**RANCANG BANGUN APLIKASI BERBASIS WEB DENGAN MENERAPKAN *GENERAL STATISTIC METHOD***

**PADA PERINGKASAN ARTIKEL BERBAHASA INDONESIA**

**Proposal Tugas Akhir**

****

**LAINATUSSIFA DALIMUNTHE**

**1122001003**

**PROGRAM STUDI INFORMATIKA**

**FAKULTAS TEKNIK DAN ILMU KOMPUTER**

**UNIVERSITAS BAKRIE**

**JAKARTA**

**2016**

# HALAMAN PENGESAHAN

Judul Penelitian : Penerapan General Statistics Method pada

Peringkasan Artikel Berbahasa Indonesia

Peneliti Utama : Lainatussifa Dalimunthe

Jenis Kelamin : Perempuan

Unit Kerja : Program Studi Informatika, Fakultas Teknik dan

Ilmu Komputer, Universitas Bakrie

Alamat Unit Kerja : Gelanggang Mahasiswa GOR Soemantri

Brojonegoro Suite GF-22 Jl. H.R. Rasuna Said

Kav C-22, Jakarta Selatan

Alamat e-mail : lainatussifa.dalimunthe@gmail.com

Lama Penelitian : 6 (enam) bulan

Usulan Penelitian Tahun : 2016

|  |  |
| --- | --- |
| Menyetujui,  Dosen Pembimbing  Yusuf Lestanto, S.T., M.Sc. | Jakarta,  Peneliti,  Lainatussifa Dalimunthe |
|  |  |

**PENERAPAN *GENERAL STATISTIC METHOD***

**PADA PERINGKASAN ARTIKEL BERBAHASA INDONESIA**

**Lainatussifa Dalimunthe**

# ABSTRAK

Peringkasan teks sudah menjadi studi yang populer dalam beberapa dekade terakhir. Berbagai metode dalam peringkasan dan telah diterapkan dalam berbagai bahasa. Namun, belum banyak metode yang menerapkan peringkasan teks dalam berbahasa Indonesia. Penelitian ini berfokus pada peringkasan teks dokumen tunggal berbahasa Indonesia dengan menerapkan *General Statistic Method* (GSM). Metode GSM menggunakan kalimat dengan skor tertinggi sebagai kandidat dalam hasil ringkasan. Jumlah kalimat yang dipilih diambil berdasarkan *compression rate.* Terdapat delapan *sentence feature* yang digunakan dalam menemukan kalimat dengan skor tertinggi. Dalam menghitung evaluasi hasil ringkasan dibandingkan dengan ringkasan ideal. Sehingga memunculkan hasil ­*f-measure*, *recall,* dan *precision*.

Kata Kunci:

peringkasan teks, dokumen tunggal, Bahasa Indonesia, *General Statistic Method,* ­*sentence feature, compression rate, f-measure*, *recall, precision*

**IMPLEMENTING GENERAL STATISTIC METHOD**

**ON INDONESIAN LANGUAGE SUMMARY AUTOMIZATION**

**Lainatussifa Dalimunthe**

# ABSTRACT

*Summary Automization has become a popular study in the last decades. Various methods in Summary Automization have become a popular study in the last decades. Various methods in peringkasan and has been applied in various languages. Summary Automization has become a popular study in the last decades. Various methods in peringkasan and has been applied in various languages. However, summarization has not been widely applied method in* Bahasa Indonesia*. This study focuses on a single document summarization Indonesian language by applying the General Statistic Method (GSM). GSM uses sentences with the highest score as a candidate in the results summary. The number of sentences that have taken based compression rate. There is eight sentence features used in finding the sentence with the highest score. In calculating the summary results of the evaluation compared with ideal summary. The results of evaluation raise the f-measure, recall and precision scores.*

*Keyword*:

*summary automization, single document, Indonesian Language, General Statistic Method, ­sentence feature, compression rate, f-measure, recall, precision*

# DAFTAR ISI

[HALAMAN PENGESAHAN i](#_Toc451943621)

[ABSTRAK iii](#_Toc451943622)

[ABSTRACT iv](#_Toc451943623)

[DAFTAR ISI v](#_Toc451943624)

[DAFTAR GAMBAR vii](#_Toc451943625)

[DAFTAR TABEL viii](#_Toc451943626)

[BAB I - PENDAHULUAN 1](#_Toc451943628)

[1.1 Latar Belakang 1](#_Toc451943629)

[1.2 Rumusan Masalah 2](#_Toc451943630)

[1.3 Pembatasan Masalah 3](#_Toc451943631)

[1.4 Tujuan Penelitian 3](#_Toc451943632)

[1.5 Manfaat Penelitian 3](#_Toc451943633)

[BAB II - TINJAUAN PUSTAKA 5](#_Toc451943635)

[2.1 Penelitian Terdahulu 5](#_Toc451943636)

[2.2 Peringkas Teks Otomatis 6](#_Toc451943637)

[2.2.1 Definisi Peringkas Teks Otomatis 6](#_Toc451943638)

[2.2.2 Pendekatan Peringkasan Teks 6](#_Toc451943639)

[2.2.3 Tahapan Membuat Ringkasan 7](#_Toc451943640)

[2.2.4 Proses Peringkasan Teks 7](#_Toc451943641)

[2.3 *Preprocessing* 8](#_Toc451943642)

[2.3.1 Pemecahan Kalimat 9](#_Toc451943643)

[2.3.2 *Tokenizing* 10](#_Toc451943644)

[2.3.3 *Stop Word Removal* 10](#_Toc451943645)

[2.3.4 *Stemming* 11](#_Toc451943646)

[2.4 *Sentence Feature* 14](#_Toc451943647)

[2.4.1 *Title Feature* 14](#_Toc451943648)

[2.4.2 *Sentence Length* 15](#_Toc451943649)

[2.4.3 *Term Weight* 16](#_Toc451943650)

[2.4.4 *Sentence Position* 16](#_Toc451943651)

[2.4.5 *Sentence to Sentence Similarity* 17](#_Toc451943652)

[2.4.6 *Proper Noun* 18](#_Toc451943653)

[2.4.7 *Thematic Word* 18](#_Toc451943654)

[2.4.8 *Numerical Data* 19](#_Toc451943655)

[2.5 *General Statistic Method* 19](#_Toc451943656)

[2.6 Metode Evaluasi 21](#_Toc451943657)

[2.6.1 *Recall* dan *Precision* 21](#_Toc451943658)

[2.6.2 F-*Score* 22](#_Toc451943659)

[BAB III - METODE PENELITIAN 24](#_Toc451943661)

[3.1 Objek Penelitian 24](#_Toc451943662)

[3.2 Alat Penelitian 24](#_Toc451943663)

[3.3 Tahapan Penelitian 25](#_Toc451943664)

[3.4 Teknik Pengumpulan Data 25](#_Toc451943665)

[3.5 Analisis Penerapan Metode 26](#_Toc451943666)

[3.6.2 *Flowchart* 27](#_Toc451943667)

[3.6.2 *Use Case* 28](#_Toc451943668)

[3.6.3 *Sequence Diagram* 29](#_Toc451943669)

[3.6.4 *Class Diagram* 31](#_Toc451943670)

[3.6.5 Relasi Basis Data 33](#_Toc451943671)

[3.6 Kerangka Penelitian 34](#_Toc451943672)

[3.7 Rencana Kerja Penelitian 35](#_Toc451943673)

[DAFTAR PUSTAKA 36](#_Toc451943674)

# DAFTAR GAMBAR

[Gambar 2.1 Tahap-tahap dalam *Preprocessing* 9](#_Toc451423374)

[Gambar 2.2 Peringkasan Teks berdasarkan *General Statistic Method* 20](#_Toc451423375)

[Gambar 2.3 Cara Menghitung *Recall* 21](#_Toc451423376)

[Gambar 2.4 Cara Menghitung *Precision* 22](#_Toc451423377)

[Gambar 3.1 *Flowchart* dari Proses Peringkasan Teks…………………………...27](#_Toc451942758)

[Gambar 3.2 *Use Case* actor Pengguna 28](#_Toc451942759)

[Gambar 3.3 *Sequence Diagram Preprocessing* 29](#_Toc451942760)

[Gambar 3.4 *Sequence Diagram Sentence Scoring* 30](#_Toc451942761)

[Gambar 3.5 *Class Diagram* 31](#_Toc451942762)

[Gambar 3.6 Relasi *Database* 33](#_Toc451942763)

[Gambar 3.7 Kerangka Penelitian 34](#_Toc451942764)

# DAFTAR TABEL

[Tabel 2.1 Contoh Pemecahan Kalimat 9](#_Toc451423969)

[Tabel 2.2 Contoh *Tokenizing* 10](#_Toc451423970)

[Tabel 2.3 Contoh *Stop Word Removal* 11](#_Toc451423971)

[Tabel 2.4 Aturan pemenggalan Awalan *Stemmer* Nazief dan Adriani 13](#_Toc451423972)

Tabel 3.1 Kerangka Kerja Penelitian ……………………………………………32

# BAB I

# PENDAHULUAN

## Latar Belakang

Luasnya penggunaan internet telah meningkatkan jumlah informasi yang dapat diakses. Terlalu banyaknya informasi yang dapat digunakan pengguna justru membuat pengguna semakin sulit. Pengguna harus menyaring banyaknya dokumen untuk dapat menemukan informasi yang dibutuhkan (Bhatia & Jaiswal, 2015). Adanya mesin pencari memang membantu pengguna dalam pencarian. Namun, hasil dari mesin pencari belum tentu memberikan informasi yang optimal. Terkadang pengguna harus membaca hasil yang terdiri dari ribuan teks yang membutuhkan waktu yang lama. Oleh karena itu *Automatic Text Summarization* (peringkas teks otomatis) diperkenalkan untuk membantu pengguna dalam memperoleh informasi yang penting (Wan & Ma, 2013).

Tujuan dari *Automatic Text Summarization* adalah untuk menghasilkan ringkasan dari satu atau lebih dokumen yang relevan untuk pengguna. Ringkasan tersebut harus mampu memberikan konten dengan jumlah kata yang lebih sedikit tanpa kehilangan konten utamanya. Pengguna mampu menghemat waktu dalam membaca teks atau dokumen tanpa kehilangan informasi utama. (Wan & Ma, 2013)

Berbagai macam metode telah diperkenalkan dalam *Automatic Text Summarization.* Beberapa contoh metode diantaranya adalah *Graph based methods*, *feature vector based*, *cluster-based* dan *evolutionary based algorithm*. Beberapa tipe metode *Evolutionary Algorithm* yang telah diperkenalkan adalah *Genetic Algorithms*, *Evolutionary Programming*, dan *Evolution Strategies Classifier Programming*. Secara umum proses dari *Evolutionary Algorithm* termasuk simulasi individu atau biasa disebut kromosom lewat proses seleksi, mutasi, dan reproduksi. Salah satu *Evolutionary Algorithm* yang populer adalah *Genetic Algorithm* (Meenaa & Gopalanib, 2015).

Peringkasan teks berbahasa Indonesia telah dilakukan oleh beberapa peneliti sebelumnya. Penelitian-penelitian tersebut mengadopsi metode-metode yang telah digunakan sebelumnya dalam peringkasan teks berbahasa Inggris. Salah satu penelitian yang menerapkan metode *Genetic Algorithm* dalam peringkasan teks berbahasa Indonesia dilakukan oleh Aristoteles pada tahun 2013. Penelitian tersebut menggunakan sebelas fitur dari *Genetic Algortihm* terhadap 150 dokumen. Evaluasi pada penelitian tersebut dihitung menggunakan *F-measure, Recall,* dan *Precision*. Penelitian tersebut menghasilkan akurasi sebesar 47.46% (Aristoteles, 2013).

Pada penelitian ini peneliti melakukan peringkasan teks dengan mengadopsi *Genetic Algorithm* dengan mengekstrak kalimat-kalimat penting yang selanjutnya akan dijadikan sebagai ringkasan. Metode yang akan digunakan pada penelitian ini akan merujuk pada penelitian yang telah dilakukan dalam peringkasan bahasa Inggris sebelumnya. Penelitian tersebut berjudul *Genetic Algorithm Based Sentence Extraction for Text Summarization* yang diperkenalkan oleh Ladda Suanmali. Penelitian tersebut berhasil menghasilkan nilai *F-measure*, *Recall*, dan *Precision* sebesar 49.80%, 44.64%, dan 46.62% (Suanmali, Salim, & Binwahlan, 2011).

## 1.2 Rumusan Masalah

Berdasarkan penjabaran isu di latar belakang, rumusan masalah pada penelitian ini adalah sebagai berikut:

1. Apa yang mendasari penulis untuk menggunakan *General Statistic Method* dalam penelitian ini? Apakah ada metode lain selain *General Statistic Method*?
2. Apa masalah mendasar yang ditemukan penulis sehingga metode *General Statistic Method* perlu diterapkan dalam peringkasan artikel berbahasa Indonesia?
3. Bagaimana mengimplementasikan *General Statistic Method* pada peringkasan artikel dengan teks berbahasa Indonesia?
4. Bagaimana proses mendapatkan artikel berbahasa Indonesia yang akan digunakan dalam pengimplementasian *General Statistic Method* dalam penelitian ini?
5. Bagaimana proses mendapatkan artikel berbahasa Indonesia yang sudah ringkas (ringkasan ideal) sehingga dapat dibandingkan dengan artikel yang telah diringkas oleh aplikasi yang menimplementasikan *General Statistic Method*?
6. Bagaimana menampilkan proses dan hasil penerapan metode *General Statistic Method* dalam bentuk aplikasi berbasis web?

## 1.3 Pembatasan Masalah

Batasan masalah di penelitian ini meliputi:

1. Aplikasi yang dikembangkan pada penelitian ini adalah aplikasi peringkasan teks berbasis web.
2. Metode peringkasan teks yang digunakan dalam penelitian ini adalah *General Statistic Method*.
3. Dokumen yang diringkas dalam penelitian ini adalah dokumen berisi artikel teks berbahasa Indonesia.

## 1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah:

1. Untuk menghasilkan ringkasan yang menerapkan metode *General Statistic Method* pada peringkasan dokumen tunggal dengan artikel berbahasa Indonesia.
2. Untuk mengembangkan aplikasi peringkas otomatis dokumen tunggal berbahasa Indonesia.

## 1.5 Manfaat Penelitian

Penelitian ini diharapkan mampu memberikan manfaat kepada berbagai pihak sebagai berikut:

1. Penelitian ini dapat membantu menyingkat waktu pembaca saat membaca artikel berbahasa Indonesia.
2. Penelitian ini dapat digunakan sebagai referensi untuk peneliti yang lain dalam mengimplementasikan *General Statistic Method*  dalam peringkasan teks pada dokumen tunggal berbahasa Indonesia.

## 1.6 Sistematika Penulisan

Sistematika penulisan proposal penelitian ini disusun untuk memberikan gambaran umum tentang penelitian yang akan dilakukan. Proposal penelitian ini terdiri dari tiga bab. Sistematika penulisan proposal penelitian ini adalah sebagain berikut :

**BAB I PENDAHULUAN**

Bab ini menjelaskan mengenai latar belakang penelitian, rumusan masalah, pembatasan masalah, tujuan penelitian dan manfaat penelitian.

**BAB II TINJAUAN PUSTAKA**

Bab ini menjelaskan mengenai penelitian terdahulu, teori yang berhubungan, dan metode yang akan digunakan.

**BAB III METODOLOGI PENELITIAN**

Bab ini membahas tentang objek penelitian, alat penelitian, tahapan penelitian, teknik pengumpulan data, dan kerangka kerja penelitian.

# BAB II

# TINJAUAN PUSTAKA

## 2.1 Penelitian Terdahulu

Tiga penelitian sebelumnya yang akan digunakan sebagai acuan adalah sebagai berikut:

1. Penelitian yang dilakukan oleh Ladda Suanmali, Naomie Salim, Mohammed Salem Binwahlan dengan judul *Genetic Algorithm Based Sentence Extraction for Text Summarization*. Pada penelitian tersebut peneliti membandingkan dua metode yaitu *Genetic Algorithm* dan *General Statistic Method.* Dua metode tersebut menerapkan delapan fitur dalam mengekstrak kalimat untuk membuat ringkasan. Delapan fitur tersebut adalah *title feature, sentence length, term weight, sentence position, sentence to sentence similarity, proper noun, thematic word,* dan *numerical data*. Penelitian tersebut berhasil menghasilkan nilai *F-measure*, *Recall*, dan *Precision* sebesar 49.80%, 44.64%, dan 46.62% untuk metode *Genetic Algorithm* dan 49.58%, 44.12%, dan 46.25% untuk metode *General Statistic Method* (Suanmali, Salim, & Binwahlan, 2011).
2. Penelitian yang dilakukan oleh Agung Auliaguntary Arif Putra yang berjudul Implementasi Text Summarization Menggunakan Metode Vector Space Model Pada Artikel Berita Berbahasa Indonesia. Pada penelitian ini peneliti menerapkan algoritma *vector space* model dalam melakukan peringkasan otomatis artikel berita berbahasa Indonesia. Penelitian tersebut menerapkan tiga tahapan dalam peringkasan teks yaitu tahap praproses, perhitungan *tf-idf*, dan penerapan *vector space model*. Penelitian tersebut menghasilkan nilai *precision* sebesar 72%, *recall* sebesar 65%, dan *f-measure* sebesar 68% dengan *compression rate* sebesar 35 (Putra, 2015).
3. Penelitian yang dilakukan oleh Kenneth Vincent yang berjudul Peringkas Temporal Dokumen Berbahasa Indonesia menggunakan *Latent Dirichlet Allocation.* Penelitian ini mengambil artikel berbahasa Indonesia untuk dirangkum dalam bentuk *timeline*. Metode yang dipakai dalam pengenalan topik berita adalah metode *Latent Dirichlet Allocation*. Penelitian ini berhasil menghasilkan ringkasan dengan akurasi sebesar 34.26% untuk masukan yang telah dihapus stopwordsnya (Vincent, 2015).

## 2.2 Peringkas Teks Otomatis

### 2.2.1 Definisi Peringkas Teks Otomatis

Berdasarkan Kamus Besar Bahasa Indonesia Departemen Pendidikan Nasional Balai Pustaka mendefinisikan ringkasan, yaitu hasil yang meringkaskan, ikhtisar, singkatan cerita, dan kependekan (Pusat Bahasa Depdiknas, 2008). Ringkasan merupakan ekspresi yang ketat dari isi utama suatu artikel, tujuannya untuk memberitahu pembaca inti dari suatu pikiran utama. Ringkasan mengambil isi yang paling penting dari sumber informasi yang kemudian menyajikannya kembali dalam bentuk yang lebih ringkas bagi penggunanya. Di sisi lain, defenisi otomatis menurut Kamus Besar Bahasa Indonesia Departemen Pendidikan Nasional Balai Pustaka yaitu dengan sendirinya, secara otomat (Pusat Bahasa Depdiknas, 2008).

Peringkas teks otomatis adalah teknik pembuatan ringkasan dari sebuah teks secara otomatis dengan memanfaatkan aplikasi yang dijalankan pada komputer untuk menghasilkan informasi yang paling penting dari dokumen aslinya (Zaman & Winarko, 2011).

### 2.2.2 Pendekatan Peringkasan Teks

Ada dua jenis pendekatan yang dilakukan untuk meringkas sebuah teks yaitu ekstraksi dan abstraksi (Nenkova & McKeown, 2011).

1. Ekstraksi (*extractive summary*)

Dalam teknik ekstraksi, hal yang dianggap paling penting adalah menyalin unit-unit teks dari teks sumber. Unit-unit teks yang disalin dapat berupa klausa utama, kalimat utama, atau paragraph utama tanpa ada penambahan kalimat-kalimat baru yang tidak terdapat pada dokumen aslinya.

1. Abstraksi (*abstractive summary*)

Teknik abstraksi menggunakan metode linguistik untuk memerika dan menafsirkan teks dokumen menjadi ringkasan. Ringkasan teks tersebut dihasilkan dengan cara menambahkan kalimat-kalimat baru yang merepresentasikan intisari teks sumber ke dalam bentuk yang berbeda dengan kalimat-kalimat yang ada pada teks sumber.

### 2.2.3 Tahapan Membuat Ringkasan

Ada tujuh langkah yang perlu dilakukan dalam meringkas sebuah teks (Sartuni, Arifin, & Hutagalung, 2015). Tujuh langkah tersebut adalah:

1. Membaca naskah atau teks asli beberapa kali.
2. Mencatat gagasan utama penulis. Dalam artikel, harus dicatat kalimat topik pada setiap paragraph.
3. Membuang paragraph yang berisi contoh, deskripsi, atau kutipan.
4. Membuang berbagai keterangan tambahan yang tidak penting dalam sebuah kalimat.
5. Mengubah dialog langsung ke dalam bentuk kalimat tidak langsung.
6. Sedapat mungkin, menggunakan kalimat tunggal.
7. Menyusun ringkasan dengan mempertahankan susunan gagasan penulis yang asli.

### 2.2.4 Proses Peringkasan Teks

Secara umum terdapat tiga tahapan dalam proses peringkasan teks, yaitu sebagai berikut (Mustaqhfiri, 2011):

1. Identifikasi Topik

Tahapan ini meliputi identifikasi faktor yang sangat penting tentang apa yang dibicarakan dalam teks tersebut. Ada beberapa teknik untuk melakukannya, diantaranya:

1. Dalam beberapa tipe teks, informasi penting terdapat dalam bagian-bagian tertentu dalam teks tersebut, seperti dalam judul, kalimat pertama, kaliamat terakhir dan lain sebagainya.
2. Beberapa kata atau frase mengidentifikasi intisari dari suatu teks.
3. Beberapa kata cenderung muncul lebih sering. Ini dijadikan faktor penentu topik dari suatu teks (*word frequency*).
4. Ada juga topik diidentifikasi dari jumlah pengertiannya atau makna dibandingkan kata.
5. Interpretasi

Tahap ini dilakukan berdasarkan metode yang akan digunakan dalam meringkas teks. Apabila menggunakan metode ekstraktif maka tahap peringkasan ini dilakukan sesuai dengan metode ekstraktif tersebut. Namun, apabila menggunakan metode abstraktif tahap ini dilakukan dengan cara penggabungan pengertian yang serupa menjadi satu, penghilangan redundansi dan lain sebagainya.

1. *Generating*

Tahapan ini adalah tahan pembentukan hasil akhir. Terdiri dari penggabungan frase, pencetakan kata atau frase dan pembentukan kalimat. Terdapat dua metode yang dapat digunakan, diantaranya:

1. *Extraction*, hasil akhir berisi kalimat atu frase yang dihasilkan setelah semua tahap proses pada *text summarization* selesai dilakukan.
2. *Topic List*, hasil akhir berisi kata-kata yang sering muncul atau penggabungan pengertian yang telah diinterpretasi.

## 2.3 *Preprocessing*

*Preprocessing* adalah tahapan untuk mempersiapkan teks menjadi data yang akan diolah di tahapan berikutnya. Inputan awal pada proses ini adalah berupa dokumen. Pada umumnya *preprocessing* memiliki beberapa tahapan. Pada penelitian ini tahap *preprocessing* memiliki empat tahap secara berurut yaitu proses pemecahan kalimat, proses *tokenizing* kata, proses *stop word removal* dan proses *stemming*.

Gambar 2.1 Tahap-tahap dalam *Preprocessing*

### 2.3.1 Pemecahan Kalimat

Tahap yang pertama kali dilakukan dalam *preprocessing* adalah tahap pemecahan kalimat. Pemecahan kalimat yaitu proses memecah string teks dokumen yang panjang menjadi kumpulan kalimat-kalimat. Fungsi split() digunakan dalam pemecahan kalimat menggunakan tanda titik “.”, tanda tanya ”?” dan tanda tanya “!” sebagai delimiter untuk memotong string dokumen.

Tabel 2.1 Contoh Pemecahan Kalimat

|  |  |
| --- | --- |
| **Artikel** | **Hasil Pemecahan Kalimat** |
| Edukasi atau pemberian materi telah diberikan kepada seluruh individu sedari kecil, entah itu dari keluarga atau dari sekolah. Kemudian, apa kabar dengan edukasi sekolah? Sekolah merupakan tempat di mana para murid mengembangkan dirinya menjadi individu yang mempunyai wawasan luas, melatih murid menjadi pribadi yang dinamis dan selektif, dan mengedukasi murid. | Edukasi atau pemberian materi telah diberikan kepada seluruh individu sedari kecil entah itu dari keluarga atau dari sekolah |
| Kemudian apa kabar dengan edukasi sekolah |
| Sekolah merupakan tempat di mana para murid mengembangkan dirinya menjadi individu yang mempunyai wawasan luas melatih murid menjadi pribadi yang dinamis dan selektif dan mengedukasi murid |

### 2.3.2 *Tokenizing*

Setelah melakukan proses tahap pemecahan kalimattahap yang dilakukan selanjutnya adalah proses *tokenizing.* *Tokenizing* adalah proses pemotongan string input berdasarkan tiap kata yang menyusunnya. Sehingga menghasilkan sejumlah kata dari satu kalimat. Pemecahan kalimat menjadi kata-kata tunggal dilakukan dengan men-*scan* kalimat dengan pemisah (*delimiter*) white space (*spasi*, *tab*, dan *newline*). Apabila ada beberapa kata sama dalam sebuah kalimat, maka hasil dari proses *tokenizing* hanya menghasilkan satu kata tersebut.

Tabel 2.2 Contoh *Tokenizing*

|  |  |  |
| --- | --- | --- |
| **Hasil Pemecahan Kalimat** | **Hasil *Tokenizing*** | |
| Aina dan Nona tidak makan mie dan juga tidak makan bakso ataupun minum-minum lain | Aina  dan  Nona  tidak  lain | makan  mie  juga  minum |

### 2.3.3 *Stop Word Removal*

Setelah melakukan proses *tokenizing* maka proses yang dilakukan selanjutnya adalah proses *stop word removal.* Penghapusan *Stopword* merupakan proses penghilangan kata *stopword*. *Stopword* adalah kata-kata yang sering kali muncul dalam dokumen namun arti dari kata-kata tersebut tidak deskriptif dan tidak memiliki keterkaitan dengan tema tertentu. Misalnya “dalam”, ”pada”, “dari”, ”sebuah”, ”karena” dan lain sebagainya. Sehingga sejumlah kata merupakan hasil dari proses *tokenizing* akan dihapus dalam proses *stop word removal*.

Tabel 2.3 Contoh *Stop Word Removal*

|  |  |  |
| --- | --- | --- |
| **Hasil Pemecahan Kalimat** | | **Hasil *Stop Word Removal*** |
| edukasi  atau  pemberian  materi  telah  diberikan | kepada  seluruh  individu  sedari  kecil | edukasi  pemberian  materi  diberikan  individu  kecil |

### 2.3.4 *Stemming*

Setelah melakukan proses *stop word removal* proses yang dilakukan selanjutnya adalah proses *stemming. Stemming* adalah proses pemetaan dari penguraian berbagai bentuk kata baik itu *prefix*, *sufix*, maupun gabungan antara *prefix* dan *sufix* (*confix*), menjadi bentuk kata dasarnya (Zaman & Winarko, 2011). Algoritma *stemmer* yang diperkenalkan Nazief dan Adriani didefinisikan sebagai berikut (Keke, Chikita, & Prayogo, 2012):

* 1. Di awal proses *stemming* dan setiap langkah yang selanjutnya dilakukan, lakukan pengecekan hasil proses stemming kata yang di-input-kan pada langkah tersebut ke kamus kata dasar. Jika kata ditemukan, berarti kata tersebut sudah berbentuk kata dasar dan proses stemming dihentikan. Jika tidak ditemukan, maka langkah selanjutnya dilakukan.
  2. Hilangkan *Inflection Suffixes* (“-lah”, “-kah”, “-ku”, “-mu”, atau “-nya”). 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 :

1. 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.
2. Akhiran yang dihapus (“-i”, “-an” atau “-kan”) dikembalikan, lanjut ke langkah 4.
   1. Hilangkan derivation prefixes.
   2. Langkah 4 berhenti jika :

i. Terjadi kombinasi awalan dan akhiran yang terlarang.

ii. Awalan yang dideteksi saat ini sama dengan awalan yang dihilangkan sebelumnya.

iii. Tiga awalan telah dihilangkan.

* 1. Identifikasikan tipe awalan dan hilangkan. Awalan terdiri dari dua tipe:

i. Standar (“di-”, “ke-”, “se-”) yang dapat langsung dihilangkan dari kata.

ii. Kompleks (“me-”, “be-”, “pe”, “te-”) adalah tipe-tipe awalan yang dapat bermorfologi sesuai kata dasar yang mengikutinya. Oleh karena itu, gunakan aturan pada Tabel 2.1 untuk mendapatkan hasil pemenggalan yang tepat.

* 1. Cari kata yang telah dihilangkan awalannya ini di dalam kamus kata dasar. Apabila tidak ditemukan, maka langkah 4 diulangi kembali. Apabila ditemukan, maka keseluruhan proses dihentikan.
  2. Apabila setelah langkah 4 kata dasar masih belum ditemukan, maka proses *recoding* dilakukan dengan mengacu pada aturan pada Tabel 2.1. *Recoding* dilakukan dengan menambahkan karakter *recoding* di awal kata yang dipenggal. Pada Tabel 2.1, karakter *recoding* adalah huruf kecil setelah tanda hubung („-‟) dan terkadang berada sebelum tanda kurung. Sebagai contoh, kata “menangkap” (aturan 15), setelah dipenggal menjadi “nangkap”. Karena tidak valid, maka recoding dilakukan dan menghasilkan kata “tangkap”.
  3. Jika semua langkah gagal, maka input kata yang diuji pada algoritma ini dianggap sebagai kata dasar.

Tabel 2.4 Aturan pemenggalan Awalan *Stemmer* Nazief dan 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 | beC1erC2... | be-C1erC2... dimana C1!={‟r‟|‟l‟} |
| 6 | terV... | ter-V... | te-rV... |
| 7 | terCerV... | ter-CerV... dimana C!=‟r‟ |
| 8 | terCP... | ter-CP... dimana C!=‟r‟ dan P!=‟er‟ |
| 9 | teC1erC2... | te-C1erC2... 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... |
| 22 | perCAP | per-CAP... dimana C!=‟r‟danP!=‟er‟ |
| 23 | perCAerV... | per-CAerV... dimana C!=‟r‟ |
| 24 | pem{b|f|V}... | pem-{b|f|V}... |
| 25 | pem{rV|V}... | pe-m{rV|V}... | pe-p{rV|V}... |
| 26 | pen{c|d|j|z}... | pen-{c|d|j|z}... |
| 27 | penV... | pe-nV... | pe-tV... |
| 28 | peng{g|h|q}... | peng-{g|h|q}... |
| 29 | pengV... | peng-V... | peng-kV... |
| 30 | penyV... | peny-sV… |
| 31 | pelV... | pe-lV... kecuali “pelajar” yang menghasilkan “ajar” |
| 32 | peCerV... | per-erV... dimana C!={r|w|y|l|m|n} |
| 33 | peCP... | pe-CP... dimana C!={r|w|y|l|m|n} dan P!=‟er‟ |
| 25 | pem{rV|V}... | pe-m{rV|V}... | pe-p{rV|V}... |
| 26 | pen{c|d|j|z}... | pen-{c|d|j|z}... |
| 27 | penV... | pe-nV... | pe-tV... |
| 28 | peng{g|h|q}... | peng-{g|h|q}... |
| 29 | pengV... | peng-V... | peng-kV... |
| 30 | penyV... | peny-sV… |
| 31 | pelV... | pe-lV... kecuali “pelajar” yang menghasilkan “ajar” |
| 32 | peCerV... | per-erV... dimana C!={r|w|y|l|m|n} |
| 33 | peCP... | pe-CP... dimana C!={r|w|y|l|m|n} dan P!=‟er‟ |
| 25 | pem{rV|V}... | pe-m{rV|V}... | pe-p{rV|V}... |
| 26 | pen{c|d|j|z}... | pen-{c|d|j|z}... |
| 27 | penV... | pe-nV... | pe-tV... |
| 28 | peng{g|h|q}... | peng-{g|h|q}... |
| 29 | pengV... | peng-V... | peng-kV... |
| 30 | penyV... | peny-sV… |
| 31 | pelV... | pe-lV... kecuali “pelajar” yang menghasilkan “ajar” |
| 32 | peCerV... | per-erV... dimana C!={r|w|y|l|m|n} |
| 33 | peCP... | pe-CP... dimana C!={r|w|y|l|m|n} dan P!=‟er‟ |
| 25 | pem{rV|V}... | pe-m{rV|V}... | pe-p{rV|V}... |
| 26 | pen{c|d|j|z}... | pen-{c|d|j|z}... |
| 27 | penV... | pe-nV... | pe-tV... |
| 28 | peng{g|h|q}... | peng-{g|h|q}... |
| 29 | pengV... | peng-V... | peng-kV... |
| 30 | penyV... | peny-sV… |
| 31 | pelV... | pe-lV... kecuali “pelajar” yang menghasilkan “ajar” |
| 32 | peCerV... | per-erV... dimana C!={r|w|y|l|m|n} |
| 33 | 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”

## 2.4 *Sentence Feature*

Setelah melakukan proses *preprocessing,* setiap kalimat akan diberikan nilai skor. Dalam memberikan nilai skor diterapkan delapan *sentence feature* terhadap masing-masing kalimat. Delapan *sentence score* yang diajukan oleh (Suanmali, Salim, & Binwahlan, 2011) adalah *title feature*, *sentence length, term weight, sentence position, sentence to sentence similarity, proper noun, thematic word,* dan *numerical data*.

### 2.4.1 *Title Feature*

Apabila ada kata dalam kalimat yang memiliki kata yang sama dengan kata yang terdapat pada judul teks maka kalimat tersebut akan mendapatkan skor yang lebih tinggi. Ini ditentukan dengan menghitung perbandingan jumlah kata yang sama dengan judul. Hasil dari fitur ini ditentukan dengan menghitung rasio jumlah kata sama dengan judul dengan jumlah kata yang ada dalam kalimat tersebut. Sehingga semakin banyak kata yang sama dengan kata yang terdapat pada judul maka semakin tinggi nilai skor fitur *title feature*.

(1)

Keterangan: *S\_F1(S)* = Skor kalimat dari fitur *title feature*

*No.Title word in S* = Jumlah kata yang sama dengan kata yang

terdapat pada judul teks

*No.Word in Title* = Jumlah kata yang terdapat pada judul teks

### 2.4.2 *Sentence Length*

Dalam sebuah teks terkadang terdapat kalimat pendek sepeti nama penulis dan tanggal kejadian. Fitur ini berguna untuk menyaring kalimat pendek tersebut. Kalimat-kalimat pendek seperti itu lebih baik tidak dicantumkan dalam hasil ringkasan. Hasil dari fitur ini ditentukan dengan menghitung perbandingan jumlah kata yang ada dengan jumlah kata terpanjang pada kalimat dalam artikel tersebut. Sehingga semakin banyak kata dalam kalimat maka semakin tinggi nilai skor dari fitur *sentence length*.

(2)

Keterangan: *S\_F2(S)* = Skor kalimat dari fitur *sentence length*

*No.Word occurring in S* = Jumlah kata dalam kalimat

= Jumlah kata yang terdapat pada kalimat terpanjang

### 2.4.3 *Term Weight*

*Term weight* adalah frekuensi munculnya sebuah katadalam dokumen. Perhitungan *term weight* sudah sering digunakan dalam perhitungan kepentingan kalimat. Skor dari kalimat dihitung lewat penjumlahan skor kata dalam kalimat. Skor *term weight*dari kata dihitung dengan menggunakan metode *tf-idf* (*term frequency, inverse document frequency*) yang diterapkan dalam kalimat. Dalam penelitian ini, metode tersebut diterapkan dalam *tf-isf* (*term frequency, inverse sentence frequency*). Dalam menghitung *tf-isf* digunakan dapat digunakan formula di bawah ini.

(3)

Keterangan: *w* = nilai *weight* dari kata

= frekuensi munculnya kata *i* dalam dokumen

*N* = jumlah total kalimat

*ni* = jumlah kalimat dimana kata *i* muncul

Fitur *term weight* dapat dihitung sebagai berikut:

(4)

Keterangan: *S\_F3(S)* = Skor kalimat dari fitur *term weight*

= total nilai *weight* dari seluruh kata yang ada dalam kalimat

= nilai tertinggi dari total nilai *weight* dari seluruh kata yang ada dalam kalimat yang ada di dalam teks

*k* = jumlah kata dalam kalimat

### 2.4.4 *Sentence Position*

Dalam sebuah teks posisi kalimat dapan menunjukkan pentingnya kalimat tersebut. Pada umumnya kalimat yang terletak di bagian awal teks cenderung lebih penting dari kalimat yang lain. Pada fitur ini lima kalimat awal dari teks akan mendapatkan skor lebih. Semakin awal kalimat terletak dalam teks semakin tinggi skor yang didapat dari fitur ini. Sementara kalimat-kalimat lainnya hanya akan mendapatkan nilai skor 0 dari fitur ini.

(5)

Keterangan: *S\_F4(S)* = Skor kalimat dari fitur *sentence position*

= Kalimat pertama dalam teks diberi skor

= Kalimat kedua dalam teks diberi skor

= Kalimat ketiga dalam teks diberi skor

= Kalimat keempat dalam teks diberi skor

= Kalimat kelima dalam teks diberi skor

= Kalimat lainnya yang ada dalam teks diberi skor 0

### 2.4.5 *Sentence to Sentence Similarity*

Kemiripan antar-kalimat merupakan kata yang muncul dalam kalimat sama dengan kata yang muncul dalam kalimat-kalimat lainnya. Skor dari fitur ini dihitung dengan membandingkan rasio dari dari jumlah kalimat mirip dari kalimat tersebut dengan jumlah maksimal dari kalimat yang mirip. Apabila sebuah kalimat memiliki kemiripan yang tinggi dengan kalimat lainnya maka kalimat tersebut akan memiliki skor yang tinggi dari fitur ini.

(6)

Keterangan: *S\_F5(S)* = Skor kalimat dari fitur *sentence similarity*

= jumlah nilai *sentence*

*similarity* dalam kalimat

= jumlah nilai *sentence*

*similarity* tertinggi yang ada dalam kalimat yang ada di teks

### 2.4.6 *Proper Noun*

*Proper noun adalah* penggolongan kata benda secara spesifik. Contohnya adalah nama institusi, organisasi, hari, bulan, bangsa, agama, dan tempat. *Proper noun* selalu menggunakan huruf kapital di awal katanya. Biasanya kalimat yang berisi kata benda yang lebih *proper* merupakan salah satu kalimat yang penting dan kemungkinan besar termasuk dalam ringkasan dokumen. Skor untuk fitur ini dihitung dengan membandingkan rasio dari jumlah *proper noun* dalam kalimat dengan panjang kalimat.

(7)

Keterangan: *S\_F6(S)* = Skor kalimat dari fitur *proper noun*

*No.Word occurring in S* = Jumlah *proper noun* dalam kalimat

= Jumlah kata dalam kalimat

### 2.4.7 *Thematic Word*

Kata yang sering muncul dalam sebuah teks biasanya adalah salah satu kata yang penting. Kata tersebut disebut dengan kata tematik. Apabila ada sebuah kalimat yang memiliki jumlah kata tematik yang tinggi ada kemungkinan bahwa kalimat tersebut adalah kalimat topik dalam dokumen. Jumlah kata tematik menunjukkan kata-kata dengan kemungkinan relativitas maksimal. Penelitian ini menggunakan sepuluh kata yang paling sering muncul untuk dipertimbangkan sebagai tematik. Skor untuk fitur ini dihitung dengan membandingkan rasio dari jumlah kata tematik dalam kalimat dengan ringkasan maksimum kata tematik dalam kalimat.

(8)

Keterangan: *S\_F7(S)* = skor kalimat dari fitur *thematic word*

= jumlah *thematic word* dalam kalimat

= nilai tertinggi dari jumlah *thematic word* dalam kalimat yang ada di dalam teks

### 2.4.8 *Numerical Data*

Sebuah kalimat yang berisi data numerik dianggap penting dan kemungkinan besar termasuk dalam ringkasan dokumen. Data numerik dianggap penting karena biasanya memberikan informasi kuantitas. Skor dari fitur ini dihitung dengan membandingkan rasio dari jumlah kata *numerical data* dalam kalimat dengan jumlah kata dari kalimat tesebut.

(9)

Keterangan: *S\_F8(S)* = skor kalimat dari fitur *numerical data*

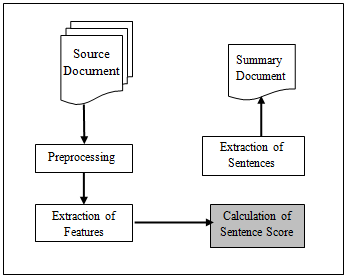
= jumlah *numerical data* dalam kalimat

= Jumlah kata dalam kalimat

## 2.5 *General Statistic Method*

Delapan fitur yang telah dijelaskan sebelumnya diterapkan kepada masing-masing kalimat. Hasil skor dari masing-masing kalimat tersebut dijadikan acuan dalam menerapkan *General Statistic Method*. *General Statistic Method* akan menentukan kalimat mana yang akan digunakan sebagai kalimat rangkuman. Jumlah kalimat yang akan ditampilkan bergantung kepada presentasi *compression rate* yang akan digunakan. Berikut langkah-langkah penting dalam *General Statistic Method*:

* + 1. Baca dokumen sumber ke dalam sistem.
    2. Lakukan tahap *preprocessing* terhadap masing-masing kalimat.
    3. Hitung skor masing-masing fitur setiap kalimat dengan menerapkan delapan fitur pada sub-bab 2.4.
    4. Hitung *sentence weight* dari masing-masing kalimat dengan mengkalkulasi hasil skor dari delapan fitur.
    5. Pilih kalimat-kalimat dengan *sentence weight* tertinggi. Jumlah kalimat ditentukan berdasarkan *compression rate.*



Gambar 2.2 Peringkasan Teks berdasarkan *General Statistic Method*

Peringkasan teks berdasarkan *General Statistic Method* mengacu pada *sentence weight* dari masing-masing kalimat. *Sentence weight* adalah jumlah dari seluruh skor fitur yang telah diterapkan pada masing-masing kalimat. Dalam menghitung *sentence weight* diterapkan fungsi di bawah ini.

Dimana *Score(S)* adalah nilai *sentence weight* dari kalimat S*.* Sementara *S\_Fk(S)* adalah skor fitur dari kalimat S.

## 2.6 Metode Evaluasi

Dalam mengevaluasi hasil ringkasan dibutuhkan metode khusus untuk menentukan kualitas hasil ringkasan. Menurut (Zaman & Winarko, 2011) metode untuk melakukan evaluasi terhadap hasil dari ringkasan secara umum dibagi dua, yaitu:

1. Ekstrinsik

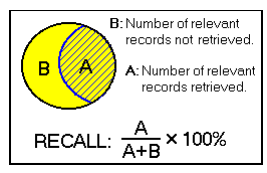
Metode evaluasi ekstrinsik adalah menghitung efektivitas dan akseptabilitas dari hasil ringkasan untuk tugas-tugas tertentu, misalnya *assessment* terhadap hasil ringkasan.

2. Intrinsik

Metode evaluasi intrinsik adalah evaluasi yang dilakukan oleh sistem peringkas itu sendiri, misalnya menggunakan *F- score*. Evaluasi ini difokuskan pada tingkat koheren dan informatif dari hasil ringkasan.

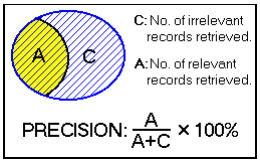
### 2.6.1 *Recall* dan *Precision*

Dalam menerapkan fungsi *recall, precision,* dan *f-score* dibutuhkan ringkasan ideal. Ringkasan ideal ringkasan yang dibuat oleh manusia dengan menerapkan tahap-tahap pembuatan ringkasan (Sartuni, Arifin, & Hutagalung, 2015). Ringkasan tersebut akan dijadikan acuan dalam menentukan nilai *recall* dan *precision*.

**

Gambar 2.3 Cara Menghitung *Recall*

*Recall* (R) adalah rasio dari jumlah kalimat dari hasil ringkasan sistem yang relevan dengan ringkasan ideal dibandingkan dengan jumlah kalimat yang ada di ringkasan ideal (Davis & Goadrich, 2006). Sehingga semakin banyak ringkasan sistem yang relevan dengan ringkasan ideal maka semakin tinggi pula nilai dari *recall.* Nilai dari *recall* ditampilkan dalam bentuk persentasi.

**

Gambar 2.4 Cara Menghitung *Precision*

*Precision* (P) adalah rasio dari jumlah kalimat dari hasil ringkasan sistem yang relevan dengan ringkasan ideal dibandingkan dengan jumlah kalimat yang ada di ringkasan sistem (Davis & Goadrich, 2006). Sehingga semakin banyak ringkasan sistem yang relevan dengan ringkasan ideal maka semakin tinggi pula nilai dari *precision.* Nilai dari *precision* ditampilkan dalam bentuk persentasi.

### 2.6.2 F-*Score*

Dalam menentukan *f-score* dibutuhkan hasil nilai dari *precision* dan *recall. F*-*score* adalah ukuran gabungan yang menggabungkan nilai dari *precision* dan *recall*. Di bawah ini merupakan fungsi yang dibutuhkan dalam menetukan *f-score.*

F = *f-score* P = *precision* R = *recall*

*F-score* merupakan rasio perbandingan dari dua kali jumlah perkalian nilai *precision* dan *recall* dibandingkan dengan jumlah dari nilai *precision* dan *recall* (Powers, 2007).

# BAB III

# METODE PENELITIAN

Bab ini menjelaskan tentang metodologi penelitian yang akan diterapkan pada penelitian ini. Bab ini membahas tentang objek penelitian, alat penelitian, tahapan penelitian, teknik pengumpulan data, dan kerangka kerja penelitian.

## 3.1 Objek Penelitian

Objek penelitian yang digunakan dalam penelitian ini adalah dokumen tunggal berbahasa Indonesia. Dokumen yang digunakan adalah dokumen yang hanya berisi artikel teks dengan penggunaan bahasa Indonesia yang baik dan benar. Dokumen tersebut harus menggunakan bahasa dengan ejaan yang telah disempurnakan agar dapat diproses oleh aplikasi peringkas teks otomatis yang akan dibangun dalam penelitian ini.

Artikel yang digunakan dalam penelitian ini adalah opini-opini yang dipublikasikan dalam situs berita populer. Situs berita populer yang dimaksud adalah situs berita berbahasa Indonesia. Ada lima situs berita yang akan dijadikan acuan yaitu Tempo, Kompas, Republika, Media Indonesia, dan Detik. Jumlah opini yang dikumpulkan adalah 50 artikel. Lima puluh artikel tersebut yang akan dijadikan *data set* penelitian. Jumlah artikel yang digunakan dalam penelitian ini diputuskan berdasarkan jumlah dokumen yang pada umumnya digunakan pada penelitian peringkasan artikel berbahasa Indonesia lainnya.

## 3.2 Alat Penelitian

Dalam melakukan penelitian ini digunakan berbagai *tools* yang mendukung peringkasan teks. Berikut daftar perangkat lunak yang digunakan:

1. Sistem Operasi *Windows* 8 64bit
2. Bahasa Pemrograman PHP versi 5.6.14
3. *Web server* XAMPP versi 3.2.2
4. *Notepad++* versi 7.3.3
5. *Web Browser Mozilla Firefox* versi 47.0.1

## 3.3 Tahapan Penelitian

Dalam melakukan penelitian ini ada empat tahap yang akan dilakukan. Empat tahap tersebut adalah sebagai berikut:

1. Studi Literatur

Tahap ini adalah tahap yang dilakukan untuk menentukan latar belakang dan rumusan masalah. Setelah latar belakang masalah ditemukan, peneliti melakukan studi literatur yang terkait dengan peringkasan teks secara otomatis dan metode pengembangan aplikasi. Pada studi literatur penelitian ini mengadopsi *General Statistic Method* dalam peringkasan dokumen tunggal berbahasa Indonesia.

1. Analisis dan Penerapan

Pada tahap ini dilakukan analisis dalam menerapkan *General Statistic Method* pada peringkasan dokumen tunggal berbahasa Indonesia.

1. Evaluasi

Pada tahap ini peneliti mengevaluasi hasil ringkasan dari sistem dengan membandingkan ringkasan ideal dengan mengitung hasil dari *precision, recall*, dan *f-measure.*

1. Finalisasi penulisan laporan

Tahapan terakhir dari penelitian ini adalah finalisasi penulisan laporan. Tahap ini dilakukan setelah penerapan metode dan pembangunan aplikasi selesai.

## 3.4 Teknik Pengumpulan Data

Data set yang diperlukan dalam penelitian ini terdiri dari dua data set yaitu data set artikel asli dan ringkasan ideal. Artikel asli adalah artikel sumber yang belum diringkas. Sementara ringkasan ideal adalah artikel asli yang sudah diringkas oleh manusia.

Artikel asli yang dikumpulkan pada penelitian ini adalah 50 artikel opini. Pengumpulan 50 artikel tersebut dilakukan secara manual. Artikel-artikel tersebut diambil melai lima situs berita populer yaitu Tempo, Kompas, Republika, Media Indonesia, dan Detik.

Setelah mengumpulkan artikel asli, penulis meringkas artikel-artikel tersebut. Artikel asli yang sudah diringkas akan menjadi ringkasan ideal. Peringkasan artikel tersebut dilakukan secara manual dengan menerapkan tahapan peringkasan. Ringkasan ideal tersebut kemudian akan dibandingkan dengan ringkasan yang telah dihasilkan aplikasi untuk menerapkan metode evaluasi. Sehingga nilai *precision, recall,* dan *f-score* dari ringkasan dapat ditemukan.

## 3.5 Analisis Penerapan Metode

Dalam menerapkan metode *General Statistic Method* dibutuhkan analisis yang mendalam. Untuk membantu menggambarkan analisis tersebut penulis merancang *flow chart, use case, sequence diagram*, dan analisis *class diagram.*

### 3.5.1 *Flowchart*



Gambar 3.1 *Flowchart* dari Proses Peringkasan Teks

Dalam menerapkan metode *General Statistical Method* dalam peringkasan teks penelitian ini menerapkan empat langkah seperti dalam Gambar 3.1. Langkah pertama adalah *preprocessing*, langkah kedua adalah pemberian skor atau *sentence scoring*, langkah ketiga adalah menjumlah nilai skor dari kalimat, dan lagkah terakhir adalah pemilihan kalimat.

### 3.5.2 *Use Case*

Berikut *use case* aktor pengguna dari proses peringkasan teks dalam penelitian ini.



Gambar 3.2 *Use Case* actor Pengguna

Gambar 3.2 menunjukkan *use case* dari aktor pengguna. Pengguna dapat menggunakan dua fitur dalam sistem ini yaitu memasukkan artikel dan melihat ringkasan. Pada saat memasukkan artikel pengguna mengunggah satu dokumen tunggal berbahasa Indonesia ke dalam sistem. Pasa saat melihat ringkasan, pengguna dapat melihat hasil ringkasan dari artikel yang sudah diringkas sistem.

### 3.5.3 *Sequence Diagram*

Dalam peringkasan teks terdapat dua tahap *sewuence diagram*. Yaitu *preprocessing* dan *sentence scoring.*



Gambar 3.3 *Sequence Diagram Preprocessing*

Gambar 3.3 menunjukkan proses *preprocessing*. Setelah pengguna melakukan input artikel, artikel dipecah menjadi beberapa kalimat. Kalimat yang telah dipecah tersebut diproses menggunakan tiga fungsi yaitu *tokenizing*, *stop word removal*, dan *stemming*.



Gambar 3.4 *Sequence Diagram Sentence Scoring*

Gambar 3.4 menunjukkan *sequence diagram sentence scoring* dimana kalimat yang sudah diproses di tahap *preprocessing* diberi skor masing-masing. Setiap kalimat akan diterapkan delapan fungsi. Yaitu, *title feature, sentence length, term weight, sentence position, sentence similarity, proper noun, thematic word,* dan *numerical data*. Jumlah dari setiap fungsi menjadi jumlah skor kalimat. Kalimat akan diurutkan berdasarkan jumlah skor. Kalimat-kalimat dengan jumlah skor tertinggi menjadi kalimat yang akan masuk dalam ringkasan.

### 3.5.4 *Class Diagram*



Gambar 3.5 *Class Diagram*

Pada Gambar 3.5 menjelaskan bahwa terdapat 3 *class* yang digunakan yaitu *Article*, *Sentence*, dan *Word*. Berikut penjelasan dari masing-masing *class*:

1. *Article*

*Class Article* merupakan *class* induk yang menyimpan artikel asli dan ringkasannya. Selain itu *class* tersebut juga menyimpan detail-detail lainnya seperti kata-kata tematik, jumlah kata terbanyak, jumlah kemiripan kalimat dan lain-lain. *Class Article* memiliki *subclass* bernama *class Sentence.*

1. *Sentence*

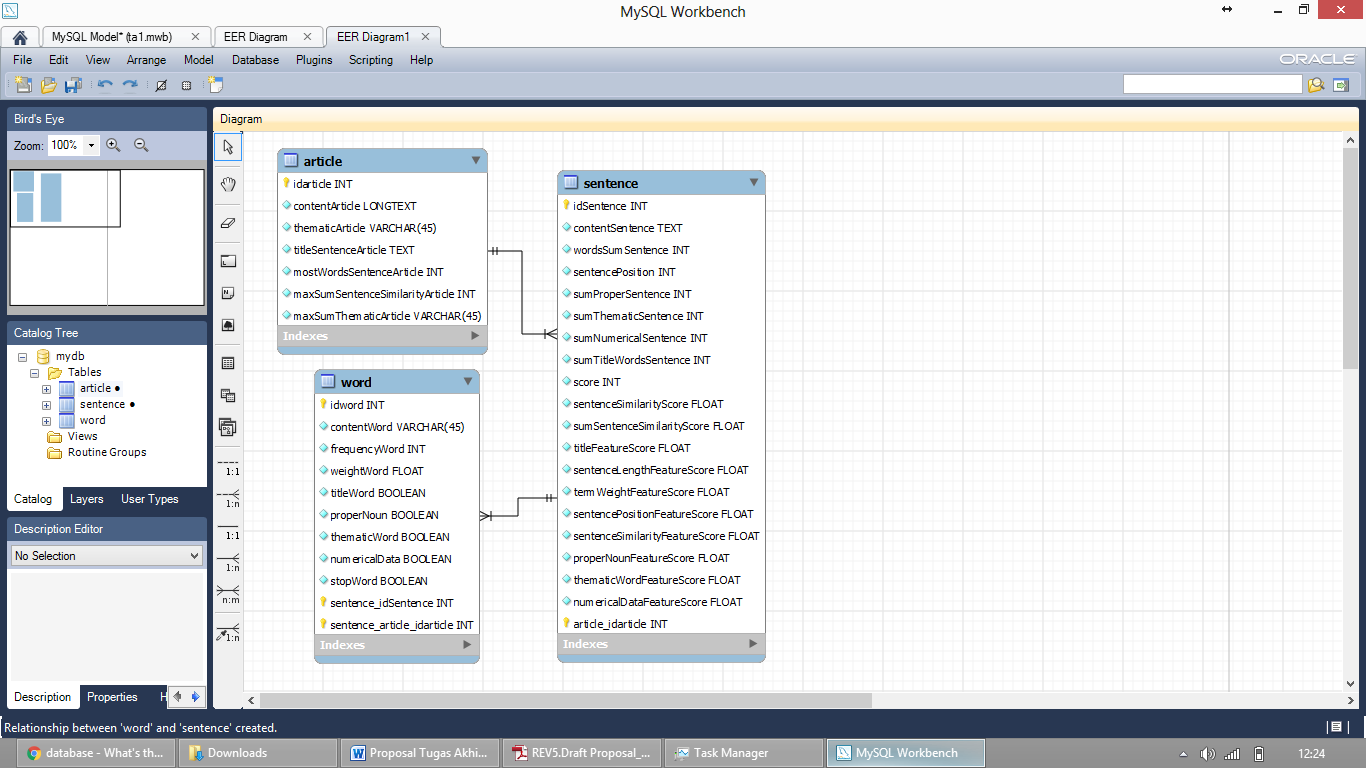
*Class* *Sentence* merupakan *class* yang berisi kalimat-kalimat yang sudah dipecah dari artikel asli. *Class* ini juga menyimpan nilai-nilai variabel yang berfungsi untuk menduku perhitungan skor kalimat. *Class Sentence* merupakan *subclass* dari *superclass class Article.*

1. *Word*

*Class Word* merupakan *class* yang terdiri darikata-kata yang terdapat dalam kalimat. Kata-kata tersebut akan memiliki karakteristik seperti *proper noun*, *thematic word, numerical data* dan lainnya yang mendukung perhitungan skor. *Class* ini menyimpan nilai skor dari setiap kalimat. Dalam *class* ini juga menyimpan fungsi-fungsi yang digunakan dalam perhitungan *score*. *Class Word* merupakan *subclass* dari *superclass class Sentence.*

### 3.5.5 Relasi Basis Data

Setiap iterasi peringkasan yang dilakukan akan disimpan dalam *database*. Terdapat empat kelas yang disimpan dalam *database* yaitu *article, sentence,* and *word*. Relasi *database* dalam sistem peringkasan dalam penelitian ini adalah sebagai berikut:



Gambar 3.6 Relasi *Database*

## 3.6 Kerangka Penelitian

Pada penelitian ini memiliki tahapan yang dilakukan dalam alur di bawah ini.

Gambar 3.7 Kerangka Penelitian

Tahap awal yang dilakukan dalam penelitian ini adalah studi kebutuhan. Pada studi kebutuhan peneliti harus menemukan mengapa penelitian dengan topik ini harus dilakukan. Setelah melakukan studi kebutuhan penulis melakukan studi literatur. Pada tahap ini peneliti mempelajari berbagai literatur yang berhubungan dengan topik penelitian. Setelah itu peneliti memutuskan metode apa yang akan diterapkan dalam penelitian ini. Setelah memutuskan metode yang digunakan peneliti mempelajari lebih dalam bagaimana metode tersebut diterapkan. Setelah itu, peneliti melakukan analisis dari penerapan metode tersebut dalam penelitian. Setelah melakukan analisis, peneliti menerapkan metode tersebut ke dalam sistem. Dalam menerapkan metode peneliti perlu melakukan testing. Ketika hasil dari sistem tidak sama dengan hasil yang seharusnya dari metode. Maka dilakukan penerapan ulang dengan melakukan perubahan-perubahan. Dua tahap tersebut dilakukan berulang-ulang sampai hasil dari sistem sesuai dengan yang diharapkan. Ketika hasil sudah sesuai tahap selanjutnya yang dilakukan adalah evaluasi. Pada tahap ini peneliti membandingan hasil ringkasan sistem dengan ringkasan ideal. Setelah menemukan nilai perbandingan maka penelitian bisa dilanjutkan ke tahap selanjutnya. Tahap terakhir adalah finalisasi pelaporan. Di tahap ini peneliti mendokumentasikan dan menuliskan laporan dari penelitian.

## 3.7 Rencana Kerja Penelitian

Adapun rencana kerja penelitian yang direncanakan penulis adalah sebagai berikut:

**Table 3.1 *Timeline* Kerja**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| No | Jenis Kegiatan | April | | | | Mei | | | | Juni | | | | Juli | | | | Agustus | | | |
| 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| 1 | Penentuan Topik |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 2 | Explorasi Topik |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 3 | Penyusunan Proposal |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 4 | Studi Literatur |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 5 | Analisis |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 6 | Implementasi |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 7 | Evaluasi |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 8 | Penyusunan Laporan TA |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

Dalam mengerjakan penelitian ini, peneliti membutuhkan 5 bulan. Pada bulan April hingga Mei peneliti melakukan penentuan topik, explorasi topik, penyusunan proposal, studi literatur. Pada bulan Mei hingga Agustus peneliti melakukan analisis, implementasi, evaluasi, penyusunan laporan tugas akhir.

# BAB IV

# IMPLEMENTASI

## 4.1 Implementasi Sistem

Pada bab ini akan dilakukan implementasi dan pengujian terhadap sistem yang dibangun, yaitu peringkas teks otomatis. Hasil perancangan pada tahap sebelumnya kemudian diimplementasikan ke dalam bahasa pemrograman. Setelah tahap implementasi selesai, maka dilakukan tahap pengujian terhadap sistem peringkas teks otomatis dimana akan dilihat kekurangan-kekurangan pada sistem yang baru untuk selanjutnya di adakan pengembangan sistem.

### 4.1.1 Implementasi Perangat Keras

Implementasi perangkat keras merupakan spesifikasi hardware yang digunakan dalam mengimplementasikan sistem yang dibuat. Berikut adalah spesifikasi hardware untuk menjalankan sistem peringkas teks otomatis :

1. Processor intel CORE i5
2. Harddisk 500 GB
3. Memory 2 GB
4. VGA 1015 MB
5. Mouse Wireless Logitech
6. Keyboard QWERTY Creative
7. USB 3 buah
8. Monitor 14 inch

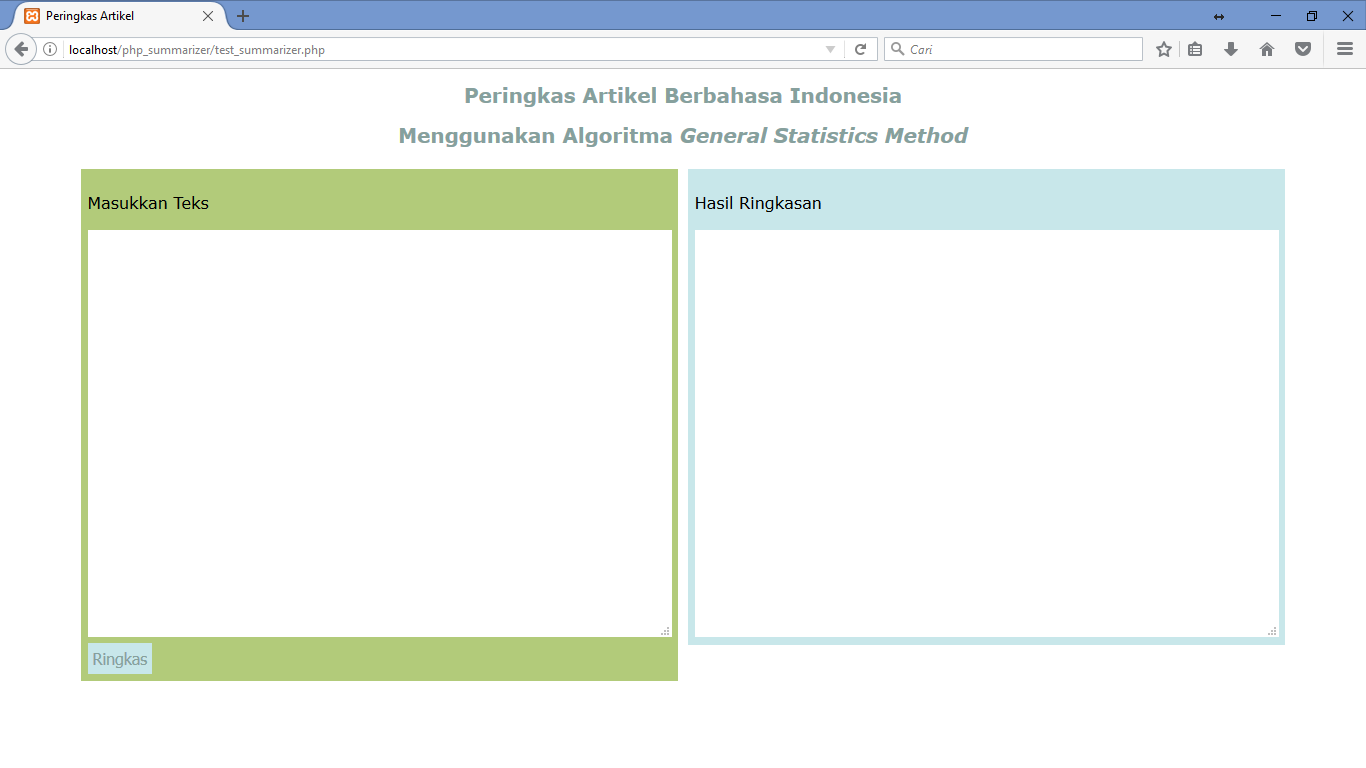
### 4.1.2 Implementasi Perangat Lunak

Implementasi perangkat lunak merupakan software yang digunakan dalam mengimplementasikan sistem yang dibuat. Berikut adalah software dalam pembangunan sistem peringkas teks otomatis :Processor intel CORE i5

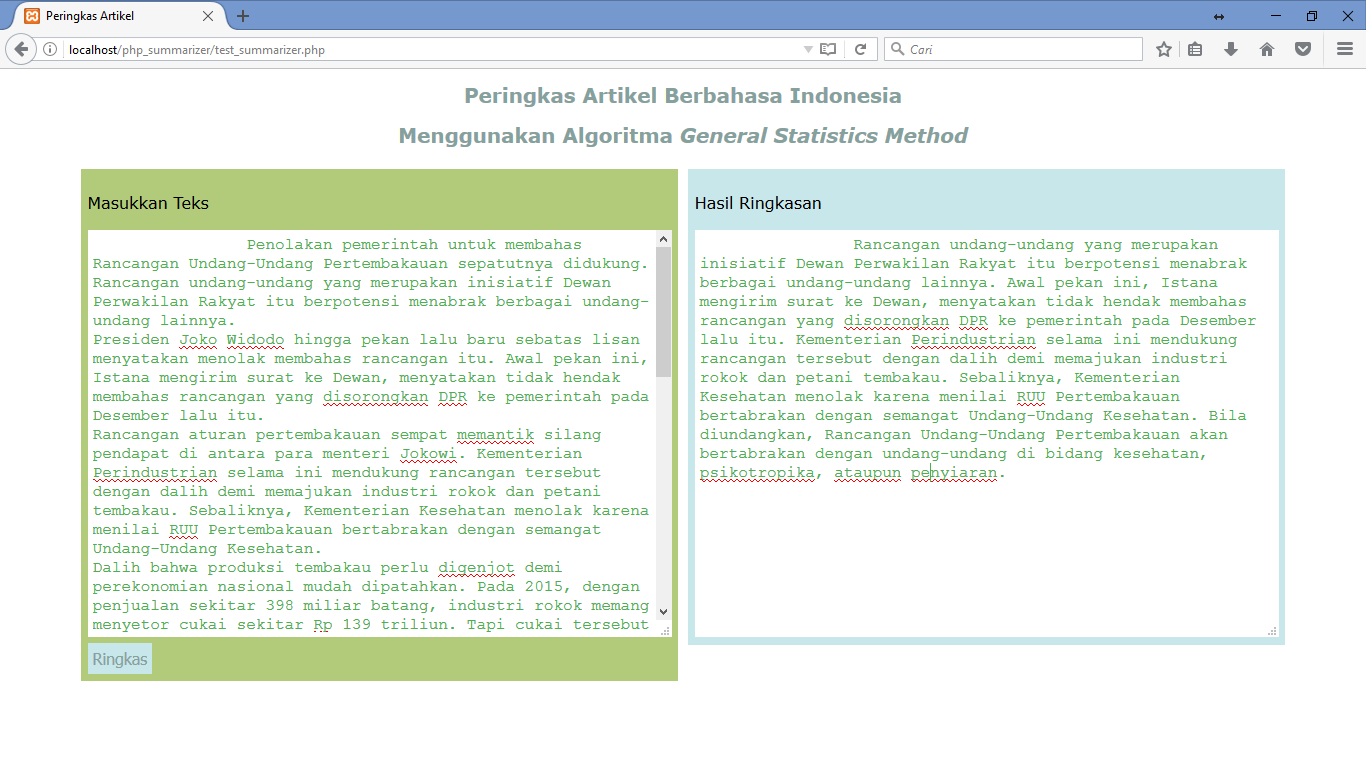
1. Windows 7 64bit
2. Web Server XAMPP ver 3.2.2
3. Bahasa Pemograman PHP ver 5.6.14
4. Notepad++ 7.3.3
5. Mozilla Firefox ver 54.0.1

### 4.1.2 Implementasi Antar Muka

Bagian ini berisi implementasi antarmuka dari penelitian ini. Antarmuka pada pembangunan perangkat lunak dalam penelitian ini hanya memiliki satu halaman. Dalam halaman tersebut, terdapat dua buah kotak teks dan sebuah tombol. Kotak teks yang pertama digunakan untuk menginput teks yang akan diringkas, sementera kotak teks kedua digunakan untuk melihat hasil teks yang telah diringkas oleh sistem. Tombol bertuliskan ringkas digunakan untuk melakukan proses peringkasan.



Gambar 4.1 Tampilan Antarmuka Awal



Gambar 4.2 Tampilan Antarmuka setelah Teks Diringkas

# LAMPIRAN

## LAMPIRAN I DATA SET

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| No | Judul | Penulis | Sumber | URL |
| 1 | Adakah Perusahaan Nasional yang Usianya di Atas 100 Tahun | Prof Rhenald Kasali | KOMPAS | http://bisniskeuangan.kompas.com/read/2015/12/16/051900626/Adakah.Perusahaan.Nasional.yang.Usianya.di.Atas.100.Tahun.?page=all |
| 2 | Hati-hati Sudden Shift, Fenomena Perubahan Abad Ke-21 | Rhenald Kasali | KOMPAS | http://bisniskeuangan.kompas.com/read/2015/08/24/054110726/Hati-hati.Sudden.Shift.Fenomena.Perubahan.Abad.Ke-21?page=all |
| 3 | Mencerdaskan Kehidupan Bangsa | Yudi Latif | KOMPAS | http://nasional.kompas.com/read/2015/05/05/15100041/Mencerdaskan.Kehidupan.Bangsa |
| 4 | Mengapa Kita Pesimistis | Anton Hendranata | KOMPAS | http://bisniskeuangan.kompas.com/read/2015/05/30/191958126/Mengapa.Kita.Pesimistis?page=all |
| 5 | Dana Desa Pacu Denyut Ekonomi | Ferry Kisihandi | Republika | http://www.republika.co.id/berita/koran/opini-koran/16/04/25/o66nne1-dana-desa-pacu-denyut-ekonomi |
| 6 | Mari Awasi RUU Penyiaran | Muhammad Dong | Republika | http://www.republika.co.id/berita/koran/opini-koran/16/04/28/o6c3k611-mari-awasi-ruu-penyiaran |
| 7 | Mengenalkan Logo Halal Sejak Dini | Eko Triyanto | Republika | http://www.republika.co.id/berita/koran/opini-koran/16/04/27/o6a64512-mengenalkan-logo-halal-sejak-dini |
| 8 | Momentum Perbaikan Dunia Pendidikan | Republika | Republika | http://www.republika.co.id/berita/koran/opini-koran/16/05/02/o6jfg312-momentum-perbaikan-dunia-pendidikan |
| 9 | Perlakuan Istimewa Koruptor | Mustakim | Republika | http://www.republika.co.id/berita/koran/opini-koran/16/04/28/o6c3k613-perlakuan-istimewa-koruptor |
| 10 | Saatnya Madrasah Jadi Pilihan | Republika | Republika | http://www.republika.co.id/berita/koran/opini-koran/16/05/02/o6jfg311-saatnya-madrasah-jadi-pilihan |
| 11 | Sekolah Negeri untuk Anak Pejabat | Sri Nurhidayah | Republika | http://www.republika.co.id/berita/koran/opini-koran/16/05/02/o6jfg310-sekolah-negeri-untuk-anak-pejabat |
| 12 | Seputar Hadis Puasa Rajab | Ali Mustafa Yaqub | Republika | http://www.republika.co.id/berita/koran/opini-koran/16/04/29/o6dy876-seputar-hadis-puasa-rajab |
| 13 | Transformasi Kerajaan Era Raja Salman | Republika | Republika | http://www.republika.co.id/berita/koran/opini-koran/16/04/29/o6dy877-transformasi-kerajaan-era-raja-salman |
| 14 | Hilangnya Rasa Aman | Ayu Ingke | Republika | http://www.republika.co.id/berita/koran/opini-koran/17/01/18/ojyrg61-hilangnya-rasa-aman |
| 15 | Menyoal Pidato Megawati | Chaya Yuliatri | Republika | http://www.republika.co.id/berita/koran/opini-koran/17/01/18/ojyrg6-menyoal-pidato-megawati |
| 16 | Peta Kemiskinan Indonesia | Yusuf Wibisono | Republika | http://www.republika.co.id/berita/koran/opini-koran/17/01/18/ojyrg62-peta-kemiskinan-indonesia |
| 17 | Penelantaran Anak | Muryani | Republika | http://www.republika.co.id/berita/koran/opini-koran/17/01/17/ojwws620-penelantaran-anak |
| 18 | Roh Jihad Ekonomi Ummat | Dahnil Anzar Simanjuntak | Republika | http://www.republika.co.id/berita/koran/opini-koran/17/01/17/ojwws619-roh-jihad-ekonomi-umat |
| 19 | Debat Pilkada dan Kapasitas Kepemimpinan | Bimo Joga Sasongko | Republika | http://www.republika.co.id/berita/koran/opini-koran/17/01/17/ojwws622-debat-pilkada-dan-kapasitas-kepemimpinan |
| 20 | Malari, Abe, dan Pesan Kemandirian | Republika | Republika | http://www.republika.co.id/berita/koran/opini-koran/17/01/17/ojwws621-malari-abe-dan-pesan-kemandirian |
| 21 | TDL Naik Rakyat Tercekik | Yulia Dwi | Republika | http://www.republika.co.id/berita/koran/opini-koran/17/01/16/ojv246-tdl-naik-rakyat-tercekik |
| 22 | Dana Haji untuk Bangun Infrastruktur | Yeni Mulati | Republika | http://www.republika.co.id/berita/koran/opini-koran/17/01/16/ojv2465-dana-haji-untuk-bangun-infrastruktur |
| 23 | Peluang Agropreneur Muda | Oeng Anwarudin | Republika | http://www.republika.co.id/berita/koran/opini-koran/17/01/16/ojv2461-peluang-agropreneur-muda |
| 24 | Antisipasi Gejolak Timteng | Republika | Republika | http://www.republika.co.id/berita/koran/opini-koran/17/01/16/ojv2463-antisipasi-gejolak-timteng |
| 25 | Orang Hilang Sudah Kembali | Wulandari | Republika | http://www.republika.co.id/berita/koran/opini-koran/17/01/14/ojrl427-orang-hilang-sudah-kembali |
| 26 | Menggugat Ekspor Mineral Mentah | Ardy Muawin | Republika | http://www.republika.co.id/berita/koran/opini-koran/17/01/14/ojrl429-menggugat-ekspor-mineral-mentah |
| 27 | Membangun Ekonomi Desa | Dewi Hutabarat | Republika | http://www.republika.co.id/berita/koran/opini-koran/17/01/14/ojrl428-membangun-ekonomi-desa |
| 28 | Homo Sapiens yang Sakit | Ahmad Baedowi | Media Indonesia | http://mediaindonesia.com/news/read/98323/homo-sapiens-yang-sakit/2017-03-27 |
| 29 | Siapakah Guru Indonesia? | Satia Prihatin Zein | Media Indonesia | http://mediaindonesia.com/news/read/98322/siapakah-guru-indonesia/2017-03-27 |
| 30 | Nyepi dan Altruisme | IBG Yudha Triguna | Media Indonesia | http://mediaindonesia.com/news/read/98045/nyepi-dan-altruisme/2017-03-25 |
| 31 | Dramaturgi Wakil Rakyat | W Wempy Hadir | Media Indonesia | http://mediaindonesia.com/news/read/98043/dramaturgi-wakil-rakyat/2017-03-25 |
| 32 | Sengkarut Taksi Online | Tulus Abadi | Media Indonesia | http://mediaindonesia.com/news/read/97932/sengkarut-taksi-online/2017-03-24 |
| 33 | Menanggulangi Tb di Masyarakat | Fainal Wirawan | Media Indonesia | http://mediaindonesia.com/news/read/97883/menanggulangi-tb-di-masyarakat/2017-03-24 |
| 34 | Akreditasi bukan Sekadar Kewajiban Regulasi | Asep Saefuddin | Media Indonesia | http://mediaindonesia.com/news/read/97690/akreditasi-bukan-sekadar-kewajiban-regulasi/2017-03-23 |
| 35 | Quo Vadis Generasi Muda? | Muhammad Farid Fellow | Media Indonesia | http://mediaindonesia.com/news/read/97689/quo-vadis-generasi-muda/2017-03-23 |
| 36 | Pemilu Timor Leste, Pertarungan Masa Depan | Ganewati Wuryandari | Media Indonesia | http://mediaindonesia.com/news/read/97526/pemilu-timor-leste-pertarungan-masa-depan/2017-03-22 |
| 37 | Mahir, Mahar, dan Murah | Lasarus Jehamat | Media Indonesia | http://mediaindonesia.com/news/read/97517/mahir-mahar-dan-murah/2017-03-22 |
| 38 | Tantangan Populisme di Eropa | Teuku Rezasyah | Media Indonesia | http://mediaindonesia.com/news/read/97352/tantangan-populisme-di-eropa/2017-03-21 |
| 39 | Perang Melawan Pedofilia | Bagong Suyanto | Media Indonesia | http://mediaindonesia.com/news/read/97347/perang-melawan-pedofilia/2017-03-21 |
| 40 | Myth of Happiness | Ahmad Baedowi | Media Indonesia | http://mediaindonesia.com/news/read/97204/myth-of-happiness/2017-03-20 |
| 41 | Membangun Pendidikan ala Finlandia | Marthunis | Media Indonesia | http://mediaindonesia.com/news/read/97203/membangun-pendidikan-ala-finlandia/2017-03-20 |
| 42 | Pemilu Timor Leste dan Respons Indonesia | Arya Sandhiyudha | Media Indonesia | http://mediaindonesia.com/news/read/96943/pemilu-timor-leste-dan-respons-indonesia/2017-03-18 |
| 43 | Di Balik Krisis Hubungan Turki-UE | Smith Alhadar | Media Indonesia | http://mediaindonesia.com/news/read/96940/di-balik-krisis-hubungan-turki-ue/2017-03-18 |
| 44 | Implikasi Hukum Kerja Sama Plurilateral | Asep N Mulyana | Media Indonesia | http://mediaindonesia.com/news/read/96840/implikasi-hukum-kerja-sama-plurilateral-1/2017-03-17 |
| 45 | Demokrasi Warga | W Wempy Hadir | Media Indonesia | http://mediaindonesia.com/news/read/110127/alarm-untuk-birokrat-dan-politisi/2017-06-22 |
| 46 | Pencurian Berkas Perkara MK | Tempo | Tempo | https://www.tempo.co/read/opiniKT/2017/03/24/13872/pencurian-berkas-perkara-mk |
| 47 | Konsistensi Presiden Ihwal Rokok | Tempo | Tempo | https://www.tempo.co/read/opiniKT/2017/03/23/13868/konsistensi-presiden-ihwal-rokok |
| 48 | Serius dengan Air Bersih | Tempo | Tempo | https://www.tempo.co/read/opiniKT/2017/03/22/13867/serius-dengan-air-bersih |
| 49 | Jaringan Pedofilia Internasional | Tempo | Tempo | https://www.tempo.co/read/opiniKT/2017/03/24/13872/pencurian-berkas-perkara-mk |
| 50 | Regulasi Angkutan Online | Tempo | Tempo | https://www.tempo.co/read/opiniKT/2017/03/27/13875/regulasi-angkutan-online |

# DAFTAR PUSTAKA

Aristoteles. (2013). *Penerapan Algoritma Genetika pada Peringkasan Teks Dokumen Bahasa Indonesia.* Lampung: Jurusan Ilmu Komputer FMIPA Universitas Lampung.

Balbin, J. P. (2011). *Automatic Summarization of Senate Bills for Dissemination in a Social Network Environtment.* Manila: De La Salle University.

Bhatia, N., & Jaiswal, A. (2015). Trends in Extractive and Abstractive Techniques in Text Summarization. *International Journal of Computer Applications*, 1-5.

Davis, J., & Goadrich, M. (2006). *The Relationship Between Precision-Recall and ROC Curves.* Madison: Department of Computer Sciences and Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison.

Fathul, A. (2012). *Sistem Temu Kembali dengan Metode Vector Space Model.* Fakultas Teknologi Informasi Universitas Stikubank: Semarang.

Hamidy, G. M. (2011). *Pengembangan lanjut agregator berita bahasa indonesia menggunakan teknik content extraction, any-of classifier, dan term counting.* Depok: Fakultas Ilmu Komputer Universitas Indonesia.

Hu, M., Sun, A., & Lim. (2007). Comments Oriented Blog Summarization by Sentence Extraction. *Proceedings of the sixteenth ACM conference on Conference on information and knowledge management*, 901-904.

Keke, D., Chikita, R., & Prayogo, A. D. (2012). *Algoritma Nazief dan Andriani.* Yogyakarta: Program Studi S1 Ilmu Komputer Jurusan Ilmu Komputer dan Elektronika Fakultas Matematika dan Ilmu Pengetahuna Alam Universitas Gadjah Mada.

Meenaa, Y. K., & Gopalanib, D. (2015). Evolutionary Algorithms for Extractive Automatic Text. *International Conference on Intelligent Computing, Communication & Convergence*, 244-249.

Mustaqhfiri, M. (2011). *Peringkasan Teks Otomatis Berita Olahraga Berbahasa Indonesia Menggunakan Metode Maximum Marginal Revelance.* Malang: Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.

Natasha. (2010). *Pengembangan sistem agregator berita bahasa indonesia dengan klasifikasi berbasis naive bayes dan clustering berbasis non-negative matrix factorization.* Depok: Fakultas Ilmu Komputer Universitas Indonesia.

Nenkova, A., & McKeown, K. (2011). Automatic Summarization. *The Essence of Knowedge*, 104-105.

Powers, D. M. (2007). *Evaluation: From Precision, Recall and F-Factor to ROC, Informedness, Markedness & Correlation.* Adelaide: School of Informatics and Engineering Flinders University of South Australia.

Pusat Bahasa Depdiknas. (2008). *Kamus Besar Bahasa Indonesia.* Jakarta: Balai Pustaka.

Putra, A. A. (2015). *Implementasi Vector Space Model pada Artikel Berita Berbahasa Indonesia.* Coblong: Fakultas Teknik dan Ilmu Komputer Universitas Komputer Indonesia.

Sartuni, R., Arifin, Z., & Hutagalung, R. (2015). *Implementasi Bahasa Akademik.* Jakarta: Pustaka Mandiri.

Suanmali, L., Salim, N., & Binwahlan, M. S. (2011). Genetic Algorithm Based Sentence Extraction for Text Summarization. *International Journal of Innovative Computing*, 1-22.

Vincent, K. (2015). *Peringkas temporal dokumen berita berbahasa Indonesia menggnakan latent direchlet allocation.* Depok: Fakultas Ilmu Komputer Universitas Indonesia.

Wan, Y., & Ma, J. (2013). A Comprehensive Method for Text Summarization Based on Latent Semantic Analysis. *Springer-Verlag Berlin Heidelberg*, 394–401.

Zaman, B., & Winarko, E. (2011). Analisis Fitur Kalimat untuk Peringkas Teks Otomatis pada Bahasa Indonesia. *IJCSS-Indonesian Journal of Computing and Cybernetics System 5*.

Zuliarso, E., & Mustofa, K. (2009). Crawling Web Berdasarkan Ontology. *Jurnal Teknologi Informasi DINAMIK*, 105-112.