

**SISTEM *AUTOMATIC TEXT SUMMARIZATION*  
MENGUNAKAN ALGORITMA *TEXTRANK***

**SKRIPSI**

Oleh :  
**MUHAMMAD ADIB ZAMZAM**  
**NIM. 15650058**



**JURUSAN TEKNIK INFORMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM  
MALANG  
2019**

**SISTEM *AUTOMATIC TEXT SUMMARIZATION*  
MENGUNAKAN ALGORITMA *TEXTRANK***

**SKRIPSI**

**Diajukan kepada:  
Universitas Islam Negeri (UIN) Maulana Malik Ibrahim Malang  
Untuk Memenuhi Salah Satu Persyaratan Dalam  
Memperoleh Gelar Sarjana Komputer (S.Kom)**

**Oleh:  
MUHAMMAD ADIB ZAMZAM  
NIM. 15650058**

**JURUSAN TEKNIK INFORMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM  
MALANG  
2019**

**LEMBAR PERSETUJUAN**  
**SISTEM AUTOMATIC TEXT SUMMARIZATION**  
**MENGGUNAKAN ALGORITMA *TEXTRANK***

**SKRIPSI**

Oleh :  
**MUHAMMAD ADIB ZAMZAM**  
NIM. 15650058

Telah Diperiksa dan Disetujui untuk Diuji  
Tanggal : 6 Desember 2019

Dosen Pembimbing I

  
Dr. Cahyo Crysdian  
NIP. 19740424 200901 1 008

Dosen Pembimbing II

  
Khadijah Fahmi Hayati, M.Kom  
NIDT. 19900626 20160801 2 077

Mengetahui,  
Ketua Jurusan Teknik Informatika  
Fakultas Sains dan Teknologi  
Universitas Islam Negeri Maulana Malik Ibrahim Malang

  
Dr. Cahyo Crysdian  
NIP. 19740424 200901 1 008

**LEMBAR PENGESAHAN**  
**SISTEM AUTOMATIC TEXT SUMMARIZATION**  
**MENGGUNAKAN ALGORITMA TEXTRANK**

**SKRIPSI**

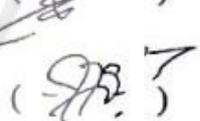
Oleh:  
**MUHAMMAD ADIB ZAMZAM**  
NIM. 15650058

Telah Dipertahankan di Depan Dewan Penguji  
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan  
untuk Memperoleh Gelar Sarjana Komputer (S.Kom)  
Pada Tanggal 6 Desember 2019

**Susunan Dewan Penguji**

- |                       |   |   |
|-----------------------|---|---|
| 1. Penguji Utama      | : | <u>Irwan Budi Santoso, M.Kom</u><br>NIP. 19770103 201101 1 004    |
| 2. Ketua Penguji      | : | <u>A'la Syauqi, M.Kom</u><br>NIP. 19771201 200801 1 007           |
| 3. Sekretaris Penguji | : | <u>Dr. Cahyo Crysdian</u><br>NIP. 19740424 200901 1 008           |
| 4. Anggota Penguji    | : | <u>Khadijah Fahmi Hayati, M.Kom</u><br>NIDT. 19900626201608012077 |

**Tanda tangan**

()  
()  
()  
()

Mengetahui,  
Ketua Jurusan Teknik Informatika  
Fakultas Sains dan Teknologi  
Universitas Islam Negeri Maulana Malik Ibrahim Malang



  
**Dr. Cahyo Crysdian**  
NIP. 19740424 200901 1 008

## PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan dibawah ini:

Nama : Muhammad Adib zamzam  
NIM : 15650058  
Fakultas/Jurusan : Sains dan Teknologi/Teknik Informatika  
Judul Skripsi : Sistem *Automatic Text Summarization* Menggunakan  
Algoritma *Textrank*

Menyatakan dengan sebenarnya bahwa Skripsi yang saya tulis ini benar-benar merupakan hasil karya sendiri, bukan merupakan pengambilalihan data, tulisan atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka.

Apabila dikemudian hari terbukti atau dapat dibuktikan Skripsi ini hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 21 November 2019

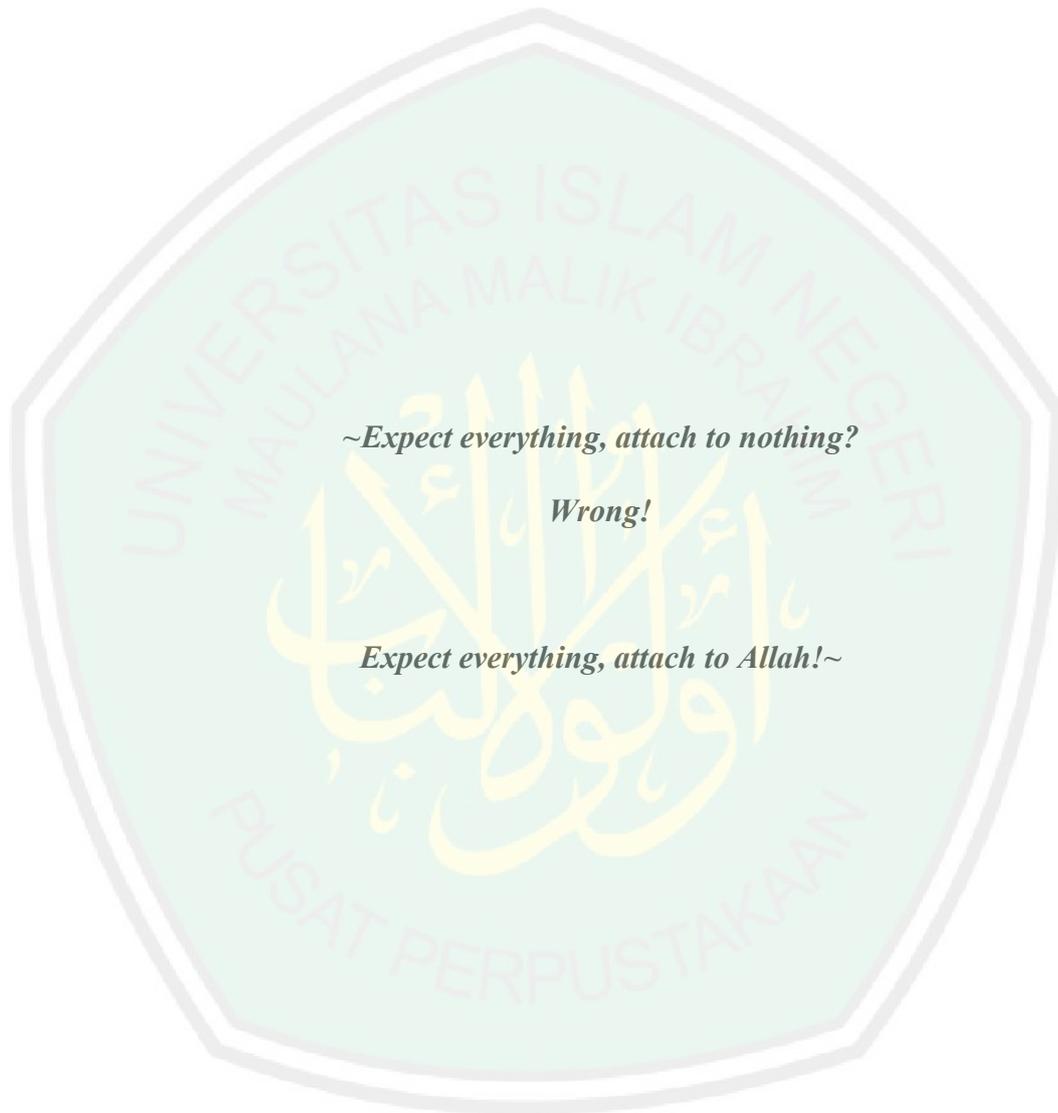
Yang membuat pernyataan,



Muhammad Adib zamzam

NIM. 15650058

## HALAMAN MOTTO



## HALAMAN PERSEMBAHAN

### الْحَمْدُ لِلَّهِ رَبِّ الْعَالَمِينَ

Untuk Bapak dan Ibuku tercinta, teman-teman Angkatan yang selalu memberikan pelajaran, semangat dan dukungan. Serta siapapun yang hendak mengembangkan pengetahuannya.



## KATA PENGANTAR

Assalamualaikum Wr. Wb.

Puji syukur kehadirat Allah SWT atas limpahan rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan studi dengan tepat waktu sekaligus menyelesaikan skripsi dengan baik dan lancar. Tujuan dari penyusunan skripsi ini guna memenuhi salah satu syarat untuk bisa menempuh ujian sarjana komputer pada Fakultas Sains dan Teknologi (FSAINTEK) Program Studi Teknik Informatika di Universitas Islam Negeri (UIN) Maulana Malik Ibrahim Malang. Didalam pengerjaan skripsi ini telah melibatkan banyak pihak yang sangat membantu dalam banyak hal. Oleh sebab itu, disini penulis sampaikan rasa terima kasih sedalam-dalamnya kepada:

1. Prof. Dr. Abdul Haris, M.Ag, selaku Rektor Universitas Islam Negeri (UIN) Maulana Malik Ibrahim Malang.
2. Dr. Sri Harini, M.Si, selaku Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri (UIN) Maulana Malik Ibrahim Malang.
3. Dr. Cahyo Crysdiyan, selaku Ketua Jurusan Teknik Informatika sekaligus Dosen Pembimbing I yang telah membimbing dalam penyusunan skripsi ini hingga selesai.
4. Khadijah Fahmi Hayati, M.Kom, selaku Dosen Pembimbing II yang telah membimbing dalam penyusunan skripsi ini hingga selesai.
5. Irwan Budi Santoso, M.Kom dan A'la Syauqi, M.Kom, selaku Dosen penguji yang telah memberikan banyak saran untuk kebaikan penulis.

6. Ayah, Ibu, dan adik tercinta yang telah banyak memberikan doa dan dukungan kepada penulis secara moril maupun materil hingga skripsi ini dapat terselesaikan.
7. Teman-teman Teknik Informatika Interface 2015 yang senantiasa memberi motivasi dan berjuang bersama selama menjadi mahasiswa.
8. Semua pihak yang telah banyak membantu dalam penyusunan skripsi ini yang tidak bisa penulis sebutkan satu persatu.

Penulis menyadari bahwa dalam penyusunan skripsi ini masih terdapat kekurangan dan penulis berharap semoga skripsi ini bisa memberikan manfaat kepada para pembaca khususnya bagi penulis secara pribadi.

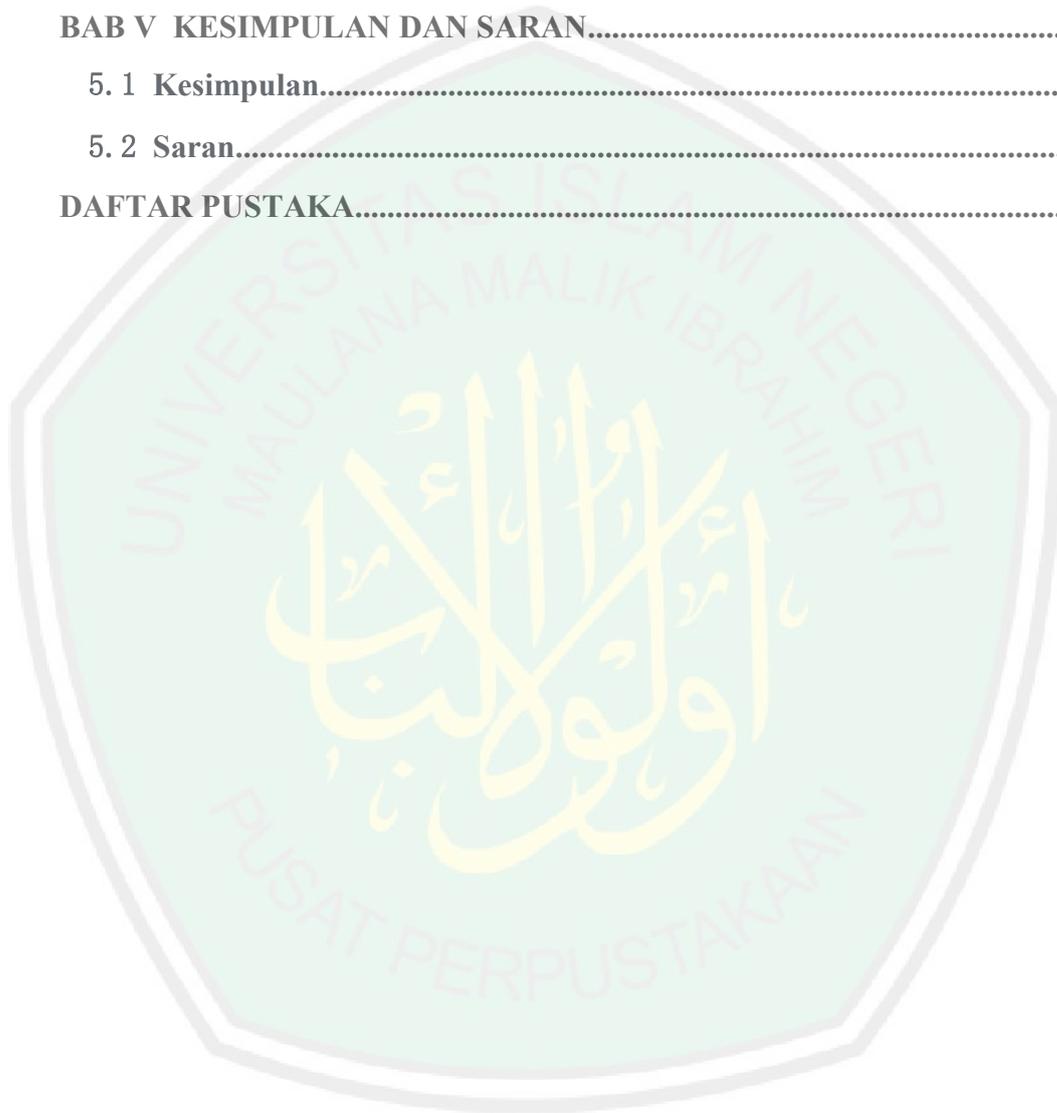
Malang, 20 Desember 2019

Penulis

## DAFTAR ISI

HALAMAN PENGANTAR.....	i
LEMBAR PERSETUJUAN.....	ii
LEMBAR PENGESAHAN.....	iii
PERNYATAAN KEASLIAN TULISAN.....	iv
HALAMAN MOTTO.....	v
HALAMAN PERSEMBAHAN.....	vi
KATA PENGANTAR.....	vii
DAFTAR ISI.....	ix
DAFTAR GAMBAR.....	xi
DAFTAR TABEL.....	xii
ABSTRAK.....	xiii
ABSTRACT.....	xiv
المُلخَص.....	xv
<b>BAB 1 PENDAHULUAN.....</b>	<b>1</b>
1.1. Latar Belakang.....	1
1.2. Pernyataan Masalah.....	3
1.3. Tujuan Penelitian.....	3
1.4. Manfaat Penelitian.....	3
1.5. Batasan Penelitian.....	4
<b>BAB II STUDI PUSTAKA.....</b>	<b>5</b>
2.1. <i>Summarization</i> .....	5
2.1.1 <i>Extraction-based summarization</i> .....	5
2.1.1.1 <i>Surface-level</i> .....	6
2.1.1.2 <i>Intermediate-level</i> .....	6
2.1.1.3 <i>Deep parsing</i> .....	7
2.1.2 <i>Abstractive summarization</i> .....	8
2.2. Algoritma <i>TextRank</i> .....	8
<b>BAB III DESAIN DAN IMPLEMENTASI.....</b>	<b>12</b>
3.1 Data .....	12
3.2 Desain dan Implementasi.....	13
3.2.1 Tahap <i>Preprocess</i> .....	15
3.2.2 Tahap <i>Main Process</i> .....	20

3.2.3 Uji coba sistem.....	28
3.2.4 <i>User Interface</i> .....	28
<b>BAB IV UJI COBA DAN PEMBAHASAN.....</b>	<b>35</b>
4.1. Skenario Uji Coba.....	35
4.2. Hasil Uji Coba.....	37
4.3. Pembahasan.....	58
<b>BAB V KESIMPULAN DAN SARAN.....</b>	<b>63</b>
5.1 Kesimpulan.....	63
5.2 Saran.....	63
<b>DAFTAR PUSTAKA.....</b>	<b>65</b>



## DAFTAR GAMBAR

Gambar 3.1 Ilustrasi pengumpulan data 1 artikel dan pembuatan rangkuman referensi.....	12
Gambar 3.2 Struktur tabel artikel.....	13
Gambar 3.3 Desain Sistem.....	14
Gambar 3.4 Contoh cuplikan kalimat.....	15
Gambar 3.5 <i>Flowchart Stemming</i> .....	16
Gambar 3.6 <i>Flowchart Tokenizing dan filtering</i> .....	17
Gambar 3.7 <i>Flowchart Preprocessing</i> .....	18
Gambar 3.8 <i>Source code preprocess</i> .....	19
Gambar 3.9 Pemisahan artikel menjadi list kalimat.....	19
Gambar 3.10 Ilustrasi kalimat menjadi simpul.....	21
Gambar 3.11 Graf dengan sisi berbobot.....	22
Gambar 3.12 <i>Flowchart</i> Representasi Kalimat menjadi Graf.....	23
Gambar 3.13 Graf dengan sisi & simpul berbobot.....	24
Gambar 3.14 <i>Flowchart</i> iterasi skor <i>TextRank</i> pada setiap simpul kalimat.....	25
Gambar 3.15 <i>Source code</i> pembentukan graf.....	26
Gambar 3.16 <i>Source code</i> Iterasi <i>TextRank</i> .....	27
Gambar 3.17 Ilustrasi alur penghitungan iterasi <i>TextRank</i> .....	27
Gambar 3.18 <i>Source code</i> evaluasi ROUGE.....	28
Gambar 3.19 Rancangan UI halaman utama.....	29
Gambar 3.20 Rancangan UI halaman artikel.....	29
Gambar 3.21 Halaman Awal.....	30
Gambar 3.22 Bagian halaman input artikel.....	31
Gambar 3.23 Bagian halaman perangkuman artikel.....	33
Gambar 4.1 Isi artikel asli (data input).....	41
Gambar 4.2 Hasil Rangkuman referensi.....	46
Gambar 4.3 Hasil Rangkuman sistem.....	47

## DAFTAR TABEL

Tabel 3.1 Pemisahan kalimat menjadi per kata.....	19
Tabel 3.2 Hasil <i>Preprocess</i> .....	20
Tabel 4.1 Daftar judul Dokumen artikel.....	37
Tabel 4.2 Pemisahan artikel menjadi kalimat-kalimat.....	41
Tabel 4.3 Tabel Data responden dan total jumlah kalimat.....	44
Tabel 4.4 Pembentukan rangkuman referensi.....	45
Tabel 4.5 Statistik artikel.....	47
Tabel 4.6 Daftar kata unik rangkuman sistem.....	47
Tabel 4.7 Daftar kata unik rangkuman referensi.....	48
Tabel 4.8 Daftar kemunculan kata yang sama.....	48
Tabel 4.9 Jumlah kata unik.....	49
Tabel 4.10 Hasil Nilai <i>recall</i> ROUGE Tiap Artikel.....	50
Tabel 4.11 Perbandingan rata-rata jumlah kata dan kalimat.....	58

## ABSTRAK

Zamzam, Muhammad Adib . 2019. **Sistem Automatic Text Summarization Menggunakan Algoritma Textrank**. Skripsi. Jurusan Teknik Informatika. Fakultas Sains dan Teknologi. Universitas Islam Negeri Maulana Malik Ibrahim Malang.

Pembimbing :(i) Dr. Cahyo Crysdian  
(ii) Khadijah Fahmi Hayati, M.Kom

---

Kata Kunci : *Summarization, Automatic Summarization, Text, TextRank, graph*

*Text summarization* (perangkuman teks) adalah pendekatan yang bisa digunakan untuk meringkas atau memadatkan teks artikel yang panjang menjadi lebih pendek dan ringkas sehingga hasil rangkuman teks yang relatif lebih pendek bisa mewakili teks yang panjang. *Automatic Text Summarization* adalah perangkuman teks yang dilakukan secara otomatis oleh komputer. Terdapat dua macam algoritma *Automatic Text Summarization* yaitu *Extraction-based summarization* dan *Abstractive summarization*. Algoritma *TextRank* merupakan algoritma *extraction-based* atau *extractive*, dimana ekstraksi di sini berarti memilih unit teks (kalimat, segmen-segmen kalimat, paragraf atau passages), lalu dianggap berisi informasi penting dari dokumen dan menyusun unit-unit (kalimat-kalimat) tersebut dengan cara yang benar. Hasil penelitian dengan input 50 artikel dan hasil rangkuman sebanyak 12,5% dari teks asli menunjukkan bahwa sistem memiliki nilai *recall* ROUGE 41,659 %. Nilai tertinggi *recall* ROUGE tertinggi tercatat pada artikel 48 dengan nilai 0,764. Nilai terendah *recall* ROUGE tercatat pada artikel 37 dengan nilai 0,167.

## ABSTRACT

Zamzam, Muhammad Adib . 2019. *Automatic Text Summarization System Using Textrank Algorithm*. Essay. Department of Informatics Engineering, Faculty of Science and Technology, Islamic State University of Maulana Malik Ibrahim of Malang.

Counselor: (i) Dr. Cahyo Crysdiان  
(ii) Khadijah Fahmi Hayati, M.Kom

---

Kata Kunci : *Summarization, Automatic Summarization, Text, TextRank, graph*

Text summarization is an approach to summarize or condense long text to shorter text, so that the shorter text is represent the real text. Automatic Text Summarization is text summarization which done automatically by computer. There are two methods, first Extraction-based summarization and Abstractive summarization. TextRank Algorithm is extraction-based or extractive, in the current context, extraction means choosing the text unit (sentences, sentence segments, paragraph or passages), then it is calculated and considered as important and finally the units is arranged at the correct manner. The result from 50 articles with summary rate of 12,5% had the ROUGE recall value 41.659%. Highest value known at 48<sup>th</sup> article with value 0,764. Lowest ROUGE recall value known at 37<sup>th</sup> with value 0,167.

## المخلص

محمد أديب زمزم. ٢٠١٩. توزيع الملمس الكثافة الضباب في المحاكاة ٢دي ضباب غير متجانس ضوضاء البييرلين. قسم هندسة المعلوماتية لكلية العلوم والتكنولوجيا في جامعة مولانا مالك إبراهيم الإسلامية الحكومية بمالانق. المشرف : (١) الدكتور جهيو كريسديان. (٢) خديجة فهمي حياتي حولي .

الكلمات الرئيسية : تلخيص ، تلخيص التلقائي ، النص ، TextRank ، الرسم البياني

تلخيص النص هو طريقة لتلخيص أو تكثيف النص الطويل للنص الأقصر ، بحيث يمثل النص الأقصر النص الحقيقي. تلخيص النص التلقائي هو تلخيص النص الذي يتم تلقائياً بواسطة الكمبيوتر. هناك طريقتان ، أولهما تلخيص قائم على الاستخراج وملخص تلخيصي. خوارزمية TextRank تعتمد على الاستخراج أو الاستخراج ، في السياق الحالي ، يعني الاستخراج اختيار وحدة النص (الجملة ، شرائح الجملة ، الفقرة أو المقاطع) ، ثم يتم حسابها واعتبارها مهمة وأخيراً يتم ترتيب الوحدات بالطريقة الصحيحة. كانت النتيجة من ٥٠ مقالاً بمعدل ملخص قدره ١٢,٥ % قيمة استدعاء ROUGE بنسبة ٤١,٦٥٩%. أعلى قيمة معروفة في المادة ٤٨ بقيمة ٠,٧٦٤. أقل قيمة استدعاء تذكر ROUGE في ٣٧ بقيمة ٠,١٦٧ .

## BAB 1

### PENDAHULUAN

#### 1.1. Latar Belakang

Pertumbuhan internet yang sangat cepat berbanding lurus dengan jumlah informasi dan dokumen online yang sangat besar pula. Di masa ini, media yang paling banyak digunakan untuk menyampaikan informasi adalah media teks (El Bazzi, 2016). Sehingga menemukan sebuah representasi yang tepat dari informasi tekstual (rangkuman) merupakan sebuah hal yang penting untuk membantu pemahaman keseluruhan teks supaya lebih cepat.

Rangkuman penting untuk telaah sekilas dari keseluruhan bagian dokumen atau bab. Penulisan rangkuman pada umumnya bisa dilakukan dengan mudah jika pembaca relatif familiar terhadap teks yang dibaca atau jika teks tersebut tidak terlalu panjang. Hasil rangkuman setiap individu pun berbeda karena apa yang ditulis juga merupakan buah pikiran individu yang bersifat subjektif sedangkan untuk menghasilkan rangkuman yang baik dan bagus perlu pertimbangan yang bersifat objektif. Merangkum juga termasuk pekerjaan berat bagi manusia jika teks sangat panjang, maka dibutuhkan sistem yang bisa melakukan pekerjaan tersebut secara otomatis dengan cara yang lebih objektif dan relatif lebih mudah. Di sebuah ayat Al - Qur'an pada surat Ar-Ra'd ayat 3 tertulis,

وَهُوَ الَّذِي مَدَّ الْأَرْضَ وَجَعَلَ فِيهَا رَوَاسِيَ وَأَنْهَارًا وَمِنْ كُلِّ الشَّجَرِ جَعَلَ فِيهَا زَوْجَيْنِ  
أُنثَيْنِ يُغْشَىٰ اللَّيْلَ النَّهَارَ إِنَّ فِي ذَلِكَ لَآيَاتٍ لِّقَوْمٍ يَتَفَكَّرُونَ ﴿٣﴾

“Dan Dia yang menghamparkan bumi dan menjadikan gunung-gunung dan

sungai-sungai di atasnya. Dan padanya Dia menjadikan semua buah-buahan berpasang-pasangan; Dia menutupkan malam kepada siang. Sungguh, pada yang demikian itu terdapat tanda-tanda (kebesaran Allah) bagi orang-orang yang berpikir.”.

Menurut Tafsir Jalalain ayat di atas dijelaskan bahwa di alam semesta terdapat bukti-bukti kemahakuasaan dan kemahaesaan Allah Swt. bagi orang-orang yang mau berfikir dan merenungkan ciptaan-Nya. Untuk mencapai keimanan yang sempurna maka manusia harus menjadi orang mau berfikir dan merenungkan ciptaan-Nya dan mengambil hikmah pada hasil renungannya. Pada pembuatan rangkuman dokumen teks juga diperlukan pikiran pembaca dan pemahaman terhadap maksud penulis yang sebenarnya. Dengan memahami dua fakta tersebut, maka sesungguhnya terdapat cara untuk mendapatkan inti dari suatu hal. Dalam konteks mendapatkan inti informasi dari sebuah teks, maka inti informasi pada umumnya diperoleh dengan membuat rangkuman (*summary*).

Salah satu pembahasan pada *Natural Language Processing* adalah *Text Summarization*. *Text Summarization* bertujuan untuk mengkaji tentang otomatisasi perangkuman teks dengan pilihan metode-metode / algoritma yang bermacam-macam. *Automatic text summarization*, sebuah reduksi teks menjadi pada isi pentingnya ialah problem yang sangat kompleks yang dalam perkembangannya sejauh ini memberikan banyak tantangan pada komunitas ilmiah. *Text summarization* memiliki aplikasi yang relevan pada informasi *online* yang diberikan masyarakat dimana pertumbuhannya bersifat eksponensial (Saggion, 2016).

Dalam kesempatan ini peneliti hendak membangun sistem *Automatic text summarization*. Salah satu algoritma dalam pembahasan *text summarization* adalah Algoritma *TextRank* yang termasuk dalam kategori *extractive summarization*. Algoritma *TextRank* merupakan metode yang sederhana karena perangkuman dilakukan dengan menilai setiap kalimat dan melakukan perangkungan tanpa mengubah isi teks asli yang dijadikan rangkuman.

### **1.2. Pernyataan Masalah**

Dengan latar belakang yang dipaparkan maka pernyataan masalah yaitu, berapa nilai relevansi hasil rangkuman penggunaan algoritma *TextRank* untuk merangkum dokumen Bahasa Indonesia jika diukur menggunakan ROUGE.

### **1.3. Tujuan Penelitian**

Mengukur relevansi hasil rangkuman penggunaan algoritma *TextRank* untuk merangkum dokumen Bahasa Indonesia yang diukur menggunakan ROUGE.

### **1.4. Manfaat Penelitian**

Penelitian ini akan membuktikan nilai relevansi rangkuman yang dihasilkan dari sebuah algoritma atau metode *TextRank*. Dengan mengetahui hal tersebut, selanjutnya metode dapat dibandingkan dengan metode-metode perangkuman teks yang lain atau diterapkan pada aplikasi atau software terkait pengolah teks. Sistem perangkum juga sangat bermanfaat bagi para pembaca buku, akademisi, jurnalis dan pekerjaan lain yang membutuhkan perangkuman teks.

### 1.5. Batasan Penelitian

Batasan yang ada pada penelitian ini adalah sebagai berikut:

1. Dokumen Bahasa Indonesia yang digunakan diambil dari media mainstream online.
2. Fokus pembahasan lebih mengutamakan pada pengukuran algoritma *TextRank* dalam melakukan perangkuman teks daripada membahas mekanisme crawling dokumen online.



## BAB II

### STUDI PUSTAKA

#### 2.1. *Summarization*

Zadbuke (2016) menerangkan *Text summarization* (perangkuman teks) adalah pendekatan yang bisa digunakan untuk meringkas atau memadatkan teks artikel yang panjang menjadi lebih pendek dan ringkas sehingga hasil rangkuman teks yang relatif lebih pendek bisa mewakili teks yang panjang. Pada dasarnya menggunakan prinsip *Natural Language Processing* dan algoritma-algoritma untuk membuat sistem mengerti artikel-artikel dan menghasilkan rangkuman yang lebih pendek dan efisien. *Automatic summarization* (otomasi perangkuman) adalah proses mengurangi dokumen teks menggunakan program komputer dalam rangka membuat rangkuman yang menyimpan titik-titik terpenting dari dokumen asli. *Automatic data summarization* adalah wilayah yang sangat penting dalam *machine learning* dan *data mining*. Di zaman ini, teknologi *summarization* digunakan pada banyak sektor industri. Sebuah contoh adalah search engine seperti Google. Contoh lain yang termasuk yaitu *document summarization*, *image collection summarization* dan *video summarization*. Ide pokok *summarization* adalah menemukan bagian yang representatif, yang memiliki informasi dari seluruh bagian. Macam-macam *summarization* dijelaskan pada bagian berikut ini.

##### 2.1.1 *Extraction-based summarization*

Ekstraksi berarti memilih unit teks (kalimat, segmen-segmen kalimat, paragraf atau passages), lalu dianggap berisi informasi penting dari dokumen dan

menyusun unit-unit ini dengan cara yang benar. Menurut Radev et al. , algoritma-algoritma berbasis ekstraksi terbagi menjadi tiga, yaitu:

#### **2.1.1.1 Surface-level**

Algoritma-algoritma yang termasuk di kategori ini tidak menggali ke dalam level linguistik dari sebuah teks; lebih pada penggunaan elemen linguistik tertentu untuk mengidentifikasi segmen-segmen yang paling relevan dari dokumen. Digunakan di awal masa-masa studi summarization, pada level ini menggunakan teknik kemunculan kata-kata sampai bobot kalimat.

Sabuna (2017) membahas perangkuman ekstraktif dengan metode *sentence scoring*. Metode ini mengambil 8 nilai dari sebuah kalimat diantaranya adalah TF/IDF, huruf besar, kata benda, frasa isyarat, data numerik, panjang kalimat, posisi kalimat dan kesamaan dengan judul yang kemudian nilainya diproses sehingga menjadi skor dari sebuah kalimat. Skor tersebut menjadi hasil data latih dimana data tersebut digunakan sebagai acuan decision tree untuk menentukan sebuah kalimat termasuk dalam rangkuman atau tidak. *Score sentence* termasuk metode supervised.

#### **2.1.1.2 Intermediate-level**

Kategori ini menggunakan informasi linguistik yang lebih canggih dari kategori sebelumnya namun masih tergolong kurang dari deep parsing. Salah satu contohnya yaitu lexical chain recognition. *Lexical chains* ialah urutan kata-kata yang terhubung oleh relasi semantik kebahasaan.

Barzilay (1997) melakukan penelitian perangkuman ekstraktif dengan menggunakan *lexical chain*. Barzilay menyatakan bahwa ekstraksi rangkuman

dapat dilakukan dengan menggunakan *lexical chain* sebagai model topik pembahasan. Algoritma baru yang dipaparkan yaitu WordNet thesaurus, sebuah algoritma penanda teks yang termasuk bagian dari percakapan atau pidato (*part-of-speech tagger*) dan sebuah algoritma segmentasi. Perangkuman dilakukan dalam empat langkah yaitu: segmentasi teks asli, pembangunan *lexical chain*, identifikasi *strong chain* dan kalimat yang cenderung memiliki *chain* yang kuat diekstrak.

### 2.1.1.3 Deep parsing

Kategori ini didasarkan pada ide yang menjelaskan bahwa di sebuah kasus summarization dibutuhkan penggunaan teknik linguistik yang dalam yang memanfaatkan struktur teks yang tidak bersambungan. Beberapa pendekatan mengacu pada *rhetorical structure theory* (RST), yang bertujuan memanfaatkan struktur sebagian teks untuk menghasilkan abstrak.

Chengcheng (2010) menggunakan metode *summarization* yang berdasarkan *Rhetorical Structure Theory* (RST). RST merupakan sebuah kerangka kerja yang dirancang untuk perhitungan struktur teks pada tingkat klausa. Sistem mengutip (*extract*) struktur retorikal dari teks dan menggabungkan relasi retorikal antara kalimat-kalimat dan memotong bagian yang kurang penting. Kemudian digunakan metode *natural language generation* berdasarkan model untuk menghasilkan sebuah rangkuman yang tersusun rapi. Rangkuman yang dibangkitkan dengan cara ini menunjukkan efektifitasnya yang tinggi.

### 2.1.2 *Abstractive summarization*

Teknik ekstraksi hanya menyalin informasi yang dianggap paling penting oleh sistem (*key clause*, kalimat atau paragraf-paragraf), teknik abstraksi melibatkan paraphrasing section dari dokumen asal. Pada umumnya, abstraksi dapat membentuk teks lebih kuat daripada ekstraksi, namun program yang mampu melakukannya lebih sulit untuk dikembangkan mengetahui syarat sistem tersebut harus menggunakan teknologi *natural language generation*, dimana teknologi tersebut masih merupakan bidang yang berkembang. Sistem yang menghasilkan rangkuman dengan abstraksi berdasarkan pemahaman teks dan membangkitkan teks yang benar secara *grammar*, ringkas dan jelas. Sangat sedikit sistem yang dibuat di kategori ini.

Menurut Mani & Maybury (1998) hasil rangkuman dari metode-metode tersebut kemudian dapat dievaluasi dengan menggunakan metode-metode untuk evaluasi rangkuman yang bisa dibagi menjadi dua kategori, yaitu intrinsik dan ekstrinsik. Evaluasi Intrinsik mengukur kualitas rangkuman secara langsung (membandingkannya dengan rangkuman yang ideal). Metode ekstrinsik mengukur seberapa baik rangkuman membantu kinerja tugas tertentu (klasifikasi dan semacamnya).

## 2.2. Algoritma *TextRank*

*TextRank* merupakan metode pemeringkatan berdasarkan graf (*graph-based ranking*) dimana graf merupakan hasil ekstraksi dari teks menggunakan *natural language text*. Graf pada *TextRank* yaitu graf berbobot tak berarah. Algoritma *graph-based ranking* merupakan sebuah cara menentukan simpul-

simpul penting pada graf, berdasarkan pada informasi global yang digambarkan secara rekursif dari keseluruhan graf. Ide pokoknya yaitu menerapkan *voting* atau rekomendasi dari simpul-simpul di graf. Ketika satu simpul  $a$  terhubung ke simpul  $b$ , maka sejatinya simpul  $a$  memberi suara (*vote*) pada simpul  $b$  tersebut. Semakin tinggi angka penilaian pada sebuah simpul, semakin penting pula simpul tersebut. Sehingga, skor yang berhubungan dengan sebuah simpul ditentukan berdasarkan suara yang diperoleh simpul dari simpul-simpul lain dan skor simpul-simpul yang memperoleh suara (Mihalcea, 2004).

Misal,  $G = (V,E)$  sebagai graf dengan  $V$  adalah sekumpulan simpul dan  $E$  adalah sekumpulan sisi, dimana  $E$  merupakan sebagian dari  $V \times V$ . Sebuah sisi memiliki bobot yang dapat ditentukan nilainya dengan rumus kesamaan (*similar formula*). Pada simpul  $V_i$ , terdapat  $In(V_i)$ , simpul-simpul yang menunjuk ke simpul tersebut (*predecessor*) dan  $Out(V_i)$ , sekumpulan simpul-simpul yang ditunjuk oleh  $V_i$  (*successor*). Skor total pada simpul  $V_i$  didefinisikan seperti berikut (Brin & Page, 1998):

$$S(V_i) = (1 - d) + d * \sum_{j \in In(V_i)} \frac{1}{|Out(V_j)|} S(V_j) \quad (2.1)$$

,dimana  $d$  adalah *damping factor* yang dapat bernilai diantara 0 - 1. Nilai  $d$  berperan dalam mengintegrasikan probabilitas perpindahan dari satu simpul ke satu simpul acak pada graf. Pada konteks *web surfing*, algoritma ini menerapkan "*random surfer model*", dimana pengguna secara acak menekan link dengan kemungkinan  $d$  dan berpindah ke halaman yang sama sekali baru dengan kemungkinan  $1-d$ . Faktor  $d$  biasanya bernilai 0,85 (Brin & Page, 1998).

Dimulai dari nilai awal yang diisikan pada tiap simpul di graf, iterasi berjalan hingga mencapai ambang batas. Ambang batas dicapai ketika nilai error untuk setiap simpul pada graf jatuh dibawah batas (*threshold*). Nilai error sebuah simpul  $V_i$  didefinisikan sebagai sebagai selisih antara  $S(V_i)$ (skor simpul) dengan skor yang dihitung pada iterasi  $k$ ,  $S^k(V_i)$ . Karena skor sebenarnya tidak diketahui, nilai error diperkirakan dengan selisih antara skor yang dihitung pada dua iterasi berikutnya:

$$S^{k+1}(V_i) - S^k(V_i). \quad (2.2)$$

Setelah menjalankan algoritma, skor dihubungkan pada tiap simpul, yang mewakili “pentingnya” simpul dalam suatu graf. Nilai akhir yang diperoleh setelah TextRank selesai tidak dipengaruhi oleh pilihan nilai awal melainkan jumlah iterasi menuju ambang bataslah yang berpengaruh.

Metode yang diterapkan merupakan graf berbobot maka perlu diketahui bobot tiap sisi yang menghubungkan simpul-simpul. Maka rumus *TextRank* adalah sebagai berikut:

$$WS(V_i) = (1 - d) + d * \sum_{V_j \in In(V_i)} \frac{w_{ji}}{\sum_{V_k \in Out(V_j)} w_{jk}} WS(V_j) \quad (2.3)$$

Rincian Rumus :

$WS(V_i)$  = Bobot simpul  $V_i$  &  $d$  = damping factor (0.85)

$V_k$  = Himpunan simpul tetangga  $V_j$

$WS(V_i)$  = Bobot simpul  $V_j$

$V_j$  = Himpunan simpul tetangga  $V_i$

$w_{ji} & w_{jk}$  = bobot sisi antara simpul  $j$  &  $i$  dan  $j$  &  $k$  secara berurutan

Menurut Kaynar (2017) *Automatic document summarization* adalah proses reduksi dokumen yang menetapkan bagian teks yang penting. Pada penelitiannya dilakukan uji coba 4 algoritma *similarity* yang berbeda untuk efisiensi algoritma *TextRank* dalam menghasilkan rangkuman. Ditemukan bahwa metode *levenshtein* memiliki skor Rouge-1 yang paling tinggi dibanding 3 algoritma *similarity* lainnya.

El Bazzi (2016) melakukan penelitian *keyword extraction* dengan dokumen Berbahasa Arab. Mengikuti cara kerja algoritma *TextRank*, graf dibentuk dengan kata-kata / istilah sebagai simpul. Representasi graf menawarkan keuntungan yang memungkinkan pemodelan dokumen yang lebih ekspresif daripada pendekatan *bag of words* dan meningkatkan kinerja klasifikasi. Hasil percobaan menunjukkan metode *indexing* berbasis graf adalah pendekatan yang menjanjikan untuk pengindeksan semantik dan kontekstual.

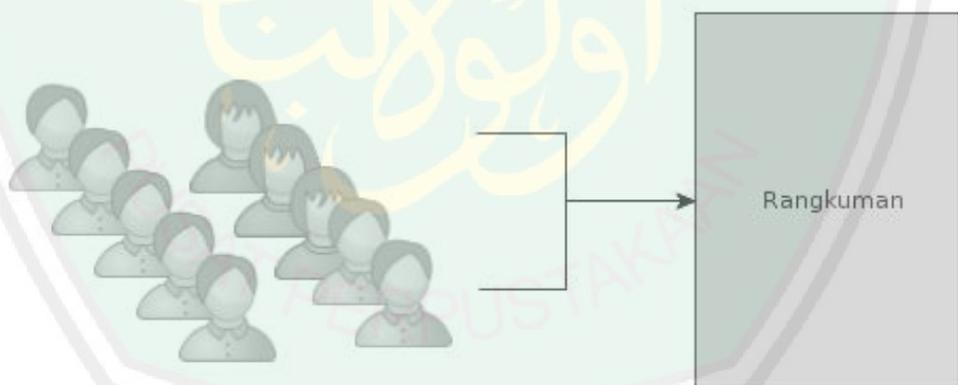
Nanos (2017) membahas perangkuman otomatis pada dokumen transkrip pertemuan (meeting) pada virtual meeting systems. Representasi struktur *TextRank* biasanya hanya menggunakan graf saja, namun pada penelitian ini ditambahkan struktur pertemuan. Hal ini membantu ekstraksi kalimat yang paling penting dari sebuah transkrip pertemuan. Hasil menunjukkan bahwa memanfaatkan struktur tersebut dapat mengarahkan pada otomatisasi perangkuman yang lebih relevan.

## BAB III

### DESAIN DAN IMPLEMENTASI

#### 3.1 Data

Sumber yang digunakan untuk penelitian ini didapatkan dari situs *tirto.id*. Dokumen / artikel diambil sebanyak 50 teks berita atau eksposisi dari situs tersebut. Rangkuman manual (Rangkuman referensi atau *ground truth*) didapatkan dari 10 orang relawan. Satu artikel dirangkum oleh 10 orang. Hasil akhir sebuah rangkuman referensi dari sebuah artikel adalah himpunan kalimat-kalimat yang sering muncul dari 10 rangkuman referensi dari 1 artikel yang telah dibuat. Pengambilan data responden menggunakan *Google form*. Ilustrasi pengumpulan data teks serta pembuatan rangkuman secara manual dapat dilihat pada Gambar 3.1 berikut.



Gambar 3.1 Ilustrasi pengumpulan data 1 artikel dan pembuatan rangkuman referensi

Artikel-artikel, rangkuman referensi dan rangkuman sistem kemudian disimpan di *database* yang telah disediakan untuk mempermudah akses data. Struktur tabel artikel di database ditampilkan pada Gambar 3.2.

id	sumber	judul	isi	rsys	rreferensi	statistik	nilairouge
int(11)	longtext	varchar(255)	longtext	longtext	longtext	text	varchar(255)

Gambar 3.2 Struktur tabel artikel

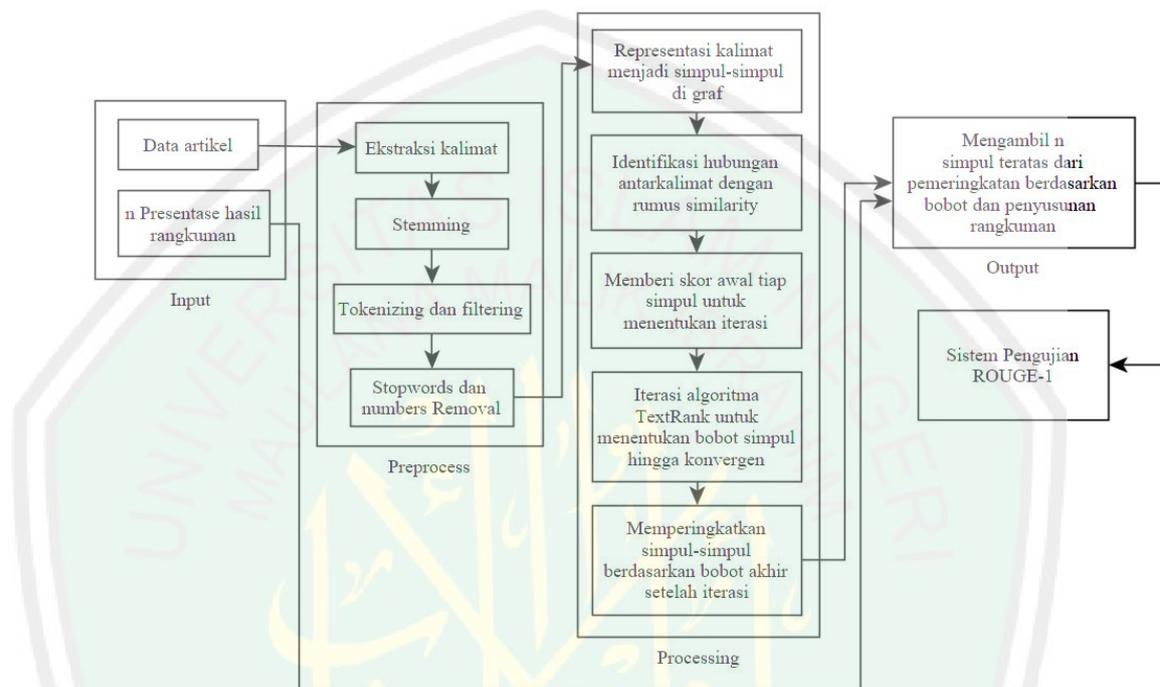
Struktur tabel database pada Gambar 3.2 memiliki 8 kolom yaitu:

1. *id*, bertipe *integer* untuk *primary key*
2. *sumber*, bertipe *longtext* untuk menyimpan URL sumber artikel.
3. *judul*, bertipe *varchar* untuk menyimpan judul artikel.
4. *isi*, bertipe *text* untuk menyimpan isi artikel.
5. *rsys*, bertipe *longtext* untuk menyimpan rangkuman hasil sistem.
6. *rreferensi*, bertipe *longtext* untuk menyimpan rangkuman referensi (*ground truth*).
7. *statistik*, bertipe *text* untuk menyimpan data perhitungan jumlah kalimat dan kata pada artikel data input, rangkuman sistem dan rangkuman referensi.
8. *nilairouge*, bertipe *varchar* untuk menyimpan nilai perhitungan ROUGE-1.

### 3.2 Desain dan Implementasi

Desain dan implementasi sistem *automatic text summarization* setiap bagian dibahas pada bagian berikut. Sistem *automatic text summarization* pada kesempatan ini terdiri dari *preprocess*, *main process*, *pengujian sistem* dan rancangan *user interface*. Tahap *preprocess* adalah tahap mengolah isi artikel asli menjadi data yang siap diolah pada *main process*. Tahap *Main Process* adalah

tahap proses memperoleh nilai. Desain sistem *automatic text summarization* ditunjukkan pada Gambar 3.3.



Gambar 3.3 Desain Sistem

Implementasi sistem yang telah dirancang menggunakan bahasa pemrograman python 3.66 berbasis web dengan *framework* Django versi 2.1.5. Aplikasi kemudian terhubung pada *database* MySQL (*Ver 15.1 Distrib 10.3.16-MariaDB, for Linux (x86\_64)*). Aplikasi dibangun pada sistem operasi Linux Ubuntu pada *hardware processor* Intel(R) Core(TM) i3-3120M CPU & RAM 4 GB.

Sebagai contoh, Gambar 3.4 berikut adalah cuplikan kalimat dari sebuah artikel yang hendak diproses.

Contoh :

Antibiotik adalah obat yang digunakan untuk mencegah dan mengobati infeksi bakteri. Sejak ditemukan, penemuan yang dipandang sebagai prestasi dunia kedokteran modern ini berfungsi sebagai landasan pengobatan. Namun demikian, penggunaan berlebihan dan penyalahgunaan antibiotik dalam layanan kesehatan manusia dan hewan telah mendorong penyebaran resistensi antibiotik. Resistensi antibiotik terjadi ketika bakteri mengubah mekanisme dalam menanggapi penggunaan obat tersebut.

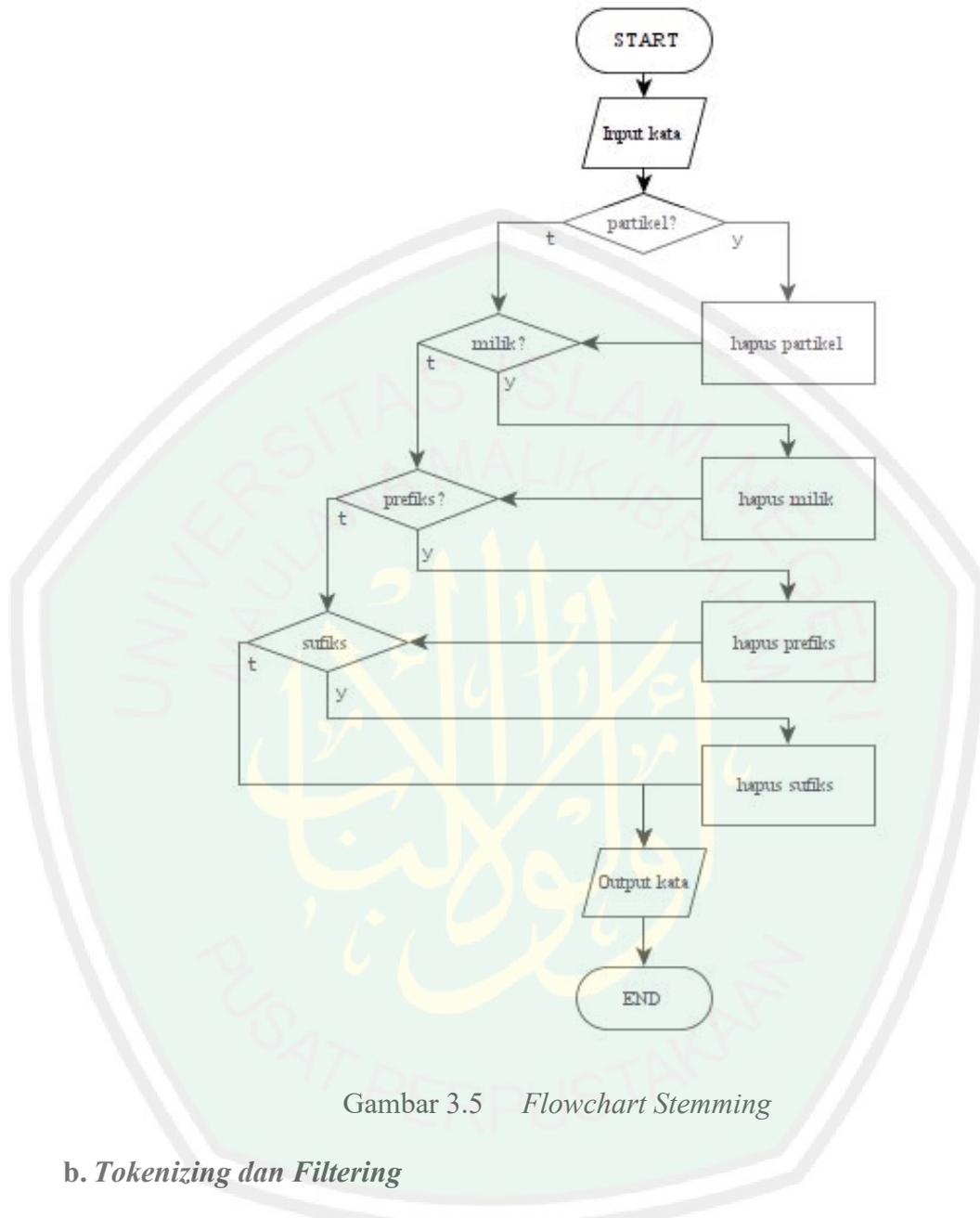
Gambar 3.4 Contoh cuplikan kalimat

### 3.2.1 Tahap *Preprocess*

*Preprocessing* dilakukan untuk meminimalkan jumlah kata / kalimat yang relatif tidak penting seperti kata hubung antar subjek atau objek dan urutan-urutan kata yang bukan merupakan ciri-ciri kalimat standar (tidak hanya berupa angka). Tahap ini penting karena mempengaruhi tahap utama secara langsung dalam perbandingan kata-kata. Setelah dibersihkan dari bagian-bagian yang kurang penting, tahap utama hanya memproses *token* atau bagian yang penting saja. Tahap *preprocessing* terdiri dari proses Ekstraksi kalimat dari teks yang kemudian setiap kalimat dilakukan *Tokenizing*, *Filtering*, *Stemming*, *Stopwords* dan *number Removal*.

#### a. *Stemming*

*Stemming* merupakan proses mengubah kata-kata yang memiliki imbuhan menjadi kata baku sehingga kata-kata yang diproses sistem adalah kata-kata dasar dalam Bahasa Indonesia. *Stemming* di kesempatan ini menggunakan modul Sastrawi yang tersedia di python. Alur *stemming* ditunjukkan pada Gambar 3.5.

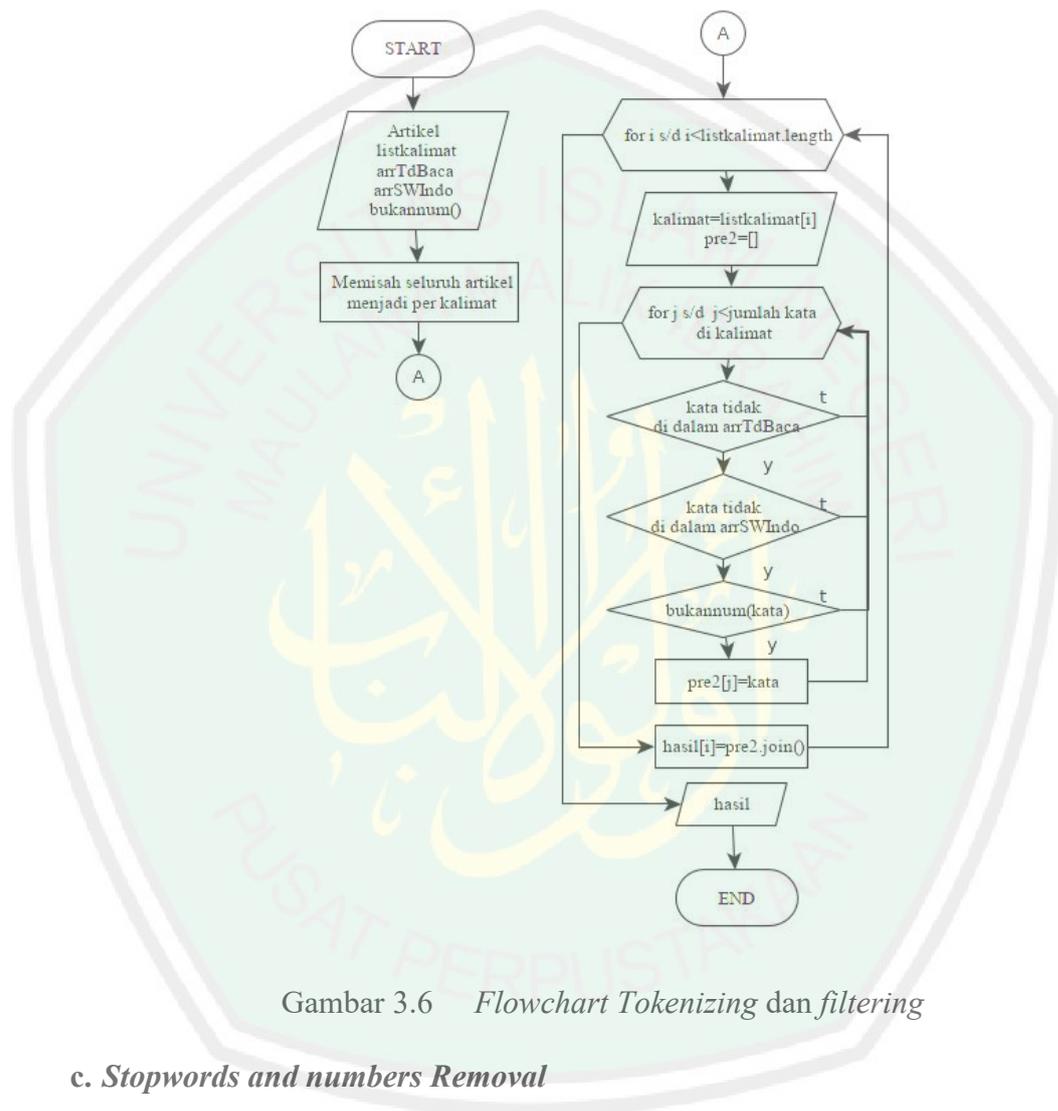


Gambar 3.5 Flowchart Stemming

### b. Tokenizing dan Filtering

*Tokenizing dan Filtering* menghilangkan tanda baca yang tidak diperlukan dan dilakukan dengan memisahkan kalimat menjadi kata-kata (Kurniawan et,al, 2014). Kalimat dipisah menjadi kata-kata dengan pemisah karakter spasi. Setiap kata-kata di kalimat kemudian dibandingkan apakah tidak termasuk pada kumpulan tanda baca. Jika tidak termasuk pada kumpulan tanda baca, maka kata-

kata masuk dalam hasil *preprocess*, Tanda baca diambil dari modul standar library python *string.punctuation*. Berikutnya mengubah semua token ke bentuk huruf kecil (*lower case*). Alur *tokenizing* dan *filtering* bisa dilihat di Gambar 3.6.



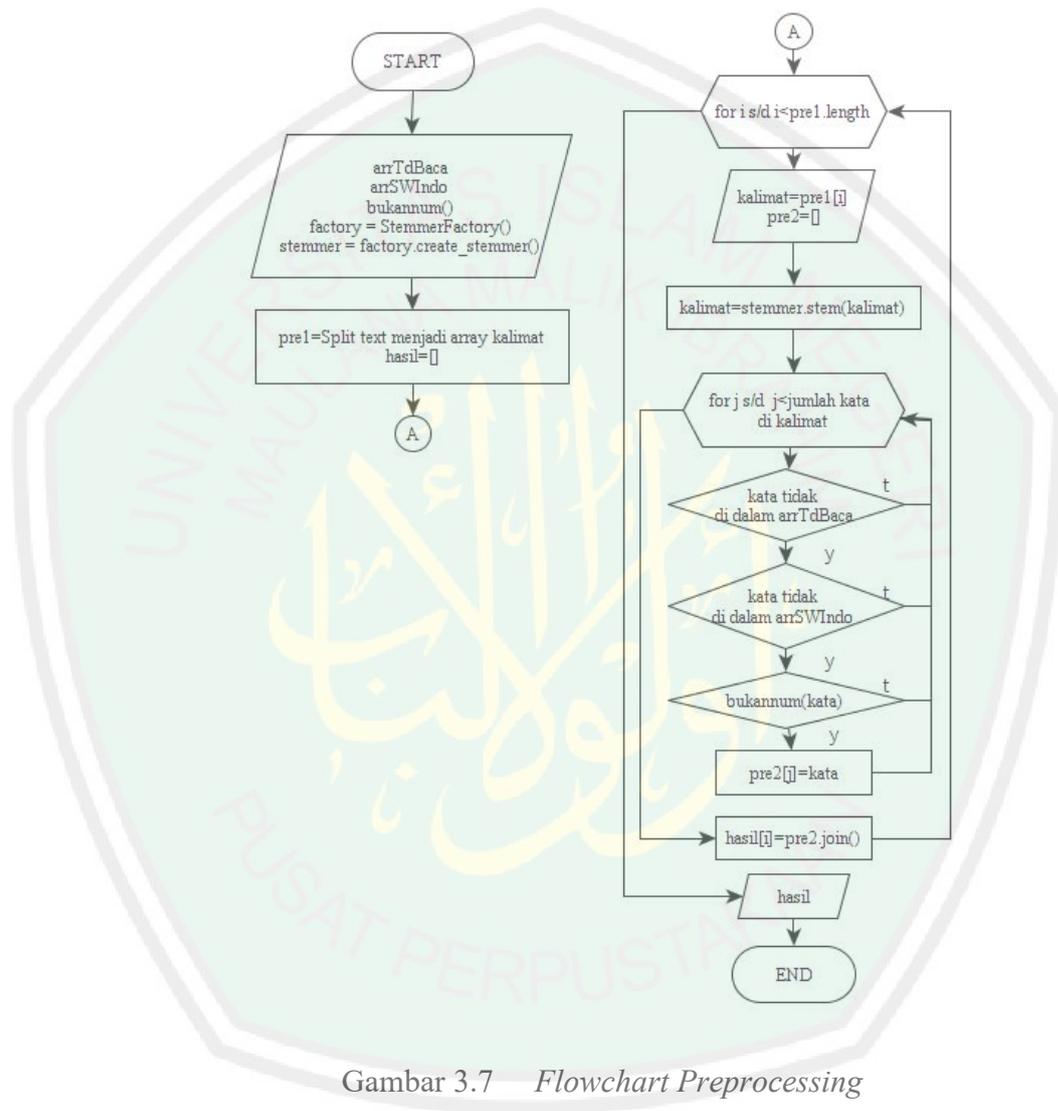
Gambar 3.6 Flowchart Tokenizing dan filtering

### c. Stopwords and numbers Removal

*Stopwords removal* merupakan proses penghilangan kata tidak penting pada deskripsi melalui pengecekan kata-kata hasil *parsing* deskripsi apakah termasuk di dalam daftar kata tidak penting (*stoplist*) atau tidak (Nurdiana, 2016). Angka dalam bentuk kata tidak dimasukkan dalam perhitungan karena dapat

menimbulkan multitafsir. Kamus *stopword* bahasa indonesia diperoleh dari modul python bernama *stop\_words*.

Keseluruhan *Preprocessing* ditunjukkan pada Gambar 3.7 berikut.



Gambar 3.7 Flowchart Preprocessing

Source code preprocess ditampilkan pada Gambar 3.8

```

def inputteks(teks):
    # pecah kalimat dengan titik spasi
    teks=re.sub("[\r\n]", " ", teks)
    pre1=re.split(r"[.!?;!][\n]", teks)
    pre1=[p.strip() for p in pre1 if p!="r"]

    print(len(pre1),"kalimat terisi")
  
```

```

return teks,pre1

def preproses(pre1):
    res=[]
    for t,i in zip(pre1,range(0,len(pre1))):
        if len(t.split())<3 :
            pre1.remove(t)
            continue
        t=stemmer.stem(t)
        pre2=[]
        n=0
        t=t.split(" ")
        for t2 in t:
            n+=1
            if (t2 not in arrTdBaca) and (t2 not in arrSwIndo) and bukannum(t2):
                pre2.append(t2.lower())
        result=" ".join(pre2).strip()
        res.append(result)
    return res

```

Gambar 3.8 *Source code preprocess*

*Source code* di atas menghasilkan output dengan urutan pada Gambar 3.9.

1. Antibiotik adalah obat yang digunakan untuk mencegah dan mengobati infeksi bakteri.
2. Sejak ditemukan, penemuan yang dipandang sebagai prestasi dunia kedokteran modern ini berfungsi sebagai landasan pengobatan.
3. Namun demikian, penggunaan berlebihan dan penyalahgunaan antibiotik dalam layanan kesehatan manusia dan hewan telah mendorong penyebaran resistensi antibiotik.
4. Resistensi antibiotik terjadi ketika bakteri mengubah mekanisme dalam menanggapi penggunaan obat tersebut.

Gambar 3.9 Pemisahan artikel menjadi list kalimat

Pemisahan kalimat menjadi kata-kata dan contoh hasil *preprocess* ditunjukkan pada Tabel 3.1 dan Tabel 3.2.

Tabel 3.1 Pemisahan kalimat menjadi per kata

Kalimat	Kata-kata				
1	Antibiotik	adalah	obat	yang	digunakan
	untuk	mencegah	dan	mengobati	infeksi
	bakteri.				
2	Sejak	ditemukan,	penemuan	yang	dipandang
	sebagai	prestasi	dunia	kedokteran	modern
	ini	berfungsi	sebagai	landasan	pengobatan.
3	Namun	demikian,	penggunaan	berlebihan	dan

	penyalahgunaan	antibiotik	dalam	layanan	kesehatan
	manusia	dan	hewan	telah	mendorong
	penyebaran	resistensi	antibiotik.		
4	Resistensi	antibiotik	terjadi	ketika	bakteri
	mengubah	mekanisme	dalam	menanggapi	penggunaan
	obat	tersebut.			

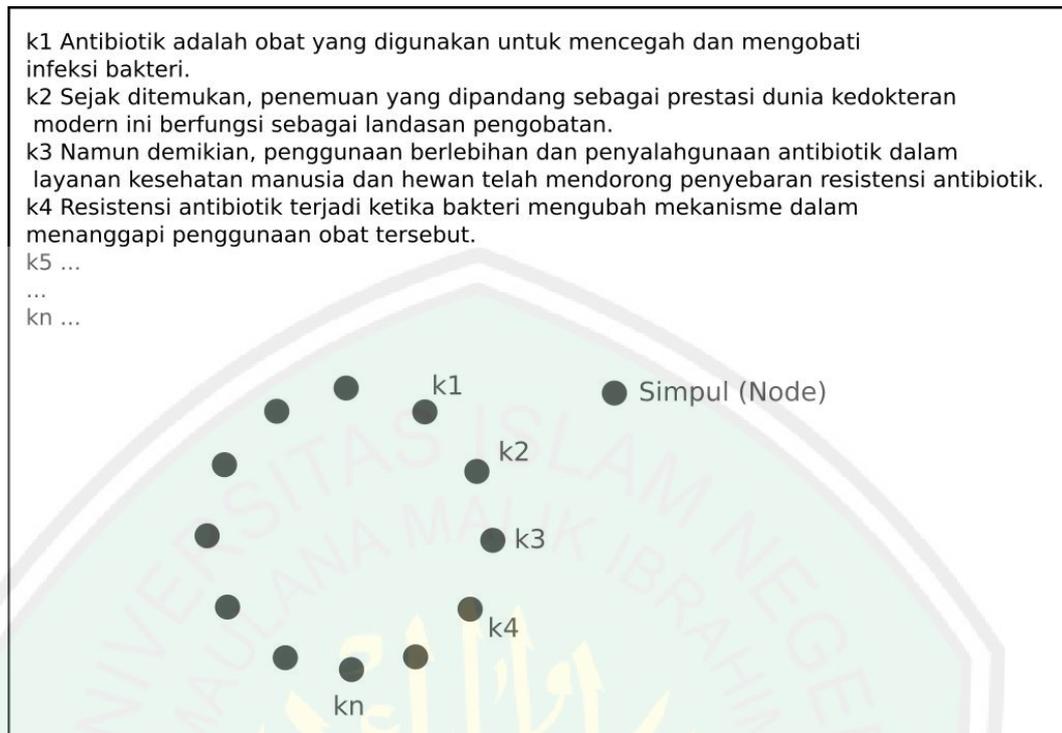
Tabel 3.2 Hasil *Preprocess*

Kalimat	Kata-kata				
1	antibiotik	obat	guna	untuk	cegah
	obat	infeksi	bakteri		
2	temu	temu	pandang	prestasi	dunia
	dokter	modern	fungsi	landas	obat
3	guna	antibiotik	layan	sehat	manusia
	hewan	dorong	sebar	resistensi	antibiotik
4	resistensi	antibiotik	jadi	bakteri	ubah
	mekanisme	tanggap	guna	obat	sebut

### 3.2.2 Tahap *Main Process*

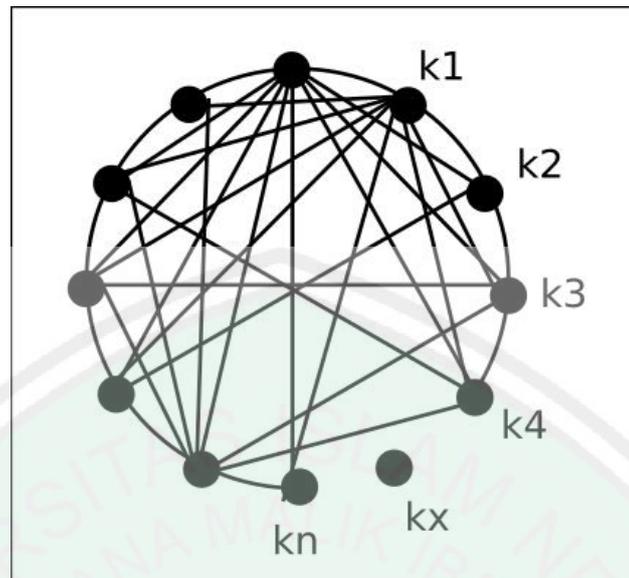
Tahap *Main Process* adalah representasi teks menjadi graf, mengidentifikasi hubungan antarkalimat, pembobotan tiap kalimat berdasarkan identifikasi hubungan antarkalimat dan perangkingan kalimat berdasarkan bobot yang telah dihitung. Representasi dilakukan sebagai berikut:

1. Ekstraksi kalimat dengan menjadikan seluruh kalimat sebagai simpul dalam graf. Graf yang dibuat adalah graf tak-berarah (*undirected graph*). Ilustrasi dapat dilihat pada Gambar 3.10.



Gambar 3.10 Ilustrasi kalimat menjadi simpul

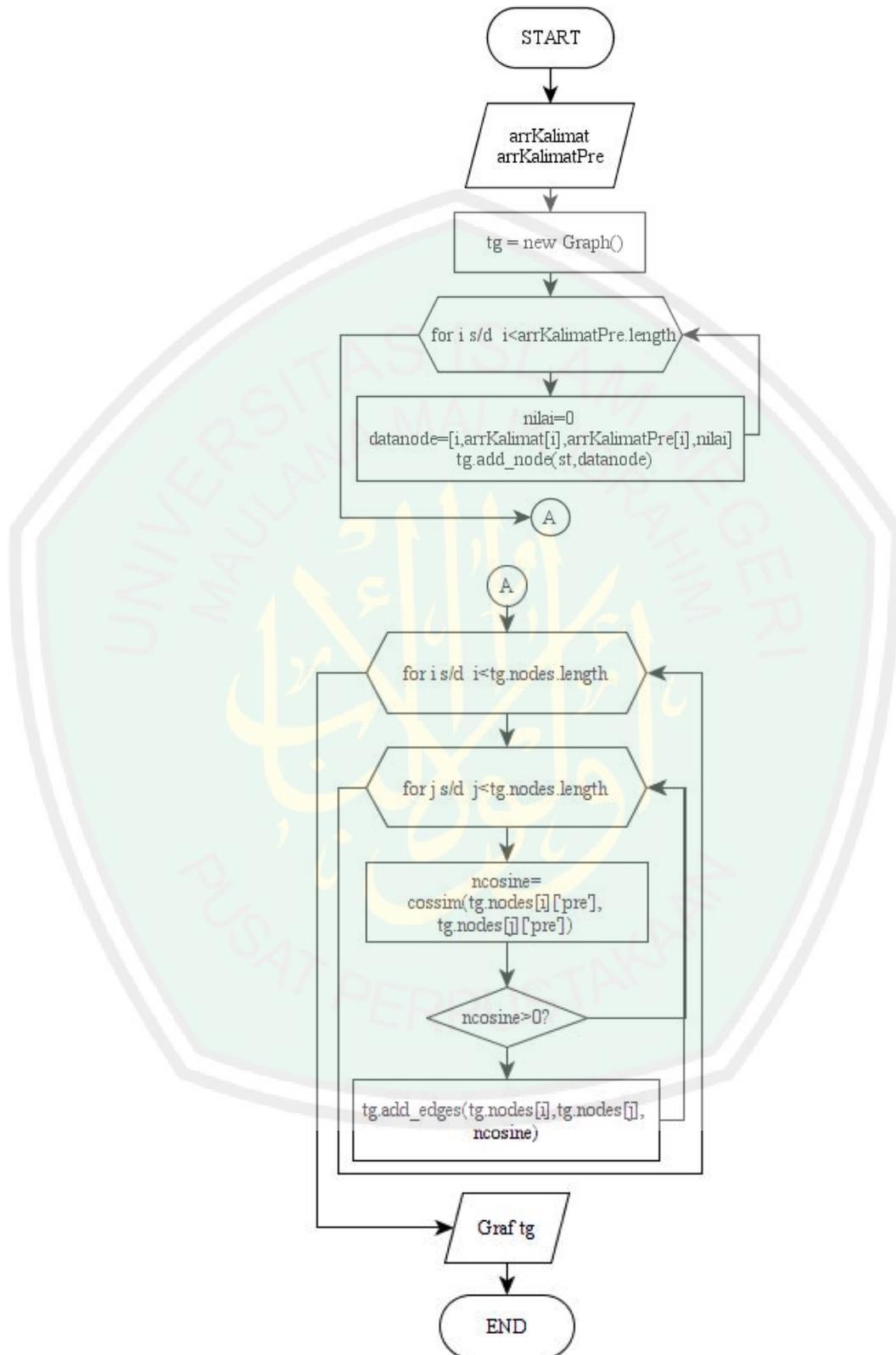
2. Identifikasi hubungan antarkalimat dengan membuat sisi antara simpul-simpul dengan menggunakan *cosine similarity*. *Cosine similarity* merupakan fungsi yang menerima dua buah objek atau lebih dan mengembalikan nilai kemiripan (*similarity*) antara kedua objek tersebut berupa bilangan riil. *Cosine Similarity* mengukur dua objek vektor yang masing-masing memiliki  $n$  dimensi dengan menemukan nilai kosinus dari sudut yang dibentuk dua vektor tersebut. Umumnya, nilai yang dihasilkan oleh fungsi *similarity* berkisar pada interval  $[0...1]$  (Zhiqiang, 2009). Ilustrasi identifikasi hubungan antarkalimat dapat dilihat pada Gambar 3.11.



Gambar 3.11 Graf dengan sisi berbobot

Jika simpul yang memiliki sisi yang terhubung pada simpul yang lain, maka simpul tersebut diketahui memiliki hubungan karena memiliki nilai *cosine similarity* lebih dari 0, yang artinya kedua kalimat tersebut memiliki keterkaitan. Pada contoh di atas terdapat simpul  $kx$  yang tidak memiliki sisi apapun yang terhubung ke simpul lain. Artinya kalimat pada simpul tersebut tidak memiliki keterkaitan antara kalimat manapun dari seluruh artikel.

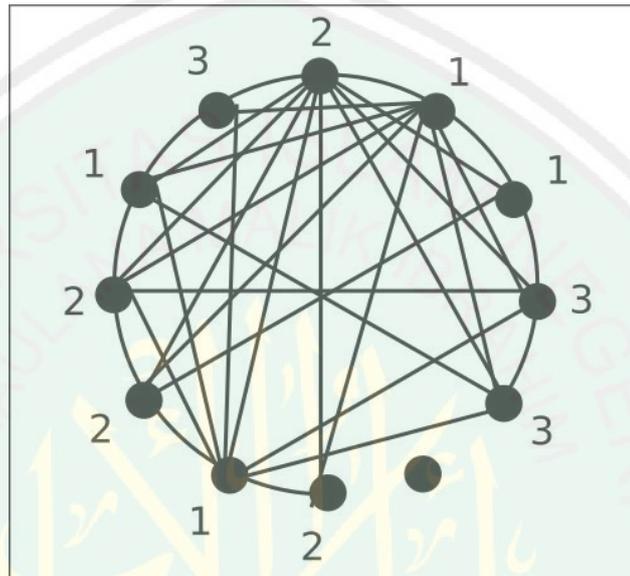
Tahap ekstraksi kalimat dan identifikasi tersebut bisa dilakukan dalam satu rangkaian pada Gambar 3.12 berikut:



Gambar 3.12 *Flowchart* Representasi Kalimat menjadi Graf

Setelah menjadi graf berbobot maka dilakukan iterasi penilaian kalimat dengan tahap berikut.

3. Memberi skor awal simpul berupa nilai acak untuk menentukan iterasi. Ilustrasi inisialisasi skor dapat dilihat pada Gambar 3.13 berikut.



Gambar 3.13 Graf dengan sisi & simpul berbobot

Dalam kasus di gambar, setiap simpul memiliki nilai *random* 1-3. Nilai *random* ini tidak mempengaruhi perhitungan iterasi *TextRank* berikutnya.

4. Melakukan iterasi algoritma *TextRank* sampai *error rate* tiap simpul konvergen atau mencapai ambang batas di bawah *threshold*. *Error Rate* disini adalah perbedaan antar dua skor simpul yang dihitung pada iterasi yang berurutan dengan rumus:

$$S^{k+1}(V_i) - S^k(V_i). \quad (3.1)$$

Iterasi *TextRank* ditunjukkan pada Gambar 3.14.



## Pembentukan graf

Source code Pembentukan graf ditampilkan pada Gambar 3.15.

```
def buatgraf(pre,pre1):
    tg = nx.Graph()
    for i,p1,kalimatpre in zip(range(len(pre1)),pre1,pre):
        tg.add_node(i,kalimat=p1,pre=kalimatpre,nilai=0)
    # set nilai sisi (w)
    for i in tg.nodes:
        for j in tg.nodes:
            if i==j: continue
            ncosine=cossim(tg.nodes[i]['pre'],tg.nodes[j]['pre'])
            if ncosine>0:
                tg.add_edge(i,j,weight=ncosine)
            else:
                if debug['buatgraf'] :
                    print(i,"x",j,"sim :",ncosine,"zero")
    return tg
```

Gambar 3.15 Source code pembentukan graf

## Iterasi TextRank

Source code Iterasi TextRank ditampilkan pada Gambar 3.16.

```
toleransi=1/10000
def textrank(tg):
    nsimpul=[]
    s=[random.randint(1,3) for x in range(len(tg.nodes))]
    iterasi=0
    ilanjut=True
    while ilanjut:
        if debug['textrank2'] : print("iterasi",iterasi)
        nsimpul=[]
        for i in tg.nodes():
            wij,wjk=0,0
            sigma=0
            for j in tg.neighbors(i):
                if debug['textrank'] : print(i,j)
                wij=tg[j][i]['weight']
                if debug['textrank'] : print("wij",wij)
                wjk=sum(tg[i][j]['weight'] for k in tg.neighbors(j))
                if debug['textrank'] : print("wjk d",wjk)
                sigma+=(wij*s[j])/wjk
                if debug['textrank'] : print("sigma",sigma)
            # sigma
            if debug['textrank'] :
                print("wij",wij,"wjk",wjk)
                print("sigma",sigma)
            if wjk>0:
                txtrank=(1-d)+d*sigma
                if debug['textrank'] :
                    print("s[" ,i,"]",s[i])
```

```

        print("txt",txtrank)
    # hitung error
    error=math.fabs(txtrank-s[i])

    if error>toleransi:
        s[i]=txtrank
    elif i==(len(tg.nodes)-1):
        ilanjut=False

    tg.nodes[i]['nilai']=txtrank
    nsimpul.append([i,tg.nodes[i]])
    iterasi+=1
    if iterasi==100:
        break
    return nsimpul

```

Gambar 3.16 *Source code* Iterasi *TextRank*

*Source code* tersebut menunjukkan hasil seperti pada Gambar 3.17.

### #3

Wk1 : 1  
Wk2 : 1  
Wk3 : 1  
...  
Wkn : 1

### #4

**Toleransi = 1/10000 ~ 0.0001**

$$WS(V_i) = (1 - d) + d * \sum_{V_j \in In(V_i)} \frac{w_{ji}}{\sum_{V_k \in Out(V_j)} w_{jk}} WS(V_j)$$

$$S^{k+1}(V_i) - S^k(V_i).$$

	iterasi 0	iterasi 1 dst.	selisih
Wk1 :	1.05	1.04	= 0.01
Wk2 :	1.08	1.019	= 0.061
Wk3 :	1.03	1.02	= 0.01
...	...	...	...
Wkn :	x0	x1	= x0-x1

if Wkx<= toleransi then break

### #5

Wk3 : 1.83  
Wk1 : 1.65  
Wk2 : 1.58  
...  
Wkn : x

Gambar 3.17 Ilustrasi alur penghitungan iterasi *TextRank*

### 3.2.3 Uji coba sistem

Menurut Lin(2004), ROUGE (*Recall-Oriented Understudy For Gisting Evaluation*) adalah salah satu metode pengukuran otomatis; rangkuman dari hasil otomatis dibandingkan dengan rangkuman yang dibuat oleh manusia. Pengukuran dilakukan dengan menghitung unit-unit yang cocok seperti *n-gram* (n-kata), urutan kata dan pasangan kata antara angkuman dari hasil otomatis dibandingkan dengan rangkuman yang dibuat oleh manusia. Berikut adalah rumus ROUGE - N.

$$ROUGE-N = \frac{(\sum_{S \in \{ReferenceSummaries\}} \sum_{gram_n \in S} Count_{match}(gram_n))}{(\sum_{S \in \{ReferenceSummaries\}} \sum_{gram_n \in S} Count(gram_n))} \quad (3.2)$$

ROUGE-N merupakan pengukuran yang berorientasi pada nilai *recall*, karena penyebut dari persamaan tersebut merupakan jumlah dari n-gram yang ada pada rangkuman manual (rangkuman referensi). Sehingga, *recall* di sini adalah nilai perbandingan *n-gram* hasil rangkuman sistem dengan n-gram hasil rangkuman referensi.

*Source code* evaluasi menggunakan ROUGE ditampilkan pada Gambar 3.18

```
def rouge(teksmanual,tekssys):
    rouge = Rouge()
    nilai = rouge.get_scores(tekssys, teksmanual)
    return nilai
```

Gambar 3.18 *Source code* evaluasi ROUGE

### 3.2.4 User Interface

Sistem perangkat dibangun pada *platform web*. Berikut adalah halaman-halaman yang hendak dibangun.

#### 1. Halaman utama

No	Sumber/Judul	operasi
1	judul 1 sumber 1	[edit] [delete]
2	judul 2 sumber 2	[edit] [delete]

Gambar 3.19 Rancangan UI halaman utama

Halaman utama pada Gambar 3.19 menampilkan *form* input artikel dan daftar artikel yang telah disimpan, serta dengan tombol operasi ubah dan hapus.

## 2. Halaman artikel

Gambar 3.20 Rancangan UI halaman artikel

Halaman artikel pada Gambar 3.20 menampilkan judul, sumber, isi, rangkuman sistem dan rangkuman manual atau rangkuman referensi artikel. Pada setiap *textarea* ditunjukkan nilai statistik teks dan pada rangkuman sistem terdapat hasil penilaian ROUGE jika tombol nilai ditekan.

Sistem *Automatic TextRank Summarization* diimplementasikan pada dua halaman yaitu :

#### a. Halaman awal

Implementasi halaman awal ditunjukkan pada Gambar 3.21 berikut ini.



Gambar 3.21 Halaman Awal

Halaman awal menampilkan daftar artikel yang telah tersimpan di *database* serta skor ROUGE yang diperoleh setelah perangkuman. Berikut adalah keterangan dari Gambar 3.21.

##### 1. Menu Lihat semua artikel

*Link* untuk menuju halaman utama yang menampilkan seluruh artikel tersimpan di *database*.

##### 2. Menu Input artikel

Link untuk menuju halaman input artikel baru.

3. Judul dan *link* ke halaman artikel

*Link* untuk menuju halaman artikel spesifik yang telah tersimpan di *database*, beserta *link* menuju halaman sumber asli.

4. *Textfield* pencarian

*Form* input untuk mencari artikel berdasarkan nama.

5. Nilai Rouge

Menampilkan nilai Rouge yang telah terhitung dan tersimpan di *database*.

6. *Button* Edit

*Link* untuk menuju halaman artikel spesifik yang telah tersimpan di *database*.

7. *Button* Hapus

Aksi untuk menghapus artikel dari *database*.

### b. Halaman Input dan Perangkuman

Implementasi halaman input ditunjukkan pada Gambar 3.22 dan Gambar 3.23 berikut ini.

The screenshot shows a web form titled "Input artikel" with the following elements:

- 1**: A text input field containing the title "Senjakala Industri Manga Jepang Akibat Pembajakan".
- 2**: A text input field containing the source URL "https://tirto.id/senjakala-industri-manga-jepang-akibat-pembajakan-c9Wn".
- 3**: A text input field containing the title "Senjakala Industri Manga Jepang Akibat Pembajakan".
- 4**: A text area containing the article body text, starting with "Pembajakan dan kisah yang kurang menggigit mengancam ketahanan industri manga Jepang. tirto.id - Siapa tak kenal One Piece? Serial komik Jepang atau manga ini begitu populer, tidak hanya di negeri asalnya Jepang, tapi juga di seluruh dunia. Sayangnya, kesuksesan One Piece tak otomatis mewakili kesuksesan industri manga secara keseluruhan." and ending with "Angka penjualan tersebut terlihat mencolok ketika dibandingkan dengan komik yang menempati peringkat kedua, yakni".
- 5**: A green button labeled "Rangkum" at the bottom left of the form.

Gambar 3.22 Bagian halaman input artikel

Berikut adalah keterangan dari Gambar 3.22.

1. *Judul Artikel*

Menampilkan judul Artikel yang telah tersimpan.

2. *Textfield Sumber*

Form input untuk memasukkan sumber artikel.

3. *Textfield Judul*

Form input untuk memasukkan judul artikel.

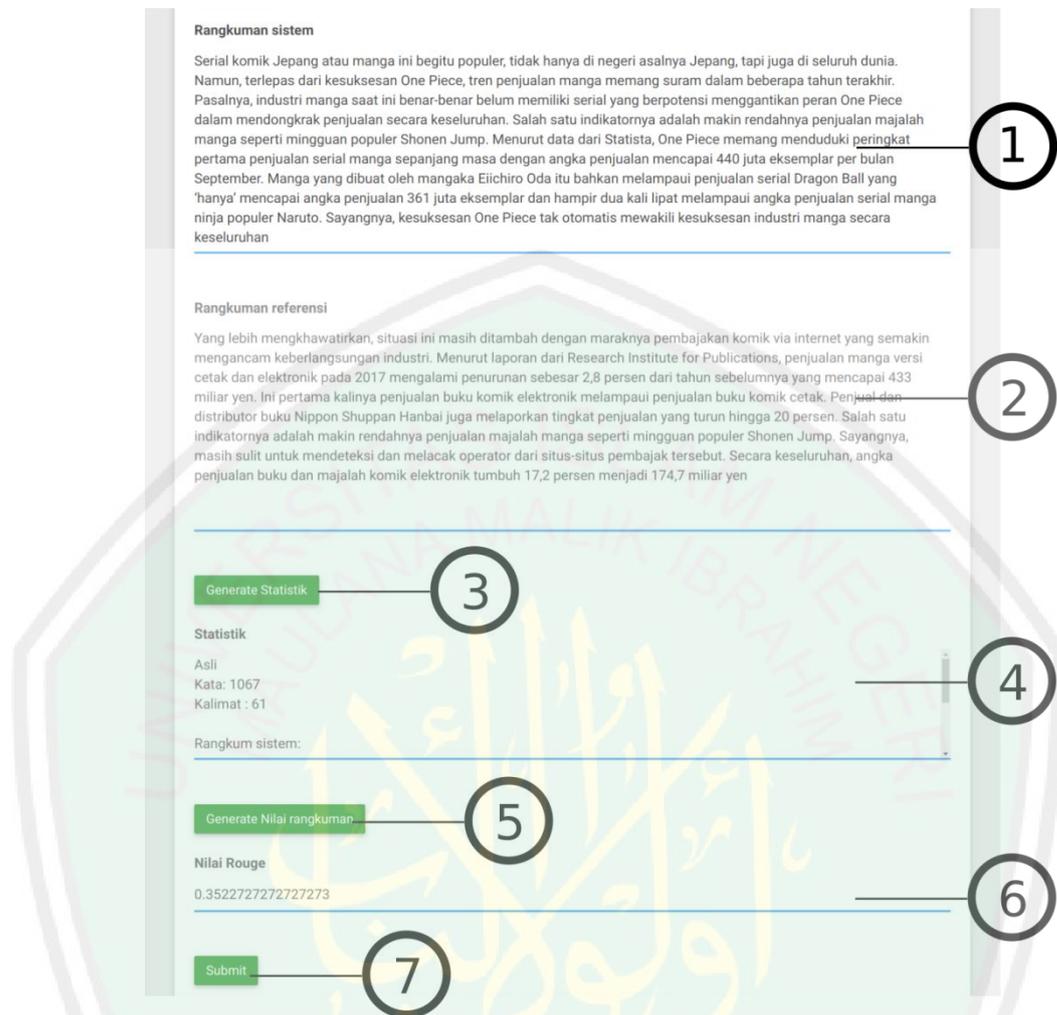
4. *TextArea Isi*

Form input untuk memasukkan isi artikel asli.

5. *Button Rangkum*

*Button* untuk memulai perangkuman teks, setelah isi artikel asli terisi dan tersimpan.





Gambar 3.23 Bagian halaman perangkuman artikel

Berikut adalah keterangan dari Gambar 3.23.

1. *Textarea* rangkuman sistem

Setelah tombol rangkum ditekan, rangkuman sistem akan muncul di sini dengan *delay* yang relatif pada panjang artikel.

2. *Textarea* rangkuman referensi

Form input untuk memasukkan rangkuman referensi dan disimpan di *database*.

3. *Button* Generate Statistik

*Button* untuk menghitung statistik teks asli, rangkuman sistem dan rangkuman referensi. Statistik yang dihitung yaitu jumlah kalimat dan kata.

#### 4. *Textarea* Statistik

Setelah *button* generate statistik ditekan, akan keluar hasil perhitungan di *field* ini.

#### 5. *Button* Generate Nilai Rangkuman

*Button* untuk menghitung nilai Rouge setelah perangkuman selesai dan tersimpan di database.

#### 6. *Textfield* Nilai Rouge

Setelah *button* generate nilai ditekan, akan keluar hasil perhitungan nilai rangkuman di *field* ini.

#### 7. *Button* Submit

*Button* untuk menyimpan semua data yang ada di semua *field*.

## BAB IV

### UJI COBA DAN PEMBAHASAN

Bab ini terdiri dari skenario uji coba, hasil uji coba dan pembahasan serta integrasi penelitian dengan pandangan Islam. Uji coba dilakukan untuk memperoleh nilai relevansi hasil rangkuman dari sistem yang telah dibangun berupa nilai *recall*.

#### 4.1. Skenario Uji Coba

Skenario uji coba dimulai dengan pengumpulan data teks serta pembuatan rangkuman referensi. Pengumpulan dokumen / artikel sebanyak 50 teks berita dari situs *tirto.id* tidak dibatasi pada sebuah topik tertentu. Artikel-artikel kemudian dikumpulkan dan disimpan di sistem melalui menu input artikel. Setelah artikel dikumpulkan mulai membuat form online untuk mengumpulkan data rangkuman referensi dari 10 responden tetap. Responden tersebut akan memilih kalimat terbaik menurut responden yang berjumlah setengah dari jumlah kalimat keseluruhan dari sebuah artikel. Misal, sebuah artikel berjumlah 60 kalimat, maka responden memilih 30 kalimat terbaik menurutnya untuk dijadikan rangkuman referensi. Setiap individu responden bisa memilih kalimat yang berbeda-beda, sehingga dilakukan perhitungan kalimat mana yang dipilih paling banyak oleh responden-responden tersebut. Kalimat-kalimat tersebut kemudian dirangking menurut banyaknya pilihan dan diambil seperdelapan dari seluruh jumlah kalimat di artikel asli dan dikumpulkan menjadi satu rangkuman referensi.

*Preprocessing (Stemming, Tokenizing, Filtering, dan Stopwords dan numbers Removal)* dilakukan di dalam sistem, sehingga hasil prerprocess tidak

disimpan. Setelah *preprocessing*, sistem melakukan penghitungan kalimat-kalimat mana saja yang menjadi kalimat terbaik untuk menjadi rangkuman. Kalimat terbaik ditunjukkan oleh hasil perhitungan skor *TextRank* hingga iterasi terakhir. Pengurutan kalimat dimulai dari yang memiliki skor *TextRank* tertinggi (*descending*). Kemudian rangkuman diambil seperdelapan dari keseluruhan kumpulan kalimat yang telah diurutkan secara *descending*, lalu disimpan di *database*.

Langkah terakhir yaitu melakukan evaluasi hasil rangkuman sistem yang dibandingkan dengan rangkuman referensi. Tingkat keberhasilan diukur dengan menggunakan nilai *Recall* dari nilai ROUGE-N (pada kesempatan ini menggunakan N=1 ~ ROUGE-1). Sistem penghitungan ini menggunakan *library python Rouge*. Hasil perhitungan kemudian juga disimpan di *database*.

Proses perancangan, pembangunan dan uji coba sistem dilakukan pada *hardware* dan *software* dengan spesifikasi sebagai berikut:

#### 1. *Hardware*

- a. Lenovo-G480 @ Intel(R) Core(TM) i3-3120M CPU @ 2.50GHz
- b. RAM 4GB
- c. VGA compatible controller, 3rd Gen Core processor Graphics Controller.  
NVIDIA 610M 1GB.
- d. HDD 500 GB

#### 2. *Software*

- a. Linux 4.15.0-47-generic #50-Ubuntu 18.04 x86\_64 GNU/Linux (uname -a)
- b. XAMPP for Linux 7.3.7

c. Visual Studio Code Versi : 1.33.1, Linux x64

d. *Browser* Google Chrome & Brave

#### 4.2. Hasil Uji Coba

Bagian ini terdiri dari hasil pengumpulan dokumen, hasil pengolahan rangkuman referensi dan hasil pengujian rangkuman sistem. Hasil pengumpulan dokumen / artikel berita dapat dilihat pada Tabel 4.1.

Tabel 4.1 Daftar judul Dokumen artikel

No	Judul Artikel
1	Senjakala Industri Manga Jepang Akibat Pembajakan
2	Esemka: Mainan 'Mobil Nasional' Jokowi yang Masih Ilusi
3	Kenapa Banyak Orang Berhenti Mencari Musik Baru?
4	Zakir Naik Tidak Mendukung ISIS, Tapi...
5	Le Minerale Versus Aqua: Bertarung di Pasar Berlanjut di Pengadilan
6	'Baku Hantam' Jokowi-Prabowo di Kandang Lawan
7	Bangkitnya Penyakit Ganas di Dunia Pasca-Antibiotik
8	Saat Banjir Tiba, Kendaraan Motor Lebih Berharga daripada Nyawa
9	Cerita Lengkap Pembunuhan Satu Keluarga di Bekasi
10	Evakuasi Korban Lion Air Makan Biaya Besar, Siapa yang Menanggung?
11	Jika Rizieq Terbukti Makar di Saudi, Indonesia Tak Bisa Apa-apa
12	Jutaan Jiwa Tinggal di Kawasan Rawan Longsor, Kok Tidak Direlokasi?
13	Kok Bisa Kiai NU Tak Masuk Survei Ulama Berpengaruh LSI?
14	Kok Bisa Pemprov DKI Jadi Provinsi Paling Informatif?
15	Kontroversi Acara Bertajuk Syiar Kekhalifahan yang Ditolak Warga
16	KPU Rentan Serangan Siber, Tapi Antisipasi Masih Minim
17	Mampukah Dirjen Hubungan Udara Baru Bikin Penerbangan Makin Aman?
18	Manuver Yusril: Membela HTI, Diabaikan Prabowo, dan Menjaga Jokowi
19	Sampai Kapan Basarnas Mencari Para Korban Pesawat Lion Air JT 610?
20	Selamat Datang di Pilpres yang Sudutkan Kaum Rentan & Disabilitas
21	Sistem di Balik Pengumpulan Nasi Kotak Reuni 212
22	UGM Tak Serius Tangani Kasus Kekerasan Seksual di Kampus?
23	Selamat Datang di Musim Paceklik Penerbangan
24	Mengapa Medan Magnet di Kutub Utara Bergeser Begitu Cepat?
25	Kondisi Psikologis Adi Saputra, dan Pengemudi yang Wajib Diwaspadai
26	Periksa Fakta Benarkah Server KPU Jebol dan Jokowi Sudah Punya Angka 57 Persen? Header Periksa Fakta.

27	Melajang: Salah Satu Cara Menuju Hidup Sehat, Santai, & Bahagia
28	Anti-Natalisme: Mengutuk Kehidupan, Melarang Kelahiran
29	Mengapa Dalih Mahal Demi Menghindari Energi Terbarukan Bermasalah?
30	Kasus Mourinho: Kenapa Penggelapan Pajak Sering Terjadi di Spanyol?
31	Memanggil Presiden Jokowi dengan Cak Jancuk, Bisakah Dijerat Hukum?
32	Kasus Agni Berakhir "Damai": Cermin Buram Kasus Pelecehan di Kampus
33	Cibiran & Stigma Jadi Alasan Masyarakat Permisif Sikapi Bunuh Diri
34	Fellaini adalah Simbol Betapa Tak Menarik Manchester United
35	Harlah NU 2019
36	Efek Domino Bagasi Berbayar: Ganggu Pariwisata, Hotel hingga UMKM
37	Menyelisik Tabloid Indonesia Barokah, Apa Isi dan Siapa di Balikny
38	Jakarta, Kota dengan 9 Kasus Pencurian Setiap Hari
39	Fadli Zon Bersilat Lidah: Bela Ahmad Dhani Tapi Enggan Ubah UU ITE
40	Wacana Motor Masuk Tol Berpotensi Tingkatkan Angka Kecelakaan
41	Rekam Jejak Kegagalan PSSI Perjuangkan Klub Indonesia di Asia
42	Jokowi Tambah Utang Rp1.809 Triliun Selama Menjabat, Berisiko?
43	Perokok Remaja Semakin Banyak, Regulasi yang Tumpul Penyebabnya
44	Menengok Langkah Anies Atasi Penurunan Tanah Agar Tak Jadi Bencana
45	Gaji Dokter Memang Masih Jauh di Bawah Rekomendasi IDI
46	Nasib Ribuan Karyawan Bank yang Disingkirkan Mesin
47	SNMPTN: Betapa Ketatnya Persaingan Masuk PTN
48	Mitos dan Fakta tentang Kertas
49	City Brain: Solusi Untuk Masalah di Jalanan?
50	Tata Ruang Indonesia Mengabaikan Potesi Bencana

Setiap individu responden bisa memilih kalimat yang berbeda-beda, sehingga dilakukan perhitungan kalimat mana yang dipilih paling banyak oleh responden-responden tersebut. Kalimat-kalimat tersebut kemudian dirangking menurut banyaknya pilihan dan diambil sebagian dari seluruh jumlah kalimat di artikel asli dan dikumpulkan menjadi satu rangkuman referensi.

Berikut ini adalah contoh penjelasan salah satu artikel, nomor tujuh dari Tabel 4.1 , dengan judul “Bangkitnya Penyakit Ganas di Dunia Pasca-Antibiotik”. Isi artikel ditunjukkan pada Gambar 4.1.

#### **Bangkitnya Penyakit Ganas di Dunia Pasca-Antibiotik**

Bakteri semakin kebal antibiotik. Penyakit seperti pneumonia, TBC, sepsis, dan gonorea kian sulit disembuhkan karena antibiotik semakin tak efektif.

tirto.id - Pada 12-18 November 2018 diselenggarakan Pekan Kewaspadaan Antibiotik Dunia (World Antibiotic Awareness Week). Momentum ini bertujuan untuk meningkatkan kesadaran global terhadap resistensi antibiotik dan mendorong praktik terbaik di kalangan masyarakat umum, tenaga kesehatan, dan pembuat kebijakan untuk menghindari resistensi antibiotik yang telah mewabah.

Apa yang perlu dilakukan?

Antibiotik adalah obat yang digunakan untuk mencegah dan mengobati infeksi bakteri. Sejak ditemukan, penemuan yang dipandang sebagai prestasi dunia kedokteran modern ini berfungsi sebagai landasan pengobatan. Namun demikian, penggunaan berlebihan dan penyalahgunaan antibiotik dalam layanan kesehatan manusia dan hewan telah mendorong penyebaran resistensi antibiotik. Resistensi antibiotik terjadi ketika bakteri mengubah mekanisme dalam menanggapi penggunaan obat tersebut.

Walhasil, justru bakterilah—bukan manusia—yang semakin kebal terhadap antibiotik. Bakteri ini kemudian dapat menginfeksi manusia lain dan lebih sulit untuk diobati daripada bakteri biasa.

Resistensi antibiotik telah menyebabkan biaya kesehatan meroket, perawatan di rumah sakit yang berkepanjangan, dan angka kematian yang meningkat. Di seluruh Uni Eropa, bakteri yang resisten terhadap antibiotik diperkirakan telah menyebabkan 25.000 kematian, kerugian karena gangguan produktivitas, dan biaya kesehatan yang mencapai lebih dari 1,5 miliar dolar AS per tahun.

Seiring meningkatnya kasus resistensi antibiotik di seluruh belahan dunia, mekanisme resistensi pun terus mengalami pembaruan sehingga mengancam keberhasilan pengobatan penyakit. Penyakit infeksi umum seperti pneumonia, TBC, sepsis, dan gonorea menjadi lebih sulit ditangani, bahkan kadang mustahil disembuhkan karena antibiotik yang tersedia semakin tak efektif.

Di negara-negara yang menjual antibiotik tanpa resep, kemunculan dan tingkat penyebaran resistensi lebih buruk. Demikian pula di negara tanpa pedoman pengobatan standar, di mana antibiotik sering diresepkan berlebihan (over-prescribed) oleh dokter

dan digunakan berlebihan (over-used) oleh masyarakat. Jika situasi ini tidak cepat-cepat ditangani, kita akan menuju era pasca-antibiotik, di mana infeksi umum dan luka ringan bisa membunuh pasien—persis di era pra-antibiotik.

"Resistensi antimikroba," kata Direktur Jenderal WHO Dr. Tedros Adhanom Ghebreyesus, "adalah keadaan darurat kesehatan global, yang secara serius akan membahayakan kemajuan pengobatan modern." Selain Tuberkulosis (TBC) yang resisten terhadap berbagai jenis obat, para ahli telah mengidentifikasi 12 kelas bakteri patogen prioritas yang semakin resisten atau kebal terhadap antibiotik yang ada di pasaran. Beberapa bakteri di antaranya menyebabkan infeksi paru-paru atau pneumonia dan infeksi saluran kemih. Penanganan penyakit-penyakit ini sangat membutuhkan obat antibiotik dengan kelas terapi yang baru.

Sialnya, di antara semua kandidat obat antibiotik baru, hanya delapan yang digolongkan oleh WHO sebagai obat inovatif yang akan memberi nilai tambah pada cadangan antibiotika hari ini. Dengan demikian, hanya ada sedikit pilihan pengobatan untuk TBC dan bakteri patogen gram negatif yang telah mengalami pertambahan resistensi secara luas, termasuk bakteri *Acinetobacter* dan *Enterobacteriaceae*. Seperti bakteri gram negatif terdahulu, yaitu *Klebsiella* dan *E.coli*, semua bakteri tersebut dapat menyebabkan infeksi parah atau sepsis yang mematikan.

Beberapa langkah penting lain harus diambil untuk mengurangi dampak resistensi antibiotik dan membatasi penyebarannya. Masyarakat umum bisa mencegah infeksi dengan teratur mencuci tangan, menjaga kebersihan makanan, menghindari kontak dekat dengan orang yang sakit, dan melakukan vaksinasi ulangan.

Antibiotik tetap bisa digunakan selama diresepkan oleh dokter. Namun, hal lain yang tak kalah penting adalah tidak mengonsumsi antibiotik sisa atau berbagi antibiotik dengan orang lain.

Para petugas sektor pertanian dapat berkontribusi mengurangi wabah resistensi antibiotik dengan cara menggunakan antibiotik untuk hewan dalam kasus penyakit menular saja dan di bawah pengawasan dokter hewan. Petugas juga bisa memberikan vaksinasi kepada hewan untuk mengurangi kebutuhan antibiotik.

Pada lini lainnya, promosi dan penerapan praktik yang baik perlu dilakukan di semua tahap produksi, mulai dari pengolahan makanan dari sumber hewan dan tumbuhan yang aman, pengadopsian sistem berkelanjutan dengan meningkatkan kebersihan, penanganan hewan agar bebas stres, hingga pemenuhan standar internasional untuk penggunaan antibiotik yang bertanggung jawab sesuai ketentuan OIE, FAO, dan WHO. Para pelaku industri bidang kesehatan dapat mengambil peran penting seperti berinvestasi untuk antibiotik, vaksin, alat diagnosis baru, serta membiayai riset terkait.

Adalah kewajiban bagi para dokter, apoteker, dan petugas kesehatan lainnya untuk mencegah infeksi dengan memastikan kebersihan tangan, instrumen medis, dan lingkungan RS. Mereka juga wajib memberikan vaksinasi terbaru kepada pasien (ketika terjadi dugaan infeksi bakteri), melakukan biakan bakteri dan pemeriksaan penunjang medik lainnya untuk konfirmasi, serta hanya meresepkan dan mengeluarkan antibiotik jika benar-benar dibutuhkan, pada dosis dan durasi pengobatan yang tepat.

Terakhir, para pejabat dan pembuat kebijakan kesehatan mesti bertindak cepat dengan menyusun rencana terpadu secara nasional, bahkan regional, meningkatkan pengawasan infeksi bakteri yang telah resisten terhadap antibiotik, memperkuat langkah-langkah pencegahan dan pengendalian infeksi, juga mengatur dan mempromosikan penggunaan obat antibiotik yang tepat dan berkualitas. Selain itu, pembuat kebijakan juga harus menyusun informasi tentang dampak resistensi antibiotik dan memberikan apresiasi atas pengembangan obat, vaksin, serta alat diagnosis baru.

Momentum Pekan Kewaspadaan Antibiotik Dunia (World Antibiotic Awareness Week) 2018 mengingatkan kita untuk berpikir ulang dan mencari saran (Think twice. Seek advice), sebelum menggunakan antibiotik. Kita juga disadarkan bahwa penyalahgunaan antibiotik akan membuat kita semua berisiko.

Waktu kita dengan antibiotik hampir habis, sehingga menjadi tugas kita bersama untuk menjaga antibiotik. Sudahkah kita sadar?

Gambar 4.1 Isi artikel asli (data input)

Artikel asli tersebut kemudian dipisah menjadi per kalimat untuk dijadikan pilihan kalimat yang bisa dipilih oleh responden. Kemudian mengambil data kalimat mana yang paling banyak dipilih oleh 10 responden. Hasil daftar kalimat dari artikel asli di atas ditunjukkan pada Tabel 4.2, hasil pilihan 10 responden dan total pilihan ditunjukkan pada Tabel 4.2.

Tabel 4.2 Pemisahan artikel menjadi kalimat-kalimat

No. kalimat	Kalimat
1	Bakteri semakin kebal antibiotik.
2	Penyakit seperti pneumonia, TBC, sepsis, dan gonorea kian sulit disembuhkan karena antibiotik semakin tak efektif.
3	tirto.id - Pada 12-18 November 2018 diselenggarakan Pekan Kewaspadaan Antibiotik Dunia (World Antibiotic Awareness Week).
4	Momentum ini bertujuan untuk meningkatkan kesadaran global

	terhadap resistensi antibiotik dan mendorong praktik terbaik di kalangan masyarakat umum, tenaga kesehatan, dan pembuat kebijakan untuk menghindari resistensi antibiotik yang telah mewabah.
5	Apa yang perlu dilakukan?
6	Antibiotik adalah obat yang digunakan untuk mencegah dan mengobati infeksi bakteri.
7	Sejak ditemukan, penemuan yang dipandang sebagai prestasi dunia kedokteran modern ini berfungsi sebagai landasan pengobatan.
8	Namun demikian, penggunaan berlebihan dan penyalahgunaan antibiotik dalam layanan kesehatan manusia dan hewan telah mendorong penyebaran resistensi antibiotik.
9	Resistensi antibiotik terjadi ketika bakteri mengubah mekanisme dalam menanggapi penggunaan obat tersebut.
10	Walhasil, justru bakterilah—bukan manusia—yang semakin kebal terhadap antibiotik.
11	Bakteri ini kemudian dapat menginfeksi manusia lain dan lebih sulit untuk diobati daripada bakteri biasa.
12	Resistensi antibiotik telah menyebabkan biaya kesehatan meroket, perawatan di rumah sakit yang berkepanjangan, dan angka kematian yang meningkat.
13	Di seluruh Uni Eropa, bakteri yang resisten terhadap antibiotik diperkirakan telah menyebabkan 25.000 kematian, kerugian karena gangguan produktivitas, dan biaya kesehatan yang mencapai lebih dari 1,5 miliar dolar AS per tahun.
14	Seiring meningkatnya kasus resistensi antibiotik di seluruh belahan dunia, mekanisme resistensi pun terus mengalami pembaruan sehingga mengancam keberhasilan pengobatan penyakit.
15	Penyakit infeksi umum seperti pneumonia, TBC, sepsis, dan gonorea menjadi lebih sulit ditangani, bahkan kadang mustahil disembuhkan karena antibiotik yang tersedia semakin tak efektif.
16	Di negara-negara yang menjual antibiotik tanpa resep, kemunculan dan tingkat penyebaran resistensi lebih buruk.
17	Demikian pula di negara tanpa pedoman pengobatan standar, di mana antibiotik sering diresepkan berlebihan (over-prescribed) oleh dokter dan digunakan berlebihan (over-used) oleh masyarakat.
18	Jika situasi ini tidak cepat-cepat ditangani, kita akan menuju era pasca-antibiotik, di mana infeksi umum dan luka ringan bisa membunuh pasien—persis di era pra-antibiotik.
19	Resistensi antimikroba,” kata Direktur Jenderal WHO Dr.
20	Tedros Adhanom Ghebreyesus, “adalah keadaan darurat kesehatan global, yang secara serius akan membahayakan kemajuan pengobatan modern. Selain Tuberkulosis (TBC) yang resisten terhadap berbagai jenis obat, para ahli telah mengidentifikasi 12 kelas bakteri patogen prioritas yang semakin resisten atau kebal terhadap antibiotik yang

	ada di pasaran.
21	Beberapa bakteri di antaranya menyebabkan infeksi paru-paru atau pneumonia dan infeksi saluran kemih.
22	Penanganan penyakit-penyakit ini sangat membutuhkan obat antibiotik dengan kelas terapi yang baru.
23	Sialnya, di antara semua kandidat obat antibiotik baru, hanya delapan yang digolongkan oleh WHO sebagai obat inovatif yang akan memberi nilai tambah pada cadangan antibiotika hari ini.
24	Dengan demikian, hanya ada sedikit pilihan pengobatan untuk TBC dan bakteri patogen gram negatif yang telah mengalami peningkatan resistensi secara luas, termasuk bakteri Acinetobacter dan Enterobacteriaceae.
25	Seperti bakteri gram negatif terdahulu, yaitu Klebsiella dan E.coli, semua bakteri tersebut dapat menyebabkan infeksi parah atau sepsis yang mematikan.
26	Beberapa langkah penting lain harus diambil untuk mengurangi dampak resistensi antibiotik dan membatasi penyebarannya.
27	Masyarakat umum bisa mencegah infeksi dengan teratur mencuci tangan, menjaga kebersihan makanan, menghindari kontak dekat dengan orang yang sakit, dan melakukan vaksinasi ulangan.
28	Antibiotik tetap bisa digunakan selama diresepkan oleh dokter.
29	Namun, hal lain yang tak kalah penting adalah tidak mengonsumsi antibiotik sisa atau berbagi antibiotik dengan orang lain.
30	Para petugas sektor pertanian dapat berkontribusi mengurangi wabah resistensi antibiotik dengan cara menggunakan antibiotik untuk hewan dalam kasus penyakit menular saja dan di bawah pengawasan dokter hewan.
31	Petugas juga bisa memberikan vaksinasi kepada hewan untuk mengurangi kebutuhan antibiotik.
32	Pada lini lainnya, promosi dan penerapan praktik yang baik perlu dilakukan di semua tahap produksi, mulai dari pengolahan makanan dari sumber hewan dan tumbuhan yang aman, pengadopsian sistem berkelanjutan dengan meningkatkan kebersihan, penanganan hewan agar bebas stres, hingga pemenuhan standar internasional untuk penggunaan antibiotik yang bertanggung jawab sesuai ketentuan OIE, FAO, dan WHO.
33	Para pelaku industri bidang kesehatan dapat mengambil peran penting seperti berinvestasi untuk antibiotik, vaksin, alat diagnosis baru, serta membiayai riset terkait.
34	Adalah kewajiban bagi para dokter, apoteker, dan petugas kesehatan lainnya untuk mencegah infeksi dengan memastikan kebersihan tangan, instrumen medis, dan lingkungan RS.
35	Mereka juga wajib memberikan vaksinasi terbaru kepada pasien (ketika terjadi dugaan infeksi bakteri), melakukan biakan bakteri dan pemeriksaan penunjang medik lainnya untuk konfirmasi, serta hanya

	meresepkan dan mengeluarkan antibiotik jika benar-benar dibutuhkan, pada dosis dan durasi pengobatan yang tepat.
36	Terakhir, para pejabat dan pembuat kebijakan kesehatan mesti bertindak cepat dengan menyusun rencana terpadu secara nasional, bahkan regional, meningkatkan pengawasan infeksi bakteri yang telah resisten terhadap antibiotik, memperkuat langkah-langkah pencegahan dan pengendalian infeksi, juga mengatur dan mempromosikan penggunaan obat antibiotik yang tepat dan berkualitas.
37	Selain itu, pembuat kebijakan juga harus menyusun informasi tentang dampak resistensi antibiotik dan memberikan apresiasi atas pengembangan obat, vaksin, serta alat diagnosis baru.
38	Momentum Pekan Kewaspadaan Antibiotik Dunia (World Antibiotic Awareness Week) 2018 mengingatkan kita untuk berpikir ulang dan mencari saran (Think twice.
39	Seek advice), sebelum menggunakan antibiotik.
40	Kita juga disadarkan bahwa penyalahgunaan antibiotik akan membuat kita semua berisiko.
41	Waktu kita dengan antibiotik hampir habis, sehingga menjadi tugas kita bersama untuk menjaga antibiotik.
42	Sudahkah kita sadar?

Tabel 4.3 Tabel Data responden dan total jumlah kalimat.

No kalimat	Responden dan pilihan kalimat										Total
	1	2	3	4	5	6	7	8	9	10	
1	1	1	1	1	1	0	1	1	1	1	9
2	1	1	1	1	1	1	1	0	1	1	9
3	1	0	1	1	1	1	1	1	0	1	8
4	1	1	1	1	1	1	1	1	0	1	9
5	0	0	0	0	1	1	1	1	1	0	5
6	1	1	1	1	1	1	1	1	1	1	10
7	0	1	1	1	1	0	1	1	0	1	7
8	1	0	1	0	1	0	1	1	1	1	7
9	1	1	0	1	1	1	1	1	1	0	8
10	1	0	0	1	0	0	1	1	0	1	5
11	1	0	0	1	0	1	1	0	0	1	5
12	0	1	1	1	1	1	1	1	1	0	8
13	0	1	1	1	0	1	0	1	0	0	5
14	1	1	1	1	1	1	1	0	0	1	8
15	1	1	1	0	1	1	1	1	0	0	7
16	0	1	0	1	1	0	0	1	1	0	5
17	0	0	0	1	0	1	0	1	0	1	4

18	1	0	0	0	0	1	1	1	1	0	5
19	0	0	0	0	1	1	1	1	0	1	5
20	0	0	1	0	0	0	0	0	1	0	2
21	1	1	1	0	1	0	1	0	1	1	7
22	1	0	0	0	1	0	0	1	0	0	3
23	1	1	1	0	0	0	1	0	0	0	4
24	1	1	1	0	0	0	0	0	0	0	3
25	0	0	0	0	0	0	0	1	0	0	1
26	1	1	0	0	0	0	0	0	0	0	2
27	1	1	0	0	0	0	0	0	0	0	2
28	0	1	1	0	0	0	0	0	0	0	2
29	0	0	1	0	0	0	0	0	0	0	1
30	1	1	0	0	0	0	0	0	0	0	2
31	0	1	1	0	0	0	0	0	0	0	2
32	1	1	1	0	0	0	0	0	0	0	3
33	1	1	1	0	0	0	0	0	0	0	3
34	0	0	1	0	0	0	0	0	0	0	1
35	1	1	1	0	0	0	0	0	0	0	3
36	1	1	1	0	0	0	0	0	0	0	3
37	1	0	0	0	0	0	0	0	0	0	1
38	0	1	1	0	0	0	0	0	0	0	2
39	0	0	0	0	0	0	0	0	0	0	0
40	1	0	0	0	0	0	0	0	0	0	1
41	1	0	0	0	0	0	0	0	0	0	1
42	0	0	0	0	0	0	0	0	0	0	0

Dari data responden di Tabel 4.3, maka dibentuk rangkuman referensi dengan cara mengurutkan kalimat berdasarkan nilai total pilihan secara *descending*, lalu diambil 12,5% kalimat dari jumlah kalimat total di artikel asli. Pembentukan rangkuman referensi ditunjukkan pada Tabel 4.4.

Tabel 4.4 Pembentukan rangkuman referensi.

No	No kalimat lama	Kalimat	Jumlah pilihan
1	6	Antibiotik adalah obat yang digunakan untuk mencegah dan mengobati infeksi bakteri.	10
2	1	Bakteri semakin kebal antibiotik.	9
3	2	Penyakit seperti pneumonia, TBC, sepsis, dan gonorea kian sulit disembuhkan karena antibiotik semakin tak efektif.	9

4	4	Momentum ini bertujuan untuk meningkatkan kesadaran global terhadap resistensi antibiotik dan mendorong praktik terbaik di kalangan masyarakat umum, tenaga kesehatan, dan pembuat kebijakan untuk menghindari resistensi antibiotik yang telah mewabah.	9
5	3	tirto.id - Pada 12-18 November 2018 diselenggarakan Pekan Kewaspadaan Antibiotik Dunia (World Antibiotic Awareness Week).	8

Dari daftar kalimat di Tabel 4.4, maka kalimat dapat dijadikan berbentuk paragraf. Bentuk paragraf kalimat ditunjukkan pada Gambar 4.2.

Antibiotik adalah obat yang digunakan untuk mencegah dan mengobati infeksi bakteri. Bakteri semakin kebal antibiotik. Penyakit seperti pneumonia, TBC, sepsis, dan gonoreasia sulit disembuhkan karena antibiotik semakin tak efektif. Momentum ini bertujuan untuk meningkatkan kesadaran global terhadap resistensi antibiotik dan mendorong praktik terbaik di kalangan masyarakat umum, tenaga kesehatan, dan pembuat kebijakan untuk menghindari resistensi antibiotik yang telah mewabah. tirto.id - Pada 12-18 November 2018 diselenggarakan Pekan Kewaspadaan Antibiotik Dunia (World Antibiotic Awareness Week)

Gambar 4.2 Hasil Rangkuman referensi

Rangkuman sistem ditunjukkan pada Gambar 4.3.

Mereka juga wajib memberikan vaksinasi terbaru kepada pasien (ketika terjadi dugaan infeksi bakteri), melakukan biakan bakteri dan pemeriksaan penunjang medik lainnya untuk konfirmasi, serta hanya meresepkan dan mengeluarkan antibiotik jika benar-benar dibutuhkan, pada dosis dan durasi pengobatan yang tepat. Momentum ini bertujuan untuk meningkatkan kesadaran global terhadap resistensi antibiotik dan mendorong praktik terbaik di kalangan masyarakat umum, tenaga kesehatan, dan pembuat kebijakan untuk menghindari resistensi antibiotik yang telah mewabah. Antibiotik adalah obat yang digunakan untuk mencegah dan mengobati infeksi bakteri. Terakhir, para pejabat dan pembuat kebijakan kesehatan mesti bertindak cepat dengan menyusun rencana terpadu secara nasional, bahkan regional, meningkatkan pengawasan infeksi bakteri yang telah resisten terhadap antibiotik, memperkuat langkah-langkah pencegahan dan pengendalian infeksi, juga mengatur dan mempromosikan penggunaan obat antibiotik yang tepat dan berkualitas. Pada lini lainnya, promosi dan penerapan praktik yang baik perlu dilakukan di semua tahap produksi, mulai dari pengolahan makanan dari sumber

hewan dan tumbuhan yang aman, pengadopsian sistem berkelanjutan dengan meningkatkan kebersihan, penanganan hewan agar bebas stres, hingga pemenuhan standar internasional untuk penggunaan antibiotik yang bertanggung jawab sesuai ketentuan OIE, FAO, dan WHO

Gambar 4.3 Hasil Rangkuman sistem

Artikel asli (data input) , rangkuman sistem dan rangkuman referensi tersebut memiliki statistik seperti pada Tabel 4.5.

Tabel 4.5 Statistik artikel

Statistik	Jumlah kata	Jumlah kalimat
Artikel asli (data input)	824	42
Rangkuman sistem	179	5
Rangkuman referensi	76	5

Hasil rangkuman sistem dan rangkuman referensi kemudian dihitung nilai *recallnya* menggunakan ROUGE-1. ROUGE-1 adalah nilai jumlah kata-kata rangkuman sistem yang cocok dengan rangkuman referensi, yang dibandingkan dengan nilai jumlah kata-kata unik pada rangkuman referensi. Rangkuman referensi pada kesempatan ini adalah *ground truth* untuk mendapatkan nilai *recall*. Contoh daftar kata-kata unik pada rangkuman sistem ditunjukkan pada Tabel 4.6.

Tabel 4.6 Daftar kata unik rangkuman sistem

dari	dengan	berkualitas	regional	jika
pengobatan	global	masyarakat	mempromosikan	infeksi
cepat	antibiotik	sumber	sesuai	mesti
menghindari	langkah	semua	berkelanjutan	ini
para	terbaru	sistem	adalah	mengatur
promosi	lini	bahkan	umum	standar
makanan	stres	tumbuhan	dibutuhkan	terbaik
meresepkan	penanganan	baik	bertujuan	pengawasan
kalangan	di	jawab	melakukan	tenaga
memberikan	fao	pasien	kesadaran	dan
pada	terakhir	tahap	hewan	pemenuhan
terhadap	pemeriksaan	ketentuan	resisten	hingga
kebijakan	nasional	secara	pejabat	digunakan

terpadu	konfirmasi	terjadi	kesehatan	aman
wajib	meningkatkan	momentum	pengadopsian	penunjang
penggunaan	mereka	mengobati	serta	hanya
yang	mulai	pengolahan	pencegahan	produksi
biakan	mendorong	mengeluarkan	vaksinasi	memperkuat
durasi	oie	obat	benar	bertindak
dosis	pengendalian	juga	bakteri	ketika
menyusun	pembuat	dugaan	praktik	medik
agar	kepada	penerapan	rencana	untuk
resistensi	kebersihan	dilakukan	mewabah	perlu
bertanggung	mencegah	bebas	who	
telah	tepat	internasional	lainnya	

Contoh daftar kata-kata pada rangkuman referensi ditunjukkan pada Tabel 4.7.

Tabel 4.7 Daftar kata unik rangkuman referensi

antibiotic	pembuat	mencegah	tenaga	pekan
meningkatkan	pada	2018	dan	umum
kesehatan	terhadap	pneumonia	telah	disembuhkan
mendorong	kebijakan	november	global	bertujuan
gonorea	bakteri	infeksi	masyarakat	di
menghindari	praktik	world	penyakit	dunia
efektif	obat	yang	semakin	kesadaran
momentum	sulit	diselenggarakan	adalah	12
karena	mewabah	ini	antibiotik	untuk
kewaspadaan	awareness	sepsis	digunakan	kian
kalangan	tirto	tak	id	
mengobati	kebal	resistensi	18	
tbc	week	terbaik	seperti	

Terdapat kesamaan kata-kata yang muncul pada daftar kata-kata rangkuman sistem pada Tabel 4.6 dan referensi pada Tabel 4.7. Contoh daftar kemunculan kata yang sama ditunjukkan pada Tabel 4.8.

Tabel 4.8 Daftar kemunculan kata yang sama

meningkatkan	pada	mewabah	tenaga	digunakan
kesehatan	pembuat	mencegah	dan	umum
mendorong	terhadap	infeksi	telah	bertujuan
menghindari	kebijakan	yang	global	di
momentum	bakteri	ini	masyarakat	kesadaran
kalangan	praktik	resistensi	adalah	untuk

mengobati	obat	terbaik	antibiotik	
-----------	------	---------	------------	--

Jumlah kata-kata unik pada Tabel 4.6, Tabel 4.7 dan Tabel 4.8 ditunjukkan pada Tabel 4.9.

Tabel 4.9 Jumlah kata unik

Kata unik pada	Jumlah
Rangkuman sistem ( $S$ )	123
Rangkuman referensi ( $R$ )	62
Kemunculan yang sama pada rangkuman sistem dan referensi ( <i>Overlapping occurrences - O</i> )	34

Nilai ROUGE diperoleh dari perbandingan nilai Kemunculan yang sama pada rangkuman sistem dan referensi (*Overlapping word - O*) dengan rangkuman referensi ( $R$ ). Nilai yang digunakan perhitungan adalah jumlah kemunculan saja, sehingga urutan kemunculan kata tidak berpengaruh. Perhitungan nilai ROUGE berdasarkan nilai pada Tabel 4.9 ditunjukkan sebagai berikut.

$$ROUGE - 1 = \frac{O}{R} = \frac{34}{62} = 0.5483870967741935$$

Uji coba sistem dilakukan dengan membangkitkan nilai ROUGE dari rangkuman sistem & rangkuman referensi setiap artikel. Untuk perbandingan, rekap & hasil perhitungan seluruh artikel dengan rangkuman sebanyak 50% (Eris .et.al, 2017) ditampilkan pada Tabel 4.10. Rekap & hasil perhitungan seluruh artikel dengan rangkuman sebanyak 12,5% dapat dilihat pada Tabel 4.11.

Tabel 4.10 Hasil Nilai *recall* ROUGE Tiap Artikel (rangkuman 50%)

No	Judul Artikel	Asli		Rangkum sistem		Rangkum referensi		Jumlah Kata		Nilai ROUGE
		Kata	Kalimat	Kata	Kalimat	Kata	Kalimat	<i>Overlapping</i>	Referensi unik	
1	Senjakala Industri Manga Jepang Akibat Pembajakan	1067	62	611	30	573	28	218	276	0,789855072
2	Esemka: Mainan 'Mobil Nasional' Jokowi yang Masih Ilusi	1251	89	651	42	681	43	264	344	0,76744186
3	Kenapa Banyak Orang Berhenti Mencari Musik Baru?	787	54	430	26	443	26	198	255	0,776470588
4	Zakir Naik Tidak Mendukung ISIS, Tapi...	1446	66	1032	31	1323	38	462	525	0,88
5	Le Minerale Versus Aqua: Bertarung di Pasar Berlanjut di Pengadilan	1077	66	640	32	620	33	254	340	0,747058824
6	'Baku Hantam' Jokowi-Prabowo di Kandang Lawan	669	49	381	23	357	22	160	216	0,740740741
7	Bangkitnya Penyakit Ganas di Dunia Pasca-Antibiotik	824	42	534	21	361	21	167	211	0,791469194
8	Saat Banjir Tiba, Kendaraan Motor Lebih Berharga daripada Nyawa	821	52	463	25	501	26	189	247	0,765182186
9	Cerita Lengkap Pembunuhan Satu Keluarga di Bekasi	645	49	393	24	360	21	207	241	0,858921162
10	Evakuasi Korban Lion Air Makan Biaya Besar, Siapa yang Menanggung?	507	35	279	15	246	15	93	155	0,6
11	Jika Rizieq Terbukti Makar di Saudi, Indonesia Tak Bisa Apa-apa	643	42	395	20	311	20	156	200	0,78
12	Jutaan Jiwa Tinggal di Kawasan Rawan Longsor, Kok Tidak Direlokasi?	728	53	420	27	453	28	171	250	0,684



27	Melajang: Salah Satu Cara Menuju Hidup Sehat, Santai, & Bahagia	920	55	526	24	529	28	224	294	0,761904762
28	Anti-Natalisme: Mengutuk Kehidupan, Melarang Kelahiran	1249	79	731	36	717	39	290	395	0,734177215
29	Mengapa Dalih Mahal Demi Menghindari Energi Terbarukan Bermasalah?	741	53	455	23	421	23	192	237	0,810126582
30	Kasus Mourinho: Kenapa Penggelapan Pajak Sering Terjadi di Spanyol?	720	44	419	19	407	21	161	209	0,770334928
31	Memanggil Presiden Jokowi dengan Cak Jancuk, Bisakah Dijerat Hukum?	776	52	434	23	361	23	155	225	0,688888889
32	Kasus Agni Berakhir "Damai": Cermin Buram Kasus Pelecehan di Kampus	1030	71	657	30	546	29	236	281	0,839857651
33	Cibiran & Stigma Jadi Alasan Masyarakat Permisif Sikapi Bunuh Diri	836	63	565	28	367	23	165	226	0,730088496
34	Fellaini adalah Simbol Betapa Tak Menarik Manchester United	710	54	429	24	348	23	166	216	0,768518519
35	Harlah NU 2019	611	40	278	16	350	18	92	224	0,410714286
36	Efek Domino Bagasi Berbayar: Ganggu Pariwisata, Hotel hingga UMKM	807	55	479	24	393	23	157	233	0,673819742
37	Menyelisik Tabloid Indonesia Barokah, Apa Isi dan Siapa di Balikinya	558	42	259	17	377	18	133	218	0,610091743
38	Jakarta, Kota dengan 9 Kasus Pencurian Setiap Hari	836	56	515	28	466	30	180	233	0,772532189
39	Fadli Zon Bersilat Lidah: Bela Ahmad Dhani Tapi Enggan Ubah UU ITE	667	56	361	24	308	22	144	183	0,786885246
40	Wacana Motor Masuk Tol Berpotensi Tingkatkan Angka Kecelakaan	733	43	497	19	333	18	169	203	0,832512315
41	Rekam Jejak Kegagalan PSSI Perjuangkan Klub Indonesia di Asia	900	51	539	23	457	23	205	261	0,785440613

42	Jokowi Tambah Utang Rp1.809 Triliun Selama Menjabat, Berisiko?	933	54	489	24	431	22	195	267	0,730337079
43	Perokok Remaja Semakin Banyak, Regulasi yang Tumpul Penyebabnya	650	57	311	24	363	24	134	225	0,595555556
44	Menengok Langkah Anies Atasi Penurunan Tanah Agar Tak Jadi Bencana	517	38	327	16	235	15	106	159	0,666666667
45	Gaji Dokter Memang Masih Jauh di Bawah Rekomendasi IDI	518	34	312	14	291	18	117	183	0,639344262
46	Nasib Ribuan Karyawan Bank yang Disingkirkan Mesin	600	41	368	19	326	19	141	191	0,738219895
47	SNMPTN: Betapa Ketatnya Persaingan Masuk PTN	1146	59	670	26	704	28	220	306	0,718954248
48	Mitos dan Fakta tentang Kertas	844	50	541	25	459	22	194	246	0,788617886
49	City Brain: Solusi Untuk Masalah di Jalanan?	1359	89	932	38	757	39	361	447	0,807606264
50	Tata Ruang Indonesia Mengabaikan Potensi Bencana	630	47	366	21	327	20	152	220	0,690909091
	Rata-rata	812,68	53,52	478,62	24,38	447,46	24,66	186,22	252,22	0,72914214

Tabel 4.11 Hasil Nilai *recall* ROUGE Tiap Artikel (rangkuman 12,5%)

No	Judul Artikel	Asli		Rangkum sistem:		Rangkum Manual:		Jumlah Kata		Nilai ROUGE
		Kata	Kalimat	Kata	Kalimat	Kata	Kalimat	<i>Overlapping</i>	Referensi unik	
1	Senjakala Industri Manga Jepang Akibat Pembajakan	1067	61	141	7	120	7	31	88	0,352272727
2	Esemka: Mainan 'Mobil Nasional' Jokowi yang Masih Illusi	1251	87	202	10	190	10	80	115	0,695652174
3	Kenapa Banyak Orang Berhenti Mencari Musik Baru?	787	53	79	6	97	6	28	75	0,373333333
4	Zakir Naik Tidak Mendukung ISIS, Tapi...	1446	63	272	7	348	9	85	227	0,374449339
5	Le Minerale Versus Aqua: Bertarung di Pasar Berlanjut di Pengadilan	1077	64	171	8	186	8	52	132	0,393939394
6	Baku Hantam' Jokowi-Prabowo di Kandang Lawan	669	48	69	5	75	5	20	62	0,322580645
7	Bangkitnya Penyakit Ganas di Dunia Pasca-Antibiotik	824	42	179	5	76	5	34	62	0,548387097
8	Saat Banjir Tiba, Kendaraan Motor Lebih Berharga daripada Nyawa	821	51	114	6	119	6	39	76	0,513157895
9	Cerita Lengkap Pembunuhan Satu Keluarga di Bekasi	645	49	141	6	106	5	63	84	0,75
10	Evakuasi Korban Lion Air Makan Biaya Besar, Siapa yang Menanggung?	507	31	77	3	53	3	8	47	0,170212766
11	Jika Rizieq Terbukti Makar di Saudi, Indonesia Tak Bisa Apa-apa	643	41	101	5	87	5	48	66	0,727272727
12	Jutaan Jiwa Tinggal di Kawasan Rawan Longsor, Kok Tidak Direlokasi?	728	55	115	6	131	7	41	105	0,39047619

13	Kok Bisa Kiai NU Tak Masuk Survei Ulama Berpengaruh LSI?	709	39	77	4	131	5	43	102	0,421568627
14	Kok Bisa Pemprov DKI Jadi Provinsi Paling Informatif?	575	35	84	4	53	4	30	51	0,588235294
15	Kontroversi Acara Bertajuk Syiar Kekhalifahan yang Ditolak Warga	786	59	89	7	125	8	19	96	0,197916667
16	KPU Rentan Serangan Siber, Tapi Antisipasi Masih Minim	465	33	78	4	367	19	69	212	0,325471698
17	Mampukah Dirjen Hubungan Udara Baru Bikin Penerbangan Makin Aman?	732	39	101	4	92	4	55	74	0,743243243
18	Manuver Yusril: Membela HTI, Diabaikan Prabowo, dan Menjaga Jokowi	916	61	182	7	107	7	48	80	0,6
19	Sampai Kapan Basarnas Mencari Para Korban Pesawat Lion Air JT 610?	525	35	124	4	53	4	16	46	0,347826087
20	Selamat Datang di Pilpres yang Sudutkan Kaum Rentan & Disabilitas	743	56	68	6	115	7	18	92	0,195652174
21	Sistem di Balik Pengumpulan Nasi Kotak Reuni 212	628	44	75	5	75	5	42	57	0,736842105
22	UGM Tak Serius Tangani Kasus Kekerasan Seksual di Kampus?	519	29	58	3	61	3	14	53	0,264150943
23	Selamat Datang di Musim Paceklik Penerbangan	868	55	149	6	111	6	30	87	0,344827586
24	Mengapa Medan Magnet di Kutub Utara Bergeser Begitu Cepat?	1164	75	214	9	172	8	62	128	0,484375
25	Kondisi Psikologis Adi Saputra, dan Pengemudi yang Wajib Diwaspadai	1316	62	278	7	168	8	30	121	0,247933884
26	Periksa Fakta Benarkah Server KPU Jebol dan Jokowi Sudah Punya Angka 57	932	70	139	8	133	7	52	93	0,559139785

	Persen? Header Periksa Fakta.									
27	Melajang: Salah Satu Cara Menuju Hidup Sehat, Santai, & Bahagia	920	49	156	6	146	7	56	107	0,523364486
28	Anti-Natalisme: Mengutuk Kehidupan, Melarang Kelahiran	1249	72	217	9	143	9	55	111	0,495495495
29	Mengapa Dalih Mahal Demi Menghindari Energi Terbarukan Bermasalah?	741	46	83	5	97	5	29	81	0,358024691
30	Kasus Mourinho: Kenapa Penggelapan Pajak Sering Terjadi di Spanyol?	720	40	123	4	111	5	21	78	0,269230769
31	Memanggil Presiden Jokowi dengan Cak Jancuk, Bisakah Dijerat Hukum?	776	48	136	5	104	5	33	77	0,428571429
32	Kasus Agni Berakhir "Damai": Cermin Buram Kasus Pelecehan di Kampus	1030	61	175	7	157	7	67	117	0,572649573
33	Cibiran & Stigma Jadi Alasan Masyarakat Permisif Sikapi Bunuh Diri	836	56	186	7	90	5	20	71	0,281690141
34	Fellaini adalah Simbol Betapa Tak Menarik Manchester United	710	50	176	6	89	5	21	75	0,28
35	Harlah NU 2019	611	35	104	4	108	4	21	93	0,225806452
36	Efek Domino Bagasi Berbayar: Ganggu Pariwisata, Hotel hingga UMKM	807	48	142	6	89	5	43	79	0,544303797
37	Menyelisik Tabloid Indonesia Barokah, Apa Isi dan Siapa di Balikny	558	35	48	4	140	4	13	78	0,166666667
38	Jakarta, Kota dengan 9 Kasus Pencurian Setiap Hari	836	56	162	7	118	7	31	82	0,37804878
39	Fadli Zon Bersilat Lidah: Bela Ahmad Dhani Tapi Enggan Ubah UU ITE	667	49	99	6	95	5	12	69	0,173913043
40	Wacana Motor Masuk Tol Berpotensi Tingkatkan Angka Kecelakaan	733	38	128	4	86	4	23	76	0,302631579

41	Rekam Jejak Kegagalan PSSI Perjuangkan Klub Indonesia di Asia	900	47	154	5	100	5	25	86	0,290697674
42	Jokowi Tambah Utang Rp1.809 Triliun Selama Menjabat, Berisiko?	933	49	143	6	96	5	29	90	0,322222222
43	Perokok Remaja Semakin Banyak, Regulasi yang Tumpul Penyebabnya	650	51	92	6	110	6	15	89	0,168539326
44	Menengok Langkah Anies Atasi Penurunan Tanah Agar Tak Jadi Bencana	517	32	77	4	56	3	24	48	0,5
45	Gaji Dokter Memang Masih Jauh di Bawah Rekomendasi IDI	518	28	93	3	77	4	12	61	0,196721311
46	Nasib Ribuan Karyawan Bank yang Disingkirkan Mesin	600	39	98	4	99	4	54	79	0,683544304
47	SNMPTN: Betapa Ketatnya Persaingan Masuk PTN	1146	53	166	6	212	7	39	142	0,274647887
48	Mitos dan Fakta tentang Kertas	844	50	159	6	151	5	78	102	0,764705882
49	City Brain: Solusi Untuk Masalah di Jalanan?	1359	77	238	9	198	9	81	154	0,525974026
50	Tata Ruang Indonesia Mengabaikan Potesi Bencana	630	42	114	5	107	5	39	90	0,433333333
	Rata-rata	812,68	49,76	132,92	5,72	122,6	6,02	37,96	91,32	0,416594005

### 4.3. Pembahasan

Di penelitian ini digunakan ROUGE untuk mendapatkan nilai *recall*, yaitu nilai perbandingan rangkuman dari hasil otomatis dengan rangkuman referensi. Nilai *recall* di sini menunjukkan nilai relevansi hasil rangkuman sistem terhadap teks asli, sehingga semakin tinggi nilai ROUGE, semakin relevan hasil rangkuman yang diproduksi. Nilai rata-rata *recall* ROUGE dari sistem *automatic text summarization* di atas yaitu 0,416594005. Nilai persen dari rata-rata tersebut adalah 41,6594005%. Nilai tertinggi *recall* ROUGE tertinggi tercatat pada artikel 48 dengan nilai 0,764705882. Nilai terendah *recall* ROUGE tercatat pada artikel 37 dengan nilai 0,166666667.

Dari percobaan tersebut diketahui bahwa perbandingan jumlah kata dan kalimat hasil sistem dengan hasil referensi memiliki perbedaan yang tidak terlalu jauh. Hal ini diketahui dengan nilai rata-rata kata & kalimat yang dihasilkan dari kedua cara tersebut. Artikel asli memiliki rata-rata 812,68 kata dan rata-rata 49,76 kalimat. Rangkuman sistem memiliki rata-rata 132,92 kata dan 5,72 kalimat, sedangkan rangkuman referensi memiliki nilai rata-rata 122,6 kata dan 6,02 kalimat. Perbandingan nilai rata-rata tersebut ditunjukkan pada Tabel 4.12.

Tabel 4.12 Perbandingan rata-rata jumlah kata dan kalimat.

	Isi Asli (a)	Rangkum sistem (rs)	Rangkum referensi (rf)	$rs : a (x)$	$rf : a (y)$	Selisih (x-y)
Kata	812,68	132,92	122,6	0,163557612	0,150858887	0,012698725
Kalimat	49,76	5,72	6,02	0,114951768	0,120980707	0,006028939

Dari tabel perbandingan di atas, diketahui bahwa hasil rangkuman sistem tidak berbeda jauh dengan rangkuman referensi dari segi kuantitas. Selisih

perbandingan rata-rata kata rangkuman sistem dengan rangkuman referensi yaitu 0,012698725 dan selisih perbandingan rata-rata kalimatnya yaitu 0,006028939.

Nilai ROUGE berbanding lurus dengan jumlah rangkuman yang dihasilkan. Hal ini diketahui dari perbandingan hasil rangkuman 50% dan 12,5 % dan evaluasi ROUGE tiap rangkuman. Perbandingan rata-rata antara rangkuman 50% dan 12,5 % ditunjukkan pada tabel 4.13.

Tabel 4.13 Perbandingan rata-rata rangkuman 50% dan 12,5%.

Rangku man	Asli		Rangku m sistem		Rangku m referensi		Jumlah Kata		Nilai ROUGE
	Kata	Kalimat	Kata	Kalimat	Kata	Kalimat	<i>Overlapp ing</i>	Referensi unik	
50%	812,68	53,52	478,62	24,38	447,46	24,66	186,22	252,22	0,729142 14
12,5%	812,68	49,76	132,92	5,72	122,6	6,02	37,96	91,32	0,416594 005

Menurut Saggion, metode paling efektif dan serbaguna yang digunakan untuk automatic summarization bergantung pada metode ekstraktif (extractive methods), yang menitikberatkan pada pemilihan kalimat yang relevan dari kumpulan dokumen asli supaya teks yang dihasilkan memberikan potongan informasi penting. Faktanya metode-metode tersebut masih jauh dari ideal, misal pada multi-document summarization, pemilihan kalimat dari dokumen yang berbeda dapat mengindikasikan redundansi, yang harusnya dieliminasi. Seringkali sebagian kalimat relevan, tetapi yang dipilih justru yang kurang operasional. Pemilihan / ekstraksi kalimat dari dokumen-dokumen yang berbeda dapat menghasilkan rangkuman yang sulit dibaca.

Dalam kesempatan ini, peneliti menggunakan ROUGE untuk mengevaluasi hasil rangkuman sistem. Menurut saggion, kalimat-kalimat dari metode perangkuman ekstraktif tersusun dari teks aslinya, yang artinya sangat

mungkin untuk mengukur kualitas hasil rangkuman dengan membandingkannya dengan rangkuman referensi. Jika tidak memungkinkan (seperti pada perangkuman abstraktif), maka hanya metode manual yang bisa digunakan.

Dari hasil sistem yang telah dirancang dan diimplementasikan, pengguna dapat memperoleh rangkuman dari sistem dengan hanya mengisi data dan beberapa kali klik. Jika dibandingkan dengan cara manual, pembuatan rangkuman menggunakan sistem yang otomatis ini memiliki perbedaan yang sangat signifikan. Dari segi waktu, perangkuman manual dengan cara memilih kandidat kalimat terbaik memakan waktu lebih banyak dari perangkuman menggunakan sistem. Melakukan perangkuman otomatis juga menggunakan sumberdaya tenaga dan pikiran manusia yang relatif lebih sedikit.

Dalam surat Ar-Ra'd. ayat 3 disampaikan :

وَهُوَ الَّذِي مَدَّ الْأَرْضَ وَجَعَلَ فِيهَا رَوَاسِيَ وَأَنْهَارًا وَمِنْ كُلِّ الثَّمَرَاتِ جَعَلَ فِيهَا زَوْجَيْنِ  
 اثْنَيْنِ يُغْشَىٰ اللَّيْلَ النَّهَارَ إِنَّ فِي ذَلِكَ لَآيَاتٍ لِّقَوْمٍ يَتَفَكَّرُونَ ﴿٣﴾

“Dan Dia yang menghamparkan bumi dan menjadikan gunung-gunung dan sungai-sungai di atasnya. Dan padanya Dia menjadikan semua buah-buahan berpasang-pasangan; Dia menutupkan malam kepada siang. Sungguh, pada yang demikian itu terdapat tanda-tanda (kebesaran Allah) bagi orang-orang yang berpikir.”.

Menurut Tafsir Jalalain ayat di atas dijelaskan bahwa di alam semesta terdapat bukti-bukti kemahakuasaan dan kemahaesaan Allah Swt. bagi orang-orang yang mau berfikir dan merenungkan ciptaan-Nya. Untuk mencapai keimanan yang sempurna maka manusia harus menjadi orang mau berfikir dan

merenungkan ciptaan-Nya dan mengambil hikmah pada hasil renungannya. Pada pembuatan rangkuman dokumen teks juga diperlukan pikiran pembaca dan pemahaman terhadap maksud penulis yang sebenarnya. Dengan memahami dua fakta tersebut, maka sesungguhnya terdapat cara untuk mendapatkan inti dari suatu hal. Dalam konteks mendapatkan inti informasi dari sebuah teks, maka inti informasi pada umumnya diperoleh dengan membuat rangkuman (*summary*).

Dalam surat Al - Insyirah disampaikan:

فَإِنَّ مَعَ الْعُسْرِ يُسْرًا ۝ إِنَّ مَعَ الْعُسْرِ يُسْرًا ۝ ٦

“Maka sesungguhnya beserta kesulitan ada kemudahan. sesungguhnya beserta kesulitan itu ada kemudahan” (QS. Al- Insyirah : 5-6)

Tafsir Ibnu Abbas radhiallahu ‘anhu lebih jauh diterangkan bahwa Allah Ta’ala menyebut “satu kesulitan di antara dua kemudahan”. Menurut penjelasan ulama, alasannya adalah karena kesulitan (*al-usr*) yang tersurat di dalam dua ayat tersebut memiliki bentuk *definitive*/tunggal atau *makrifah*. Jadi, walaupun disebut dua kali, cuma satu kesulitannya. Sementara itu, kemudahan (*yusr*) diekspresikan dengan *indefinite article* atau *naqirah* yang mengindikasikan bentuk jamak atau lebih dari satu.

Menurut Lutfi (2016), makna yang terkandung adalah, bahwa ada satu kesulitan dan disusul dengan dua kemudahan. Inti nya kalau ada satu kesulitan kita akan beruntung. Bahkan dalam salah satu hadist disebutkan bahwa Rasulullah salallahu ‘alayhi wa sallam, sebagaimana dikutip Ibnu Katsir dari riwayat Anas bin Malik, pernah mengilustrasikannya saat duduk bersama para sahabat di depan sebuah batu. “Saat kesulitan datang dan masuk batu ini, maka kemudahan pasti akan datang dan masuk pula menghilangkan kesulitan tersebut.”

Semua kejadian yang ada di dunia adalah karena Allah swt. termasuk masalah, kesulitan dan kemudahan. Semua hal yang tercipta selalu seimbang karena Allah maha pencipta, menciptakan semuanya berpasang-pasangan. Jika pasangan dari masalah adalah solusi, maka kesulitan adalah kemudahan. Sebuah masalah terdapat beberapa solusi. Dari beberapa solusi tersebut bisa dipilih yang terbaik.

Dari Abu Hurairah ra, Nabi SAW, bersabda:

مَنْ نَفَسَ عَنْ مُؤْمِنٍ كَرْبَةً مِنْ كَرْبِ الدُّنْيَا نَفَسَ اللَّهُ عَنْهُ كَرْبَةً مِنَ  
 الْقِيَامَةِ وَمَنْ يَسِّرَ عَلَى مَعْسِرٍ يَسِّرَ اللَّهُ عَلَيْهِ فِي الدُّنْيَا وَالْآخِرَةِ وَمَنْ سَتَرَ  
 مُسْلِمًا سَتَرَهُ اللَّهُ فِي الدُّنْيَا وَالْآخِرَةِ وَاللَّهُ فِي عَوْنِ الْعَبْدِ مَا كَانَ الْعَبْدُ فِي عَوْنِ خِيهِ

“Barang siapa yang melepaskan satu kesusahan seorang mukmin, pasti Allah akan melepaskan darinya satu kesusahan pada hari kiamat. Barang siapa yang menjadikan mudah urusan orang lain, pasti Allah akan memudahkannya di dunia dan di akhirat. Barang siapa yang menutupi aib seorang muslim, pasti Allah akan menutupi aibnya di dunia dan di akhirat. Allah senantiasa menolong hamba Nya selama hamba Nya itu suka menolong saudaranya”. (HR. Muslim, no. 2699).

Hasil implementasi berupa aplikasi diharapkan dapat bermanfaat. Selanjutnya metode dapat dibandingkan dengan metode-metode perangkuman teks yang lain atau diterapkan pada aplikasi atau software terkait pengolahan teks. Sistem perangkum juga dapat dimanfaatkan bagi para pembaca buku, akademisi, jurnalis dan pekerjaan lain yang membutuhkan perangkuman teks.

## BAB V

### KESIMPULAN DAN SARAN

#### 5.1 Kesimpulan

Dari hasil dan pembahasan, didapatkan kesimpulan bahwa nilai relevansi hasil rangkuman penggunaan algoritma *Textrank*, jika dinilai dengan ROUGE adalah 41,659 %. Nilai tertinggi *recall* ROUGE tertinggi tercatat pada artikel 48 dengan nilai 0,764. Nilai terendah *recall* ROUGE tercatat pada artikel 37 dengan nilai 0,167. Dari hasil sistem yang telah dirancang dan diimplementasikan, pengguna dapat memperoleh rangkuman dari sistem dengan waktu yang lebih sedikit karena melibatkan proses perhitungan komputer jika dibandingkan dengan merangkum secara manual.

#### 5.2 Saran

Dari kesimpulan yang didapatkan, untuk meningkatkan hasil rangkuman sistem maka bisa dilakukan hal sebagai berikut :

- a. Melibatkan pemahaman sistem terhadap konteks teks atau kalimat. Setiap kalimat yang diproses sistem dengan pemahaman konteks kalimat akan meningkatkan akurasi, relevansi dan kualitas hasil rangkuman sehingga relatif lebih mudah dibaca.
- b. Melibatkan kamus sinonim dari sebuah bahasa atau berbagai bahasa. Kata-kata yang diproses di dalam perhitungan perbandingan kalimat tidak dicek sinonimnya sehingga kata-kata yang berarti sama masih dianggap berbeda.

- c. Algoritma *TextRank* tidak tergantung pada sebuah bahasa, namun kamus *stopwords* dan *stemming* masih belum mencakup bahasa lain secara lebih dalam. Penelitian berikutnya diharapkan menambah cakupan bahasa lain.



## DAFTAR PUSTAKA

- Barzilay, R., Elhadad, M. 1997. Using Lexical Chains for Text Summarization. Mathematics and Computer Science Dept. Ben Gurion University in the Negev Beer-Sheva, Israel.
- Brin, S. Page, L. 1998. The Anatomy of a Large-Scale Hypertextual Web Search Engine. Stanford : Stanford University.
- Chengcheng, L. 2010. Automatic Text Summarization Based On Rhetorical Structure Theory. ICCASM 2010. Inner Mongolia Normal University, China.
- El Bazzi, M.S., Mammass, D., Zaki, T. 2016. A graph based method for Arabic documen indexing. Ibn Zohr University, Agadir, Maroko.
- Eris, E. Christanti, V. M. Pragantha, Jeanny. 2017. Penerapan Algoritma Textrank Untuk Automatic Summarization Pada Dokumen Berbahasa Indonesia. Universitas Tarumanegara, Jakarta.
- Kaynar, O. Isik, Y.E., Gormez, Y. 2017. Graph Based Automatic Document Summarization with Different Similarity Methods. Cumhuriyet Universitesi, Sivas, Turki.1
- Kurniawan, A. Solihin, F., dan Hastarita, F. 2014. Perancangan dan Pembuatan Aplikasi Pencarian Informasi Beasiswa dengan Menggunakan Cosine Similarity. Jurnal SimanteC Volume 4, Nomor 2.
- Lin, C.Y. ROUGE : A Package for Automatic Evaluation of Summaries. 2004. California : Information Sciences Institute University of Southern California.
- Lutfi Firdaus Munawar, Lc. MA.. 2016. Tafsir Surat Al Insyirah di <http://manhajuna.com/tafsir-surat-al-insyirah/> (diakses 8 Oktober 2019)
- Mani, I., Maybury, M.T. 1999. Advances in Automatic Text Summarization. MIT Press.
- Mihalcea, R. Tarau, P. 2004. TextRank: Bringing Order into Texts. Texas : University of North Texas.
- Nanos, A.G., James, A.E., Iqbal, R., Hedley, Y.L. 2017. Content Summarisation of Conversation in the Context of Virtual Meetings: An Enhanced TextRank Approach. Distributed Systems Research Group, Coventry University, Coventry, United Kingdom
- Nurdiana, O., Jumadi., dan Nursantika, D. 2016. Perbandingan Metode Cosine Similarity dengan Metode Jaccard Similarity pada Aplikasi Pencarian Terjemahan Al-Qur'an dalam Bahasa Indonesia. Jurnal Online Informatika Volume 1, Nomor 1.
- Sabuna, P. M., Setyohadi, D.B.. 2017. Summarizing Indonesian Text Automatically By Using Sentence Scoring And Decision Tree. 2nd International Conferences on Information Technology, Information Systems and Electrical Engineering (ICITISEE). Universitas Atma Jaya Yogyakarta.
- Saggion, H. Poibeau, T. 2012. Automatic Text Summarization : Past, Present and Future .
- Zadbuke, A. Pimenta, S. Padwal, D. Wangikar, V. 2016. Automatic Summarization of News Articles using TextRank. IJARCSSE. University of Mumbai, Maharashtra, India.

Zhiqiang, L., Werimin, S., Zhenhua, Y. 2009. Measuring Semantic Similarity between Words Using Wikipedia. IEEE. 251-255.3

<https://www.codecogs.com/cnedit.php>

