0^2} \times \sqrt{0^2 + 0,81093^2 + 0^2 + 0^2 + 0^2 + 1,098612^2 + 1,098612^2 + 1,098612^2 + 0^2 + 0^2}} \end{aligned}$$

### 2.1.6 Konsep Dasar *Maximum Marginal Relevance* (MMR)

*Summarization* (peringkasan) bertujuan untuk menghasilkan ringkasan sebuah dokumen atau sekelompok dokumen. *Teks summarization* dapat dikategorikan dalam peringkasan *single-document* atau *multi-document*.

Peringkasan pada *single-document*, Carbonell (1998 : 336) mengusulkan metode *Maximum Marginal Relevance* (MMR) untuk menghasilkan ringkasan. Metode ini diusulkan pertama kali oleh Carbonell dan Goldstein pada tahun 1998.

*Maximum Marginal Relevance* (MMR) adalah salah satu dari sekian metode ekstraksi teks yang dapat diterapkan untuk meringkas dokumen tunggal maupun multi dokumen dengan cara melakukan ranking ulang dan membandingkan *similarity* antar dokumen. Menurut Toth, B. (2010), *Maximum Marginal Relevance* (MMR) adalah teknik peringkasan yang bertujuan untuk menangkap informasi yang relevan dan tidak mengandung redundansi. MMR berfungsi untuk menghitung kesamaan antara bagian teks. Tujuannya adalah dicapainya skor kalimat berdasarkan kesamaan (*similarity*) *query* yang diberikan.

Metode MMR sering digunakan untuk peringkasan teks karena metode MMR sederhana dan efisien (Xie, S dan Liu, Y., 2008 : 4985). Jika kesamaan (*similarity*) antara satu kalimat dengan kalimat yang lain tinggi, maka kemungkinan terjadi redundansi. Rumus untuk menghitung nilai MMR yang dapat mengurangi redundansi adalah :

$$MMR(S_i) = \lambda \cdot Sim_1(S_i, Q) - (1 - \lambda) \cdot \max Sim_2(S_i, S_j)$$

Keterangan :

$\lambda$  = parameter bobot untuk mengatur tingkat relevansi

$S_i$  = vektor bobot kata yang menjadi kandidat



$S_j$  = vektor bobot kata selain yang menjadi kandidat

$Q$  = vektor bobot kata dari *query user* (judul berita)

$Sim_1$  = nilai *similarity* antara *query* dengan tiap kalimat

$Sim_2$  = nilai *similarity* antara kalimat

Sim adalah kosinus kesamaan antara dua vektor fitur.  $\lambda$  adalah koefisien untuk mengatur relevansi kalimat dan mengurangi redundansi. Nilai parameter  $\lambda$  adalah 1 atau 0 atau diantaranya ( $0 < \lambda < 1$ ). Pada saat parameter  $\lambda = 1$  maka nilai MMR yang diperoleh cenderung relevan terhadap dokumen asli. Ketika  $\lambda = 0$  maka nilai MMR yang diperoleh akan cenderung relevan terhadap kalimat yang diekstrak sebelumnya yang akan dibandingkan. Oleh sebab itu, sebuah kombinasi linear dari kedua kriteria dioptimalkan ketika nilai  $\lambda$  terdapat pada interval  $0 < \lambda < 1$ . Untuk peringkasan dengan dokumen yang kecil, seperti artikel berita akan menghasilkan hasil ringkasan yang baik, jika nilai parameter  $\lambda = 0,7$  atau  $\lambda = 0,8$  (Goldstein, 2008).

Penggunaan rumus MMR dalam perangkian ulang adalah untuk mendapatkan ringkasan dengan *similarity query* kalimat tinggi, sedangkan *similarity* antara kalimat rendah. Pada rumus dibawah ini merupakan rumus yang memperhitungkan relevansi kalimat dengan *query*. Jadi, rumus tersebut merupakan benih untuk menentukan kalimat yang akan dipilih selanjutnya untuk menjadi ringkasan. Kalimat yang akan dipilih sebagai ringkasan berikutnya adalah kalimat yang berdekatan dengan kalimat benih ini.

$$MMR (S_i) = Sim (S_i, Query)$$

### 2.1.7 Evaluasi Peringkasan Teks

Menurut Nedunchelian (2011:5), proses evaluasi hasil *text summarization* dilakukan menggunakan tiga parameter yaitu *precision*, *recall*, dan *F-measure*. Cara mengevaluasi dengan membandingkan ringkasan otomatis hasil sistem dengan ringkasan manual.

#### 1. *Precision*

Merupakan perbandingan jumlah informasi relevan yang didapatkan sistem dengan jumlah seluruh informasi yang terambil oleh sistem baik yang relevan maupun tidak. Persamaan *precision* ditunjukkan pada persamaan berikut :

$$P = \frac{\textit{correct}}{(\textit{correct} + \textit{wrong})}$$

#### 2. *Recall*

Merupakan perbandingan jumlah informasi relevan yang didapatkan sistem dengan jumlah seluruh informasi relevan yang ada dalam koleksi informasi (baik yang terambil atau tidak terambil oleh sistem).

$$R = \frac{\textit{correct}}{(\textit{correct} + \textit{missed})}$$

Keterangan:

*Correct* : jumlah kalimat yang diekstrak oleh sistem dan manusia.

*Wrong* : jumlah kalimat yang diekstrak oleh sistem tetapi tidak diekstrak oleh manusia.

*Missed* : jumlah kalimat yang diekstrak oleh manusia tetapi tidak diekstrak oleh sistem.

### 3. *F-measure*

Merupakan hubungan antara *recall* dan *precision* yang mempresentasikan akurasi sistem. Persamaan *F-measure* seperti berikut :

$$F = \frac{2 * R * P}{(R + P)}$$

#### **2.1.8 Peringkasan Teks Otomatis Menggunakan Metode TF-IDF-DF dan Metode MMR**

Pada penelitian ini, peringkasan teks otomatis dengan menggunakan metode TF-IDF-DF untuk pembobotan kata dan menggunakan metode MMR untuk peringkasannya. Inputan teks berupa artikel berita sebagai *single dokument* yang merupakan bahan mentah untuk menghasilkan ringkasan (*summary*). Untuk menghasilkan ringkasan, artikel berita tersebut harus menjalankan proses *text preprocessing*, yaitu tahap untuk mengubah artikel berita yang merupakan bahan mentah menjadi teks yang siap untuk dihitung bobot katanya. *Text preprocessing* melewati beberapa proses seperti segmentasi kalimat, *case folding*, *tokenizing*, *filtering*, dan *stemming*.

Peringkasan teks otomatis dengan metode TF-IDF-DF dan MMR terdiri dari tahap-tahap berikut :

(1) Artikel berita diinput dengan memasukkan *query* berupa judul dan teks beritanya.

(2) Segmentasi kalimat

Memecah paragraf menjadi kalimat-kalimat. Pemecahan dilakukan berdasarkan tanda baca berupa tanda titik (.), tanda tanya (?) dan tanda seru (!). Pemisahannya menggunakan fungsi *split()*.

(3) *Case folding*

Proses mengubah huruf kapital menjadi huruf kecil dan membuang semua tanda baca, angka dan simbol.

(4) *Tokenizing*

Memecah kalimat menjadi kata berdasarkan spasi antara kata.

(5) *Filtering*

Proses pembuangan kata yang tidak berpengaruh terhadap proses peringkasan. Kumpulan kata tersebut berupa *stopword*, sehingga kata tersebut tidak mengganggu proses pembobotan kata nantinya.

(6) *Stemming*

Merupakan proses pencarian kata dasar dengan cara membuang imbuhan yang terdapat pada kata (kembali dalam bentuk akarnya).

(7) Perhitungan pembobotan kata dengan metode TF-IDF-DF

Pada tahap ini, penghitungan bobot kata dimulai dengan mencari nilai TF (*Term Frequency*), yaitu mencari nilai banyaknya kata yang muncul dalam suatu kalimat. Berikutnya mencari nilai IDF (*Invers Document Frequency*) merupakan perhitungan jumlah kata (*term*) dalam seluruh kalimat pada dokumen. Terakhir menghitung nilai DF (*Document Frequency*) yaitu nilai jumlah kalimat yang mengandung suatu kata.

(8) Perhitungan *cosine similarity*

Menghitung kesamaan antara satu kalimat dengan seluruh kalimat lain dan antara *query* (judul) dengan seluruh kalimat.

(9) Perhitungan MMR

Tahap ini menghitung nilai relevansi antara nilai *cosine similarity query* dengan seluruh kalimat dan kalimat dengan seluruh kalimat (Mustaqhfiri, *et al.*, 2011:144).

(10) Perankingan kalimat

Ringkasan artikel berita diperoleh dari memilih tiga kalimat (ukuran ringkasan yang diinginkan) dengan skor MMR yang tertinggi (Mustaqhfiri, *et al.*, 2011:135).

(11) Menentukan nilai *precision*, *recall*, dan *F-measure*.

Sebuah sistem informasi dikatakan baik jika tingkat *precision*, *recall*, dan *F-measure*-nya tinggi.

## 2.2 Penelitian Terdahulu

Beberapa penelitian terdahulu yang berkaitan dengan penelitian peringkasan teks otomatis, seperti :

Luthfan Hadi Pramono, Arief Syaichu Rohman, dan Hilwadi Hindersah dari Institut Teknologi Bandung tahun 2013 dalam penelitiannya yang berjudul “*Modified Weighting Method in TF\*IDF Algorithm for Extracting User Topic Based on Email and Social Media in Integrated Digital Assistant*”, menyatakan bahwa mengekstraksi topik dengan memodifikasi TF\*IDF yaitu TF\*IDF\*DF diperoleh peningkatan pada jumlah informasi berita dan terdapat perbedaan informasi berita. Untuk algoritma TF\*IDF diperoleh 39 berita, sedangkan algoritma TF\*IDF\*DF diperoleh 46 berita.

Muchammad Mustaqhfiri, Zainal Abidin dan Ririen Kusumawati dari Universitas Islam Negeri Maulana Malik Ibrahim Malang tahun 2011 dalam penelitiannya yang berjudul “Peringkasan Teks Otomatis Berita Berbahasa Indonesia Menggunakan Metode *Maximum Marginal Relevance*”, membahas tentang metode MMR yang dapat digunakan untuk meringkas *single document* secara otomatis. Dokumen diperoleh dari portal berita *online* Tempo Interaktif dengan hasil rata-rata *recall* 60%, *precision* 77% dan *f-measure* 60% berdasarkan perbandingan dengan ringkasan manual.

Jaime G. Carbonell dan Jade Golstein dari *Carnegie Mellon University* tahun 1998 dalam penelitian yang berjudul “*The Use of MMR and Diversity-Based Reranking in Document Reranking and Summarization*”, membahas tentang perangkaian dengan metode MMR menghasilkan kemanfaatan dan merupakan cara yang berguna untuk menyediakan informasi kepada *user* dengan membiarkan *user* meminimalkan redundansi. Penelitian ini diterapkan pada kasus peringkasan multi dokumen.

Octarina Nur Hidayah, Warih Maharani, dan Kusuma Ayu Laksitowening dari Institut Teknologi Telkom Bandung tahun 2013 dalam penelitiannya yang berjudul “*Opinion Summarization Review Produk dalam Bahasa Indonesia dengan Metode Maximum Marginal Relevance (MMR)*”, membahas tentang sistem *opinion summarization* yang dapat meringkas *review* produk dalam bahasa Indonesia dan mengelompokkannya dalam kategori positif dan negatif. Berdasarkan pengujian yang dilakukan, peringkasan *review*

produk berbahasa Indonesia menghasilkan *precision* tertinggi 88,89% dengan *recall* 88,89%.

Zaman, B. & Winarko, E. pada tahun 2011 dalam penelitiannya yang berjudul “Analisis Fitur Kalimat untuk Peringkasan Teks Otomatis pada Bahasa Indonesia”, menghasilkan kesimpulan bahwa tingkat akurasi peringkasan teks otomatis menggunakan pembobotan fitur lebih baik dibandingkan dengan tanpa fitur pembobotan. Dokumen uji cobanya diambil dari [kompas.com](http://kompas.com).

Aristoteles dari Universitas Lampung dalam penelitiannya yang berjudul “Penerapan Algoritma Genetika pada Peringkasan Teks Dokumen Bahasa Indonesia”, membahas tentang meringkas dokumen bahasa Indonesia yang berjenis file teks dengan menggunakan algoritma genetika. Sumber dokumen uji cobanya diambil dari harian [kompas online](http://kompas.com). Dalam penelitian ini menerapkan sebelas fitur teks, yaitu posisi kalimat, *positive keyword*, *negative keyword*, kemiripan antar kalimat, kalimat menyerupai judul, kalimat yang mengandung nama entiti, kalimat yang mengandung data numerik, koneksi antar-kalimat, penjumlahan bobot antar-kalimat, dan kalimat semantik. Hasil pengujian menunjukkan bahwa akurasi dengan pemampatan 30%, 20%, 10% sebesar 47,46%, 4,29% dan 35,01%.

### **2.3 Perangkat Pembangun Aplikasi**

XAMMP v3.2.1 terdiri dari beberapa aplikasi sebagai berikut :

#### (1) PHP

PHP (*Hypertext Preprocessor*) merupakan bahasa *script server-side* yang digunakan untuk mempermudah dan meningkatkan keefisiensi dalam

pembuatan Web. PHP merupakan software *open-source* yang disisipkan pada dokumen HTML (Peranginangin, 2006:2). PHP juga merupakan *script* pemograman yang terletak dan dieksekusi di *server* dan berfungsi sebagai pengolah data pada *server* sehingga website akan lebih interaktif dan dinamis.

## (2) MySQL

MySQL adalah salah satu program yang dapat digunakan sebagai *database* dan merupakan salah satu *software* untuk *database server* yang banyak digunakan. MySQL bersifat *Open Source* dan menggunakan SQL. Kelebihan My SQL diantaranya adalah dapat digunakan oleh beberapa *user* dalam waktu bersamaan tanpa mengalami masalah, memiliki kecepatan yang bagus dalam menangani *query* sederhana, mampu menangani basis data dalam skala besar, memiliki keamanan yang baik, memiliki operator dan fungsi secara penuh dan mendukung perintah *Select* dan *Where* dalam perintah *query*, dll (MADCOMS, 2011:140).

## (3) phpMyAdmin

Merupakan sekumpulan *script* PHP yang dipergunakan untuk mengelola *database* MySQL. phpMyAdmin ditujukan untuk menangani administrasi MySQL berbasis web. Artinya interaksi pemeliharaan dilakukan oleh *client* dengan menggunakan antar muka *browser*, sedangkan *script* PHP sendiri akan dijalankan oleh suatu web *server*. Keuntungan dengan basis web di sini adalah fleksibel penggunaan dan kemudahan interaksi *user*,



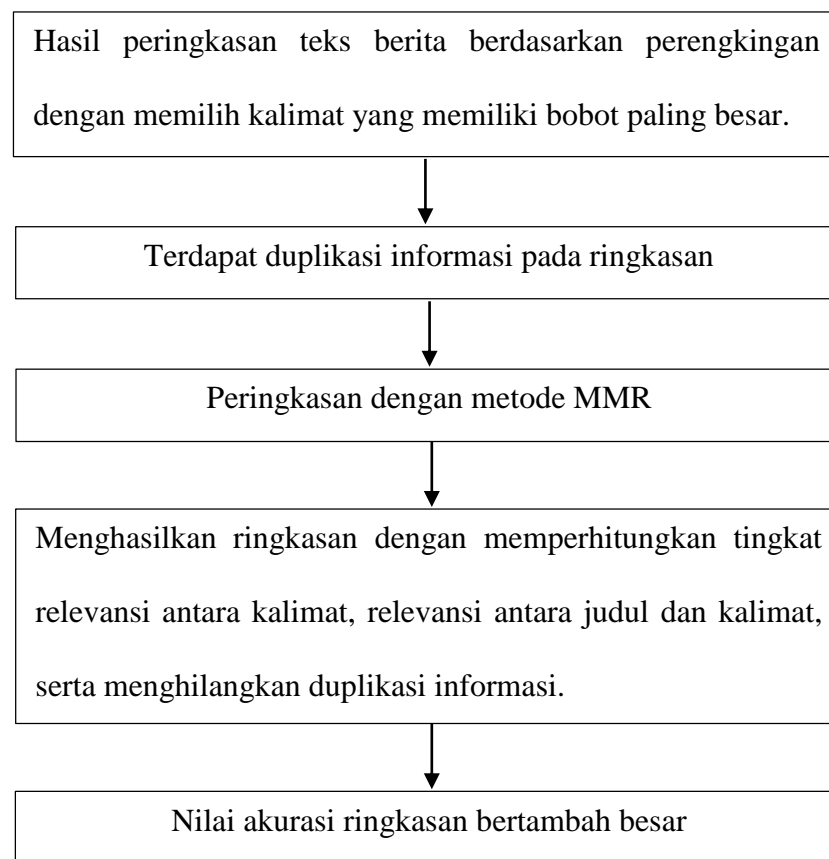
tidak tergantung pada platform sistem operasinya (Utdirartatmo, 2002:119).

## 2.4 Kerangka Pikir

Media *online* di Indonesia berkembang dengan pesat, terlihat dari jumlah portal berita *online* yang meningkat. Berdasarkan konten berita, portal berita *online* ada yang berskala nasional dan lokal. Struktur kalimat pada portal berskala nasional lebih dominan menggunakan kalimat yang panjang, sedangkan berita pada portal berskala lokal lebih banyak menggunakan kalimat yang lebih ringkas. Kalimat yang panjang memiliki informasi yang lebih lengkap jika dibandingkan dengan kalimat yang ringkas. Meningkatnya jumlah portal berita *online* menimbulkan tingginya jumlah berita yang disuguhkan kepada masyarakat. Oleh karena itu, kebutuhan peringkasan teks semakin diperlukan untuk kemudahan dan kehematan waktu manusia dalam mengetahui inti berita.

Penelitian mengenai peringkasan teks otomatis telah banyak dilakukan oleh peneliti terdahulu, tetapi sebagian besar hasil ringkasan ditentukan berdasarkan perangkangan saja. Kelemahan yang ditimbulkan yaitu terdapat duplikasi informasi, karena proses perangkangan hanya mengambil kalimat dengan bobot yang tertinggi tanpa adanya pengecekan kesamaan informasi antara satu kalimat dengan kalimat yang lain, sehingga dibutuhkan metode yang dapat mengatasi permasalahan tersebut seperti metode MMR (*Maximum Marginal Relevance*).

Metode MMR memiliki beberapa kelebihan yaitu, sederhana, efisien dan dapat mengurangi redundansi. Metode MMR menghasilkan ringkasan dengan memperhitungkan tingkat relevansi antara kalimat serta relevansi antara judul dan kalimat, sehingga tidak terjadi duplikasi informasi dan tingkat akurasi ringkasan berita semakin besar. Pada gambar 2.2 disajikan kerangka pikir penelitian dalam bentuk skema.



Gambar 2.2 Kerangka Pikir

## **BAB III**

### **METODOLOGI PENELITIAN**

#### **3.1 Metode Pengumpulan Data**

Metode pengumpulan data yang diterapkan dalam memperoleh data yang dibutuhkan yaitu:

##### **3.1.1 Observasi**

Observasi atau pengamatan dilakukan pada beberapa portal berita *online* berskala nasional dan lokal. Observasi pada portal berita *online* berskala nasional dilakukan pada antaranews.com., sedangkan observasi portal berita *online* berskala lokal dilakukan pada radarmagelang.com dan tribunjateng.com. Berdasarkan pengamatan yang peneliti lakukan, peneliti menemukan struktur kalimat pada portal berskala nasional lebih dominan menggunakan kalimat yang panjang, sedangkan berita pada portal berskala lokal lebih banyak menggunakan kalimat yang lebih pendek. Kalimat yang panjang memiliki informasi yang lebih lengkap jika dibandingkan dengan kalimat yang pendek. Kalimat pertama pada berita ada yang telah menggambarkan penjelasan dari judul dan ada juga yang belum menjelaskan judul berita. Berita yang digunakan sebagai uji coba diambil dari bulan Januari 2015 sampai dengan Maret 2015.

##### **3.1.2 Studi Pustaka**

Studi pustaka meliputi : pengumpulan materi yang diperoleh dari buku-buku dan jurnal-jurnal penelitian mengenai peringkasan teks otomatis

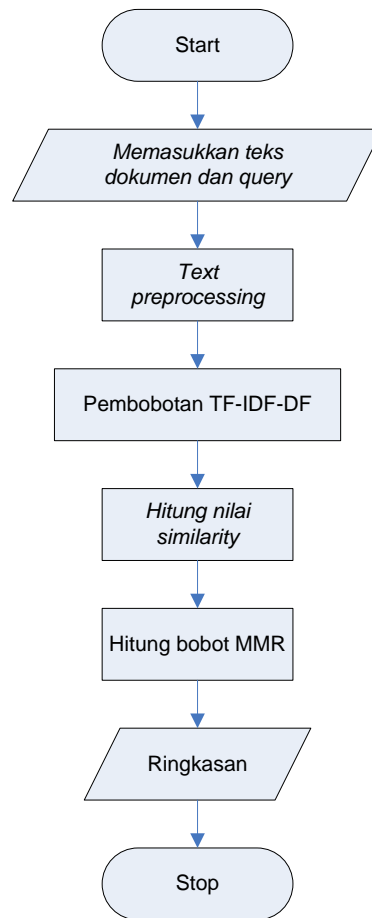
yang menggunakan metode *Maximum Marginal Relevance* (MMR) dan metode *Term Frequency-Inverse Document Frequency-Document Frequency* (TF-IDF-DF) .

## 3.2 Pengembangan Sistem

Pengembangan sistem peringkasan teks otomatis yang digunakan dalam penelitian ini melalui tahapan sebagai berikut :

### 3.2.1 Arsitektur Sistem

Sistem peringkasan dokumen berita yang dikembangkan dalam penelitian ini memiliki arsitektur seperti yang ditunjukkan pada gambar 3.1. Tahap sistem peringkasan teks dimulai dari penerimaan input berupa teks dokumen berita hingga melakukan *text preprocessing*. Hasil dari *text preprocessing* akan dihitung bobot tiap *term* (kata) dengan menggunakan algoritma TF-IDF-DF dan menghitung nilai bobot *similarity*. Tahap berikutnya menghitung bobot MMR kalimat sehingga memperoleh hasil ringkasan.



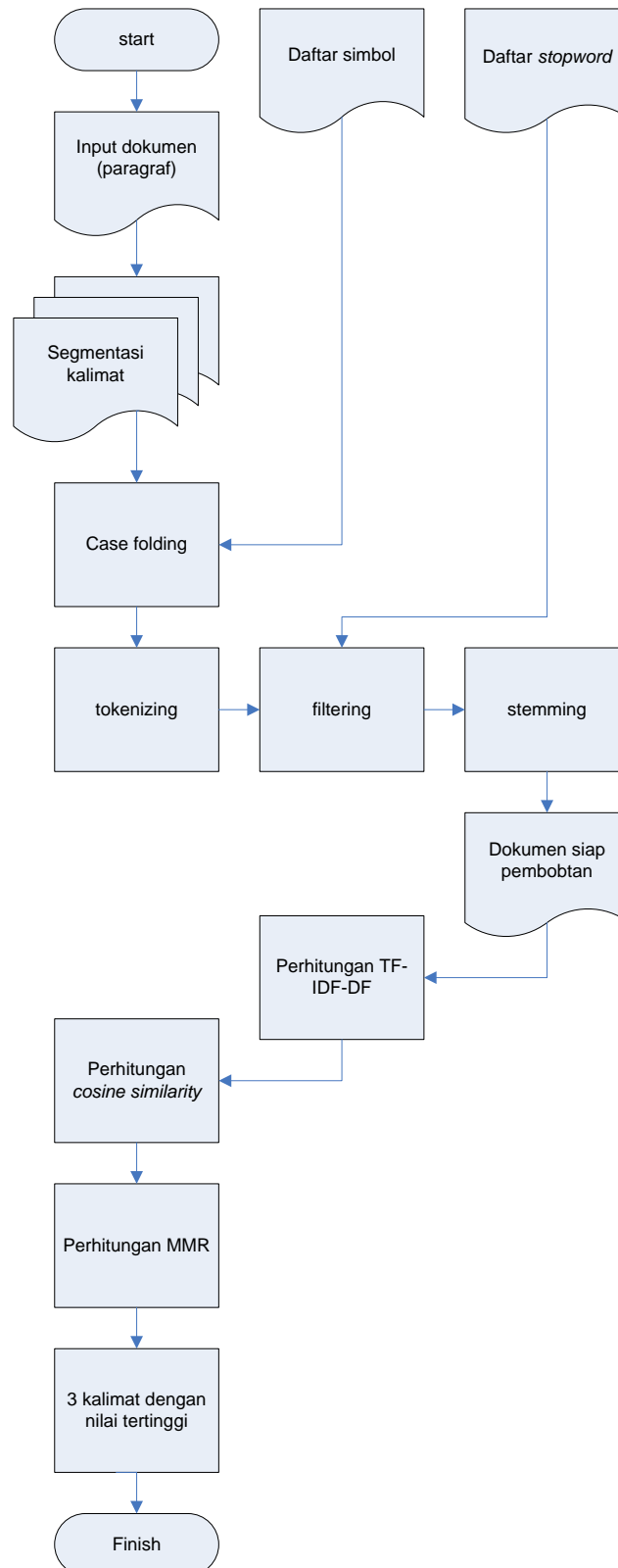
Gambar 3.1 Arsitektur sistem peringkasan teks otomatis

### 3.2.2 Desain

Desain perangkat lunak merupakan suatu proses yang terdiri dari tahapan-tahapan sehingga perangkat lunak yang dibuat berjalan sesuai dengan tujuan perangkat lunak dibangun.

#### a) Alur Proses Pengembangan Sistem

Penjabaran lebih jelas untuk rancangan alur peringkasan teks dapat dilihat pada Gambar 3.2.



Gambar 3.2 Alur Proses Pengembangan Sistem

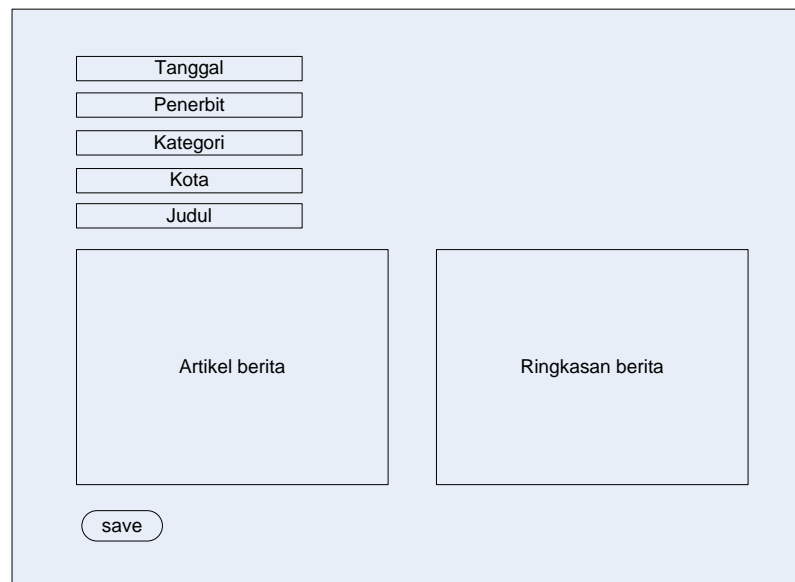
**b) Perancangan *Interface***

Berikut ini desain *interface* peringkasan teks berita dapat dilihat pada gambar 3.3 dan 3.4.



The interface for the news summarization process consists of a light blue background. At the top left, there are five stacked rectangular input fields labeled 'Tanggal', 'Penerbit', 'Kategori', 'Kota', and 'Judul'. Below these fields are two large rectangular boxes: 'Artikel berita' on the left and 'Ringkasan berita' on the right. At the bottom left, there are two rounded buttons labeled 'proses' and 'reset'.

Gambar 3.3 *Interface* proses peringkasan



The interface for saving the news summarization consists of a light blue background. At the top left, there are five stacked rectangular input fields labeled 'Tanggal', 'Penerbit', 'Kategori', 'Kota', and 'Judul'. Below these fields are two large rectangular boxes: 'Artikel berita' on the left and 'Ringkasan berita' on the right. At the bottom left, there is a single rounded button labeled 'save'.

Gambar 3.4 *Interface* menyimpan ringkasan

### 3.2.3 Pengkodean

Dalam pengkodean menggunakan bahasa Pemrograman PHP. Berikut *pseudocode* fungsi-fungsi utama dalam proses peringkasan teks berita, yaitu:

#### 1) *Text Preprocessing*

- a) Pemecahan paragraf menjadi kalimat dan pemecahan kalimat menjadi kata (*tokenizing*)

```

Input paragraf
Kalimat ← split (paragraf) dengan tanda baca titik
Ulang ← count (Kalimat)
FOR a = 0 to Ulang
    Kalimat1 ← split (Kalimat[a]) dengan tanda tanya
    Ulang1 ← count (Kalimat1)
    FOR b = 0 to Ulang1
        Kalimat2 ← split (Kalimat1[b]) dengan tanda seru
        Ulang2 ← count (Kalimat2)
        FOR c = 0 to Ulang2
            IF Kalimat2[c] ≠ " " Hasil ← Kalimat2[c] ENDIF
            Kata ← Kalimat2[c] dengan tanda spasi
            Ulang3 ← count (Kata)
            FOR d = 0 to Ulang3
                Hasil_kata ← Kata[d]
            ENDFOR
        ENDFOR
    ENDFOR
ENDFOR

```

Gambar 3.5 *Pseudocode* Pemecahan Paragraf dan *Tokenizing*



b) *Filtering*

```

Input kata
String ← kata
Ignore ← array (1,2,3,4,5,6,7,8,9,0,%,-,/)
Wordlist ← array (atas, bukan, tiap, yang, untuk, pada, ke, para,
namun, menurut, antara, dia, dua, ia, seperti, jika, jika, sehingga,
kembali, dan, tidak, ini, karena, kepada, oleh, saat, harus,
sementara, setelah, belum, kami, sekitar, bagi, serta, di, dari,
telah, sebagai, masih, hal, ketika, adalah, itu, dalam, bisa,
bahwa, atau, hanya, kita, dengan, akan, juga, ada, mereka,
sudah, saya, terhadap, secara, agar, lain, anda, begitu, mengapa,
kenapa, yaitu, yakni, daripada, itulah, lagi, maka, tentang, demi,
dimana, kemana, pula, sambil, sebelum, sesudah, supaya, guna,
kah, pun, sampai, sedangkan, selagi, sementara, tetapi, apakah,
kecuali, sebab, selain, seolah, seraya, seterusnya, tanpa, agak,
boleh, dapat, dsb, dst, dll, dahulu, dulunya, anu, demikian, tapi,
ingin, juga, nggak, mari, nanti, melainkan, oh, ok, seharusnya,
sebetulnya, setiap, setidaknya, sesuatu, pasti, saja, toh, ya,
walau, tolong, tentu, amat, apalagi, bagaimanapun)
FOREACH (Wordlist AS Word)
    Word ← preg_quote (Word, '/')
ENDFOREACH
Filter ← str_replace (Ignore, "", Text_kecil)
RETURN String

```

Gambar 3.6 *Pseudocode* Penyaringan simbol dan *stopword*

- 2) Pembobotan dengan menggunakan metode TF-IDF-DF
- a) Menghitung nilai TF dan DF

```

FOREACH collection AS docID → doc
  doc ← strtolower(doc)
  term ← explode(doc) dengan tanda spasi
  docCount[docID] ← count (terms)
  FOREACH (terms AS term)
    IF !ISSET (dictionary[term]) THEN
      DISPLAY dictionary[term] ← array( 'df' → 0, 'posting' →
      array())
    ENDIF
    IF !ISSET (dictionary[term]['posting'][docID]) THEN
      DISPLAY dictionary[term]['df']++
      DISPLAY dictionary[term]['posting'][docID] ← array ('tf'
      ← 0)
    ENDIF
    DISPLAY dictionary[term]['posting'][docID]['tf']++
  ENDFOREACH
ENDFOREACH

```

Gambar 3.7 Fungsi Penghitungan Nilai TF dan DF

## b) Menghitung nilai TF-IDF-IDF

```

docCount ← count (index['docCount'])
FOREACH (index['dictionary'] AS vocab → entry)
  FOREACH (entry['posting'] AS docID → value)
    df ← index['dictionary'][vocab]
    COMPUTE jml sebagai hasil perkalian antara value['tf'] dan
    log(count/df['df']) dan df['df']
    index['dictionary'][vocab]['posting'][docID]['tf'] → jml
  ENDFOREACH
ENDFOREACH

```

Gambar 3.8 Pseudocode Penghitungan Nilai TF-IDF-IDF

3) Perhitungan *cosine similarity*

```

Input index
a ← COMPUTE jumlah index['docCount']
FOR i = 0 to a
  FOREACH (index['docCount'] AS key → value)
    IF key = i THEN
      FOR j=1 to a
        CALL score(key, j, index['dictionary'])
      ENDFOR
    ENDIF
  ENDFOREACH
ENDFOR

```

```
Function score(docA, docB, doct) {  
    Cosine_score ← CALL cos2(docA, docB, doct)  
    DISPLAY docA dan docB dan Cosine_score  
}  
Function cos2(docA, docB, doct){  
    Result ← 0  
    FOREACH (dict AS term → entry)  
        tf[0] ← CALL cosineSim(docA, docB, doct)  
        tf[1] ← CALL cosineSim1(docA, docB, doct)  
        result ← COMPUTE pembagian tf[0] dengan tf[1]  
    ENDFOREACH  
    RETURN result  
}
```

Gambar 3.9 *Pseudocode* Penghitungan Nilai *cosine similarity*

## 4) Perhitungan MMR

```

Function cosine3(index, kalimatcoy){
  a ← COMPUTE jumlah index['docCount']
  FOR j= 0 to a
    max ← 0
    q ← 0
    FOR i = 1 to a
      hasil ← in_array(i, split_x)
      IF hasil ≠ 1 THEN
        IF j = 0 THEN
          tf[1] ← 0
        ELSE
          tf[1] ← cos2(i,k,index['dictionary'])
        ENDIF
      tf[0] ← cos2(q,i, index['dictionary'])
      result1 ← COMPUTE 0.7*tf[0]-0.3*tf[1]
      DISPLAY i dan result1
      IF result1 ≥ max THEN
        max ← result1
        l ← i
      ENDIF
    ENDIF
  ENDFOR
  DISPLAY max sebagai nilai tertinggi
}

```

Gambar 3.10 *Pseudocode* Penghitungan Nilai MMR

### 3.3 Perancangan Pengujian

Pada subbab ini akan dilakukan perancangan uji coba dari sistem peringkasan teks otomatis, baik pengujian terhadap sistem maupun evaluasi ringkasan yang dihasilkan. Hasil ringkasan akan dievaluasi berdasarkan hasil ringkasan ideal yaitu ringkasan responden / manusia.

#### 3.3.1 Bahan Pengujian

Bahan yang akan digunakan pada proses pengujian ini, yaitu 16 teks berita berbahasa Indonesia dengan rincian 8 teks berita berskala nasional dan 8 teks berita berskala lokal. Teks berita diambil secara acak dari sekumpulan berita antaranews.com, radarjateng.com dan tribunjateng.com yang diterbitkan pada Januari 2015 sampai dengan Maret 2015.

#### 3.3.2 Tujuan Pengujian

Beberapa hal yang menjadi tujuan dari pelaksanaan pengujian terhadap sistem peringkasan teks otomatis, yaitu :

1. Memeriksa perangkat lunak apakah telah berjalan baik (tidak terjadi *error*).
2. Mengevaluasi ringkasan hasil sistem dengan menghitung nilai *precision*, *recall*, dan *F-measure*.

#### 3.3.3 Skenario dan Kriteria Pengujian

Pengujian yang dilaksanakan pada skripsi ini dibagi 2 bagian, yaitu pengujian fungsionalitas perangkat lunak dan evaluasi hasil ringkasan.

### 1. Pengujian fungsionalitas perangkat lunak

Sesuai dengan tujuan pengujian pertama maka pengujian bagian pertama ini berfungsi untuk memeriksa fungsionalitas perangkat lunak.

### 2. Evaluasi hasil ringkasan

Evaluasi hasil ringkasan terhadap sistem peringkasan teks otomatis dilakukan dengan memberikan kuesioner kepada responden karena kesesuaian ringkasan tidak dapat diperoleh secara eksakta (pasti) tetapi sifatnya mendekati. Mengetahui kualitas hasil ringkasan sistem peringkasan teks dengan cara hasil ringkasan akan dibandingkan dengan ringkasan ideal (manual). Untuk evaluasi secara manual, teks berita terlebih dahulu diringkaskan oleh beberapa responden. Berikut pihak responden yang digunakan dalam penelitian ini :

- Responden : Enam orang guru Bahasa Indonesia dengan pembagian tugas tiga guru meringkas berita berskala nasional, sedangkan tiga guru yang lain meringkas berita berskala lokal.
- Metode : Para guru diberikan 8 buah artikel berita yang diambil dari portal berita *online* berskala nasional dan lokal, kemudian guru tersebut akan memilih 3 kalimat yang paling sesuai dengan judul berita.

Dari hasil ringkasan yang dilakukan para responden akan dihitung tingkat akurasinya dengan menggunakan tiga parameter, yaitu *precision*, *recall*, dan *F-measure*.

## BAB IV

### HASIL DAN PEMBAHASAN

#### 4.1 Spesifikasi *Hardware* dan *Software*

Sistem Peringkasan Teks Otomatis merupakan aplikasi yang berbasis web. Berikut adalah spesifikasi *software* yang digunakan dalam sistem ini:

##### a. Server

Server	: 127.0.0.1
Versi Apache	: Apache/2.4.10
Versi PHP	: 5.6.3
Versi MySQL	: 5.0.11
Versi phpMyadmin	: 4.2.11

##### b. Browser

Browser	: Google Chrome Version 41.0.2272.101 m
---------	-----------------------------------------

Berikut adalah spesifikasi *hardware* yang digunakan dalam sistem ini:

Prosesor : Intel® Core™ i3-2330M CPU @ 2.20GHz 2.20 GHz

RAM : 2,00 GB

*Operating System* : Windows 7 Ultimate



## 4.2 Hasil Pengujian dan Analisis

Pada bagian ini akan dijelaskan hasil pengujian dan analisis pengaruh metode MMR dan TF-IDF-DF terhadap hasil ringkasan. Pengujian dilakukan dengan membandingkan hasil ringkasan oleh sistem dengan ringkasan referensi (ringkasan manual). Dari perbandingan tersebut dapat menunjukkan nilai rata-rata *recall*, *precision* dan *F-measure*. Nilai akurasi akan semakin baik jika nilai *recall*, *precision* dan *F-measure* semakin besar.

### 4.2.1 Hasil Responden

Hasil peringkasan teks manual yang dilakukan responden dikelompokkan berdasarkan artikel berita yang bersumberkan portal berita *online* berskala nasional dan lokal. Hasil ringkasan manual ditunjukkan pada tabel 4.1 dan tabel 4.2.

Tabel 4.1 Ringkasan manual portal berita *online* berskala nasional

No	Artikel	Ringkasan Teks
1.	<p>Judul : Samsung awali tahun baru dengan TV platform baru</p> <p>Isi :</p> <p>[1] Seoul (ANTARA News) - Raksasa elektronik Korea Selatan, Samsung Electronics, mengungkapkan hari ini bahwa tahun baru ini akan meluncurkan televisi pintar berplatform baru dengan sistem operasi Tizen, demi mengurangi ketergantungan kepada Google.</p> <p>[2] Samsung mengatakan semua dari</p>	<p>[1] Seoul (ANTARA News) - Raksasa elektronik Korea Selatan, Samsung Electronics, mengungkapkan hari ini bahwa tahun baru ini akan meluncurkan televisi pintar berplatform baru dengan sistem operasi</p>

No	Artikel	Ringkasan Teks
	<p>televisi tersambung internetnya itu akan dinyawai sistem operasi Tizen, dan akan segera mengenalkan televisi pintar Tizen untuk pertama kalinya kepada konsumen pada pameran elektronik di Las Vegas pekan depan. [3]"Membangun Platform Pintar kami dengan Tizen adalah langkah terobosan ke arah sistem yang lebih pintar dan terintegrasi," kata Lee Won-Jin, wakil presiden eksekutif Samsung seperti dikutip AFP. [4]"Tizen tidak hanya memperkaya pengalaman para konsumen kami saat ini, namun juga membuka potensi besar pada masa depan <i>home entertainment</i>," kata dia. [5] Samsung, produsen ponsel pintar terbesar di dunia, telah fokus mengembangkan sistem operasi Tizen, bersama perusahaan-perusahaan pimpinan Intel untuk mengurangi ketergantungan pada sistem operasi buatan Google, Android. [6] Tizen bisa diaplikasikan pada perangkat-perangkat mobile, kamera digital dan perangkat rumahan. [7] Samsung melepas ke pasar kamera digital dan jam tangan pintar berbasis Tizen tahun ini, demikian AFP.</p> <p><b>Sumber :</b>  <a href="http://www.antaraneews.com/berita/47179">http://www.antaraneews.com/berita/47179</a></p>	<p>Tizen, demi mengurangi ketergantungan kepada Google.  . [3]"Membangun Platform Pintar kami dengan Tizen adalah langkah terobosan ke arah sistem yang lebih pintar dan terintegrasi," kata Lee Won-Jin, wakil presiden eksekutif Samsung seperti dikutip AFP.  [7] Samsung melepas ke pasar kamera digital dan jam tangan pintar berbasis Tizen tahun ini, demikian AFP.</p>

No	Artikel	Ringkasan Teks
	<a href="#">3/samsung-awali-tahun-baru-dengan-tv-plaftorm-baru</a>	
2.	<p>Judul : El Shaarawy menangkan Milan atas Real Madrid</p> <p><b>Isi :</b></p> <p>[1] Dubai (ANTARA News) - Klub sepak bola Serie A AC Milan "mengakhiri" 22 kemenangan Real Madrid pada laga Selasa, ketika striker Stephan El Shaarawy mencetak dua gol untuk kemenangan 4-2 pada laga persahabatan di Dubai. [2] Pelatih Real Carlo Ancelotti tidak menurunkan pemain bintang Gareth Bale, Karim Benzema dan James Rodriguez, tetapi Cristiano Ronaldo turun lapangan dan mencetak gol pertama pada babak awal. [3] Tim juara Eropa itu sudah menang 22 pertandingan dalam berbagai kompetisi tapi laga Selasa merupakan pertandingan pesahabatan, bukan laga resmi. [4] Real, menang Piala Klub Dunia Minggu setelah unggul 2-0 atas klub dari Argentina, San Lorenzo, tetapi El Shaarawy tampil merusak rekor klub Italia itu setelah menggoyangkan jala gawang mereka pada tiap babak, sedangkan pemain dari Prancis Jeremy</p>	<p>[1] Dubai (ANTARA News) - Klub sepak bola Serie A AC Milan "mengakhiri" 22 kemenangan Real Madrid pada laga Selasa, ketika striker Stephan El Shaarawy mencetak dua gol untuk kemenangan 4-2 pada laga persahabatan di Dubai.</p> <p>[4] Real, menang Piala Klub Dunia Minggu setelah unggul 2-0 atas klub dari Argentina, San Lorenzo, tetapi El Shaarawy tampil merusak rekor klub Italia itu setelah menggoyangkan jala gawang mereka pada tiap babak, sedangkan pemain dari Prancis Jeremy Menez dan Giampaolo Pazzini juga sebagai pemain target.</p>

No	Artikel	Ringkasan Teks
	<p>Menez dan Giampaolo Pazzini juga sebagai pemain target.</p> <p><b>Sumber :</b></p> <p><a href="http://www.antaraneews.com/berita/47158/5/el-shaarawy-menangkan-milan-atas-real-madrid">http://www.antaraneews.com/berita/47158/5/el-shaarawy-menangkan-milan-atas-real-madrid</a></p>	
3.	<p>Judul : El Shaarawy tegaskan setia di San Siro</p> <p><b>Isi :</b></p> <p>[1] Dubai (ANTARA News) - Stephan El Shaarawy menegaskan tidak ada keinginan untuk meninggalkan AC Milan dan berharap bisa membawa kesuksesan untuk klub di tahun baru. [2] Pemain 22 tahun itu ditimpa masalah cedera sepanjang 12 bulan terakhir, tetapi membuktikan bisa kembali ke puncak penampilannya dengan menggandakan kemenangan saat berhadapan dengan Real Madrid 4-2 di laga persahabatan. [3] Dengan mendekatnya jendela transfer Januari, El Shaarawy menyangkal spekulasi bahwa dirinya kemungkinan hengkang dari San Siro. [4] "Saya telah katakan, saya bertahan di Milan dan saya akan berusaha dengan baik," katanya setelah laga persahabatan. [5] "Pada pertandingan, kami melakukannya dengan baik. Itu hari yang besar untuk</p>	<p>[1] Dubai (ANTARA News) - Stephan El Shaarawy menegaskan tidak ada keinginan untuk meninggalkan AC Milan dan berharap bisa membawa kesuksesan untuk klub di tahun baru.</p> <p>[3] Dengan mendekatnya jendela transfer Januari, El Shaarawy menyangkal spekulasi bahwa dirinya kemungkinan hengkang dari San Siro. [4] "Saya telah katakan, saya bertahan di Milan dan saya akan berusaha dengan baik," katanya setelah laga persahabatan.</p>

No	Artikel	Ringkasan Teks
	<p>saya." [6]"Pertandingan seperti ini memberikan saya kepercayaan diri. [7]Gol penting untuk mental seorang penyerang, jadi saya senang mendapat dua gol melawan Madrid." [8]"Itu juga sangat memuaskan bagi tim, kami garis bawah kami melakukan dengan baik setelah pertandingan melawan Roma dan Napoli." [9] "Sekarang kami harus melanjutkan seperti ini. [10] Kami ingin memulai liga 2015 dengan baik," tambahnya lagi seperti dikutip di goal.com.</p> <p><b>Sumber :</b>  <a href="http://www.antaraneews.com/berita/471612/el-shaarawy-tegaskan-setia-di-san-siro">http://www.antaraneews.com/berita/471612/el-shaarawy-tegaskan-setia-di-san-siro</a></p>	
4.	<p>Judul : Presiden Jokowi lantik Kasau baru  Isi :  [1] Jakarta (ANTARA News) - Presiden Joko Widodo, Jumat siang, melantik Marsekal Madya Agus Supriyatna sebagai Kepala Staf Angkatan Udara menggantikan Marsekal Ida Bagus Putu Dunia. [2]Acara yang berlangsung di Istana Negara tersebut dihadiri oleh Wakil Presiden Jusuf Kalla dan pimpinan TNI serta para kepala staf angkatan. [3] Hadir juga para menteri Kabinet Kerja serta Kepala Staf Kepresidenan Luhut B</p>	<p>[1] Jakarta (ANTARA News) - Presiden Joko Widodo, Jumat siang, melantik Marsekal Madya Agus Supriyatna sebagai Kepala Staf Angkatan Udara menggantikan Marsekal Ida Bagus Putu Dunia.  [2]Acara yang berlangsung di Istana Negara tersebut dihadiri oleh Wakil Presiden</p>

No	Artikel	Ringkasan Teks
	<p>Pandjaitan. [4] Marsma Agus Supriyatna dilantik menjadi Kasau berdasarkan Keputusan Presiden nomor 1 TNI 2015. [5] Agus Supriyatna merupakan lulusan Akabri 1983 sebelum menjadi Kasau menjabat Kepala Staf Umum TNI. [6] Ia juga pernah menjabat sebagai Wakil Irjen TNI dan Pangkoopsau II.</p> <p><b>Sumber :</b>  <a href="http://www.antarane.ws.com/berita/471946/presiden-jokowi-lantik-kasau-baru">http://www.antarane.ws.com/berita/471946/presiden-jokowi-lantik-kasau-baru</a></p>	<p>Jusuf Kalla dan pimpinan TNI serta para kepala staf angkatan. [4] Marsma Agus Supriyatna dilantik menjadi Kasau berdasarkan Keputusan Presiden nomor 1 TNI 2015.</p>
5.	<p>Judul : Penguatan radar penting tingkatkan pertahanan</p> <p>Isi :</p> <p>[1] Jakarta (ANTARA News) - Kepala Staf Angkatan Udara Marsekal Madya Agus Supriyatna menilai salah satu langkah yang penting ke depan adalah memperkuat radar nasional demi meningkatkan kemampuan pertahanan.</p> <p>[2] "Semua (terkait pengembangan-red) untuk Angkatan Udara prioritas, yang terutama kita menginginkan pertahanan udara yaitu radar-radarnya," kata Agus usai dilantik Presiden Joko Widodo di Istana Negara Jakarta, Jumat. [3] Kasau mengatakan dengan kemampuan radar yang baik maka selain memaksimalkan upaya pertahanan nasional juga bisa</p>	<p>[1] Jakarta (ANTARA News) - Kepala Staf Angkatan Udara Marsekal Madya Agus Supriyatna menilai salah satu langkah yang penting ke depan adalah memperkuat radar nasional demi meningkatkan kemampuan pertahanan.</p> <p>[2] "Semua (terkait pengembangan-red) untuk Angkatan Udara prioritas, yang terutama kita menginginkan pertahanan udara yaitu</p>

No	Artikel	Ringkasan Teks
	<p>berguna dalam membantu keamanan penerbangan nasional. [4] "Begitu juga pesawat apa saja yang masuk ke wilayah kita, karena kita mempunyai radar dimana-mana, tidak akan ada (pesawat asing-red) yang masuk (tanpa ijin-red)," katanya. [5] Dengan kemampuan deteksi yang kuat, kata Agus, maka pesawat asing yang melanggar kedaulatan wilayah RI dapat segera dicegah dan diusir. [6] Selain penguatan radar, Kasau juga menyoroti pentingnya penambahan jumlah maupun kemampuan pesawat angkut yang dapat membantu operasi tidak hanya operasi terkait pertahanan negara namun juga operasi kemanusiaan terkait bencana alam maupun kecelakaan transportasi. [7] Presiden Joko Widodo melantik Marsekal Madya Agus Supriyatna sebagai Kepala Staf Angkatan Udara menggantikan Marsekal Ida Bagus Putu Dunia berdasarkan keppres nomor 1 TNI tahun 2015 yang ditandatangani Presiden pada 2 Januari 2015. [8] Sebelum menjadi Kasau, Agus Supriyatna menjabat sebagai Kepala Staf Umum TNI. [9] Agus yang lulusan Akabri 1983 tersebut juga pernah menjabat sebagai Pangkoopsau II.</p>	<p>radar-radarnya," kata Agus usai dilantik Presiden Joko Widodo di Istana Negara Jakarta, Jumat. [4] "Begitu juga pesawat apa saja yang masuk ke wilayah kita, karena kita mempunyai radar dimana-mana, tidak akan ada (pesawat asing-red) yang masuk (tanpa ijin-red)," katanya.</p>

No	Artikel	Ringkasan Teks
	<p><b>Sumber :</b>  <a href="http://www.antaraneews.com/berita/471958/penguatan-radar-penting-tingkatkan-pertahanan">http://www.antaraneews.com/berita/471958/penguatan-radar-penting-tingkatkan-pertahanan</a></p>	
6.	<p><b>Judul :</b> Harga minyak naik dalam perdagangan di tahun baru</p> <p><b>Isi :</b>  [1] Singapura (ANTARA News) - Harga minyak dunia naik di perdagangan Asia selama hari pertama Tahun Baru pada Jumat. [2] AFP Melaporkan, minyak mentah light sweet atau West Texas Intermediate (WTI) untuk pengiriman Februari, naik 56 sen menjadi 53,83 dolar AS dalam perdagangan sore. [3] Minyak mentah Brent, patokan London, untuk pengiriman Februari naik 12 sen menjadi 57,45 dolar AS. [4] Di wilayah Asia-Pasifik, pasar di Tiongkok, Jepang, Selandia Baru, Filipina, Taiwan dan Thailand ditutup pada Jumat untuk hari libur umum. [5] "Kami melihat volume tipis karena musim perayaan masih berlangsung," Daniel Ang, analis investasi pada Phillip Futures di Singapura, mengatakan kepada AFP. [6] "Keuntungan di perdagangan Asia kemungkinan karena data stok minyak mentah AS positif yang dirilis pada Rabu</p>	<p>[1] Singapura (ANTARA News) - Harga minyak dunia naik di perdagangan Asia selama hari pertama Tahun Baru pada Jumat.</p> <p>[2] AFP Melaporkan, minyak mentah light sweet atau West Texas Intermediate (WTI) untuk pengiriman Februari, naik 56 sen menjadi 53,83 dolar AS dalam perdagangan sore.</p> <p>[9] Penurunan stok minyak mentah AS biasanya menunjukkan permintaan yang kuat di konsumen minyak utama dunia itu, sehingga mendorong harga dunia naik.</p>



No	Artikel	Ringkasan Teks
	<p>(31/12)," tambah Ang. [7] Cadangan minyak mentah AS turun 1,8 juta barel dalam pekan yang berakhir 26 Desember, Badan Informasi Energi AS mengatakan dalam laporan minyak terakhirnya untuk 2014 dirilis pada Rabu. [8] Angka tersebut membalikkan ekspektasi untuk peningkatan 900.000 barel yang diproyeksikan dalam survei Bloomberg. [9] Penurunan stok minyak mentah AS biasanya menunjukkan permintaan yang kuat di konsumen minyak utama dunia itu, sehingga mendorong harga dunia naik. [10] Produksi minyak mentah AS mencapai lebih dari sembilan juta barel per hari, tingkat tertinggi dalam lebih dari 30 tahun.</p> <p><b>Sumber :</b>  <a href="http://www.antaraneews.com/berita/47198/1/harga-minyak-naik-dalam-perdagangan-di-tahun-baru">http://www.antaraneews.com/berita/47198/1/harga-minyak-naik-dalam-perdagangan-di-tahun-baru</a></p>	
7.	<p>Judul : Cisse minta maaf sudah menyikut Coleman</p> <p><b>Isi :</b></p> <p>[1] London (ANTARA News) - Penyerang Newcastle United asal Senegal Papiss Cisse mengungkapkan permintaan maafnya pada Rabu setelah dia diskors</p>	<p>[1] London (ANTARA News) - Penyerang Newcastle United asal Senegal Papiss Cisse mengungkapkan permintaan maafnya pada Rabu setelah dia</p>

No	Artikel	Ringkasan Teks
	<p>tiga pertandingan karena menyikut pemain Everton Seamus Coleman. [2] Wasit Craig Pawson tidak melihat insiden yang terjadi ketika Newcastle menang 3-2 pada pertandingan kandang di Liga Utama Inggris pada Minggu lalu, tapi Cisse menerima dakwaan melakukan pelanggaran pada Selasa setelah didakwa yang berlaku surut. [3] Cisse tidak akan bermain bagi Newcastle sebelum bertandang ke Equatorial Guinea untuk Piala Afrika dan mengungkapkan penyesalannya karena perilakunya itu melalui pesan di dalam laman Newcastle. [4] "Saya menyesal karena melakukan sesuatu seperti ini. [5] Saya bukan tipe pemain yang melakukan hal seperti itu," katanya. [6] "Itu bukan saya. [7] Saya bukan pemain yang agresif, tapi dalam sepak bola anda tidak dapat melakukan sesuatu seperti itu. [8] "Cisse, yang mencetak gol lima kali dalam enam pertandingan terakhirnya, akan absen pada pertandingan kandang hari Kamis depan melawan Burnley, pertandingan tandang putaran ketiga Piala FA pada Sabtu ke Leicester City, dan pertandingan tandang melawan Chelsea pada 10 Januari. [9] Demikian laporan AFP.</p>	<p>diskors tiga pertandingan karena menyikut pemain Everton Seamus Coleman. [3] Cisse tidak akan bermain bagi Newcastle sebelum bertandang ke Equatorial Guinea untuk Piala Afrika dan mengungkapkan penyesalannya karena perilakunya itu melalui pesan di dalam laman Newcastle. [4] "Saya menyesal karena melakukan sesuatu seperti ini.</p>

No	Artikel	Ringkasan Teks
8.	<p>Judul : Survei: pertumbuhan pekerjaan sector swasta AS Juli melambat</p> <p><b>Isi :</b></p> <p>[1]Washington (ANTARA News) - Pertumbuhan lapangan pekerjaan sektor swasta AS melambat pada bulan lalu namun masih cenderung meningkat, kata sebuah survei swasta yang dirilis Kamis.</p> <p>[2]Perusahaan swasta menambahkan 185.000 lapangan pekerjaan pada Juli, turun dari 229.000 lapangan pekerjaan pada Juni, kata Laporan Ketenagakerjaan Nasional yang dirilis bersama oleh Automatic Data Processing (ADP) dan Moodys Analytics, berdasarkan survei bulanan, lapor Xinhua.</p> <p>[3]"Pertumbuhan lapangan kerja Juli lebih lambat dari Juni, tetapi masih sejalan dengan apa yang telah kita lihat sejak awal tahun ini," kata Carlos Rodrguez, presiden dan CEO dari ADP.</p> <p>[4]"Khususnya, perusahaan besar dengan lebih dari 500 karyawan telah menambah lapangan pekerjaan terkuat mereka sejak Desember lalu dan hampir dua kali lipat jumlah Juni."</p> <p>[5]Usaha kecil dan menengah menyumbang sebagian besar kenaikan lapangan pekerjaan dalam satu bulan</p>	<p>[1]Washington (ANTARA News) - Pertumbuhan lapangan pekerjaan sektor swasta AS melambat pada bulan lalu namun masih cenderung meningkat, kata sebuah survei swasta yang dirilis Kamis.</p> <p>[4]"Khususnya, perusahaan besar dengan lebih dari 500 karyawan telah menambah lapangan pekerjaan terkuat mereka sejak Desember lalu dan hampir dua kali lipat jumlah Juni."</p> <p>[6]Lapangan pekerjaan di sektor jasa naik 178.000 pekerjaan pada Juli, turun dari kenaikan 216.000 pekerjaan pada Juni.</p>

No	Artikel	Ringkasan Teks
	<p>terakhir, menambahkan 121.000 pekerjaan, sedangkan usaha besar dengan lebih dari 500 karyawan menambahkan 64.000 pekerjaan, jauh lebih tinggi dari 34.000 pada Juni, menurut laporan tersebut.</p> <p>[6]Lapangan pekerjaan di sektor jasa naik 178.000 pekerjaan pada Juli, turun dari kenaikan 216.000 pekerjaan pada Juni.</p> <p>[7]Sementara itu, sektor manufaktur menambahkan 2.000 pekerjaan jauh lebih sedikit dari 9.000 pada Juni.</p> <p>[8]"PHK di industri energi dan kenaikan lapangan pekerjaan yang lebih lemah di bidang manufaktur berada di belakang pelambatan," kata Mark Zandi, kepala ekonom Moodys Analytics.</p> <p><b>Sumber :</b></p> <p><a href="http://www.antaraneews.com/berita/510821/survei-pertumbuhan-pekerjaan-sektor-swasta-as-juli-melambat">http://www.antaraneews.com/berita/510821/survei-pertumbuhan-pekerjaan-sektor-swasta-as-juli-melambat</a></p>	

Tabel 4.2 Ringkasan manual portal berita *online* berskala lokal

No	Artikel	Ringkasan Teks
1.	<p>Judul :DPU Magelang Janjikan Perluasan TPSA Aman</p> <p><b>Isi :</b></p> <p><b>[1] MAGELANG</b> - Dinas Pekerjaan Umum Energi dan Sumber Daya Mineral (DPU ESDM) Kabupaten Magelang, menjamin perluasan area tempat pembuangan sampah akhir (TPSA) diperbatasan Desa Pasuruhan dan Deyangan di Kecamatan Mertoyudan, tidak akan mengganggu lingkungan. [2] Pasalnya, pengelolaan sampah di sana akan dilakukan dengan sistem control landfill atau mengubur sampah. [3] Dengan sistem Landfill nanti, lanjutnya, sampah yang datang akan langsung dikubur. [4] Setelah itu, limbah air yang dikeluarkan akan ditampung dan dikelola untuk hal yang bermanfaat. [5] ”Dengan sistem ini, pencemaran udara atau lingkungan bisa diminimalisir. [6] Sejauh ini, produksi sampah setiap hari di TPSA Pasuruhan itu bisa mencapai 50 ton. [7] Sehingga sudah over kapasitas dan diperlukan penambahan ruang baru sambil menunggu realisasi TPA regional yang direncanakan baru tahun 2018</p>	<p><b>[1] MAGELANG</b></p> <p>- Dinas Pekerjaan Umum Energi dan Sumber Daya Mineral (DPU ESDM) Kabupaten Magelang, menjamin perluasan area tempat pembuangan sampah akhir (TPSA) diperbatasan Desa Pasuruhan dan Deyangan di Kecamatan Mertoyudan, tidak akan mengganggu lingkungan. [3] Dengan sistem Landfill nanti, lanjutnya, sampah yang datang akan langsung dikubur. [4] Setelah itu, limbah air yang dikeluarkan akan ditampung dan dikelola untuk hal yang bermanfaat.7</p>

No	Artikel	Ringkasan Teks
	<p>mendatang,” kata Kepala DPU ESDM Sutarno, Jumat (23/01/2015). [8] “Sistem ini sudah banyak dikembangkan di daerah lain. [9] Dengan sistem ini, sudah banyak yang berhasil meminimalisir pencemaran lingkungan. [10] Selama ini, pembuangan sampah masih menggunakan sistem ‘open dumping’. [11] Sehingga muncul pencemaran lingkungan sekitar TPSA. [12]Datangnya penghujan, juga menyebabkan meningkatnya jumlah cairan dari tumpukan sampah,” imbuhnya. [13] Sebelumnya diberitakan, warga Pasuruhan dan Deyangan yang tergabung dalam Forum Pemuda Deyangan yang terdampak TPSA Pasuruhan, menggelar aksi unjuk rasa Minggu (18/01/2015). [14] Mereka menolak dilakukan perluasan pembuangan sampah itu. [15]Alasannya, warga akan semakin terganggu aktifitasnya.</p> <p><b>Sumber :</b>  <a href="http://radarmagelang.com/dpu-magelang-janjikan-perluasan-tpsa-aman/">http://radarmagelang.com/dpu-magelang-janjikan-perluasan-tpsa-aman/</a></p>	

No	Artikel	Ringkasan Teks
2.	<p>Judul : Petani Malaysia Studi Banding Ke Petani Salak Magelang</p> <p>Isi :</p> <p>[1] <b>MAGELANG</b>– Perwakilan petani dari Malaysia melakukan kunjungan kerja ke Desa Kaliurang, Kecamatan Srumbung, Kabupaten Magelang untuk belajar bertani salak. [2] Puluhan petani dari negeri Jiran itu tiba di Desa Kaliurang bersama para pengusaha bidang pertanian Malaysia, Kamis (22/1/2015). [3] Mereka tampak antusias melihat proses panen buah salak hingga pengepakannya. [4]Nina (42), salah satu petani Malaysia mengaku kagum dengan salah satu buah eksotis asli Indonesia itu. [5] Nina mengatakan bahwa salak Nglumut lebih khas dan segar dibanding salak Malaysia. [6] “Tekstur daging salak di sini lebih keras sehingga rasanya juga lebih khas. [7] Mungkin karena ditanam secara organik. [8] Tapi salak di negara kami sedikit lebih besar dari salak ini,” jelas Nina di sela-sela kunjungan. [9] Para petani Malaysia tidak hanya kagum pada kualitas salak Nglumut, tapi juga soal manajemen organisasi petaninya yang terwadahi dalam Gabungan Kelompok Tani (Gapoktan) Ngudi Luhur,</p>	<p>[1] <b>MAGELANG</b>– Perwakilan petani dari Malaysia melakukan kunjungan kerja ke Desa Kaliurang, Kecamatan Srumbung, Kabupaten Magelang untuk belajar bertani salak.</p> <p>[2] Puluhan petani dari negeri Jiran itu tiba di Desa Kaliurang bersama para pengusaha bidang pertanian Malaysia, Kamis (22/1/2015).</p> <p>[4]Nina (42), salah satu petani Malaysia mengaku kagum dengan salah satu buah eksotis asli Indonesia itu.</p>

No	Artikel	Ringkasan Teks
	<p>Desa Kaliurang. [10] Pimpinan rombongan, Mokhdzir bin Hj Lias yang juga mantan pejabat di Kementrian Pertanian Malaysia, menyatakan bahwa kunjungannya ketempat ini, dalam rangka ziarah tani atau kunjungan kerja.</p> <p>[11]“Selain silaturahmi karena beberapa dari leluhur kami berasal dari Indonesia, kedatangan kami ke sini adalah untuk belajar. [12]Terus terang, kami terkesan dan kagum dengan kemajuan pertanian disini,” katanya.</p> <p>Sumber:  <a href="http://radarmagelang.com/petani-malaysia-studi-banding-ke-petani-salak-magelang/">http://radarmagelang.com/petani-malaysia-studi-banding-ke-petani-salak-magelang/</a></p>	
3.	<p>Judul : Polres Magelang Tangkap Bandar Ganja</p> <p><b>Isi :</b></p> <p><b>[1] MAGELANG</b> - Kepolisian Resor (Polres) Magelang menangkap Prayoga, seorang pengedar ganja yang telah lama menjadi buronan. [2] Pria berusia 28 tahun warga kampung Kebondalemn II, RT 1/3 Kelurahan Potrobangsari, Magelang Utara itu sempat melakukan perlawanan dan menolak untuk ditangkap jajaran Satuan Narkoba Polres Magelang.[3] Pengedar ganja tersebut</p>	<p><b>[1] MAGELANG</b> - Kepolisian Resor (Polres) Magelang menangkap Prayoga, seorang pengedar ganja yang telah lama menjadi buronan. [3] Pengedar ganja tersebut ditangkap polisi di Jalan Buton Cemara 7, Kelurahan Kedungsari, Magelang Utara, Kota Magelang,</p>



No	Artikel	Ringkasan Teks
	<p>ditangkap polisi di Jalan Buton Cemara 7, Kelurahan Kedungsari, Magelang Utara, Kota Magelang, Jumat (16/1/2015) malam sekitar pukul 21.30 WIB. [4]Saat itu, dia sedang bertransaksi ganja. [5] Prayoga sempat menyebut polisi melanggar hak asasi manusia (HAM) saat ditangkap.[6]Namun, Prayoga baru tidak berlutut setelah polisi menggeledah celananya dan menemukan ganja. [7] Dia tertangkap basah saat sedang transaksi. [8] Bahkan, anggota polisi sempat melepaskan tembakan peringatan agar Daftar Pencarian Orang (DPO) ganja itu menyerah.[9] “Barang bukti yang ditemukan diantaranya ganja satu garis, satu paket besar dan tiga linting serta uang Rp4.250.000 yang diduga hasil jual beli ganja,” jelas Kapolres Magelang, AKBP Rifki, kemarin.[10] Kasat Narkoba, AKP Eko Sumbodo mengatakan, total ganja yang dibawa Paroyogo seberat 30,65 gram. [11]Menurutnya, Prayoga menerima pasokan ganja seberat 1 kilogram per transaksi dari bandar ganja Jakarta.[12] “Dia tidak hanya mengedarkan namun juga pemakai ganja. [13] Hal itu berdasarkan hasil THC Rapid Test Cassete dia positif memakai,” papar</p>	<p>Jumat (16/1/2015) malam sekitar pukul 21.30 WIB. [10] Kasat Narkoba, AKP Eko Sumbodo mengatakan, total ganja yang dibawa Paroyogo seberat 30,65 gram.</p>

No	Artikel	Ringkasan Teks
	<p>Eko.[14] Menurut Eko selain ganja dan uang hasil transaksi, polisi juga mengamankan dua ponsel Blackberry dan Honda Beat. [15] Saat ini, tersangka Prayogo dilimpahkan ke Polres Magelang Kota karena lokasi transaksi di Kota Magelang.</p> <p><b>Sumber :</b>  <a href="http://radarmagelang.com/polres-magelang-tangkap-bandar-ganja/">http://radarmagelang.com/polres-magelang-tangkap-bandar-ganja/</a></p>	
4.	<p><b>Judul :</b> Tiga Bulan Terakhir Pupuk Subsidi Di Magelang Sulit Didapat</p> <p>Isi :</p> <p><b>[1] MAGELANG -</b> Puluhan petani di Desa Sucen, Kecamatan Srumbung, Kabupaten Magelang mengaku resah karena kesulitan untuk mendapatkan pupuk bersubsidi. [2] Kondisi ini sudah terjadi selama tiga bulan terakhir. [3] Akibatnya, petani terpaksa membeli pupuk non subsidi karena diduga ada permainan oleh pengecer. [4]Ketua Gabungan Kelompok Tani (Gapoktan) “Succes Abadi” Desa Sucen, Fatkhul Mujib menjelaskan, pihaknya sejauh ini sudah mengajukan rencana definitif kebutuhan pupuk bersubsidi. [5]Namun pada kenyataanya petani tetap kesulitan</p>	<p><b>[1] MAGELANG -</b> Puluhan petani di Desa Sucen, Kecamatan Srumbung, Kabupaten Magelang mengaku resah karena kesulitan untuk mendapatkan pupuk bersubsidi. [2] Kondisi ini sudah terjadi selama tiga bulan terakhir. [3] Akibatnya, petani terpaksa membeli pupuk non subsidi karena diduga ada permainan oleh pengecer.</p>

No	Artikel	Ringkasan Teks
	<p>mendapatkan pupuk bersubsidi. [6]  “Sehingga untuk mendapatkan pupuk, petani harus membayar harga non subsidi,” katanya, Jumat (23/1). [7]  Mujib menjelaskan, biasanya gapoktan diberitahu oleh penyalur kapan pupuk turun dan kemudian gapoktan menginformasikan kepada kelompok tani untuk menebus sesuai jatah yang disepakati bersama. [8] Hanya saja, akhir-akhir ini petani tidak diberitahu lagi oleh penyalur, padahal pupuk tersebut telah didistribusi oleh distributor selama dua kali.</p> <p><b>Sumber :</b></p> <p><b><a href="http://radarmagelang.com/tiga-bulan-terakhir-pupuk-subsidi-di-magelang-sulit-didapat/">http://radarmagelang.com/tiga-bulan-terakhir-pupuk-subsidi-di-magelang-sulit-didapat/</a></b></p>	
5.	<p>Judul : Polrestabes Masih Fokus Gali Keterangan dari 12 Orang</p> <p><b>Isi :</b></p> <p>[1] TRIBUNJATENG.COM, SEMARANG - Kapolrestabes Semarang, Kombes Pol Djihartono, mengatakan pihaknya saat ini masih fokus memintai keterangan 12 orang yang sebelumnya telah dimintai keterangan. [2] "Belum ada tambahan, masih 12 orang yang kami</p>	<p>[1]  TRIBUNJATENG.CO  M, SEMARANG  - Kapolrestabes  Semarang, Kombes Pol  Djihartono, mengatakan  pihaknya saat ini masih  fokus memintai  keterangan 12 orang</p>

No	Artikel	Ringkasan Teks
	<p>mintai keterangan," kata Djihartono, Kamis (26/3/2015). [3]12 orang yang dimintai keterangan itu terdiri dari DPKAD Kota Semarang, pihak BTPN dan AU. [4] Terkait SH selaku Kepala UPTD Kasda Kota Semarang, Djihartono mengatakan pihaknya belum menetapkan tersangka. [5] "Belum ada tersangka, masih pemeriksaan semua. [6] Nanti kami ekspose semua kalau sudah ada progres baru," kata Djihartono.</p> <p><b>Sumber :</b></p> <p><a href="http://jateng.tribunnews.com/2015/03/26/polrestabes-masih-fokus-gali-keterangan-dari-12-orang">http://jateng.tribunnews.com/2015/03/26/polrestabes-masih-fokus-gali-keterangan-dari-12-orang</a></p>	<p>yang sebelumnya telah dimintai keterangan.</p> <p>[2] "Belum ada tambahan, masih 12 orang yang kami mintai keterangan," kata Djihartono, Kamis (26/3/2015).</p> <p>[4] Terkait SH selaku Kepala UPTD Kasda Kota Semarang, Djihartono mengatakan pihaknya belum menetapkan tersangka.</p>
6.	<p><b>Judul :</b> Rumah Bripka Cahyo Polisi Pekalongan Dibobol Maling</p> <p><b>Isi :</b></p> <p>[1] TRIBUNJATENG.COM, PEKALONGAN - Tanpa pandang bulu, pencurian dapat menimpa siapa pun tak terkecuali penegak hukum. [2] Seorang anggota Polres Pekalongan, saat terlelap tidur, rumahnya disatroni maling pada hari Kamis (26/3) dini hari. [3]Kasubbag Humas Polres Pekalongan, AKP Gunthur Tri Harjanti mengakui rumah yang</p>	<p>[1] TRIBUNJATENG.COM, PEKALONGAN - Tanpa pandang bulu, pencurian dapat menimpa siapa pun tak terkecuali penegak hukum.</p> <p>[2] Seorang anggota Polres Pekalongan, saat terlelap tidur, rumahnya disatroni maling pada</p>

No	Artikel	Ringkasan Teks
	<p>dibobol pencuri itu adalah milik Bripka Cahyo A Dinata (36) warga Desa Langensari, Kecamatan Kesesi, Kabupaten Pekalongan. [4] "Saat korban dan istrinya terbangun pada subuh sekitar pukul 05.00. [5] Sejumlah barang sudah hilang," katanya, Kamis (26/3). [6] Setelah mengetahui kejadian tersebut, korban kemudian melaporkannya ke Polsek Kesesi. [7] Hingga saat ini, kepolisian masih melakukan penyelidikan. [8] "Rumah korban memang sepi, sebelah kanan dan kirinya kebun tebu. [9]Kemungkinan pelaku masuk dari belakang atau depan," terangnya.</p> <p><b>Sumber :</b></p> <p><a href="http://jateng.tribunnews.com/2015/03/26/rumah-bripka-cahyo-polisi-pekalongan-dibobol-maling">http://jateng.tribunnews.com/2015/03/26/rumah-bripka-cahyo-polisi-pekalongan-dibobol-maling</a></p>	<p>hari Kamis (26/3) dini hari.</p> <p>[3]Kasubbag Humas Polres Pekalongan, AKP Gunthur Tri Harjanti mengakui rumah yang dibobol pencuri itu adalah milik Bripka Cahyo A Dinata (36) warga Desa Langensari, Kecamatan Kesesi, Kabupaten Pekalongan.</p>
7.	<p><b>Judul :</b> PT KAI Daop IV Kini Dipimpin Apriyono Wedi Chresnanto</p> <p><b>Isi :</b></p> <p>[1] TRIBUNJATENG.COM, SEMARANG - Kepala daerah operasi (Daop) PT KAI 4 Semarang diganti. [2] Pejabat sebelumnya, Wawan Ariyanto</p>	<p>[2] Pejabat sebelumnya, Wawan Ariyanto digantikan pejabat baru, Apriyono Wedi Chresnanto.</p> <p>[5] Sedangkan pejabat pegantinya, Apriyono Wedi Chresnanto</p>

No	Artikel	Ringkasan Teks
	<p>digantikan pejabat baru, Apriyono Wedi Chresnanto. [3]Serah terima jabatan dilakukan pada Rabu (25/3/2015).</p> <p>[4]Manajer humas PT kAI Daop IV, Suprpto, mengatakan Wawan Ariyanto pindah tugas ke wilayah PT KAI Divre 3 Sumatera Selatan. [5] Sedangkan pejabat pegantinya, Apriyono Wedi Chresnanto sebelumnya menjabat sebagai kepala daerah operasi 3 Cirebon. [6]"Diharapkan dengan pergantian pimpinan ini, program-program kerja PT KAI Daop 4 Semarang akan terus berjalan dan bisa segera terealisasi," tuturnya dalam rilis yang diterima tribun. [7] Adapun program-program kerja yang akan diteruskan oleh EVP PT KAI Daop 4 semarang yang baru di antaranya pelaksanaan reaktivasi jalur Kedungjati - Tuntang, reaktivasi Semarang Tawang - Pelabuhan Tanjung Emas, pengembangan revitalisasi stasiun Semarang Tawang dan Semarang Poncol, serta program-program strategis pengembangan perkeretaapian lainnya.</p> <p><b>Sumber :</b></p>	<p>sebelumnya menjabat sebagai kepala daerah operasi 3 Cirebon.</p> <p>[6]"Diharapkan dengan pergantian pimpinan ini, program-program kerja PT KAI Daop 4 Semarang akan terus berjalan dan bisa segera terealisasi," tuturnya dalam rilis yang diterima tribun.</p>

No	Artikel	Ringkasan Teks
	<a href="http://jateng.tribunnews.com/2015/03/26/pt-kai-daop-iv-kini-dipimpin-apriyono-wedi-chresnanto">http://jateng.tribunnews.com/2015/03/26/pt-kai-daop-iv-kini-dipimpin-apriyono-wedi-chresnanto</a>	
8.	<p><b>Judul</b> : Sidang Pembacaan Tuntutan Bripka Lulus Rahardi Ditunda</p> <p><b>Isi</b> :</p> <p>[1] TRIBUNJATENG.COM, KUDUS - Sidang lanjutan terhadap oknum Polres Kudus, Bripka Lulus Rahardi, dilanjutkan pada Rabu (25/3) siang. [2] Sekitar pukul 12.00 WIB sidang dengan agenda pembacaan tuntutan pun dimulai. [3] Lulus pun duduk di kursi pesakitan depan majelis hakim yang diketuai Rudi Ananta Wijaya dan dua hakim anggota Ikha Tina dan Edwin Pudyono Marwiyanto. [4] Dengan memakai rompi berwarna oranye dan peci hitam, dia langsung menundukkan kepala. [5] Hakim ketua kemudian mempersilahkan Jaksa Penuntut Umum dari Kejari Kudus untuk membacakanuntutannya. [6] JPU Kharis Rohman mengatakan belum siap dengan tuntutan. [7] Dia pun meminta waktu sepekan untuk menyiapkan tuntutan. [8] "JPU belum siap membacakan tuntutan karena belum konfirmasi ke Kejati Jawa Tengah. [9] Sehingga, sidang</p>	<p>[6] JPU Kharis Rohman mengatakan belum siap dengan tuntutan. [8] "JPU belum siap membacakan tuntutan karena belum konfirmasi ke Kejati Jawa Tengah. [9] Sehingga, sidang ditunda pekan depan.</p>

No	Artikel	Ringkasan Teks
	<p>ditunda pekan depan. [10] Untuk terdakwa, tetap berada dalam tahanan," kata hakim ketua, Rudi Ananta.</p> <p><b>Sumber :</b></p> <p><a href="http://jateng.tribunnews.com/2015/03/25/sidang-pembacaan-tuntutan-bripka-lulus-rahardi-ditunda">http://jateng.tribunnews.com/2015/03/25/sidang-pembacaan-tuntutan-bripka-lulus-rahardi-ditunda</a></p>	

Pada tabel 4.1 dan tabel 4.2 merupakan salah satu hasil ringkasan yang dilakukan oleh 6 responden. Para responden disuguhkan dengan artikel berita sebanyak 8 artikel. Pada setiap artikel berita terdapat judul dan isi berita berupa kumpulan kalimat yang telah diberi nomor pada setiap kalimat. Hasil ringkasan referensi (ringkasan manual) diperoleh dengan memilih 3 kalimat yang paling sesuai dengan judul. Dari contoh hasil peringkasan teks manual yang dilakukan oleh responden, artikel berita yang diambil dari portal berita *online* berskala nasional memperoleh ringkasan yang salah satunya adalah kalimat 1.

#### 4.2.2 Hasil Sistem

Peringkasan teks dengan sistem dilakukan dengan cara menginput *query* berupa judul berita dan isi berita ke dalam sistem. Selanjutnya, sistem memproses inputan judul dan isi berita, sehingga sistem akan mengeluarkan hasil ringkasan dengan menampilkan maksimal 3 kalimat yang memiliki nilai



MMR tertinggi. Salah satu hasil peringkasan teks otomatis yang dilakukan oleh sistem terlihat pada tabel 4.3.

Tabel 4.3 Ringkasan sistem portal berita *online* berskala nasional

No	Artikel	Ringkasan Teks
1.	<p>Judul : Samsung awali tahun baru dengan TV platform baru</p> <p>Isi :</p> <p>[1] Seoul (ANTARA News) - Raksasa elektronik Korea Selatan, Samsung Electronics, mengungkapkan hari ini bahwa tahun baru ini akan meluncurkan televisi pintar berplatform baru dengan sistem operasi Tizen, demi mengurangi ketergantungan kepada Google.</p> <p>[2] Samsung mengatakan semua dari televisi tersambung internetnya itu akan dinyawai sistem operasi Tizen, dan akan segera mengenalkan televisi pintar Tizen untuk pertama kalinya kepada konsumen pada pameran elektronik di Las Vegas pekan depan. [3]"Membangun Platform Pintar kami dengan Tizen adalah langkah terobosan ke arah sistem yang lebih pintar dan terintegrasi," kata Lee Won-Jin, wakil presiden eksekutif Samsung seperti dikutip AFP. [4]"Tizen tidak hanya memperkaya pengalaman para konsumen kami saat ini, namun juga</p>	<p>[1] Seoul (ANTARA News) - Raksasa elektronik Korea Selatan, Samsung Electronics, mengungkapkan hari ini bahwa tahun baru ini akan meluncurkan televisi pintar berplatform baru dengan sistem operasi Tizen, demi mengurangi ketergantungan kepada Google.</p> <p>[7] Samsung melepas ke pasar kamera digital dan jam tangan pintar berbasis Tizen tahun ini, demikian AFP.</p>

No	Artikel	Ringkasan Teks
	<p>membuka potensi besar pada masa depan <i>home entertainment</i>," kata dia. [5] Samsung, produsen ponsel pintar terbesar di dunia, telah fokus mengembangkan sistem operasi Tizen, bersama perusahaan-perusahaan pimpinan Intel untuk mengurangi ketergantungan pada sistem operasi buatan Google, Android. [6] Tizen bisa diaplikasikan pada perangkat-perangkat mobile, kamera digital dan perangkat rumahan. [7] Samsung melepas ke pasar kamera digital dan jam tangan pintar berbasis Tizen tahun ini, demikian AFP.</p> <p><b>Sumber :</b>  <a href="http://www.antaraneews.com/berita/471793/samsung-awali-tahun-baru-dengan-tv-plaftorm-baru">http://www.antaraneews.com/berita/471793/samsung-awali-tahun-baru-dengan-tv-plaftorm-baru</a></p>	
2.	<p>Judul : El Shaarawy menangkan Milan atas Real Madrid</p> <p><b>Isi :</b></p> <p>[1] Dubai (ANTARA News) - Klub sepak bola Serie A AC Milan "mengakhiri" 22 kemenangan Real Madrid pada laga Selasa, ketika striker Stephan El Shaarawy mencetak dua gol untuk kemenangan 4-2 pada laga persahabatan di Dubai. [2] Pelatih Real</p>	<p>[1] Dubai (ANTARA News) - Klub sepak bola Serie A AC Milan "mengakhiri" 22 kemenangan Real Madrid pada laga Selasa, ketika striker Stephan El Shaarawy mencetak dua gol untuk kemenangan 4-2 pada</p>

No	Artikel	Ringkasan Teks
	<p>Carlo Ancelotti tidak menurunkan pemain bintang Gareth Bale, Karim Benzema dan James Rodriguez, tetapi Cristiano Ronaldo turun lapangan dan mencetak gol pertama pada babak awal. [3] Tim juara Eropa itu sudah menang 22 pertandingan dalam berbagai kompetisi tapi laga Selasa merupakan pertandingan pesahabatan, bukan laga resmi. [4] Real, menang Piala Klub Dunia Minggu setelah unggul 2-0 atas klub dari Argentina, San Lorenzo, tetapi El Shaarawy tampil merusak rekor klub Italia itu setelah menggoyangkan jala gawang mereka pada tiap babak, sedangkan pemain dari Prancis Jeremy Menez dan Giampaolo Pazzini juga sebagai pemain target.</p> <p><b>Sumber :</b>  <a href="http://www.antaraneews.com/berita/47158/5/el-shaarawy-menangkan-milan-atas-real-madrid">http://www.antaraneews.com/berita/47158/5/el-shaarawy-menangkan-milan-atas-real-madrid</a></p>	<p>laga persahabatan di Dubai.</p> <p>[4] Real, menang Piala Klub Dunia Minggu setelah unggul 2-0 atas klub dari Argentina, San Lorenzo, tetapi El Shaarawy tampil merusak rekor klub Italia itu setelah menggoyangkan jala gawang mereka pada tiap babak, sedangkan pemain dari Prancis Jeremy Menez dan Giampaolo Pazzini juga sebagai pemain target.</p> <p>[3] Tim juara Eropa itu sudah menang 22 pertandingan dalam berbagai kompetisi tapi laga Selasa merupakan pertandingan pesahabatan, bukan laga resmi.</p>
3.	<p>Judul : El Shaarawy tegaskan setia di San Siro</p> <p><b>Isi :</b></p>	<p>[3] Dengan mendekatnya jendela transfer Januari, El Shaarawy menyangkal</p>

No	Artikel	Ringkasan Teks
	<p>[1] Dubai (ANTARA News) - Stephan El Shaarawy menegaskan tidak ada keinginan untuk meninggalkan AC Milan dan berharap bisa membawa kesuksesan untuk klub di tahun baru. [2] Pemain 22 tahun itu ditimpa masalah cedera sepanjang 12 bulan terakhir, tetapi membuktikan bisa kembali ke puncak penampilannya dengan menggandakan kemenangan saat berhadapan dengan Real Madrid 4-2 di laga persahabatan. [3] Dengan mendekatnya jendela transfer Januari, El Shaarawy menyangkal spekulasi bahwa dirinya kemungkinan hengkang dari San Siro. [4] "Saya telah katakan, saya bertahan di Milan dan saya akan berusaha dengan baik," katanya setelah laga persahabatan. [5] "Pada pertandingan, kami melakukannya dengan baik. Itu hari yang besar untuk saya." [6] "Pertandingan seperti ini memberikan saya kepercayaan diri. [7] Gol penting untuk mental seorang penyerang, jadi saya senang mendapat dua gol melawan Madrid." [8] "Itu juga sangat memuaskan bagi tim, kami garis bawahi kami melakukan dengan baik setelah pertandingan melawan Roma dan Napoli." [9] "Sekarang kami harus melanjutkan seperti ini. [10] Kami ingin</p>	<p>spekulasi bahwa dirinya kemungkinan hengkang dari San Siro.</p> <p>[1] Dubai (ANTARA News) - Stephan El Shaarawy menegaskan tidak ada keinginan untuk meninggalkan AC Milan dan berharap bisa membawa kesuksesan untuk klub di tahun baru.</p>

No	Artikel	Ringkasan Teks
	<p>memulai liga 2015 dengan baik," tambahnya lagi seperti dikutip di goal.com.</p> <p><b>Sumber :</b></p> <p><a href="http://www.antaraneews.com/berita/471612/el-shaarawy-tegaskan-setia-di-san-siro">http://www.antaraneews.com/berita/471612/el-shaarawy-tegaskan-setia-di-san-siro</a></p>	
4.	<p>Judul : Presiden Jokowi lantik Kasau baru</p> <p>Isi :</p> <p>[1] Jakarta (ANTARA News) - Presiden Joko Widodo, Jumat siang, melantik Marsekal Madya Agus Supriyatna sebagai Kepala Staf Angkatan Udara menggantikan Marsekal Ida Bagus Putu Dunia. [2]Acara yang berlangsung di Istana Negara tersebut dihadiri oleh Wakil Presiden Jusuf Kalla dan pimpinan TNI serta para kepala staf angkatan. [3] Hadir juga para menteri Kabinet Kerja serta Kepala Staf Kepresidenan Luhut B Pandjaitan. [4] Marsma Agus Supriyatna dilantik menjadi Kasau berdasarkan Keputusan Presiden nomor 1 TNI 2015. [5]Agus Supriyatna merupakan lulusan Akabri 1983 sebelum menjadi Kasau menjabat Kepala Staf Umum TNI. [6] Ia juga pernah menjabat sebagai Wakil Irjen TNI dan Pangkoopsau II.</p>	<p>[4] Marsma Agus Supriyatna dilantik menjadi Kasau berdasarkan Keputusan Presiden nomor 1 TNI 2015.</p> <p>[1] Jakarta (ANTARA News) - Presiden Joko Widodo, Jumat siang, melantik Marsekal Madya Agus Supriyatna sebagai Kepala Staf Angkatan Udara menggantikan Marsekal Ida Bagus Putu Dunia.</p> <p>[3] Hadir juga para menteri Kabinet Kerja serta Kepala Staf Kepresidenan Luhut B Pandjaitan.</p>

No	Artikel	Ringkasan Teks
	<p><b>Sumber :</b>  <a href="http://www.antaraneews.com/berita/471946/presiden-jokowi-lantik-kasau-baru">http://www.antaraneews.com/berita/471946/presiden-jokowi-lantik-kasau-baru</a></p>	
5.	<p>Judul : Penguatan radar penting tingkatkan pertahanan</p> <p>Isi :</p> <p>[1] Jakarta (ANTARA News) - Kepala Staf Angkatan Udara Marsekal Madya Agus Supriyatna menilai salah satu langkah yang penting ke depan adalah memperkuat radar nasional demi meningkatkan kemampuan pertahanan.</p> <p>[2] "Semua (terkait pengembangan-red) untuk Angkatan Udara prioritas, yang terutama kita menginginkan pertahanan udara yaitu radar-radarnya," kata Agus usai dilantik Presiden Joko Widodo di Istana Negara Jakarta, Jumat. [3] Kasau mengatakan dengan kemampuan radar yang baik maka selain memaksimalkan upaya pertahanan nasional juga bisa berguna dalam membantu keamanan penerbangan nasional. [4] "Begitu juga pesawat apa saja yang masuk ke wilayah kita, karena kita mempunyai radar dimana-mana, tidak akan ada (pesawat asing-red) yang masuk (tanpa ijin-red)," katanya. [5] Dengan kemampuan deteksi yang kuat, kata Agus, maka pesawat</p>	<p>[1] Jakarta (ANTARA News) - Kepala Staf Angkatan Udara Marsekal Madya Agus Supriyatna menilai salah satu langkah yang penting ke depan adalah memperkuat radar nasional demi meningkatkan kemampuan pertahanan.</p> <p>[6] Selain penguatan radar, Kasau juga menyoroti pentingnya penambahan jumlah maupun kemampuan pesawat angkut yang dapat membantu operasi tidak hanya operasi terkait pertahanan negara namun juga operasi kemanusiaan terkait bencana alam maupun kecelakaan transportasi.</p>

No	Artikel	Ringkasan Teks
	<p>asing yang melanggar kedaulatan wilayah RI dapat segera dicegah dan diusir.</p> <p>[6]Selain penguatan radar, Kasau juga menyoroti pentingnya penambahan jumlah maupun kemampuan pesawat angkut yang dapat membantu operasi tidak hanya operasi terkait pertahanan negara namun juga operasi kemanusiaan terkait bencana alam maupun kecelakaan transportasi. [7] Presiden Joko Widodo melantik Marsekal Madya Agus Supriyatna sebagai Kepala Staf Angkatan Udara menggantikan Marsekal Ida Bagus Putu Dunia berdasarkan keppres nomor 1 TNI tahun 2015 yang ditandatangani Presiden pada 2 Januari 2015. [8] Sebelum menjadi Kasau, Agus Supriyatna menjabat sebagai Kepala Staf Umum TNI. [9] Agus yang lulusan Akabri 1983 tersebut juga pernah menjabat sebagai Pangkoopsau II.</p> <p><b>Sumber :</b></p> <p><a href="http://www.antaraneews.com/berita/471958/penguatan-radar-penting-tingkatkan-pertahanan">http://www.antaraneews.com/berita/471958/penguatan-radar-penting-tingkatkan-pertahanan</a></p>	<p>[3] Kasau mengatakan dengan kemampuan radar yang baik maka selain memaksimalkan upaya pertahanan nasional juga bisa berguna dalam membantu keamanan penerbangan nasional.</p>

No	Artikel	Ringkasan Teks
6.	<p><b>Judul :</b> Harga minyak naik dalam perdagangan di tahun baru</p> <p><b>Isi :</b></p> <p>[1] Singapura (ANTARA News) - Harga minyak dunia naik di perdagangan Asia selama hari pertama Tahun Baru pada Jumat. [2] AFP Melaporkan, minyak mentah light sweet atau West Texas Intermediate (WTI) untuk pengiriman Februari, naik 56 sen menjadi 53,83 dolar AS dalam perdagangan sore. [3] Minyak mentah Brent, patokan London, untuk pengiriman Februari naik 12 sen menjadi 57,45 dolar AS. [4] Di wilayah Asia-Pasifik, pasar di Tiongkok, Jepang, Selandia Baru, Filipina, Taiwan dan Thailand ditutup pada Jumat untuk hari libur umum. [5] "Kami melihat volume tipis karena musim perayaan masih berlangsung," Daniel Ang, analis investasi pada Phillip Futures di Singapura, mengatakan kepada AFP. [6] "Keuntungan di perdagangan Asia kemungkinan karena data stok minyak mentah AS positif yang dirilis pada Rabu (31/12)," tambah Ang. [7] Cadangan minyak mentah AS turun 1,8 juta barel dalam pekan yang berakhir 26 Desember, Badan Informasi Energi AS mengatakan dalam laporan minyak terakhirnya untuk</p>	<p>[1] Singapura (ANTARA News) - Harga minyak dunia naik di perdagangan Asia selama hari pertama Tahun Baru pada Jumat.</p> <p>[2] AFP Melaporkan, minyak mentah light sweet atau West Texas Intermediate (WTI) untuk pengiriman Februari, naik 56 sen menjadi 53,83 dolar AS dalam perdagangan sore.</p> <p>[9] Penurunan stok minyak mentah AS biasanya menunjukkan permintaan yang kuat di konsumen minyak utama dunia itu, sehingga mendorong harga dunia naik.</p>



No	Artikel	Ringkasan Teks
	<p>2014 dirilis pada Rabu. [8] Angka tersebut membalikkan ekspektasi untuk peningkatan 900.000 barel yang diproyeksikan dalam survei Bloomberg. [9] Penurunan stok minyak mentah AS biasanya menunjukkan permintaan yang kuat di konsumen minyak utama dunia itu, sehingga mendorong harga dunia naik. [10] Produksi minyak mentah AS mencapai lebih dari sembilan juta barel per hari, tingkat tertinggi dalam lebih dari 30 tahun.</p> <p><b>Sumber :</b>  <a href="http://www.antaranews.com/berita/47198/1/harga-minyak-naik-dalam-perdagangan-di-tahun-baru">http://www.antaranews.com/berita/47198/1/harga-minyak-naik-dalam-perdagangan-di-tahun-baru</a></p>	
7.	<p>Judul : Cisse minta maaf sudah menyikut Coleman</p> <p><b>Isi :</b></p> <p>[1] London (ANTARA News) - Penyerang Newcastle United asal Senegal Papiss Cisse mengungkapkan permintaan maafnya pada Rabu setelah dia diskors tiga pertandingan karena menyikut pemain Everton Seamus Coleman. [2] Wasit Craig Pawson tidak melihat insiden yang terjadi ketika Newcastle menang 3-2 pada pertandingan kandang di Liga Utama</p>	<p>[1] London (ANTARA News) - Penyerang Newcastle United asal Senegal Papiss Cisse mengungkapkan permintaan maafnya pada Rabu setelah dia diskors tiga pertandingan karena menyikut pemain Everton Seamus Coleman.</p>

No	Artikel	Ringkasan Teks
	<p>Inggris pada Minggu lalu, tapi Cisse menerima dakwaan melakukan pelanggaran pada Selasa setelah didakwa yang berlaku surut. [3] Cisse tidak akan bermain bagi Newcastle sebelum bertandang ke Equatorial Guinea untuk Piala Afrika dan mengungkapkan penyesalannya karena perilakunya itu melalui pesan di dalam laman Newcastle. [4] "Saya menyesal karena melakukan sesuatu seperti ini. [5] Saya bukan tipe pemain yang melakukan hal seperti itu," katanya. [6] "Itu bukan saya. [7] Saya bukan pemain yang agresif, tapi dalam sepak bola anda tidak dapat melakukan sesuatu seperti itu. [8] "Cisse, yang mencetak gol lima kali dalam enam pertandingan terakhirnya, akan absen pada pertandingan kandang hari Kamis depan melawan Burnley, pertandingan tandang putaran ketiga Piala FA pada Sabtu ke Leicester City, dan pertandingan tandang melawan Chelsea pada 10 Januari. [9] Demikian laporan AFP.</p> <p><b>Sumber :</b>  <a href="http://www.antaraneews.com/berita/471717/cisse-minta-maaf-sudah-menyikut-coleman">http://www.antaraneews.com/berita/471717/cisse-minta-maaf-sudah-menyikut-coleman</a></p>	<p>[2] Wasit Craig Pawson tidak melihat insiden yang terjadi ketika Newcastle menang 3-2 pada pertandingan kandang di Liga Utama Inggris pada Minggu lalu, tapi Cisse menerima dakwaan melakukan pelanggaran pada Selasa setelah didakwa yang berlaku surut.</p> <p>[3] Cisse tidak akan bermain bagi Newcastle sebelum bertandang ke Equatorial Guinea untuk Piala Afrika dan mengungkapkan penyesalannya karena perilakunya itu melalui pesan di dalam laman Newcastle.</p>

No	Artikel	Ringkasan Teks
8.	<p>Judul : Survei: pertumbuhan pekerjaan sector swasta AS Juli melambat</p> <p><b>Isi :</b></p> <p>[1]Washington (ANTARA News) - Pertumbuhan lapangan pekerjaan sektor swasta AS melambat pada bulan lalu namun masih cenderung meningkat, kata sebuah survei swasta yang dirilis Kamis.</p> <p>[2]Perusahaan swasta menambahkan 185.000 lapangan pekerjaan pada Juli, turun dari 229.000 lapangan pekerjaan pada Juni, kata Laporan Ketenagakerjaan Nasional yang dirilis bersama oleh Automatic Data Processing (ADP) dan Moodys Analytics, berdasarkan survei bulanan, lapor Xinhua.</p> <p>[3]"Pertumbuhan lapangan kerja Juli lebih lambat dari Juni, tetapi masih sejalan dengan apa yang telah kita lihat sejak awal tahun ini," kata Carlos Rodrguez, presiden dan CEO dari ADP.</p> <p>[4]"Khususnya, perusahaan besar dengan lebih dari 500 karyawan telah menambah lapangan pekerjaan terkuat mereka sejak Desember lalu dan hampir dua kali lipat jumlah Juni."</p> <p>[5]Usaha kecil dan menengah menyumbang sebagian besar kenaikan lapangan pekerjaan dalam satu bulan</p>	<p>[1]Washington (ANTARA News) - Pertumbuhan lapangan pekerjaan sektor swasta AS melambat pada bulan lalu namun masih cenderung meningkat, kata sebuah survei swasta yang dirilis Kamis.</p> <p>[6]Lapangan pekerjaan di sektor jasa naik 178.000 pekerjaan pada Juli, turun dari kenaikan 216.000 pekerjaan pada Juni.</p> <p>[4]"Khususnya, perusahaan besar dengan lebih dari 500 karyawan telah menambah lapangan pekerjaan terkuat mereka sejak Desember lalu dan hampir dua kali lipat jumlah Juni."</p>

No	Artikel	Ringkasan Teks
	<p>terakhir, menambahkan 121.000 pekerjaan, sedangkan usaha besar dengan lebih dari 500 karyawan menambahkan 64.000 pekerjaan, jauh lebih tinggi dari 34.000 pada Juni, menurut laporan tersebut.</p> <p>[6]Lapangan pekerjaan di sektor jasa naik 178.000 pekerjaan pada Juli, turun dari kenaikan 216.000 pekerjaan pada Juni.</p> <p>[7]Sementara itu, sektor manufaktur menambahkan 2.000 pekerjaan jauh lebih sedikit dari 9.000 pada Juni.</p> <p>[8]"PHK di industri energi dan kenaikan lapangan pekerjaan yang lebih lemah di bidang manufaktur berada di belakang pelambatan," kata Mark Zandi, kepala ekonom Moodys Analytics.</p> <p><b>Sumber :</b></p> <p><a href="http://www.antaraneews.com/berita/510821/survei-pertumbuhan-pekerjaan-sektor-swasta-as-juli-melambat">http://www.antaraneews.com/berita/510821/survei-pertumbuhan-pekerjaan-sektor-swasta-as-juli-melambat</a></p>	

Untuk memperoleh hasil ringkasan artikel berita seperti tabel 4.3, sistem harus melalui beberapa tahap seperti *text preprocessing*, pembobotan kata dengan metode TF-IDF-DF, menghitung *cosine similarity* dan MMR.

#### 4.2.2.1. *Text preprocessing*

Tujuan dari tahap *text preprocessing* yaitu merubah artikel berita menjadi kata-kata yang siap diproses untuk perhitungan bobot kata. Beberapa proses dari *text preprocessing*, yaitu segmentasi kalimat, *case folding*, *tokenizing*, *filtering*, dan *stemming*. Berikut ini adalah salah satu contoh dokumen yang diinputkan dalam proses *text preprocessing* disertai tahapan proses *text preprocessing* :

Judul : El Shaarawy menangkan Milan atas Real Madrid

**Isi :**

Klub sepak bola Serie A AC Milan "mengakhiri" 22 kemenangan Real Madrid pada laga Selasa, ketika striker Stephan El Shaarawy mencetak dua gol untuk kemenangan 4-2 pada laga persahabatan di Dubai. Pelatih Real Carlo Ancelotti tidak menurunkan pemain bintang Gareth Bale, Karim Benzema dan James Rodriguez, tetapi Cristiano Ronaldo turun lapangan dan mencetak gol pertama pada babak awal. Tim juara Eropa itu sudah memenangi 22 pertandingan dalam berbagai kompetisi tapi laga Selasa merupakan pertandingan persahabatan, bukan laga resmi. Real, memenangi Piala Klub Dunia Minggu setelah unggul 2-0 atas klub dari Argentina, San Lorenzo, tetapi El Shaarawy tampil merusak rekor klub Italia itu setelah menggoyangkan jala gawang mereka pada tiap babak, sedangkan pemain dari Prancis Jeremy Menez dan Giampaolo Pazzini juga sebagai pemain target.

a. Segmentasi kalimat

Merupakan pemecahan paragraf menjadi kalimat. Pemecahan dilakukan dengan memisahkan berdasarkan tanda baca titik (.), tanda tanya (?) dan tanda seru (!). Hasil dari proses segmentasi kalimat terlihat pada tabel 4.4.

Tabel 4.4 Hasil segmentasi kalimat

No.	Kalimat
0.	<b>El Shaarawy menangkan Milan atas Real Madrid</b>
1.	Klub sepak bola Serie A AC Milan "mengakhiri" 22 kemenangan Real Madrid pada laga Selasa, ketika striker Stephan El Shaarawy mencetak dua gol untuk kemenangan 4-2 pada laga persahabatan di Dubai
2.	Pelatih Real Carlo Ancelotti tidak menurunkan pemain bintang Gareth Bale, Karim Benzema dan James Rodriguez, tetapi Cristiano Ronaldo turun lapangan dan mencetak gol pertama pada babak awal
3.	Tim juara Eropa itu sudah memenangi 22 pertandingan dalam berbagai kompetisi tapi laga Selasa merupakan pertandingan persahabatan, bukan laga resmi
4.	Real, memenangi Piala Klub Dunia Minggu setelah unggul 2-0 atas klub dari Argentina, San Lorenzo, tetapi El Shaarawy tampil merusak rekor klub Italia itu setelah menggoyangkan jala gawang mereka pada tiap babak, sedangkan pemain dari Prancis Jeremy Menez dan Giampaolo Pazzini juga sebagai pemain target

b. *Case folding*

Paragraf berita yang telah dipotong menjadi kalimat akan menjalankan proses *case folding*. *Case folding* adalah mengubah semua teks menjadi karakter dengan huruf kecil dan membuang semua karakter selain a-z. Selain itu, tanda baca, bilangan numerik dan simbol juga dihilangkan. Tabel 4.5 merupakan hasil proses *case folding*.

Tabel 4.5 Hasil *case folding*

No.	Kalimat
0.	<b>el shaarawy menangkan milan atas real Madrid</b>
1.	klub sepak bola serie a ac milan mengakhiri kemenangan real madrid pada laga selasa ketika striker stephan el shaarawy mencetak dua gol untuk kemenangan pada laga persahabatan di dubai
2.	pelatih real carlo ancelotti tidak menurunkan pemain bintang gareth bale karim benzema dan james rodriguez tetapi cristiano ronaldo turun lapangan dan mencetak gol pertama pada babak awal
3.	tim juara eropa itu sudah memenangi pertandingan dalam berbagai kompetisi tapi laga selasa merupakan pertandingan persahabatan bukan laga resmi
4.	real memenangi piala klub dunia minggu setelah unggul atas klub dari argentina san lorenzo tetapi el shaarawy tampil merusak rekor klub italia itu setelah menggoyangkan jala gawang mereka pada tiap babak sedangkan pemain dari prancis jeremy menez dan giampaolo pazzini juga sebagai pemain target

c. *Tokenizing* kata

Merupakan proses pemotongan kalimat menjadi kata-kata. Pemotongan kalimat berdasarkan *delimiter* yang menyusunnya, yaitu spasi (" ").

Tabel 4.6 Hasil *tokenizing*

No.	Kata	No.	Kata	No.	Kata	No.	Kata
1	El	33	persahabatan	65	eropa	97	el
2	shaarawy	34	di	66	itu	98	shaarawy
3	menangkan	35	dubai	67	sudah	99	tampil
4	Milan	36	pelatih	68	memenangi	100	merusak
5	Atas	37	real	69	pertandingan	101	rekor
6	Real	38	carlo	70	dalam	102	klub
7	madrid	39	ancelotti	71	berbagai	103	italia
8	Klub	40	tidak	72	kompetisi	104	itu
9	Sepak	41	menurunkan	73	tapi	105	setelah
10	Bola	42	pemain	74	laga	106	menggoyangkan
11	Serie	43	bintang	75	selasa	107	jala
12	A	44	gareth	76	merupakan	108	gawang
13	Ac	45	bale	77	pertandingan	109	mereka
14	Milan	46	karim	78	persahabatan	110	pada
15	mengakhiri	47	benzema	79	bukan	111	tiap
16	kemenangan	48	dan	80	laga	112	babak
17	real	49	james	81	resmi	113	sedangkan
18	madrid	50	rodriguez	82	real	114	pemain
19	pada	51	tetapi	83	memenangi	115	dari
20	laga	52	cristiano	84	piala	116	prancis
21	selasa	53	ronaldo	85	klub	117	jeremy
22	ketika	54	turun	86	dunia	118	menez
23	striker	55	lapangan	87	minggu	119	dan
24	stephan	56	dan	88	setelah	120	giampaolo
25	el	57	mencetak	89	unggul	121	pazzini
26	shaarawy	58	gol	90	atas	122	juga
27	mencetak	59	pertama	91	klub	123	sebagai
28	dua	60	pada	92	dari	124	pemain
29	gol	61	babak	93	argentina	125	target
30	untuk	62	awal	94	san		
31	kemenangan	63	tim	95	lorenzo		
32	laga	64	juara	96	tetapi		



d. *Filtering*

Dalam tahap *filtering* ini melakukan pembuangan *stopword*. *Stopword* adalah kata-kata yang tidak memiliki makna atau kata yang kurang berarti dan sering muncul dalam kumpulan kata. Cara membuang kata-kata yang tidak penting dengan mengecek pada kamus *stopword*. Jika kata ada yang sama dengan *stopword*, maka kata akan dibuang atau dihapus.

Tabel 4.7 Hasil *filtering*

No.	Kata	No.	Kata	No.	Kata	No.	Kata
1	el	25	gol	49	pertama	73	unggul
2	shaarawy	26	kemenangan	50	babak	74	klub
3	menangkan	27	laga	51	awal	75	argentina
4	milan	28	persahabatan	52	tim	76	san
5	real	29	dubai	53	juara	77	lorenzo
6	madrid	30	pelatih	54	eropa	78	el
7	klub	31	real	55	memenangi	79	shaarawy
8	sepak	32	carlo	56	pertandingan	80	tampil
9	bola	33	ancelotti	57	berbagai	81	merusak
10	serie	34	menurunkan	58	kompetisi	82	rekor
11	a	35	pemain	59	laga	83	klub
12	ac	36	bintang	60	selasa	84	italia
13	milan	37	gareth	61	merupakan	85	menggoyangkan
14	mengakhiri	38	bale	62	pertandingan	86	jala
15	kemenangan	39	karim	63	persahabatan	87	gawang
16	real	40	benzema	64	bukan	88	babak
17	madrid	41	james	65	laga	89	pemain
18	laga	42	rodriguez	66	resmi	90	prancis
19	selasa	43	cristiano	67	real	91	jeremy
20	striker	44	ronaldo	68	memenangi	92	menez
21	stephan	45	turun	69	piala	93	giampaolo
22	el	46	lapangan	70	klub	94	pazzini
23	shaarawy	47	mencetak	71	dunia	95	pemain
24	mencetak	48	gol	72	minggu	96	target

e. *Stemming*

*Stemming*, yaitu mengembalikan suatu kata ke bentuk akarnya (*root word*), sehingga setiap kata memiliki representasi yang sama. Dalam metode ini hanya menangani afiks (imbuhan) prefiks (awalan) dan sufiks (akhiran) saja. Hal ini disebabkan oleh jarang terjadi kasus penambahan imbuhan infiks (sisipan) dalam bahasa Indonesia. Hasil *stemming* terlihat pada tabel 4.8.

Tabel 4.8 Hasil *stemming*

No.	Kata	No.	Kata	No.	Kata	No.	Kata
1	a	18	eropa	35	lorenzo	52	rupa
2	ac	19	gareth	36	madrid	53	rusak
3	akhir	20	gawang	37	main	54	sahabat
4	ancelotti	21	giampaolo	38	menang	55	san
5	argentina	22	gol	39	menez	56	selasa
6	awal	23	goyang	40	milan	57	sepak
7	babak	24	italia	41	minggu	58	serie
8	bale	25	jala	42	bagai	59	shaarawy
9	benzema	26	james	43	pazzini	60	stephan
10	bintang	27	jeremy	44	pertama	61	striker
11	bola	28	juara	45	piala	62	tampil
12	carlo	29	karim	46	prancis	63	tanding
13	cetak	30	klub	47	real	64	target
14	cristiano	31	kompetisi	48	rekor	65	tim
15	dunia	32	laga	49	resmi	66	turun
16	dubai	33	lapang	50	rodriguez	67	unggul
17	el	34	latih	51	ronaldo		

#### 4.2.2.2. Algoritma TF-IDF-DF

Setelah proses *text preprocessing*, tahap selanjutnya yaitu penghitungan bobot kata dengan algoritma TF-IDF-DF. Matriks kata untuk penghitungan bobot kata disajikan pada tabel 4.9. Berikut adalah proses perhitungan bobot kata pada term "a" :

$$w_{i,j} = \left( tf_{i,j} \times \log \left( \frac{N}{df_i} \right) \right) \times df_i$$

$$w_{1,1} = \left( 0 \times \log \left( \frac{5}{1} \right) \right) \times 1 = 0 \quad \rightarrow \text{bobot term "a" pada query}$$

$$w_{1,2} = \left( 1 \times \log \left( \frac{5}{1} \right) \right) \times 1 = 1.60944 \rightarrow \text{bobot term "a" pada dokumen 1 (D1)}$$

$$w_{1,3} = \left( 0 \times \log \left( \frac{5}{1} \right) \right) \times 1 = 0 \quad \rightarrow \text{bobot term "a" pada dokumen 2 (D2)}$$

$$w_{1,4} = \left( 0 \times \log \left( \frac{5}{1} \right) \right) \times 1 = 0 \quad \rightarrow \text{bobot term "a" pada dokumen 3 (D3)}$$

$$w_{1,5} = \left( 0 \times \log \left( \frac{5}{1} \right) \right) \times 1 = 0 \quad \rightarrow \text{bobot term "a" pada dokumen 4 (D4)}$$

Untuk hasil perhitungan bobot kata pada semua *term* dengan metode TF-IDF-DF, dapat dilihat pada table 4.9 berikut :

Tabel 4.9 Hasil penghitungan bobot masing-masing kata

term	Tf					Df	Idf	W				
	Q	D1	D2	D3	D4			Q	D1	D2	D3	D4
a	0	1	0	0	0	1	1,6094	0	1,60944	0	0	0
ac	0	1	0	0	0	1	1,6094	0	1,60944	0	0	0
akhir	0	1	0	0	0	1	1,6094	0	1,60944	0	0	0
ancelotti	0	0	1	0	0	1	1,6094	0	0	1,6094	0	0
argentina	0	0	0	0	1	1	1,6094	0	0	0	0	1,6094
awal	0	0	1	0	0	1	1,6094	0	0	1,6094	0	0

term	Tf					Df	ldf	W				
	Q	D1	D2	D3	D4			Q	D1	D2	D3	D4
babak	0	0	1	0	1	2	0,9163	0	0	1,8326	0	1,8326
bale	0	0	1	0	0	1	1,6094	0	0	1,6094	0	0
benzema	0	0	1	0	0	1	1,6094	0	0	1,6094	0	0
bintang	0	0	1	0	0	1	1,6094	0	0	1,6094	0	0
bola	0	1	0	0	0	1	1,6094	0	1,60944	0	0	0
carlo	0	0	1	0	0	1	1,6094	0	0	1,6094	0	0
cetak	0	1	1	0	0	2	0.9163	0	1,83258	1,8326	0	0
cristiano	0	0	1	0	0	1	1.6094	0	0	1,6094	0	0
dunia	0	0	0	0	1	1	1.6094	0	0	0	0	1,6094
dubai	0	1	0	0	0	1	1.6094	0	1,60944	0	0	0
el	1	1	0	0	1	3	0.5108	1,5325	1,53248	0	0	1,5325
eropa	0	0	0	1	0	1	1.6094	0	0	0	1,6094	0
gareth	0	0	1	0	0	1	1.6094	0	0	1,6094	0	0
gawang	0	0	0	0	1	1	1.6094	0	0	0	0	1,6094
giampaolo	0	0	0	0	1	1	1.6094	0	0	0	0	1,6094
gol	0	1	1	0	0	2	0.9163	0	1,83258	1,8326	0	0
goyang	0	0	0	0	1	1	1.6094	0	0	0	0	1,6094
italia	0	0	0	0	1	1	1.6094	0	0	0	0	1,6094
jala	0	0	0	0	1	1	1.6094	0	0	0	0	1,6094
james	0	0	1	0	0	1	1.6094	0	0	1,6094	0	0
jeremy	0	0	0	0	1	1	1.6094	0	0	0	0	1,6094
juara	0	0	0	1	0	1	1.6094	0	0	0	1,6094	0
karim	0	0	1	0	0	1	1.6094	0	0	1,6094	0	0
klub	0	1	0	0	3	2	0.9163	0	1,83258	0	0	5,4977
kompetisi	0	0	0	1	0	1	1.6094	0	0	0	1,6094	0
laga	0	2	0	2	0	2	0.9163	0	3,66516	0	3,6652	0
lapang	0	0	1	0	0	1	1.6094	0	0	1,6094	0	0
latih	0	0	1	0	0	1	1.6094	0	0	1,6094	0	0
lorenzo	0	0	0	0	1	1	1.6094	0	0	0	0	1,6094
<b>madrid</b>	1	1	0	0	0	2	0.9163	1,8326	1,83258	0	0	0
main	0	0	1	0	2	2	0.9163	0	0	1,8326	0	3,6652
<b>menang</b>	1	2	0	1	1	4	0.2231	0,8926	1,78515	0	0,8926	0,8926
menez	0	0	0	0	1	1	1.6094	0	0	0	0	1,6094
<b>milan</b>	1	1	0	0	0	2	0.9163	1,8326	1,83258	0	0	0
minggu	0	0	0	0	1	1	1.6094	0	0	0	0	1,6094
bagai	0	0	0	1	0	1	1.6094	0	0	0	1,6094	0
pazzini	0	0	0	0	1	1	1.6094	0	0	0	0	1,6094
pertama	0	0	1	0	0	1	1.6094	0	0	1,6094	0	0
piala	0	0	0	0	1	1	1.6094	0	0	0	0	1,6094

term	Tf					Df	Idf	W				
	Q	D1	D2	D3	D4			Q	D1	D2	D3	D4
prancis	0	0	0	0	1	1	1.6094	0	0	0	0	1,6094
real	1	1	1	0	1	4	0.2231	0,8926	0,89257	0,8926	0	0,8926
rekor	0	0	0	0	1	1	1.6094	0	0	0	0	1,6094
resmi	0	0	0	1	0	1	1.6094	0	0	0	1,6094	0
rodriguez	0	0	1	0	0	1	1.6094	0	0	1,6094	0	0
ronaldo	0	0	1	0	0	1	1.6094	0	0	1,6094	0	0
rupa	0	0	0	1	0	1	1.6094	0	0	0	1,6094	0
rusak	0	0	0	0	1	1	1.6094	0	0	0	0	1,6094
sahabat	0	1	0	1	0	2	0.9163	0	1,83258	0	1,8326	0
san	0	0	0	0	1	1	1.6094	0	0	0	0	1,6094
selasa	0	1	0	1	0	2	0.9163	0	1,83258	0	1,8326	0
sepak	0	1	0	0	0	1	1.6094	0	1,60944	0	0	0
serie	0	1	0	0	0	1	1.6094	0	1,60944	0	0	0
shaarawy	1	1	0	0	1	3	0.5108	1,5325	1,53248	0	0	1,5325
stephan	0	1	0	0	0	1	1.6094	0	1,60944	0	0	0
striker	0	1	0	0	0	1	1.6094	0	1,60944	0	0	0
tampil	0	0	0	0	1	1	1.6094	0	0	0	0	1,6094
tanding	0	0	0	2	0	1	1.6094	0	0	0	3,2189	0
target	0	0	0	0	1	1	1.6094	0	0	0	0	1,6094
tim	0	0	0	1	0	1	1.6094	0	0	0	1,6094	0
turun	0	0	2	0	0	1	1.6094	0	0	3,2189	0	0
unggul	0	0	0	0	1	1	1.6094	0	0	0	0	1,6094

Keterangan:

Q : *query*

tf : *term frequency*

df : *document frequency*

idf: *inverse document frequency* ( $\text{Log}_{10}(N/DF)$ )

W : *Bobot Kata* (TF \* IDF \* DF)

Pada kolom  $tf$  terdapat angka nol (0) artinya dalam suatu kalimat tidak terdapat kata tersebut, sedangkan angka selain nol (0) menandakan banyaknya kata tersebut didalam suatu kalimat.

#### 4.2.2.3. Algoritma *cosine similarity*

Jika bobot kata telah diperoleh, selanjutnya mencari nilai *cosine similarity*. Perhitungan *cosine similarity* dibagi menjadi 2 tahap, yaitu :

- a. Perhitungan relevansi antara dokumen dan *query* (judul)

Menghitung cosinus sudut dari dua vektor, yaitu  $W$  (bobot) dari tiap dokumen atau kalimat dengan  $W$  (bobot) dari *query* (judul).

- b. Perhitungan *similarity* antara dokumen

Menghitung *cosine* sudut vektor  $W$  (bobot) suatu kalimat dengan vektor  $W$  (bobot) kalimat yang lain.

Perhitungan *cosine similarity* secara lengkap dapat dilihat pada lampiran 1.

Hasil perhitungan *cosine similarity* dapat dilihat pada tabel 4.10 dan tabel 4.11.

Tabel 4.10 *Relevansi* antara judul dan semua kalimat

	<b>D1</b>	<b>D2</b>	<b>D3</b>	<b>D4</b>
<b>Q</b>	0,460986	0,027733	0,031417	0,170121

Tabel 4.11 *Similarity* antara kalimat

	<b>D1</b>	<b>D2</b>	<b>D3</b>	<b>D4</b>
<b>D1</b>	1	0,11361	0,372452	0,201615
<b>D2</b>	0,11361	1	0	0,133128
<b>D3</b>	0,372452	0	1	0,011051
<b>D4</b>	0,201615	0,133128	0,011051	1

Keterangan :

Q : *query* (judul dalam artikel berita)

D<sub>i</sub> : Dokumen/kalimat (i=1,2,3...)

#### 4.2.2.4. Metode MMR

Setelah perhitungan *cosine similarity* diperoleh, maka tahap berikutnya menghitung nilai MMR. Algoritma *maximum marginal relevance* digunakan untuk merangking kalimat-kalimat sebagai tanggapan terhadap *query* yang diberikan user. Perhitungan MMR dilakukan dengan iterasi mengkombinasikan 2 matrik *cosine similarity*, yaitu relevansi antara *query* terhadap keseluruhan kalimat dan *similarity* antara kalimat.

Prinsip perhitungan metode MMR adalah mengambil kalimat dengan nilai tertinggi dari setiap perhitungan iterasi. Iterasi akan berhenti, jika nilai hasil MMR maksimum sama dengan nol (0). Adapun nilai parameter  $\lambda$  yang digunakan pada perhitungan MMR adalah  $\lambda = 0,7$  (Carbonell dan Goldstein, 1998 :335). Proses perhitungan MMR sebagai berikut dengan catatan  $Sim_1(S_i, Q)$  adalah *relevance query*. Sedangkan  $Sim_1(S_i, S')$  adalah *similarity* kalimat terhadap kalimat yang diekstrak :

$$MMR(S_i) = \lambda \cdot Sim_1(S_i, Q) - (1 - \lambda) \cdot \max Sim_2(S_i, S')$$

## 1. Perhitungan iterasi ke- 1

Nilai  $Sim_1(S_i, S')$  adalah kosong (0).

$$\begin{aligned} MMR(1) &= 0,7 * Sim_1(S_1, Q) - (1 - 0,7) * \max Sim_2(S_1, S') \\ &= 0,7 * 0,460986 - (1 - 0,7) * 0 = 0,32269 \end{aligned}$$

$$\begin{aligned} MMR(2) &= 0,7 * Sim_1(S_2, Q) - (1 - 0,7) * \max Sim_2(S_2, S') \\ &= 0,7 * 0,027733 - (1 - 0,7) * 0 = 0,019413 \end{aligned}$$

$$\begin{aligned} MMR(3) &= 0,7 * Sim_1(S_3, Q) - (1 - 0,7) * \max Sim_2(S_3, S') \\ &= 0,7 * 0,031417 - (1 - 0,7) * 0 = 0,021992 \end{aligned}$$

$$\begin{aligned} MMR(4) &= 0,7 * Sim_1(S_4, Q) - (1 - 0,7) * \max Sim_2(S_4, S') \\ &= 0,7 * 0,170121 - (1 - 0,7) * 0 = 0,119084 \end{aligned}$$

Dari hasil perhitungan pada iterasi ke-1, diperoleh nilai maximum MMR = 0,32269 pada D1 atau pada kalimat 1. Oleh karena itu, kalimat 1 akan dipilih sebagai ringkasan. Tabel 4.12 merupakan hasil perhitungan MMR pada iterasi ke-1.

Tabel 4.12 Perhitungan MMR iterasi ke-1

	<b>D1</b>	<b>D2</b>	<b>D3</b>	<b>D4</b>
<b>Iterasi 1</b>	<b>0,322690</b>	0,019413	0,021992	0,119084
<b>Iterasi 2</b>				
<b>Iterasi 3</b>				
<b>Iterasi 4</b>				



## 2. Perhitungan iterasi ke-2

Pada iterasi ke- 2, nilai maximum MMR pada iterasi ke- 1 akan digunakan untuk menghitung *similarity* pada  $\max Sim_2(S_i, S_j)$  yaitu  $\max Sim_2(S_i, S_1)$ . Tabel 4.13 ditunjukkan nilai  $S_i$  yang digunakan.

Tabel 4.13 Nilai  $S_i$  untuk perhitungan MMR iterasi ke-2

	<b>D1</b>	<b>D2</b>	<b>D3</b>	<b>D4</b>
<b>D1</b>				
<b>D2</b>	0,11361			
<b>D3</b>	0,372452			
<b>D4</b>	0,201615			

$$\begin{aligned} MMR(2) &= 0,7 * Sim_1(S_2, Q) - (1 - 0,7) * \max Sim_2(S_2, S_1) \\ &= 0,7 * 0,027733 - (1 - 0,7) * 0,11361 = -0,01467 \end{aligned}$$

$$\begin{aligned} MMR(3) &= 0,7 * Sim_1(S_3, Q) - (1 - 0,7) * \max Sim_2(S_3, S_1) \\ &= 0,7 * 0,031417 - (1 - 0,7) * 0,372452 = -0,08974 \end{aligned}$$

$$\begin{aligned} MMR(4) &= 0,7 * Sim_1(S_4, Q) - (1 - 0,7) * \max Sim_2(S_4, S_1) \\ &= 0,7 * 0,170121 - (1 - 0,7) * 0,201615 = 0,0586 \end{aligned}$$

Dari hasil perhitungan pada iterasi ke-2, diperoleh nilai maximum MMR = 0,0586 pada D4 atau pada kalimat 4. Oleh karena itu, kalimat 4 akan dipilih sebagai ringkasan. Tabel 4.14 menunjukkan MMR iterasi ke-2.

Tabel 4.14 Nilai MMR iterasi ke-2

	<b>D1</b>	<b>D2</b>	<b>D3</b>	<b>D4</b>
<b>Iterasi 1</b>	<b>0,322690</b>	0,019413	0,021992	0,119084
<b>Iterasi 2</b>		-0,014670	-0,089744	<b>0,058600</b>
<b>Iterasi 3</b>				
<b>Iterasi 4</b>				

## 3. Perhitungan iterasi ke- 3

Pada iterasi ke- 3, untuk menghitung *similarity* pada  $\max Sim_2(S_i, S_j)$ , dicari dengan membandingkan nilai maksimum *similarity* antara D1 (S1) dengan D4 (S4), lihat tabel 4.15. Dengan mencari terlebih dahulu *similarity* maksimum D1 (kalimat ke 1) terhadap semua dokumen kecuali D1 dan D4 dan *similarity* maksimum D4 (kalimat ke 4) terhadap semua dokumen kecuali D1 dan D4. Setelah itu dicari nilai maksimum keduanya untuk menghitung  $\max Sim_2(S_i, S_j)$ .

Tabel 4.15 Nilai *similarity* yang digunakan

	<b>D1</b>	<b>D2</b>	<b>D3</b>	<b>D4</b>
<b>D1</b>				
<b>D2</b>	0,113610			0,133128
<b>D3</b>	0,372452			0,011051
<b>D4</b>				

Perhitungan perbandingan nilai *similarity* maksimum ( $S_i, S_1$ ) dan *similarity* maksimum ( $S_i, S_4$ ) sebagai berikut :

*Similarity* maksimum ( $S_i, S_1$ ) :

$$Sim_2(S_2, S_1) = 0,11361$$

$$Sim_2(S_3, S_1) = 0,372452$$

Nilai *similarity* maksimum ( $S_i, S_1$ ) adalah 0,372452.

*Similarity* maksimum ( $S_i, S_4$ ) :

$$Sim_2(S_2, S_4) = 0,113128$$

$$Sim_2(S_3, S_4) = 0,011051$$

Nilai *similarity* maksimum ( $S_i, S_4$ ) adalah 0,113128.

Jadi, *similarity* maksimum ( $S_i, S_1$ ) dan ( $S_i, S_4$ ) adalah 0,372452 pada D1.

$$MMR(2) = 0,7 * Sim_1(S_2, Q) - (1 - 0,7) * \max Sim_2(S_2, S_1)$$

$$= 0,7 * 0,027733 - (1 - 0,7) * 0,11361 = -0,01467$$

$$MMR(3) = 0,7 * Sim_1(S_3, Q) - (1 - 0,7) * \max Sim_2(S_3, S_1)$$

$$= 0,7 * 0,031417 - (1 - 0,7) * 0,372452 = -0,08974$$

Dari hasil perhitungan pada iterasi ke-3, diperoleh nilai maximum MMR < 0 pada D2 dan D3, sehingga tidak ada kalimat yang dipilih sebagai ringkasan dan iterasi perhitungan akan berhenti. Hasil perhitungan MMR terlihat pada tabel 4.16 :

Tabel 4.16 Hasil perhitungan MMR

	<b>D1</b>	<b>D2</b>	<b>D3</b>	<b>D4</b>
<b>Iterasi 1</b>	<b>0,32269</b>	0,019413	0,021992	0,119084
<b>Iterasi 2</b>	-	-0,01467	-0,08974	<b>0,0586</b>
<b>Iterasi 3</b>	-	-0,01467	-0,08974	-
<b>Iterasi 4</b>	-	-	-	-

Keterangan :

Iterasi<sub>i</sub> : *query* / judul dalam artikel berita (i=1,2,3...)

D<sub>i</sub> : Dokument/kalimat (i=1,2,3...)

Pada tabel 4.16 dipaparkan bahwa pada iterasi 1 kalimat (dokumen) yang tertinggi terdapat pada kalimat 1 (D1), sehingga kalimat 1 (D1) menjadi ringkasan. Pada iterasi 2, yang menjadi ringkasannya berikutnya adalah kalimat 4 (D4). Kalimat 1 (D1) pada iterasi 2 tidak terdapat nilai MMR, karena kalimat 1 (D1) telah dipilih sebelumnya menjadi ringkasan. Iterasi 3 tidak ada kalimat yang menjadi ringkasan, karena nilai  $\max \text{MMR} \leq 0$ . Perangkingan hasil ringkasan terlihat seperti tabel 4.17 :

Tabel 4.17 Hasil perangkingan kalimat

Rangking	Kalimat (D) ke	Max MMR
1	1	0,32269
2	4	0,0586

Pada Tabel 4.17, kalimat yang menjadi ringkasan adalah kalimat 1 (D1) dengan nilai maksimal MMR adalah 0,32296 dan kalimat 4 (D4) dengan nilai maksimal MMR adalah 0,0586.

#### 4.2.2.5. Evaluasi sistem

Mengukur tingkat akurasi hasil ringkasan oleh sistem terhadap hasil ringkasan manual dapat menggunakan tiga parameter yaitu *precision*, *recall*, dan *f-measure*. Mencari nilai *precision*, *recall*, dan *F-measure*, maka harus mencari terlebih dahulu *precision*, *recall*, dan *F-measure* setiap responden dengan system, kemudian mencari rata-rata *precision*, *recall*, dan *F-measure* pada artikel berita. Berikut perhitungan *precision*, *recall*, dan *F-measure* :

1. Responden 1 (R1) dengan sistem pada artikel 1:

Perhitungan nilai *precision* (P) :

$$P1 = \frac{\textit{correct}}{(\textit{correct} + \textit{wrong})}$$

$$= \frac{2}{(2 + 0)} = 1$$

Sedangkan perhitungan nilai *recall* (R) :

$$R1 = \frac{\textit{correct}}{(\textit{correct} + \textit{missed})}$$

$$= \frac{2}{(2 + 1)} = 0,666667$$

2. Responden 2 (R2) dengan sistem pada artikel 1:

Perhitungan nilai *precision* (P) :

$$P2 = \frac{\textit{correct}}{(\textit{correct} + \textit{wrong})}$$

$$= \frac{1}{(1 + 1)} = 0,5$$

Sedangkan perhitungan nilai *recall* (R) :

$$R2 = \frac{\textit{correct}}{(\textit{correct} + \textit{missed})}$$

$$= \frac{1}{(1 + 2)} = 0,333333$$

3. Responden 3 (R3) dengan sistem pada artikel 1:

Perhitungan nilai *precision* (P) :

$$P3 = \frac{\text{correct}}{(\text{correct} + \text{wrong})}$$

$$= \frac{1}{(1 + 1)} = 0,5$$

Sedangkan perhitungan nilai *recall* (R) :

$$R3 = \frac{\text{correct}}{(\text{correct} + \text{missed})}$$

$$= \frac{1}{(1 + 2)} = 0,333333$$

Langkah berikutnya adalah mencari rata-rata *precision*, *recall*, dan *F-measure* pada artikel 1.

$$P = \frac{P1 + P2 + P3}{3} = \frac{1 + 0,5 + 0,5}{3} = 0,666667$$

$$R = \frac{R1 + R2 + R3}{3} = \frac{0,666667 + 0,333333 + 0,333333}{3} = 0,444444$$

Selanjutnya mencari nilai *F-measure* (F) :

$$F = \frac{2 * R * P}{(R + P)}$$

$$= \frac{2 * 0,444444 * 0,666667}{0,444444 + 0,666667} = 0,533333$$

Keterangan :

*Correct* : jumlah kalimat yang diekstrak oleh sistem dan manusia.

*Wrong* : jumlah kalimat yang diekstrak oleh sistem tetapi tidak diekstrak oleh manusia.

*Missed* : jumlah kalimat yang diekstrak oleh manusia tetapi tidak diekstrak oleh sistem.

## 1.2. Perbandingan Hasil Ringkasan Portal Berita *Online* Berskala Nasional dan Lokal

Hasil ringkasan artikel berita yang dilakukan sistem terhadap artikel berita dari portal berita *online* berskala nasional dengan portal berita *online* berskala lokal memperoleh hasil yang berbeda. Hasil peringkasan artikel berita dari portal berita *online* berskala nasional dan lokal yang dilakukan oleh sistem terlihat pada tabel 4.18 dan tabel 4.19.

Tabel 4.18 Nilai akurasi pada portal berita *online* nasional

Artikel ke	R1	R2	R3	Sistem	Precision	Recall	F-Measure
1	1,3,7	1,2,3	1,2,5	1,7	0,666667	0,444444	0,5333333
2	1,4	1,2,5	1,3,4	1,4	0,666667	0,666667	0,7619048
3	1,3,4	1,4,3	1,3,4	3,1	1	0,666667	0,8000002
4	1,2,4	1,4,5	1,4,5	4,1,3	0,666667	0,666667	0,666667
5	1,2,4	1,2,3	1,2,3	1,6,3	0,555556	0,555556	0,5555557
6	1,2,9	1,2,3	1,2,3	1,2,9	0,888889	0,777778	0,8296298
7	1,3,4	1,3,4	1,5,7	1,2,3	0,555556	0,555556	0,5555557
8	1,4,6	1,4,6	1,3,4	1,6,4	0,888889	0,888889	0,888889
<b>Rata-Rata</b>					0,763889	0,652778	0,7039762

Tabel 4.19 Nilai akurasi pada portal berita *online* lokal

Artikel ke	R1	R2	R3	Sistem	Precision	Recall	F-Measure
1	7,3,5	1,3,4	1,4,9	1,11,14	0,222222	0,222222	0,222222
2	1,2,9	1,2,4	1,2,5	9,10,2	0,444444	0,444444	0,4444443
3	1,3,8	1,3,10	1,3,6	3,11,2	0,333333	0,333333	0,333333
4	1,5,3	1,2,3	3,7,8	5,2,1	0,444445	0,444445	0,4444447
5	1,3,5	1,2,4	1,2,3	1,2,3	0,777778	0,777778	0,777778
6	1,2,3	1,2,3	1,3,6	3,7,2	0,555556	0,555556	0,5555557
7	2,4,9	2,5,6	1,2,4	4,5,6	0,444444	0,444444	0,4444443
8	1,9,7	6,8,9	8,6,9	2,1,9	0,444444	0,444444	0,4444443
<b>Rata-Rata</b>					0,458333	0,458333	0,4583333

Keterangan :

$R_i$  : hasil ringkasan yang dilakukan oleh responden ( $i=1,2,3$ )

Sistem : hasil ringkasan yang dilakukan oleh sistem

*Precision* : kemampuan sistem memanggil dokumen yang relevan

*Recall* : kemampuan sistem memanggil dokumen yang tidak relevan

*F-measure* : nilai akurasi

Menurut hasil pengujian pada tabel 4.14 dan 4.15 juga dapat diketahui bahwa peringkasan artikel berita dari portal berita *online* nasional menghasilkan nilai *precision* 76,39 %, *recall* 65,28 %, dan *f-measure* 70,4 %, sedangkan peringkasan artikel berita dari portal berita *online* lokal menghasilkan nilai *precision* 45,83 %, *recall* 45,83 %, dan *f-measure* 45,83 %.

Berdasarkan hasil *precision*, *recall*, dan *f-measure* dari portal berita *online* nasional dan lokal, diketahui bahwa metode MMR dengan pembobotan kata menggunakan metode TF-IDF-DF lebih baik diterapkan pada portal berita *online* nasional karena menghasilkan nilai akurasi yang lebih tinggi yaitu 70,4% dibandingkan pada portal berita *online* lokal dengan akurasi 45,83%. Hal ini disebabkan berita pada portal *online* berskala lokal terdapat kata yang berlainan dengan kata pada judul (*query*), tetapi memiliki makna yang sama, sehingga sistem akan memilih kalimat lain yang lebih banyak mengandung kata seperti kata pada *query* (judul). Selain itu, teks berita tersusun dari banyak kalimat-kalimat yang pendek, sehingga akan mempengaruhi hasil ringkasan. Nilai MMR akan besar jika dalam kalimat tersebut terdapat banyak kata yang sama dengan *query* dan tersusun dari kalimat yang pendek. Contoh sederhananya dapat dilihat pada tabel 4.20.



Tabel 4.20 Pengaruh panjang kalimat

<b>Kalimat</b>	<b>MMR</b>
Hobi Ani <b>membaca</b>	0,252743
Terutama novel, <b>buku</b> yang Ani sukai	0,124753

*Query* : buku bacaan

Pada tabel 4.20 terlihat bahwa pada kalimat pertama terdapat 3 kata dengan 1 *query* menghasilkan nilai MMR 0,252743. Pada kalimat kedua terdapat 5 kata (*stopword* dihilangkan) dengan 1 *query* menghasilkan nilai MMR 0,124753. Jadi, jumlah *query* yang sama, tetapi dengan panjang kalimat yang berbeda akan menghasilkan nilai MMR yang berbeda. Nilai MMR akan semakin besar pada kalimat yang pendek. Contoh pengaruhnya banyak *query* dalam kalimat terlihat pada tabel 4.21.

Tabel 4.21 Pengaruh banyak *query*

<b>Kalimat</b>	<b>MMR</b>
Ani memiliki <b>hobi</b> <b>membaca</b>	0,339428
Terutama novel <b>buku</b> kesukaannya	0,158445

*Query* : hobi membaca buku

Pada tabel 4.21 terbukti bahwa pada kalimat pertama terdapat 2 *query* yaitu "hobi" dan "baca" (bentuk kata dasar) dengan nilai MMR 0,339428, sedangkan pada kalimat kedua terdapat 1 *query* yaitu "buku" dengan nilai MMR 0,158445. Jadi, semakin banyak *query* dalam suatu kalimat, maka kalimat akan memiliki nilai MMR yang tinggi, sehingga kemungkinan untuk

Judul : Daftar sekolah tahun ini mulai "online" di Balikpapan

**Isi :** [1] Balikpapan (ANTARA News) - Penerimaan Peserta Didik Baru (PPDB) secara "online" mulai dilaksanakan pada tahun 2015 di Kota Balikpapan. [2] "Pada bulan Januari sampai Februari dilaksanakan pembuatan petunjuk teknis dan *Standard Operational Procedure* (SOP). [3] Petunjuk teknis tersebut merupakan dasar pembuatan sistem aplikasi PPDB *online* yang akan dibuat oleh pihak PT. Telkom," kata Wali Kota Balikpapan, Rizal Effendi, di Balikpapan, Kamis. [4] Pada tahun 2014 Pemerintah Kota (Pemkot) Balikpapan bekerjasama dengan PT Telkom untuk mempersiapkan PPDB *online* 2015. [5] Persiapannya meliputi persiapan jaringan *online*, mempersiapkan operator pelaksana yang bertugas untuk menginput data calon siswa, serta sosialisasi ke sekolah-sekolah, katanya. [6] "Rencana alur PPDB *online* awalnya yang mendaftar mendatangi sekolah dengan membawa administrasi dan menuju operator untuk menginput data pendaftaran. [7] Kemudian operator data berupa nama, alamat, nomor ujian jenjang sebelumnya kecuali SD. [8] Setelah data terinput peserta akan menerima nomor registrasi," kata Rizal. [9] Terkait carut marutnya PPDB pada tahun 2014 dengan banyaknya "surat sakti" yang diduga dilakukan calo yang mengatasnamakan anggota DPRD Balikpapan, Walikota meminta agar masyarakat jangan percaya bila diminta sejumlah uang agar dapat masuk sekolah tertentu. [10] "Masyarakat jangan percaya bila ditawarkan calo, sampai sekarang pun saya masih menerima laporan ada yang sudah bayar lima juta tapi gak jadi masuk SMK Negeri 2. [11] Bahkan saya rencananya akan dimintai keterangan oleh Polres Balikpapan, terkait pelaksanaan PPDB," demikian Rizal.

**Sumber :** <http://www.antaranews.com/berita/471761/daftar-sekolah-tahun-ini-mulai-online-di-balikpapan>

Pada berita tersebut, sistem dan ketiga responden memilih kalimat 1, kalimat 4 dan kalimat 6 sebagai hasil ringkasan. Hal ini membuktikan bahwa sistem dengan metode MMR dapat menghasilkan ringkasan yang baik yang sesuai dengan hasil ringkasan yang dilakukan oleh responden.

Terdapat beberapa penelitian lain yang melakukan proses pengujian dengan cara yang sama, antara lain penelitian yang dilakukan oleh Muchammad Mustaqhri, Zainal Abidin dan Ririen Kusumawati dari Universitas Islam Negeri Maulana Malik Ibrahim Malang yang berjudul "Peringkasan Teks Otomatis Berita Berbahasa Indonesia Menggunakan Metode *Maximum Marginal Relevance*". Penelitian tersebut melakukan proses pengujian terhadap 30 berita yang diambil dari koran berita *online* Tempo Interaktif. Pengujian dan evaluasi hasil ringkasan menggunakan metode *recall*, *precision*, dan *f-measure*. Pada hasil evaluasi antara ringkasan sistem dengan ringkasan manual, terdapat beberapa berita yang nilai persentase *f-measure* rendah dibawah 50%.

Rendahnya nilai *f-measure* pada penelitian terdahulu diatas juga terjadi pada pengujian ini yaitu peringkasan teks berita yang diambil dari portal berita *online* berskala lokal. Hal ini tidak hanya disebabkan oleh sedikitnya jumlah kalimat yang sama dan *query* yang dimasukkan tidak menggambarkan isi, tetapi dalam penelitian ini peneliti menemukan jumlah *query* dalam satu kalimat juga mempengaruhi nilai MMR. Selain itu, besarnya nilai MMR juga dipengaruhi oleh jumlah kata dalam satu kalimat.

## **BAB V**

### **PENUTUP**

Bab ini menjelaskan kesimpulan dan saran dari penelitian yang dilakukan untuk meringkas dokumen berita menggunakan metode *Maximum Marginal Relevance* (MMR) dan pembobotan kata menggunakan metode TF-IDF-DF.

#### **5.1 Kesimpulan**

Peringkasan teks dengan metode MMR dan TF-IDF-DF pada berita yang diambil dari portal berita *online* berskala nasional menghasilkan ringkasan yang lebih baik dibandingkan dengan berita yang diambil dari portal berita *online* berskala lokal. Hasil dari uji coba yang dilakukan menghasilkan nilai *precision* 76,39 %, *recall* 65,28 %, dan *f-measure* 70,4 % pada berita dari portal berita *online* nasional, sedangkan peringkasan artikel berita dari portal berita *online* lokal menghasilkan nilai *precision* 45,83 %, *recall* 45,83 %, dan *f-measure* 45,83 %. Hal ini disebabkan pada portal *online* berskala lokal terdapat kata yang berlainan dengan kata pada judul (*query*), tetapi memiliki makna yang sama, sehingga sistem akan memilih kalimat lain yang lebih banyak mengandung kata-kata yang sama dengan *query* (judul). Selain itu, teks berita banyak menggunakan kalimat yang pendek, sehingga akan mempengaruhi hasil ringkasan. Nilai MMR akan besar jika dalam kalimat tersebut terdapat banyak kata yang sama dengan *query* dan tersusun dari kalimat yang pendek.

## 5.2 Saran

Ringkasan berita dari portal berita *online* berskala lokal menghasilkan nilai *f-measure* yang rendah dan *query* yang digunakan tidak sesuai sehingga menyebabkan kesalahan dalam menentukan ringkasan berita. Pengembangan lebih lanjut disarankan untuk memasukkan kalimat pertama sebagai bahan pertimbangan dalam menentukan ringkasan karena secara umum pada berita, kalimat pertama telah menggambarkan isi berita.

## DAFTAR PUSTAKA

- Aristoteles.2013.Penerapan Algoritma Genetika pada Peringkasan Teks Dokumen Bahasan Indonesia.*Prosiding Semirata FMIPA Universitas Lampung* :29-33.
- Barzilay, R. dan M. Elhadad. 1997. Using Lexical Chains for Text Summarization. *Proceedings of the ACL Workshop on Intelligent Scalable Text Summarization*.
- Carbonell, J.G. dan J. Goldstein. 1998. The Use of MMR and Diversity-Based Reranking in Document Reranking and Summarization. *IEEE*. 12:335-336.
- Chaer, A.2008. *Morfologi Bahasa Indonesia (Pendekatan Proses)*. Cetakan Pertama. PT. Rineka Cipta. Jakarta.
- Goldstein, J. 2008. Genre Oriented Summarization. *Thesis*. Pittsburgh : Language Technologies Institute School of Computer science Carnegie Mellon University.
- Harviyanto, N.A. 2013. Peningkatan Keterampilan Menyimak Berita Menggunakan Metode Listening Action Dan Teknik Rangsang Teks Rumpang Melalui Media Audio Pada Siswa Kelas VIII B SMP N 2 Boja. *Skripsi*. Program Studi Pendidikan Bahasa dan Sastra Indonesia Universitas Negeri Semarang. Semarang.
- Hovy, E. dan C.Y. Lin. 1997. Automated Text Summarization in SUMMARIST.
- Jezek, K. dan J. Steinberger. 2008. Automatic Text Summarization (The state of the art 2007 and new challenges). : 1-12.
- Judhita, C.2013.Akurasi Berita dalam Jurnalisme Online (Kasus Dugaan Korupsi Mahkamah Konstitusi di Portal Berita Detiknews).*Pekomnas*.16(3):145-154.
- Lahari, E.P., D.V.N.S. Kumar, dan S.S. Prasad.2014. Automatic Text Summarization with Statistical and Linguistic Features Using Successive Thresholds. *IEEE*. : 1519.
- Luhn, H.P. 1958. The Automatic Creation of Literature Abstracts. *IBM JOURNAL*. :159-165.
- MADCOMS. 2011. *Aplikasi Web Database dengan Dreamweaver dan PHP-MySQL*. Cetakan Pertama. PT ANDI. Yogyakarta.

- Mulyana, I., S. Ramadhona, dan Herfina. 2012. Penerapan Terms Frequency-Inverse Document Frequency Sistem Peringkasan Teks Otomatis Dokumen Tunggal Berbahasa Indonesia. *KNASTIK* : 1-8.
- Mustaqhfiri, M., Z. Abidin, dan R. Kusumawati. 2011. Peringkasan Teks Otomatis Berita Berbahasa Indonesia Menggunakan Metode *Maximum Marginal Relevance*. *MATICS*, 4(4): 134-147.
- Nazief, B. A. A. dan M. Adriani. 1996. Confix-Stripping : Approach to Stemming Algorithm for Bahasa Indonesia. *International Conference on Information and Knowledge Management*, : 560-563.
- Nedunchelian, R., R. Muthucumarasamy, dan E. Saranathan. 2011. Comparison of Multi Document Summarization Techniques. *International Journal of Computer Applications*. 11(3) : 155-160.
- Nugroho, Y., D.A. Putri, dan S. Laksmi. 2012. Mapping The Landscape of The Media Industry in Contemporary Indonesia. *Centre for Innovation Policy and Governance*.
- Patidar, A. K., J. Agrawal dan N. Mishra. 2012. Analysis of Different Similarity Measure Functions and Their Impacts on Shared Nearest Neighbor Clustering Approach. *International Journal of Computer Applications*. 40(16) : 1-5.
- Peranginangin, Kasiman. 2006. *Aplikasi WEB dengan PHP MySQL*. Cetakan Pertama. CV. ANDI OFFSET. Yogyakarta.
- Pramono, L.H., A.S. Rohman, dan H. Hindersah. 2013. Modified Weighting Method in TF\*IDF Algorithm for Extracting User Topic Based on Email and Social Media in Integrated Digital Assistant. [\*Rural Information & Communication Technology and Electric-Vehicle Technology\*](#). :1-6.
- Shahab, A.A. 2008. *Cara Mudah Menjadi Jurnalis*. Cetakan Pertama. Diwan Publishing. Jakarta.
- Triwibowo, A., Adiwijaya, dan M. Arifbijaksana. Penggunaan Metode Relevance Measure dan Latent Semantic Analysis (LSA) dalam Membuat Ikhtisar Dokumen Berita.
- Toth, B., D. Hakkani-Tur, dan S. Yaman. 2010. Summarization- and Learning-Based Approaches to Information Distillation. *Acoustics Speech and Signal Processing*, :5306-5309.
- Utdirartatmo, Firrar. 2002. *Mengelola Database Server MySQL di Linux dan Windows*. Cetakan Pertama. PT ANDI. Yogyakarta.

Xie, S. dan Y. Liu. 2008. Using Corpus and Knowledge-Based Similarity Measure in Maximum Marginal Relevance for Meeting Summarization. *Acoustics Speech and Signal Processing*, 8:4985 - 4988.

Yusintan, B. P., Y. Firdaus, dan W. 2010. Maharani. Perangkingan Ulang Dokumen Teks dengan Metode Maximal Marginal Relevance untuk Menghasilkan Ringkasan Teks dengan Redundansi Minimum. *Tel-U Collection*, 1-2.



## Lampiran 1

### Perhitungan *Cosine Similarity*

Rumus :

$$\text{sim}(S_1, S_2) = \frac{\sum_i t_{1i} t_{2i}}{\sqrt{\sum_i t_{1i}^2} \times \sqrt{\sum_i t_{2i}^2}}$$

Keterangan :

$S_1$  = vektor bobot kata yang menjadi kandidat

$S_2$  = vektor bobot kata selain kandidat

Dimana  $t_i$  merupakan bobot kata dari kata  $w_i$ . Berikut adalah proses perhitungan *cosine similarity* judul (*query*) dengan semua kalimat didalam teks berita :

1. Mencari  $\sum_i t_{1i} t_{2i}$  dan  $\sqrt{\sum_i t_{1i}^2}$  serta  $\sqrt{\sum_i t_{2i}^2}$

Term	W(Q)*W(Di)				Panjang Vektor				
	D1	D2	D3	D4	Q^2	D1^2	D2^2	D3^2	D4^2
a	0	0	0	0	0	2,5903	0	0	0
ac	0	0	0	0	0	2,5903	0	0	0
akhir	0	0	0	0	0	2,5903	0	0	0
ancelotti	0	0	0	0	0	0	2,5903	0	0
argentina	0	0	0	0	0	0	0	0	2,5903
awal	0	0	0	0	0	0	2,5903	0	0
babak	0	0	0	0	0	0	3,3584	0	3,3584
bale	0	0	0	0	0	0	2,5903	0	0
benzema	0	0	0	0	0	0	2,5903	0	0
bintang	0	0	0	0	0	0	2,5903	0	0
bola	0	0	0	0	0	2,5903	0	0	0
carlo	0	0	0	0	0	0	2,5903	0	0
cetak	0	0	0	0	0	3,3584	3,3584	0	0
cristiano	0	0	0	0	0	0	2,5903	0	0
dua	0	0	0	0	0	0	0	0	2,5903
dubai	0	0	0	0	0	2,5903	0	0	0
el	2,3485	0	0	2,3485	2,3485	2,3485	0	0	2,3485

Term	W(Q)*W(Di)				Panjang Vektor				
	D1	D2	D3	D4	Q^2	D1^2	D2^2	D3^2	D4^2
eropa	0	0	0	0	0	0	0	2,5903	0
gareth	0	0	0	0	0	0	2,5903	0	0
gawang	0	0	0	0	0	0	0	0	2,5903
giampaolo	0	0	0	0	0	0	0	0	2,5903
gol	0	0	0	0	0	3,3584	3,3584	0	0
goyang	0	0	0	0	0	0	0	0	2,5903
italia	0	0	0	0	0	0	0	0	2,5903
jala	0	0	0	0	0	0	0	0	2,5903
james	0	0	0	0	0	0	2,5903	0	0
jeremy	0	0	0	0	0	0	0	0	2,5903
juara	0	0	0	0	0	0	0	2,5903	0
karim	0	0	0	0	0	0	2,5903	0	0
klub	0	0	0	0	0	3,3584	0	0	30,225
kompetisi	0	0	0	0	0	0	0	2,5903	0
laga	0	0	0	0	0	13,433	0	13,433	0
lapang	0	0	0	0	0	0	2,5903	0	0
latih	0	0	0	0	0	0	2,5903	0	0
lorenzo	0	0	0	0	0	0	0	0	2,5903
<b>madrid</b>	3,3584	0	0	0	3,3584	3,3584	0	0	0
main	0	0	0	0	0	0	3,3584	0	13,433
<b>menang</b>	1,5934	0	0,7967	0,7967	0,7967	3,1868	0	0,7967	0,7967
menez	0	0	0	0	0	0	0	0	2,5903
<b>milan</b>	3,3584	0	0	0	3,3584	3,3584	0	0	0
minggu	0	0	0	0	0	0	0	0	2,5903
news	0	0	0	0	0	0	0	2,5903	0
pazzini	0	0	0	0	0	0	0	0	2,5903
pertama	0	0	0	0	0	0	0,5903	0	0
piala	0	0	0	0	0	0	0	0	2,5903
prancis	0	0	0	0	0	0	0	0	2,5903
<b>real</b>	0,7967	0,7967	0	0,7967	0,7967	0,7967	0,7967	0	0,7967
rekor	0	0	0	0	0	0	0	0	2,5903
resmi	0	0	0	0	0	0	0	2,5903	0
rodriguez	0	0	0	0	0	0	2,5903	0	0
ronaldo	0	0	0	0	0	0	2,5903	0	0
rupa	0	0	0	0	0	0	0	2,5903	0
rusak	0	0	0	0	0	0	0	0	2,5903
sahabat	0	0	0	0	0	3,3584	0	3,3584	0
san	0	0	0	0	0	0	0	0	2,5903
selasa	0	0	0	0	0	3,3584	0	3,3584	0

Term	W(Q)*W(Di)				Panjang Vektor				
	D1	D2	D3	D4	Q^2	D1^2	D2^2	D3^2	D4^2
sepak	0	0	0	0	0	2,5903	0	0	0
serie	0	0	0	0	0	2,5903	0	0	0
<b>shaarawy</b>	2,3485	0	0	2,3485	2,3485	2,3485	0	0	2,3485
stephan	0	0	0	0	0	2,5903	0	0	0
striker	0	0	0	0	0	2,5903	0	0	0
tampil	0	0	0	0	0	0	0	0	2,5903
tanding	0	0	0	0	0	0	0	10,361	0
target	0	0	0	0	0	0	0	0	2,5903
tim	0	0	0	0	0	0	0	2,5903	0
turun	0	0	0	0	0	0	10,361	0	0
unggul	0	0	0	0	0	0	0	0	2,5903
<b>JUMLAH</b>	<b>13,804</b>	<b>0,7967</b>	<b>0,7967</b>	<b>6,2903</b>	<b>13,007</b>	<b>68,935</b>	<b>63,446</b>	<b>49,44</b>	<b>105,11</b>
Akar dari jumlah					3,6065	8,3027	7,9653	7,0314	10,252

## 2. Mencari nilai *cosine similarity*

*Cosine similarity* judul (*query*) dengan dokumen 1 (D1) :

$$\text{sim}(S_q, S_1) = \frac{\sum_i t_{qi} t_{1i}}{\sqrt{\sum_i t_{qi}^2} \times \sqrt{\sum_i t_{1i}^2}} = \frac{13,804}{3,6065 \times 8,3027} = 0,460986$$

*Cosine similarity* judul (*query*) dengan dokumen 2 (D2) :

$$\text{sim}(S_q, S_2) = \frac{\sum_i t_{qi} t_{2i}}{\sqrt{\sum_i t_{qi}^2} \times \sqrt{\sum_i t_{2i}^2}} = \frac{0,7967}{3,6065 \times 7,9653} = 0,027733$$

*Cosine similarity* judul (*query*) dengan dokumen 3 (D3) :

$$\text{sim}(S_q, S_3) = \frac{\sum_i t_{qi} t_{3i}}{\sqrt{\sum_i t_{qi}^2} \times \sqrt{\sum_i t_{3i}^2}} = \frac{0,7967}{3,6065 \times 7,0314} = 0,031417$$

*Cosine similarity* judul (*query*) dengan dokumen 4 (D4) :

$$\text{sim}(S_q, S_4) = \frac{\sum_i t_{qi} t_{4i}}{\sqrt{\sum_i t_{qi}^2} \times \sqrt{\sum_i t_{4i}^2}} = \frac{6,2903}{3,6065 \times 10,252} = 0,170121$$



KEMENTERIAN PENDIDIKAN DAN KEBUDAYAAN  
UNIVERSITAS NEGERI SEMARANG  
FAKULTAS TEKNIK

Gedung E6 Lt 2, Kampus Sekaran, Gunungpati, Semarang 50229  
Telepon: 8508104  
Laman: www.te.unnes.ac.id, surel:

No. : 5018/UN37-1.5/DT/2015  
Lamp. :  
Hal : Surat Tugas Panitia Ujian Sarjana

Dengan ini kami tetapkan bahwa ujian Sarjana Fakultas Teknik UNNES untuk jurusan Teknik Elektro adalah sebagai berikut:

I. Susunan Panitia Ujian:

a. Ketua : Drs. Suryono, M.T.  
b. Sekretaris : FEDDY SETIO PRIBADI, S.Pd., MT.  
c. Pembimbing Utama : FEDDY SETIO PRIBADI, S.Pd., MT.  
d. Penguji : 1. Dr. Hari Wibawanto, M.T.  
: 2. ARYO BASKORO UTOMO, S.T., M.T.

II. Calon yang diuji:

Nama : WINDA YULITA  
NIM/Jurusan/Program Studi : 5302411139/Teknik Elektro  
/Pendidikan Teknik Informatika dan Komputer, S1  
Judul Skripsi : implementasi metode maximum marginal relevance pada peringkasan teks otomatis artikel berita

II. Waktu dan Tempat Ujian:

Hari/Tanggal : Kamis / 25 Juni 2015  
Jam : 11:00:00  
Tempat : E6 377  
Pakaian : .....

Tembusan

1. Ketua Jurusan Teknik Elektro
2. Calon yang diuji



Semarang, 30 Juni 2015

Dekan,

Drs. Muhammad Harlanu, M.Pd.

NIP. 196602151991021001



**KEPUTUSAN  
DEKAN FAKULTAS TEKNIK  
UNIVERSITAS NEGERI SEMARANG**

Nomor: 230/F1-UNNES/2015  
Tentang

**PENETAPAN DOSEN PEMBIMBING SKRIPSI/TUGAS AKHIR SEMESTER  
GASAL/GENAP  
TAHUN AKADEMIK 2014/2015**

- Menimbang : Bahwa untuk memperlancar mahasiswa Jurusan/Prodi Teknik Elektro/Pend. Teknik Informatika dan Komputer Fakultas Teknik membuat Skripsi/Tugas Akhir, maka perlu menetapkan Dosen-dosen Jurusan/Prodi Teknik Elektro/Pend. Teknik Informatika dan Komputer Fakultas Teknik UNNES untuk menjadi pembimbing.
- Mengingat : 1. Undang-undang No.20 Tahun 2003 tentang Sistem Pendidikan Nasional (Tambahan Lembaran Negara RI No.4301, penjelasan atas Lembaran Negara RI Tahun 2003, Nomor 78)  
2. Peraturan Rektor No. 21 Tahun 2011 tentang Sistem Informasi Skripsi UNNES  
3. SK. Rektor UNNES No. 164/O/2004 tentang Pedoman penyusunan Skripsi/Tugas Akhir Mahasiswa Strata Satu (S1) UNNES;  
4. SK Rektor UNNES No.162/O/2004 tentang penyelenggaraan Pendidikan UNNES;
- Menimbang : Usulan Ketua Jurusan/Prodi Teknik Elektro/Pend. Teknik Informatika dan Komputer Tanggal 6 Februari 2015

**MEMUTUSKAN**

Menetapkan :  
PERTAMA :

Menunjuk dan menugaskan kepada:

Nama : FEDDY SETIO PRIBADI, S.Pd., MT.  
NIP : 197808222003121002  
Pangkat/Golongan : III/C  
Jabatan Akademik : Lektor  
Sebagai Pembimbing

Untuk membimbing mahasiswa penyusun skripsi/Tugas Akhir :

Nama : WINDA YULITA  
NIM : 5302411139  
Jurusan/Prodi : Teknik Elektro/Pend. Teknik Informatika dan Komputer  
Topik : IMPLEMENTASI METODE MAXIMUM MARGINAL RELEVANCE PADA PERINGKASAN TEKS OTOMATIS DOKUMEN

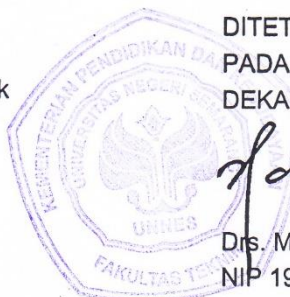
KEDUA : Keputusan ini mulai berlaku sejak tanggal ditetapkan.

Tembusan  
1. Pembantu Dekan Bidang Akademik  
2. Ketua Jurusan  
3. Petinggal



5302411139

....: FM-03-AKD-24/Rev. 00 :....



DITETAPKAN DI : SEMARANG  
PADA TANGGAL : 9 Februari 2015  
DEKAN

Drs. Muhammad Harlanu, M.Pd.  
NIP 196602151991021001