

**PERINGKASAN TEKS BAHASA INDONESIA PADA CERPEN  
MENGUNAKAN METODE *LATENT SEMANTIC ANALYSIS*  
(LSA)**

*Diajukan Sebagai Syarat Untuk Menyelesaikan  
Pendidikan Program Strata-1 Pada Jurusan Teknik Informatika*



Oleh :

**NABILAH THAHIRAH KASIM**

09021281722040

**Jurusan Teknik Informatika**

**FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA**

**2022**

**LEMBAR PENGESAHAN SKRIPSI**

**PERINGKASAN TEKS BAHASA INDONESIA PADA CERPEN  
MENGUNAKAN METODE *LATENT SEMANTIC ANALYSIS*  
(LSA)**

Oleh :

**NABILAH THAHIRAH KASIM**

**NIM : 09021281722040**

**Palembang, 27 Juli 2022**

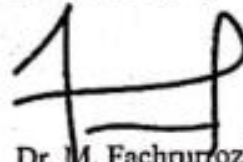
**Pembimbing I**



**Novi Yustiani, M.T.**

**NIP. 198211082012122001**

**Pembimbing II,**



**Dr. M. Fachrurrozi, M.T**

**NIP. 198005222008121002**

**Mengetahui,**

**Ketua Jurusan Teknik Informatika**



**Alvi Syahrini Utami, M.Kom.**

**NIP. 197812222006042003**

## TANDA LULUS UJIAN KOMPREHENSIF SKRIPSI

Pada hari Rabu tanggal 27 Juli 2022 telah dilaksanakan ujian komprehensif skripsi oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Nabilah Thahirah Kasim  
NIM : 09021281722040  
Judul : Peringkasan Teks Bahasa Indonesia pada Cerpen Menggunakan Metode Latent Semantic Analysis (LSA)

dan dinyatakan **LULUS**.

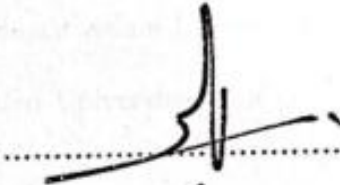
1. Ketua

Alvi Syahrini Utami, M.Kom.  
NIP. 197812222006042003



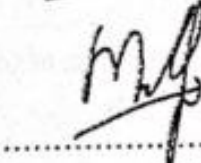
2. Penguji I

Dr. Abdiansah, S.Kom, M.Cs.  
NIP. 198410012009121005



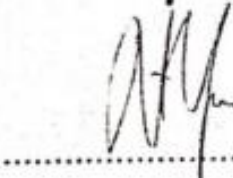
3. Penguji II

M. Qurhanul Rizqie, M.T., Ph.D.  
NIP. 1671060312870008



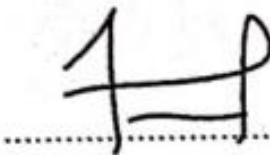
4. Pembimbing I

Novi Yusliani, M.T.  
NIP. 198211082012122001




5. Pembimbing II

Dr. M. Fachrurrozi, M.T.  
NIP. 198005222008121002



Mengetahui,

Ketua Jurusan Teknik Informatika

  
Alvi Syahrini Utami, M.Kom.  
NIP. 197812222006042003

## HALAMAN PERNYATAAN BEBAS PLAGIAT

Yang bertanda tangan dibawah ini :

Nama : Nabilah Thahirah Kasim  
NIM : 09021281722040  
Program Studi : Teknik Informatika Reguler  
Judul Skripsi : Peringkasan Teks Bahasa Indonesia pada Cerpen Menggunakan Metode *Latent Semantic Analysis* (LSA)

Hasil Pengecekan Software *iThenticate/Turnitin* : 9%

Menyatakan bahwa laporan proyek saya merupakan hasil karya sendiri dan bukan hasil plagiat. Apabila ditemukan unsur plagiat dalam laporan proyek ini, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya sesuai dalam ketentuan yang berlaku.

Demikian, pernyataan ini saya buat dengan sebenarnya dan tidak ada paksaan oleh siapapun.



Palembang, 20 Juli 2022



Nabilah Thahirah Kasim

NIM. 09021281722040

## **MOTTO DAN PERSEMBAHAN**

### **MOTTO:**

- After Allah, Family are Everything for me.
- If at first you don't succeed, try, try, try again.
- Don't be afraid to fall, be afraid not to try.

Kupersembahkan Karya Tulis ini kepada:

- Kedua Orang Tuaku Tercinta
- Saudara & Saudariku Tersayang
- Keluarga Besar
- Seluruh Dosen dan Guru
- Sahabat dan Teman Seperjuangan
- Fakultas Ilmu Komputer  
Universitas Sriwijaya
- Diri Sendiri

# **PERINGKASAN TEKS BAHASA INDONESIA PADA CERPEN MENGUNAKAN METODE *LATENT SEMANTIC ANALYSIS* (LSA)**

Oleh :

Nabilah Thahirah Kasim

09021281722040

## **ABSTRACT**

Cerpen or short stories are entertaining readings for readers of all ages. In the presentation, often the category of naming the title and presenting does not represent the content of the written story, so it will cause problems when finding the desired poem. Finally, to find out from a reader's knowledge it is necessary to understand the whole of the short story which of course takes a very long time. Another alternative to get information and understand a value-based quickly is to read the summary or synopsis of the short story. Automated text summarization is research in the field of assisting natural language to enable readers to summarize a book more efficiently to get to the heart of the story. In this study, text summarization was carried out using the Latent Semantic Analysis (LSA) algorithm. The test was carried out using test data of 30 short stories texts and the level of compression of the summary sentences was as much as 30% of the number of short stories text sentences. In this study, there are three stages, first text preprocessing, next calculating the weight of TF-IDF, and the last determining the sentence that will be a summary by calculating Latent Semantic Analysis (LSA). The level of text summary results is measured by the calculation of recall, precision, and f-measure. This study resulted in the value of the text in the short story having an average precision of 74.93%, recall of 68.4%, and f-measure of 71.43%. Based on the accuracy it can be said that it gives summary results that resemble the results of a manual summary quite well in describing the contents of the entire short story.

Keywords : *Automatic Text Summarization, TF-IDF, Latent Semantic Analysis.*

# PERINGKASAN TEKS BAHASA INDONESIA PADA CERPEN MENGUNAKAN METODE *LATENT SEMANTIC ANALYSIS* (LSA)

Oleh :

Nabilah Thahirah Kasim

09021281722040

## ABSTRAK

Cerpen atau cerita pendek menjadi suatu bacaan yang dapat menghibur bagi para pembaca dari berbagai kalangan usia. Dalam penyajian cerpen, seringkali penamaan judul dan pemberian kategori kurang mewakili isi dari cerita yang ditulis, sehingga akan menimbulkan masalah ketika akan mencari sebuah cerpen yang diinginkan. Akhirnya untuk mengetahui isi dari sebuah cerpen pembaca perlu memahami keseluruhan dari cerpen tersebut yang tentunya membutuhkan waktu yang sangat lama. Alternatif lainnya untuk mendapatkan informasi dan memahami sebuah cerpen secara cepat yaitu dengan membaca ringkasannya atau sinopsis cerpen. Peringkasan teks otomatis adalah penelitian yang ada di bidang pemrosesan bahasa alami yang bertujuan dapat membantu para pembaca dalam meringkas sebuah cerpen agar lebih efisien dalam mendapatkan inti pokok cerita. Pada penelitian ini peringkasan teks dilakukan menggunakan algoritma *Latent Semantic Analysis* (LSA). Pengujian dilakukan menggunakan data uji sebanyak 30 teks cerpen dan *compression rate* kalimat hasil ringkasan sebanyak 30% dari jumlah kalimat teks cerpen. Pada penelitian ini terdapat tiga tahap, yaitu *text preprocessing*, perhitungan bobot *TF-IDF*, dan menentukan kalimat yang akan menjadi ringkasan dengan perhitungan *Latent Semantic Analysis* (LSA). Tingkat akurasi hasil ringkasan teks diukur dengan perhitungan *recall*, *precision*, dan *f-measure*. Penelitian ini menghasilkan ringkasan teks pada cerpen memiliki rata – rata nilai *precision* sebesar 74,93%, *recall* 68,4% dan *f-measure* 71,43%. Berdasarkan tingkat akurasi yang dihasilkan dapat dikatakan hasil ringkasan dapat menyerupai hasil ringkasan manual dengan cukup baik dalam menggambarkan isi keseluruhan cerpen.

Kata Kunci : Peringkasan Teks Otomatis, *TF-IDF*, *Latent Semantic Analysis*.

## KATA PENGANTAR

Puji syukur kepada Allah SWT atas berkat dan rahmat-Nya yang telah diberikan kepada Penulis sehingga dapat menyelesaikan Skripsi ini dengan baik. Skripsi ini dengan judul **“Peringkasan Teks Bahasa Indonesia pada Cerpen Menggunakan Metode Latent Semantic Analysis (LSA)”** disusun untuk memenuhi salah satu syarat guna menyelesaikan pendidikan program Strata-1 pada Fakultas Ilmu Komputer Program Studi Teknik Informatika di Universitas Sriwijaya.

Penulis ingin mengucapkan terima kasih kepada seluruh pihak yang telah memberikan banyak bantuan dan dukungan baik materil dan moril selama proses penyusunan skripsi ini. Oleh karena itu, penulis ingin menyampaikan secara khusus kepada pihak yang telah membantu sebagai berikut:

1. Allah SWT yang telah memberikan nikmat hijrah, kelancaran dan kemudahan untuk penulis dalam menyelesaikan skripsi ini.
2. Kedua orang tua saya tercinta, saudara dan saudariku tersayang, dan seluruh keluarga besar dari “Petta Tompo” dan “Petta Punna” yang telah memberikan dukungan materil dan moril serta doa dan restu yang terus menerus demi kelancaran penyusunan skripsi ini.
3. Bapak Jaidan Jauhari, M.T selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya, Ibu Alvi Syahrini Utami, M.Kom. selaku Ketua



Jurusan Teknik Informatika, dan Ibu Mastura Diana Merieska, M.T. selaku Sekretaris Jurusan Teknik Informatika.

4. Ibu Novi Yusliani, M.T. selaku dosen pembimbing I dan Bapak Dr. M. Fachrurrozi, M.T. selaku selaku pembimbing II yang telah membimbing, mengarahkan, dan memberikan kemudahan penulis dalam proses perkuliahan serta pengerjaan skripsi.
5. Bapak Kanda Januar Miraswan, M.T. selaku dosen pembimbing akademik yang telah membimbing, mengarahkan, dan memberikan motivasi penulis selama masa perkuliahan.
6. Seluruh Dosen Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya yang telah memberikan ilmu pengetahuan kepada penulis selama kegiatan akademik berlangsung.
7. Kak Ricy dan seluruh staff tata usaha yang telah membantu dalam kelancaran proses administrasi dan akademik selama masa perkuliahan.
8. Grup WA “Masih anak IT” dan Grup Skype “Belajar” yang telah berbagi keluh kesah, motivasi, semangat, dan canda tawa selama masa perkuliahan.
9. Teman-teman IF Reguler B, dan teman teman jurusan Teknik Informatika reguler angkatan 2017 yang tidak dapat disebutkan satu persatu yang telah membantu penulis dalam menyelesaikan skripsi ini.
10. Semua teman – teman organisasi baik di BEM KM Fasilkom, NAC, dan LDF WIFI yang telah memberikan banyak pelajaran dan pengalaman berharga.

11. Teman - teman dari SMA N 1 Kota Jambi – MIPA 2 Terkhusus Alin dan Ikhsan yang saling menyemangati dalam pengerjaan skripsi ini sehingga dapat terselesaikan.

Penulis menyadari dalam penyusunan Skripsi ini masih terdapat banyak kekurangan disebabkan keterbatasan pengetahuan dan pengalaman, oleh karena itu kritik dan saran yang membangun sangat diharapkan untuk kemajuan penelitian selanjutnya. Akhir kata semoga Skripsi ini dapat berguna dan bermanfaat bagi kita semua.

Indralaya, 27 Juli 2022

Nabilah Thahirah Kasim

## DAFTAR ISI

	Halaman
LEMBAR PENGESAHAN SKRIPSI .....	ii
TANDA LULUS UJIAN SIDANG SKRIPSI .....	iii
HALAMAN PERNYATAAN BEBAS PLAGIAT .....	iv
MOTTO DAN PERSEMBAHAN .....	v
ABSTRACT .....	vi
ABSTRAK .....	vii
KATA PENGANTAR .....	viii
DAFTAR ISI .....	xii
DAFTAR TABEL .....	xiv
DAFTAR GAMBAR .....	xvii
BAB I PENDAHULUAN .....	I-1
1.1.    Pendahuluan .....	I-1
1.2.    Latar Belakang Masalah .....	I-1
1.3.    Rumusan Masalah .....	I-4
1.4.    Tujuan Penelitian .....	I-4
1.5.    Manfaat Penelitian .....	I-5
1.6.    Batasan Masalah .....	I-5
1.7.    Sistematika Penulisan .....	I-5
1.8.    Kesimpulan .....	I-7
BAB II KAJIAN LITERATUR .....	II-1
2.1    Pendahuluan .....	II-1
2.2    Landasan Teori .....	II-1
2.2.1    Cerpen .....	II-1
2.2.2    Peringkasan Teks Otomatis .....	II-2
2.2.3    Preprocessing .....	II-3
2.2.4    Algoritma TF-IDF .....	II-13
2.2.5    Metode Latent Semantic Analysis (LSA) .....	II-15

2.2.6	Tipe Evaluasi .....	II-21
2.2.7	Rational Unified Process (RUP).....	II-23
2.3	Penelitian Lain yang Relevan.....	II-25
2.4	Kesimpulan.....	II-27
 <b>BAB III METODOLOGI PENELITIAN .....</b>		<b>III-1</b>
3.1	Pendahuluan .....	III-1
3.2	Unit Penelitian.....	III-1
3.3	Pengumpulan Data .....	III-1
3.3.1	Jenis dan Sumber Data.....	III-1
3.3.2	Metode Pengumpulan Data.....	III-2
3.4	Tahapan Penelitian .....	III-2
3.4.1.	Kerangka Kerja.....	III-3
3.4.2.	Kriteria Pengujian.....	III-4
3.4.3.	Format Data Pengujian .....	III-5
3.4.4.	Alat yang Digunakan dalam Pelaksanaan Penelitian.....	III-6
3.4.5.	Pengujian Penelitian .....	III-6
3.4.6.	Analisis Hasil Pengujian dan Membuat Kesimpulan .....	III-7
3.5	Metode Pengembangan Perangkat Lunak .....	III-7
3.5.1	Fase Insepsi.....	III-7
3.5.2	Fase Elaborasi.....	III-8
3.5.3	Fase Konstruksi .....	III-8
3.5.4	Fase Transisi .....	III-8
 <b>BAB IV PENGEMBANGAN PERANGKAT LUNAK .....</b>		<b>IV-1</b>
4.1.	Pendahuluan .....	IV-1
4.2.	Fase Insepsi .....	IV-1
4.2.1.	Pemodelan Bisnis.....	IV-1
4.2.2.	Kebutuhan sistem.....	IV-2
4.2.3.	Analisis dan Desain.....	IV-4
4.3.	Fase Elaborasi .....	IV-26
4.3.1	Pemodelan Bisnis.....	IV-26

4.3.2	Perancangan Data.....	IV-26
4.3.3	Perancangan Antarmuka .....	IV-26
4.3.4	Kebutuhan Sistem .....	IV-27
4.3.5	Diagram Aktivitas ( <i>Activity Diagram</i> ).....	IV-28
4.3.6	Diagram Kelas Analisis ( <i>Analysis Class Diagram</i> ).....	IV-28
4.3.7	Diagram Alur ( <i>Sequence Diagram</i> ) .....	IV-31
4.4.	Fase Konstruksi .....	IV-33
4.4.1	Kebutuhan Sistem .....	IV-33
4.4.2	Diagram Kelas .....	IV-34
4.4.3	Implementasi.....	IV-34
4.5.	Fase Transisi.....	IV-37
4.5.1	Pemodelan Bisnis.....	IV-37
4.5.2	Kebutuhan Sistem .....	IV-37
4.5.3	Rencana Pengujian.....	IV-37
4.5.4	Implementasi.....	IV-40
4.6.	Kesimpulan.....	IV-43
BAB V HASIL DAN ANALISIS PENELITIAN .....		V-1
5.1.	Pendahuluan .....	V-1
5.2.	Data Hasil Percobaan/Penelitian .....	V-1
5.2.1.	Konfigurasi Percobaan.....	V-1
5.2.2.	Data Hasil Konfigurasi .....	V-1
5.3.	Analisis Hasil Penelitian .....	V-12
5.4.	Kesimpulan.....	V-17
BAB VI KESIMPULAN DAN SARAN .....		VI-1
6.1.	Pendahuluan .....	VI-1
6.2.	Kesimpulan.....	VI-1
6.3.	Saran.....	VI-2
DAFTAR PUSTAKA .....		xvii

## DAFTAR TABEL

	Halaman
Tabel II-1. Ilustrasi Proses Case Folding .....	II-4
Tabel II-2. Ilustrasi Proses Tokenizing Kalimat .....	II-5
Tabel II-3. Ilustrasi Proses Tokenizing Kata .....	II-6
Tabel II-4. Ilustrasi Proses Stopword Removal .....	II-7
Tabel II-5. Ilustrasi proses Stemming .....	II-8
Tabel II-6. Kombinasi Awalan Akhiran yang tidak diijinkan .....	II-10
Tabel II-7. Aturan Stemming Prefiks.....	II-11
Tabel II-8. Ilustrasi Hasil Penghitungan Bobot masing-masing Kata .....	II-15
Tabel II- 9. Ilustrasi Hasil Ringkasan .....	II-21
Tabel III-1, Rancangan Data Masukan .....	III-5
Tabel III-2. Rancangan Format Data Pengujian .....	III-5
Tabel III-3. Rancangan Hasil Pengujian Pengujian .....	III-7
Tabel IV-1. Kebutuhan Fungsional.....	IV-2
Tabel IV-2. Kebutuhan Non Fungsional.....	IV-3
Tabel IV-3. Teks Cerpen Asli .....	IV-6
Tabel IV-4. Tahapan Case Folding .....	IV-7
Tabel IV-5. Tahapan Tokenisasi Kalimat .....	IV-8
Tabel IV-6. Tahapan Tokenisasi Kata .....	IV-9
Tabel IV-7. Tahapan Stopwords Removal.....	IV-10
Tabel IV-8. Tahap Stemming .....	IV-10
Tabel IV-9. Tahap Pembobotan Kata TF-IDF .....	IV-11
Tabel IV-10. Tahap Pembobotan Kata TF-IDF .....	IV-11
Tabel IV-11. Hasil Tahapan Perhitungan SVD .....	IV-18
Tabel IV-12. Perbandingan Kalimat Ringkasan Sistem dan Manual .....	IV-19
Tabel IV-13. Perhitungan Recall, Precision, dan F-measure .....	IV-19
Tabel IV-14. Definisi Aktor.....	IV-21
Tabel IV-15. Definisi Use Case .....	IV-21

Tabel IV-16. Skenario Use Case Mengunggah File .....	IV-22
Tabel IV-17. Skenario Use Case Meringkas Teks .....	IV-23
Tabel IV-18. Skenario Use Case Menghitung Akurasi.....	IV-24
Tabel IV-19. Daftar Implementasi Kelas .....	IV-35
Tabel IV-20. Rencana Pengujian Use Case Mengunggah Teks .....	IV-38
Tabel IV-21. Rencana Pengujian Use Case Meringkas Teks .....	IV-38
Tabel IV-22. Rencana Pengujian Use Case Menghitung Akurasi .....	IV-38
Tabel IV-23. Pengujian Use Case Menunggah Teks .....	IV-40
Tabel IV-24. Pengujian Use Case Meringkas Teks .....	IV-41
Tabel IV-25. Pengujian Use Case Mengukur Tingkat Akurasi .....	IV-42
Tabel V-1. Data Cerpen dan Jumlah Kalimat Cerpen .....	V-2
Tabel V-2. Hasil pengujian Ringkasan “Bunny Helper” .....	V-7
Tabel V-3. Perbandingan Kalimat Ringkasan Sistem dan manual. ....	V-7
Tabel V-4. Hasil Akurasi Cerita Pendek.....	V-10
Tabel V-5. Nilai Rata-Rata Precision, Recall, dan F-Measure .....	V-12
Tabel V-6. Hasil Responden Terhadap Hasil Ringkasan Sistem.....	V-14

## DAFTAR GAMBAR

Halaman

Gambar II-1. Alur Proses Penelitian (Gotami et al., 2018).....	II-3
Gambar II-2. Tahapan Preprocessing.....	II-4
Gambar II-3. Gambar Matriks Singular Value Decomposiion (SVD) .....	II-17
Gambar II-4. Fase Rational Unified Process (RUP) .....	II-24
Gambar III-1. Kerangka Kerja Peringkasan Cerpen Bahasa Indonesia.....	III-3
Gambar IV-1. Diagram Use Case .....	IV-20
Gambar IV-2. Rancangan Antarmuka .....	IV-27
Gambar IV-3. Diagram Kelas Mengunggah Teks .....	IV-29
Gambar IV-4. Diagram Kelas Meringkas Teks .....	IV-29
Gambar IV-5. Diagram Kelas Menghitung Akurasi.....	IV-30
Gambar IV-6. Sequence Diagram Mengunggah Teks .....	IV-31
Gambar IV-7. Sequence Diagram Meringkas Teks .....	IV-32
Gambar IV-8. Sequence Diagram Menghitung Akurasi.....	IV-33
Gambar IV-9. Diagram Kelas Keseluruhan.....	IV-34
Gambar IV-10. Tampilan Antarmuka Perangkat Lunak.....	IV-36
Gambar V-1. Tampilan Form Peringkasan “Bunny Helper” .....	V-4
Gambar V-2. Hasil Ringkasan Manual “Bunny Helper” .....	V-5
Gambar V-3. Hasil ringkasan sistem “Bunny Helper” .....	V-6
Gambar V-4. Grafik Nilai Rata-Rata Precision, Recall, dan F-Measure.....	V-12
Gambar V-5. Grafik Akurasi Berdasarkan Jumlah Kalimat .....	V-13



# **BAB I**

## **PENDAHULUAN**

### **1.1. Pendahuluan**

Bagian ini membahas latar belakang, rumusan masalah, tujuan dan manfaat dari penelitian, batasan-batasan masalah yang dibentuk dan sistematika dalam penulisan laporan tugas akhir ini. Bagian ini berisi tentang penjelasan secara umum tentang penggunaan metode *Latent Semantic Analysis* (LSA) pada peringkasan teks dari sebuah cerpen.

### **1.2. Latar Belakang Masalah**

Kegemaran membaca merupakan aktivitas yang menunjukkan kesukaan dan kesenangan seseorang terhadap suatu bacaan seperti cerpen. Cerpen termasuk dalam teks sastra, yaitu teks yang disusun untuk tujuan artistik dan imajiner dengan memakai bahasa lisan dan tulisan (Wicaksono *et al.*, 2018).

Cerpen menurut Sumardjo (1987) adalah cerita yang terbatas membahas salah satu unsur fiksi dalam aspek terkecilnya. Cerpen menjadi pendek bukan karena bentuk dan karakternya sebagai karya sastra pendek, tetapi karena aspek masalah yang sangat terbatas dalam membahas dan menyelesaikan konflik. Tidak ada ukuran baku tentang jumlah kata minimal dan maksimal dari sebuah cerpen. Panjang atau pendeknya sebuah cerpen pada umumnya ditulis sekitar 500 hingga 30.000 kata (Narti, 2018).

Dalam penyajian cerpen, seringkali penamaan judul dan pemberian kategori kurang mewakili isi dari cerita yang ditulis, sehingga akan menimbulkan masalah ketika akan mencari sebuah cerpen yang diinginkan. Akhirnya untuk mengetahui isi dari sebuah cerpen pembaca perlu memahami keseluruhan dari cerpen tersebut yang tentunya membutuhkan waktu yang tidak singkat (Dewi *et al.*, 2020). Alternatif lainnya yaitu dengan membaca ringkasan atau sinopsis dari sebuah cerpen.

Ringkasan dibutuhkan untuk mengetahui kalimat-kalimat ide pokok dari sebuah teks cerpen atau sinopsis tanpa perlu membaca teks secara keseluruhan terlebih dahulu. Sehingga, pembaca dapat lebih cepat mengetahui ide pokok dari cerita yang disajikan. Dalam situasi seperti itu, penerapan sistem ringkasan teks dapat memanfaatkan teknologi untuk mengurangi waktu membaca seluruh teks dengan hanya membaca hasil ringkasan (Mustaqhfiri *et al.*, 2012).

Peringkasan teks, atau lebih tepatnya, peringkasan teks otomatis, adalah proses di mana komputer menghasilkan teks yang lebih ringkas daripada teks asli, tetapi tetap menyimpan informasi penting dalam teks asli (Nagwani *and* Verma, 2011). Bidang penelitian peringkasan teks telah dipelajari sejak pertengahan abad ke-20, yang pertama kali dibahas oleh Luhn (1958) (Mustaqhfiri *et al.*, 2012). Beberapa metode peringkasan teks yang pernah digunakan oleh peneliti sebelumnya seperti metode *Maximum Marginal Relevance* (MMR) (Yulita *et al.*, 2019), metode *Term frequency inverse document frequency* (TF-IDF) (Setiawan, 2015) dan metode *Latent Semantic Analysis* (LSA) (Ozsoy *et al.*, 2010).

Dalam penelitian oleh Yulita *et al.* (2019) menjelaskan kelemahan pada metode *Maximum Marginal Relevance* (MMR) adalah hasil ringkasan terdapat bagian terpisah dari hasil ringkasan yang tidak terkait secara semantik. Hal ini diakibatkan karena metode MMR tidak mempergunakan teknik *natural language comprehension*. Penelitian terkait lainnya juga dilakukan oleh Setiawan (2015) dengan menggunakan metode TF-IDF. Skema pembobotan TF-IDF merangkum teks dengan memberi bobot pada hubungan kata (*term*) dengan dokumen dan memilih dokumen dengan bobot tertinggi sebagai ringkasan. Namun, metode TF-IDF masih mempunyai kelemahan. Yaitu, hasil dari rangkumannya tidak mencerminkan isi teks aslinya dan tata bahasanya masih kurang baik.

Metode *Latent Semantic Analysis* (LSA) adalah kombinasi dari metode aljabar dan statistik. Teknik ini memunculkan struktur kata yang tersembunyi, antara kata, perasaan atau dokumen (Dokun *and* Celebi, 2015). Metode *Latent Semantic Analysis* (LSA) pernah digunakan pada peringkasan sebuah teks diantaranya peringkasan teks berita bahasa indonesia yang dilakukan oleh (Saputra *et al.*, 2017). Selain itu penelitian Xiong *and* Luo (2015) melakukan peringkasan teks dengan membandingkan metode LSA dan metode MMR. Hasil penelitian tersebut menunjukkan bahwa metode LSA lebih baik. Kelebihan peringkasan dokumen dengan metode LSA menggunakan *Singular Value Decomposition* (SVD) menurut Ozsoy *et al.* (2010) adalah untuk menemukan kata dan kalimat yang mirip secara semantik. SVD memiliki kemampuan pengurangan *noise*, yang membantu untuk meningkatkan akurasi.

Berdasarkan uraian tersebut, maka penelitian ini menggunakan metode *Latent Semantic Analysis* (LSA) dengan pendekatan aljabar linear *singular value decomposition* (SVD) yang diharapkan dapat meringkas sebuah cerpen agar lebih efisien dalam mendapatkan inti pokok cerita.

### 1.3. Rumusan Masalah

Fokus permasalahan pada penelitian ini adalah menghasilkan sebuah ringkasan teks pada cerpen menggunakan metode *Latent Semantic Analysis* (LSA). Berikut dibuat pertanyaan penelitian di bawah ini:

1. Bagaimana mengembangkan sebuah sistem peringkasan teks bahasa Indonesia pada cerpen menggunakan metode *Latent Semantic Analysis* (LSA)?
2. Bagaimana kinerja metode *Latent Semantic Analysis* (LSA) pada sistem peringkasan teks bahasa Indonesia pada cerpen berdasarkan nilai *precision*, *recall*, dan *f-measure*?

### 1.4. Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

1. Menghasilkan sebuah sistem peringkasan teks bahasa Indonesia pada cerpen menggunakan metode *Latent Semantic Analysis* (LSA).
2. Mengetahui kinerja metode *Latent Semantic Analysis* (LSA) dalam peringkasan teks bahasa Indonesia pada cerpen berdasarkan nilai *precision*, *recall*, dan *f-measure*.

### **1.5. Manfaat Penelitian**

Manfaat yang diperoleh dari penelitian ini adalah:

1. Memberikan solusi bagi para pembaca agar mempermudah untuk mengetahui isi cerita melalui ringkasan cerpen.
2. Dapat digunakan sebagai referensi untuk penelitian selanjutnya.

### **1.6. Batasan Masalah**

Batasan masalah dari penelitian ini adalah:

1. Dokumen yang digunakan pada penelitian ini berupa dokumen dalam bahasa Indonesia.
2. Sistem Peringkasan teks hanya memproses kata dan kalimat, tidak untuk rumus, simbol dan tabel.
3. *Dataset* yang digunakan pada penelitian ini berjumlah 30 data cerpen yang diambil secara acak.
4. Dokumen yang digunakan berupa teks dengan format \*.txt.
5. Tidak menangani kesalahan penulisan pada teks dokumen yang akan diringkas dan struktur kalimat yang tidak sesuai kaidah Bahasa.

### **1.7. Sistematika Penulisan**

Sistematika penulisan skripsi ini mengikuti standar penulisan tugas akhir Fakultas Ilmu Komputer Universitas Sriwijaya sebagai berikut:

## **BAB I. PENDAHULUAN**

Dijelaskan pokok-pokok pikiran yang mendasari penelitian, seperti latar belakang dari permasalahan, rumusan masalah, tujuan penelitian, keunggulan penelitian, batasan masalah, kerangka penulisan, dan kesimpulan.

## **BAB II. KAJIAN LITERATUR**

Membahas dasar-dasar teoritis yang dipakai dalam melakukan penelitian, seperti definisi cerita pendek, peringkasan teks otomatis, *text preprocessing*, Algoritma *TF-IDF*, *Latent Semantic Analysis (LSA)*, dll. Serta dibahas beberapa tinjauan pustaka dari penelitian lainnya yang berkaitan dengan penelitian yang akan dilakukan.

## **BAB III. METODOLOGI PENELITIAN**

Dijelaskan langkah-langkah yang akan di kerjakan dalam melakukan penelitian. Setiap langkah dari rencana studi dirinci sesuai dengan kerangka kerja. Bagian ini mencakup perancangan manajemen proyek dalam praktik penelitian.

## **BAB IV. PENGEMBANGAN PERANGKAT LUNAK**

Membahas proses pengembangan perangkat lunak penelitian dengan metode RUP. Terdiri dari proses analisis kebutuhan, perancangan, implementasi, dan pengujian perangkat lunak.

## **BAB V. HASIL DAN ANALISA PENELITIAN**

Dijelaskan hasil pengujian sesuai pola yang telah direncanakan. Hasil tersebut digunakan sebagai dasar penarikan kesimpulan dalam penelitian yang telah dilakukan.

## **BAB VI. KESIMPULAN DAN SARAN**

Berisi kesimpulan dari seluruh tahapan dalam penelitian dan saran berdasarkan temuan ringkasan cerita pendek.

### **1.8. Kesimpulan**

Pada bab 1 ini dapat disimpulkan bahwa teknologi dapat membantu para pembaca dalam meringkas sebuah cerpen agar lebih efisien dalam mendapatkan inti pokok cerita.

## DAFTAR PUSTAKA

- Agusta, L., Teknologi, F., Universitas, I., & Wacana, K. S. (2009). PERBANDINGAN ALGORITMA STEMMING PORTER DENGAN ALGORITMA NAZIEF & ADRIANI UNTUK STEMMING DOKUMEN TEKS BAHASA INDONESIA. In *Konferensi Nasional Sistem dan Informatika*.
- Alguliev, R., & Aliguliyev, R. (2009). Evolutionary Algorithm for Extractive Text Summarization. *Intelligent Information Management*, 1, 128–138. <https://doi.org/10.4236/iim.2009.12019>
- Arifin, T., & Dewi, K. E. (2017). Peringkasan Teks Otomatis Pada Multi Dokumen Menggunakan Relevance Vector Machine. *Skripsi, Perpustakaan UNIKOM*, 6, 1.
- Aristoteles. (2013). Penerapan Algoritma Genetika pada Peringkasan Teks Dokumen Bahasa Indonesia. *Semirata FMIPA Universitas Lampung*, 29–33. <http://jurnal.fmipa.unila.ac.id/index.php/semirata/article/download/703/523>
- Asian, J. (2007). Effective Techniques for Indonesian Text Retrieval. *Ph.D Thesis*.
- Asian, J., Williams, H. E., & Tahaghoghi, S. M. M. (2005). Stemming Indonesian. *Conferences in Research and Practice in Information Technology Series*, 38(4), 307–314. <https://doi.org/10.1145/1316457.1316459>
- Babar, S. A., & Patil, P. D. (2015). Improving performance of text summarization. *Procedia Computer Science*, 46, 354–363. <https://doi.org/10.1016/j.procs.2015.02.031>
- Badry, R. M., Eldin, A. S., & Saad Elzanfally, D. (2013). Text Summarization within the Latent Semantic Analysis Framework: Comparative Study. In *International Journal of Computer Applications* (Vol. 81, Issue 11).



- Dalal, V., & Malik, L. (2013). A survey of extractive and abstractive text summarization techniques. *International Conference on Emerging Trends in Engineering and Technology, ICETET*, 109–110. <https://doi.org/10.1109/ICETET.2013.31>
- Dewi, M. D., Nugraha, B. S. D., & Rahutomo, F. (2020). Penerapan Algoritma Score-based pada Peringkasan Teks Cerpen Otomatis. *Seminar Informatika Aplikatif Polinema*, 293–297. <http://jurnalti.polinema.ac.id/index.php/SIAP/article/view/788>
- Dokun, O., & Celebi, E. (2015). *Single-Document Summarization Using Latent Semantic Analysis*. 1(2), 0–13.
- Efendi Pohan, A. (2020). *Literacy Goes To School: GERAKAN LITERASI NASIONAL - Albert Efendi Pohan, S. Pd., M. Pd - Google Buku*. [https://books.google.co.id/books?id=VVHcDwAAQBAJ&printsec=frontcover&hl=id&source=gbs\\_ge\\_summary\\_r&cad=0#v=onepage&q&f=false](https://books.google.co.id/books?id=VVHcDwAAQBAJ&printsec=frontcover&hl=id&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false)
- Gong, Y., & Liu, X. (2001). Generic text summarization using relevance measure and latent semantic analysis. *SIGIR Forum (ACM Special Interest Group on Information Retrieval)*, 19–25. <https://doi.org/10.1145/383952.383955>
- Gotami, N. S. W., Indriati, I., & Dewi, R. K. (2018). *Peringkasan Teks Otomatis Secara Ekstraktif Pada Artikel Berita Kesehatan Berbahasa Indonesia Dengan Menggunakan Metode Latent Semantic Analysis*. <http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/2430>
- Grossman, D. A. (1998). *Information Retrieval: Algorithm and Heuristics*. <https://my.b-ok.as/book/2098407/54fa2e>
- Hanifa, F. R. N. (2019). *Coreference Resolution Bahasa Indonesia Pada Kata Ganti Kepemilikan Menggunakan Recurrent Neural Network*.

<http://elibrary.unikom.ac.id>

- Jayadianti, H., Damayanti, R., & Juwairiah. (2020). Latent Semantic Analysis ( Lsa ) Dan Automatic Text Summarization ( Ats ) Dalam Optimasi Pencarian Artikel Covid 19. *Seminar Nasional Informatika 2020, 2020*(Semnasif), 52–59.
- Kruchten, P., & Wesley, P. A. (2000). *The Rational Unified Process: An Introduction (2nd Edition)*.
- Mustaqhfiri, M., Abidin, Z., & Kusumawati, R. (2012). Peringkasan Teks Otomatis Berita Berbahasa Indonesia Menggunakan Metode Maximum Marginal Relevance. *Matics*. <https://doi.org/10.18860/mat.v0i0.1578>
- Nagwani, N. K., & Verma, S. (2011). Software Bug Classification using Suffix Tree Clustering (STC) Algorithm. *Undefined*.
- Narti. (2018). *Peningkatan Kemampuan Memahami Unsur Intrinsik Cerpen melalui Metode Kooperatif Tipe Student Teams Achievement Division (STAD) pada Siswa Kelas VIII C SMP Ta'mirul Islam Surakarta Tahun Pelajaran 2017/2018*. *Jurnal Pendidikan Empirisme*. [https://books.google.co.id/books?id=AQ6DDwAAQBAJ&pg=PA30&lpg=PA30&dq=Peningkatan+Kemampuan+Memahami+Unsur+Intrinsik+Cerpen+melalui+Metode+Kooperatif+Tipe+Student+Teams+Achievement+Division+\(STAD\)+pada+Siswa+Kelas+VIII+C+SMP+Ta'mirul+Islam+Surakarta+Tahun+](https://books.google.co.id/books?id=AQ6DDwAAQBAJ&pg=PA30&lpg=PA30&dq=Peningkatan+Kemampuan+Memahami+Unsur+Intrinsik+Cerpen+melalui+Metode+Kooperatif+Tipe+Student+Teams+Achievement+Division+(STAD)+pada+Siswa+Kelas+VIII+C+SMP+Ta'mirul+Islam+Surakarta+Tahun+)
- Nurdiana, O., Jumadi, J., & Nursantika, D. (2016). Perbandingan Metode Cosine Similarity Dengan Metode Jaccard Similarity Pada Aplikasi Pencarian Terjemah Al-Qur'an Dalam Bahasa Indonesia. *Jurnal Online Informatika*, 1(1), 59. <https://doi.org/10.15575/join.v1i1.12>
- Ozsoy, M. G., Cicekli, I., & Alpaslan, F. N. (2010). Text summarization of turkish texts using latent semantic analysis. *Coling 2010 - 23rd International Conference on*

*Computational Linguistics, Proceedings of the Conference.*

Pardede, J. (2015). *Aplikasi Peringkasan Dokumen Menggunakan Algoritma Interactive Graph-Based dan Similarity*. Jurusan Teknik Informatika Institut Teknologi Nasional.

Saputra, J., Fachrurrozi, M., & Yunita. (2017). Peringkasan Teks Berita Berbahasa Indonesia Menggunakan Metode Latent Semantic Analysis (LSA) dan Teknik Steinberger & Jezek. *Computer Science and ICT*, 3(1), 215–219.

Setiawan, D. (2015). *Peringkasan teks berita secara Otomatis menggunakan TF.IDF*.  
<http://repository.usu.ac.id/handle/123456789/46978>

Setyadi, I. W. A., Khrisne, D. C., & Suyadnya, I. M. A. (2018). Automatic Text Summarization Menggunakan Metode Graph dan Metode Ant Colony Optimization. *Majalah Ilmiah Teknologi Elektro*, 17(1), 124–130.  
<https://doi.org/10.24843/MITE.2018.v17i01.P17>

Wicaksono, A., Emzir, & Rohman, S. (2018). *Tentang Sastra: Orkestrasi Teori dan Pembelajarannya - Google Buku*. Garudhawaca.  
[https://books.google.co.id/books?id=cwhTDwAAQBAJ&hl=id&source=gbs\\_navlinks\\_s](https://books.google.co.id/books?id=cwhTDwAAQBAJ&hl=id&source=gbs_navlinks_s)

Xiong, S., & Luo, Y. (2015). A new approach for multi-document summarization based on latent semantic analysis. *Proceedings - 2014 7th International Symposium on Computational Intelligence and Design, ISCID 2014*, 1, 177–180.  
<https://doi.org/10.1109/ISCID.2014.27>

Yulita, W. (2015). *Implementasi Metode Maximum Marginal Relevance pada Peringkasan Teks*.

Yulita, W., Priyanta, S., & SN, A. (2019). Automatic Text Summarization Based on Semantic Networks and Corpus Statistics. *IJCCS (Indonesian Journal of*

*Computing and Cybernetics Systems*), 13(2), 137.  
<https://doi.org/10.22146/ijccs.38261>

Zamzam, M. A. (2019). *SISTEM AUTOMATIC TEXT SUMMARIZATION SKRIPSI MENGGUNAKAN ALGORITMA TEXTRANK.*