

**TEXT SUMMARIZATION UNTUK KORPUS DOKUMEN
DENGAN METODE LATENT SEMANTIC ANALYSIS**

Skripsi



oleh:

Stevani Dwi Utomo

71180258

**PROGRAM STUDI INFORMATIKA FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS KRISTEN DUTA WACANA**

2022

**TEXT SUMMARIZATION UNTUK KORPUS DOKUMEN
DENGAN METODE LATENT SEMANTIC ANALYSIS**

Skripsi



Diajukan kepada Program Studi Informatika Fakultas Teknologi Informasi
Universitas Kristen Duta Wacana
Sebagai Salah Satu Syarat dalam Memperoleh Gelar
Sarjana Komputer

Disusun oleh

Stevani Dwi Utomo

71180258

**PROGRAM STUDI INFORMATIKA FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS KRISTEN DUTA WACANA**

2022

HALAMAN PENGESAHAN

TEXT SUMMARIZATION UNTUK KORPUS DOKUMEN DENGAN METODE LATENT SEMANTIC ANALYSIS

Oleh: STEVANI DWI UTOMO / 71180258

Dipertahankan di depan Dewan Penguji Skripsi
Program Studi Informatika Fakultas Teknologi Informasi
Universitas Kristen Duta Wacana - Yogyakarta
Dan dinyatakan diterima untuk memenuhi salah satu syarat memperoleh gelar
Sarjana Komputer
pada tanggal 25 Oktober 2022

Yogyakarta, 3 November 2022
Mengesahkan,

Dewan Penguji:

1. Antonius Rachmat C., S.Kom.,M.Cs.
2. Budi Susanto, SKom.,M.T.
3. Willy Sudiarto Raharjo, S.Kom.,M.Cs.
4. Yuan Lukito, S.Kom., M.Cs.

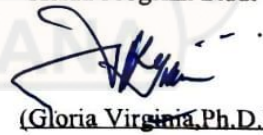


Dekan



(Restyandito, S.Kom.,MSIS.,Ph.D.)

Ketua Program Studi



(Gloria Virginia, Ph.D.)

PERNYATAAN KEASLIAN SKRIPSI

Saya menyatakan dengan sesungguhnya bahwa skripsi dengan judul:

TEXT SUMMARIZATION UNTUK KORPUS DOKUMEN DENGAN METODE LATENT SEMANTIC ANALYSIS

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

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

Yogyakarta, 25 November 2022



Stevani Dwi Utomo

71180258

HALAMAN PERSETUJUAN

Judul Skripsi : TEXT SUMMARIZATION UNTUK KORPUS
DOKUMEN DENGAN METODE LATENT
SEMANTIC ANALYSIS

Nama Mahasiswa : STEVANI DWI UTOMO

N I M : 71180258

Matakuliah : Skripsi (Tugas Akhir)

Kode : TI0366

Semester : Gasal


Tahun Akademik : 2022/2023

Telah diperiksa dan disetujui di
Yogyakarta,
Pada tanggal 3 November 2022

Dosen Pembimbing I

Dosen Pembimbing II


Antonius Rachmat C., S.Kom.,M.Cs.


Budi Susanto, SKom.,M.T.

**HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS
AKHIR UNTUK KEPENTINGAN AKADEMIS SECARA ONLINE
UNIVERSITAS KRISTEN DUTA WACANA YOGYAKARTA**

Saya yang bertanda tangan di bawah ini:

NIM : 71180258
Nama : Stevani Dwi Utomo
Prodi / Fakultas : Teknologi Informasi / Informatika
Judul Tugas Akhir : Text Summarization Untuk Korpus Dokumen
dengan Metode Latent Semantic Analysis

bersedia menyerahkan Tugas Akhir kepada Universitas melalui Perpustakaan untuk keperluan akademis dan memberikan **Hak Bebas Royalti Non Eksklusif** (*Non-exclusive Royalty-free Right*) serta bersedia Tugas Akhirnya dipublikasikan secara online dan dapat diakses secara lengkap (*full access*).

Dengan Hak Bebas Royalti Noneklusif ini Perpustakaan Universitas Kristen Duta Wacana berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk *database*, merawat, dan mempublikasikan Tugas Akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta. Demikian pernyataan ini saya buat dengan sebenar-benarnya.

Yogyakarta, 24 November 2022

Yang menyatakan,



(71180258-Stevani Dwi Utomo)



Karya sederhana ini dipersembahkan
kepada Tuhan, Keluarga Tercinta,
dan Kedua Orang Tua



Segala sesuatu indah pada waktu-Nya

Anonim

Perjalanan ribuan mil dimulai dari langkah satu mil

(Pepatah Kuno)

KATA PENGANTAR

Segala puji dan syukur kepada Tuhan yang maha kasih, karena atas segala rahmat, bimbingan, dan bantuan-Nya maka akhirnya Skripsi dengan judul Text Summarization Untuk Korpus Dokumen dengan Metode Latent Semantic Analysis ini telah selesai disusun.

Penulis memperoleh banyak bantuan dari kerja sama baik secara moral maupun spiritual dalam penulisan Skripsi ini, untuk itu tak lupa penulis ucapkan terima kasih yang sebesar-besarnya kepada:

1. Tuhan yang maha kasih,
2. Orang tua penulis, yang selalu mendukung, dan berdoa untuk perjalanan penulis menyelesaikan skripsi ini.
3. Bapak Restyandito, S. Kom, M. SIS, Ph. D. selaku Dekan Fakultas Teknologi Informasi Universitas Kristen Duta Wacana.
4. Ibu Gloria Virginia, S. Kom., MAI., Ph. D. selaku Kepala Program Studi Informatika Fakultas Teknologi Informasi Universitas Kristen Duta Wacana.
5. Antonius Rachmat C., S.Kom.,M.Cs selaku Dosen Pembimbing 1, yang telah memberikan ilmunya dan dengan penuh kesabaran membimbing penulis,
6. Budi Susanto, SKom.,M.T. selaku Dosen Pembimbing 2 yang telah memberikan ilmu dan kesabaran dalam membimbing penulis,
7. Keluarga tercinta: yang selalu mendukung penulis menyelesaikan skripsi ini.
8. Kezia Natasya, Jesslyn Septhia selaku sahabat penulis yang selalu menemani penulis dalam lika-liku kehidupan.
9. Segala pihak yang telah mendukung moral, spiritual, dan dana untuk belajar selama ini.

Laporan proposal/skripsi ini tentunya tidak lepas dari segala kekurangan dan kelemahan, untuk itu segala kritikan dan saran yang bersifat membangun guna kesempurnaan skripsi ini sangat diharapkan. Semoga proposal/skripsi ini dapat

bermanfaat bagi pembaca semua dan lebih khusus lagi bagi pengembangan ilmu komputer dan teknologi informasi.

Yogyakarta, 24 November 2022



71180258 - Stevani Dwi Utomo



DAFTAR ISI

HALAMAN PENGESAHAN.....	iii
PERNYATAAN KEASLIAN SKRIPSI.....	iv
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS SECARA ONLINE UNIVERSITAS KRISTEN DUTA WACANA YOGYAKARTA.....	vi
KATA PENGANTAR	ix
DAFTAR ISI.....	xi
DAFTAR TABEL.....	xiii
DAFTAR GAMBAR	xiv
DAFTAR LAMPIRAN.....	xvi
INTISARI.....	xvii
ABSTRACT.....	xviii
PENDAHULUAN	1
1.1. Latar Belakang Masalah.....	1
1.2. Perumusan Masalah.....	2
1.3. Batasan Masalah.....	3
1.4. Tujuan Penelitian.....	3
1.5. Manfaat Penelitian.....	3
1.6. Sistematika Penulisan.....	4
BAB II.....	5
TINJAUAN PUSTAKA DAN DASAR TEORI	5
2.1 Tinjauan Pustaka	5
2.2 Landasan Teori	9
BAB III	16
METODOLOGI PENELITIAN.....	16
3.1 Analisis Kebutuhan Sistem	16
3.1.3 Kebutuhan Fungsional	18
3.2 Perancangan Penelitian	18
3.2.1 Pengambilan data	19
3.2.2 Labeling Unique Token.....	19
3.2.3 Preprocessing	21

3.2.4 Processing	22
3.2.5 Evaluasi	23
3.2.6 Validasi	24
3.3 Diagram Alir	25
3.4 Use Case Diagram.....	25
3.5 Perancangan Struktur Data.....	26
3.6 Perancangan Antarmuka Pengguna.....	28
3.7 Perancangan Pengujian Sistem	31
BAB IV	33
IMPLEMENTASI DAN PEMBAHASAN.....	33
4.1 Implementasi Awal	33
4.2 Implementasi Sistem	33
4.2.1 Pengambilan Data	34
4.2.2 Labeling Unique Token.....	34
4.2.3 Pre processing	35
4.2.5 Evaluasi	38
4.3 Pengujian dan Analisis	40
BAB V.....	42
KESIMPULAN DAN SARAN.....	42
5.1 Kesimpulan	42
5.2 Saran.....	42
Daftar Pustaka	43
LAMPIRAN A	46
KODE SUMBER PROGRAM	46
LAMPIRAN B	59
KARTU KONSULTASI DOSEN 1.....	59
LAMPIRAN C	60
KARTU KONSULTASI DOSEN 2.....	60
LAMPIRAN D.....	61
LAMPIRAN LAIN-LAIN	61

DAFTAR TABEL

Tabel 2.1: Hasil perhitungan dataset turki.....	5
Tabel 2.2: Hasil perhitungan dataset inggris.....	6
Tabel 2.3: Tabel hasil compression rate 50%.....	7
Tabel 2.4: Tabel Pengujian.....	8
Tabel 2.5: Tabel Penilaian Term-Document.....	11
Tabel 2.6: Jenis-jenis Frasa.....	12
Tabel 4.1: Hasil pengujian.....	39



DAFTAR GAMBAR

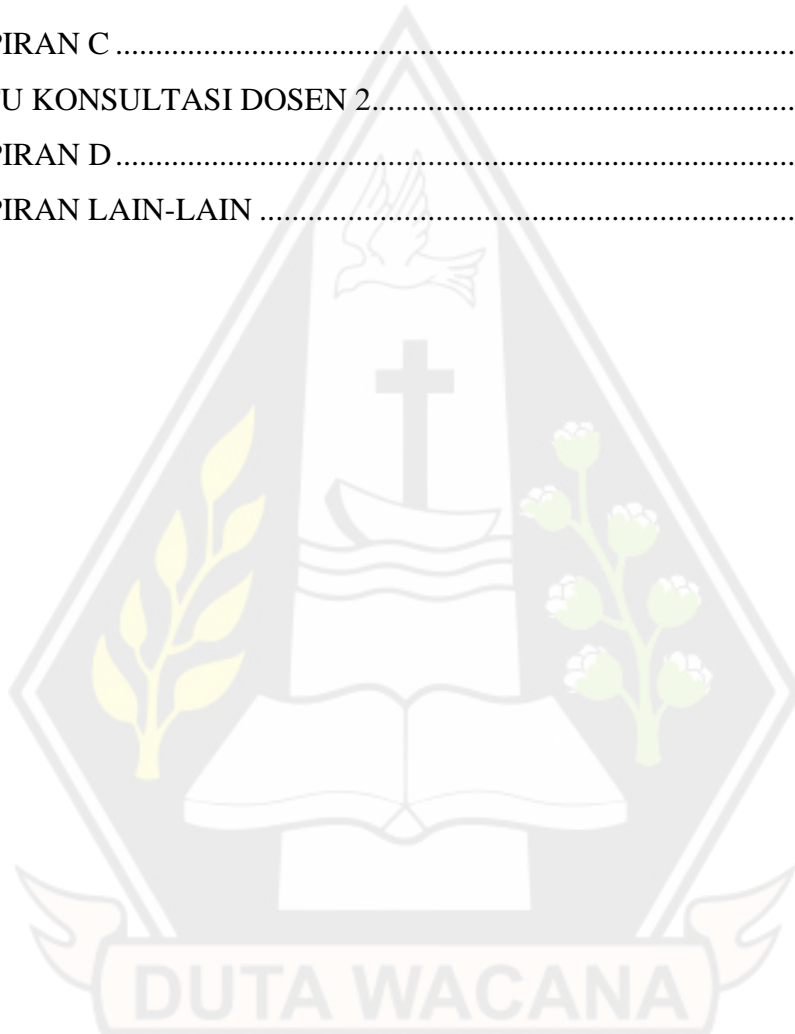
Gambar 2.1: Hasil text summarization pada teks berita.....	10
Gambar 3.1: Dataset.....	16
Gambar 3.2: Lanjutan dataset	16
Gambar 3.3 : Dataset CSV.....	17
Gambar 3.4 Website manajemen korpus.....	18
Gambar 3.5: Contoh pelabelan NLP.....	19
Gambar 3.6 : Contoh pelabelan pada label studio.....	19
Gambar 3.7: Output hasil pelabelan.....	19
Gambar 3.8: Original teks.....	21
Gambar 3.9 : Teks hasil summary sistem.....	22
Gambar 3.10 : Teks hasil summary sistem.....	22
Gambar 3.11 : Hasil summary responden.....	23
Gambar 3.12 : Diagram alir.....	23
Gambar 3.13: Sequence Diagram.....	24
Gambar 3.14 :Dataset.....	25
Gambar 3.15 : Lanjutan dataset.....	25
Gambar 3.16: Dataset csv.....	26
Gambar 3.17: Labeling json.....	26
Gambar 3.18 : Unique phrase.....	27
Gambar 3.19 : Tampilan awal masuk.....	27
Gambar 3.20 : Tampilan setelah menekan summarizer.....	28
Gambar 3.21: Tampilan original text.....	28
Gambar 3.22 : Hasil summary.....	29
Gambar 3.23 : Hasil evaluasi.....	29
Gambar 3.24 : Chart summary.....	30
Gambar 4.1: Fungsi text_to_string	32
Gambar 4.2 : Output text_to_string.....	32
Gambar 4.3 : List of token.....	33
Gambar 4.4 : Code list of stopwords.....	33
Gambar 4.5: Code base summarizer.....	34

Gambar 4.6: Code fungsi <code>_create_matrix</code>	34
Gambar 4.7: Code fungsi <code>'compute_term frequency</code>	34
Gambar 4.8: Code fungsi <code>_compute_ranks'</code>	35
Gambar 4.9: Code fungsi <code>lsa_summarizer</code>	35
Gambar 4.10: Output hasil pemrosesan Isa.....	35
Gambar 4.11: Code fungsi evaluasi.....	36
Gambar 4.12 : Output evaluasi ROUGE Gambar 4.12 : Code validasi.....	36
Gambar 4.13: Code fungsi <code>_compute_ranks</code> Gambar 4.8: Code fungsi <code>'Isa summarizer</code>	37



DAFTAR LAMPIRAN

LAMPIRAN A	46
KODE SUMBER PROGRAM	46
LAMPIRAN B	59
KARTU KONSULTASI DOSEN 1	59
LAMPIRAN C	60
KARTU KONSULTASI DOSEN 2	60
LAMPIRAN D	61
LAMPIRAN LAIN-LAIN	61



INTISARI

TEXT SUMMARIZATION UNTUK KORPUS DOKUMEN DENGAN METODE LATENT SEMANTIC ANALYSIS

Oleh

Stevani Dwi Utomo

71180258

Di masa modern ini kecepatan dalam memahami suatu dokumen dalam waktu singkat sangatlah diperlukan. Sehingga untuk mewujudkan hal tersebut dibutuhkan sebuah sistem yang digunakan untuk mengelola data teks. Salah satu sistem pengelolaan text yang cukup sering dipakai adalah text summarization. Text summarization merupakan metode peringkasan teks untuk mendapatkan inti dari keseluruhan isi dokumen. Pada penelitian ini peneliti menggunakan pendekatan metode latent semantic analysis untuk mendapatkan hasil text summarization dengan akurasi terbaik. Metode latent semantic analysis ini menggunakan pendekatan aljabar linear singular value decomposition (SVD) dan proses perhitungan TF-IDF untuk memecahkan permasalahan text summarization dokumen yang tersedia. Selain itu pada penelitian ini juga memperhatikan proses preprocessing yang dilakukan untuk mendukung hasil pemrosesan dengan metode LSA. Metode preprocessing yang ditambahkan adalah phrase detection dan labeling data. Setelah itu dilakukan pemrosesan LSA untuk menghasilkan output teks yang akan dievaluasi. Metode evaluasi yang digunakan adalah Evaluasi Kappa untuk melakukan penilaian responden terhadap output yang dihasilkan oleh pemrosesan LSA. Sehingga metode ini sangat efektif untuk menyelesaikan permasalahan text summarization.

Kata-kata kunci : Peringkasan teks, Pendeteksi Frase, Metode LSA

ABSTRACT

TEXT SUMMARIZATION FOR CORPUS DOCUMENT USING LATENT SEMANTIC ANALYSIS METHOD

By

Stevani Dwi Utomo

71180258

In this modern era, speed in understanding a document in a short time is needed. So to realize this, a system is needed that is used to manage text data. One text management system that is quite often used is text summarization. Text summarization is a text summarization method to get the gist of the whole document. In this study, the researcher used a latent semantic analysis approach to obtain a summary of the text with the best accuracy. This latent semantic analysis method uses a singular value decomposition (SVD) linear algebra approach and the TF-IDF calculation process to solve the problem of summarizing the available text. In addition, this study also pays attention to the preprocessing process carried out to support the processing results using the LSA method. The added preprocessing method is phrase detection and labeling data. After that, LSA processing is carried out to produce a promising output text. The evaluation method used is Kappa Evaluation to evaluate the outputs carried out by LSA. So this method is very effective to overcome the problem of text summarization.

Keywords : Text Summarization, Phrase Detection, LSA Method

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Dengan berkembangnya teknologi informasi dan komunikasi, data teks dalam format digital semakin meningkat, dan diperlukan manajemen untuk menyediakan data tersebut. Salah satu bidang ilmu dalam teknologi komputer yang dapat digunakan untuk mengelola data teks yaitu *text mining*. *Text mining* adalah proses mengeksplorasi dan menganalisis sejumlah besar data teks tidak terstruktur untuk mengidentifikasi konsep, pola, topik, kata kunci, dan atribut lainnya dalam data dengan bantuan perangkat lunak. (Arni, 2018) Salah satu pemanfaatan teks mining ini digunakan untuk meringkas teks pada sebuah korpus dokumen, sehingga peringkasan korpus dokumen ini tidak hanya dilakukan secara manual namun juga secara digital. Proses peringkasan text atau bisa disebut dengan *text summarization* ini merupakan pendekatan yang digunakan untuk meringkas atau memadatkan text dokumen namun tetap dapat mewakili isi dari dokumen tersebut (Zamzam et al., 2020).

Terdapat beberapa penelitian yang sudah dilakukan untuk melakukan text summarization. Rozi et al (2021) dan Gotami et al (2018) dalam penelitiannya menggunakan metode dan pendekatan yang sama yaitu Metode LSA pada umumnya mengaplikasikan aljabar *linear singular value decomposition* (SVD) dan proses perhitungan TF-IDF untuk memecahkan permasalahan text summarization namun pada jenis dokumen yang berbeda. Rozi et al (2021) menggunakan metode tersebut untuk peringkasan text pada dokumen hukum sedangkan Gotami et al (2018) menggunakannya untuk peringkasan teks pada artikel berita kesehatan. Penelitian Rozi et al (2021) mendapatkan hasil compression rate 50% yaitu 54%, 56%, 55% dan 75% serta Gotami et al (2018) memperoleh compression rate yang sama yaitu 50% dengan hasil 0.668, 0.743, 0.700 dan 0.690. Begitu pula dari penelitian yang dilakukan oleh Pratiwi untuk text Summarization pada teks hadits menggunakan metode LSA digabungkan dengan metode *Fuzzy Logic Scoring* menggunakan pendekatan SVD dan *Sentence Selection* dan menggunakan metode evaluasi ROUGE-1 untuk menentukan nilai akurasi. (Pratiwi, 2021)

Berdasarkan uraian penelitian terdahulu, terbukti bahwa metode *Latent Semantic Analysis* dapat diterapkan pada permasalahan *text summarization*. Pada penelitian ini, penulis juga mengimplementasikan metode *Latent Semantic Analysis* (LSA) untuk korpus dokumen bahasa Indonesia. Dilihat dari penelitian terdahulu yang dilakukan oleh Rozi et al (2021) dan Gotami et al (2018) hanya melakukan preprocessing yang umum dilakukan pada sebuah teks sehingga tidak memperhatikan labeling otomatis pada frasa jika menemui nama orang atau instansi yang penulisannya tidak dapat dipisahkan. Pada penelitian ini diharapkan satuan pada dokumen berupa statement dan dapat membedakan frasa. Metode LSA yang ditentukan oleh penulis menggunakan NLTK Indonesia untuk *preprocessing* dokumen dilanjutkan dengan pembuatan matriks yang nantinya diterapkan pada pendekatan SVD. Proses tersebut akan menghasilkan hasil teks ringkasan yang baik dan dapat mewakili keseluruhan isi dokumen. Untuk membuktikan apakah hasil yang memperhatikan labeling frasa ini lebih baik, maka akan dilakukan evaluasi kappa untuk membandingkan hasil *text summarization* yang memperhatikan frasa dengan hasil *text summarization* yang tidak memperhatikan frasa.

1.2. Perumusan Masalah

Rumusan masalah pada penelitian ini berdasarkan latar belakang yang sudah ditulis sebelumnya adalah sebagai berikut:

1. Bagaimana menerapkan *phrase detection* untuk *preprocessing* dokumen korpus yang sebelumnya belum memperhatikan *labeling* frasa?
2. Bagaimana performa LSA untuk melakukan *text summarization* didasarkan dengan *preprocessing phrase detection* yang dibandingkan dengan yang tidak menggunakan *phrase detection*?

1.3. Batasan Masalah

Dengan dilaksanakannya penelitian ini, maka perlu ditentukan batasan masalah sebagai berikut :

1. Sumber data diperoleh dari data teks yang sudah dikumpulkan untuk pembangunan website alun-alun (<https://app.alunalun.info/>)
2. Menggunakan nltk indonesia untuk *preprocessing* agar menghasilkan data yang siap diolah (<https://www.nltk.org/>)
3. Mengimplementasi *Library* LSA yang sudah disediakan dengan menambahkan *phrase detection*.
4. Hanya menggunakan 3 kategori korpus dokumen yang sudah ditentukan yaitu wayang, tarian tradisional dan dalang

1.4. Tujuan Penelitian

Tujuan dari penelitian ini adalah menerapkan teks *summarization* pada sebuah korpus dokumen dengan menggunakan metode *Latent Semantic Analysis* untuk meningkatkan akurasi dari *teks summarization* dengan memperhatikan *phrase detection*.

1.5. Manfaat Penelitian

Manfaat penelitian ini untuk kedepannya adalah dapat menghasilkan sistem yang dapat menghasilkan teks pendek atau ringkasan dari sebuah dokumen panjang yang kedepannya juga dapat diterapkan pada kategori lainnya dalam korpus dokumen website alun-alun UKDW.

Manfaat bagi user website Alun-alun FTI UKDW :

1. Dapat memahami isi tanpa perlu membaca keseluruhan teks.
2. Mempermudah user memilih dokumen yang diperlukan.

1.6. Sistematika Penulisan

Laporan penelitian ini terdiri dari lima bab, yaitu Bab 1, dan berisi tentang latar belakang, uraian masalah, batasan masalah, tujuan penelitian, manfaat penelitian, pendekatan penelitian dan kerangka penelitian, serta ringkasan penelitian yang harus penulis lengkapi dan manfaat bagi semua pihak yang terlibat dalam penelitian ini. Lalu dilanjutkan ke bab 2 yang berisi mengenai tinjauan pustaka dan landasan teori. Dimana pada tinjauan pustaka menjelaskan beberapa

penelitian serupa yang pernah dilakukan sebagai referensi penulis dalam melakukan penelitian ini. Sedangkan untuk landasan teori berisi teori dan definisi mengenai metode yang dipakai untuk membantu menyelesaikan penelitian ini. Bab selanjutnya adalah bab 3, bab ini berisi tentang langkah-langkah yang dilakukan mulai dari preprocessing, process hingga evaluasi yang akan diimplementasikan oleh penulis. Berikutnya bab 4 yang berisi mengenai implementasi dan analisis yang telah dibuat berdasarkan metode yang telah dijelaskan di bab sebelumnya. Bagian terakhir, Bab 5 yang berisi kesimpulan hasil yang telah didapatkan dari implementasi metode serta diberikan saran dan masukan untuk dapat mengembangkan penelitian ini.



BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Dari hasil penelitian ini didapatkan kesimpulan bahwa:

1. Penulis menggunakan 60 artikel pada 3 kategori yang ada yaitu dalang, tari tradisional dan wayang yang telah diproses, dievaluasi dan divalidasi menunjukkan bahwa rata-rata hasil evaluasi pada kategori wayang jauh lebih baik dari kategori yang lain karena teks wayang berisi cerita pendek pewayangan yang lebih umum untuk dilakukan peringkasan dibandingkan kategori yang lain.
2. Dari proses perbandingan penggunaan metric antara occurrence, tf dan tf-idf untuk setiap kategori dengan metode evaluasi ROUGE pada LSA-Phrase memiliki nilai yang relatif kecil karena panjang kalimat yang diproses setiap artikel hanya sedikit sehingga metode LSA-Phrase tidak dapat bekerja secara optimal .
3. Pada penelitian ini kurang efektif melakukan pemrosesan teks dengan memperhatikan frasa karena jumlah data yang sedikit tidak dapat bekerja secara optimal pada semantic.

5.2 Saran

Setelah dilakukan penelitian ini, terdapat beberapa hal yang dapat dikembangkan lagi kedepannya antara lain:

1. Pelabelan unique data kedepannya dapat menggunakan machine learning supaya tidak dilakukan manual.
2. Bisa mencari 1 responden yang memang ahli dalam meringkas agar hasilnya dapat lebih baik lagi.
3. Kategori yang digunakan dapat lebih banyak lagi agar lebih beragam serta jumlah data dari setiap kategorinya dapat lebih banyak lagi.

Daftar Pustaka

- Alpasian, F. N., & Cicekli, I. (2011). Text summarization using Latent Semantic Analysis. *Journal of Information Science*, 4(37), 405-417. DOI: 10.1177/0165551511408848
- Arni, U. D. (2018, Oktober 25). *Apa itu Text Mining*. Garuda Cyber Indonesia. <https://news.garudacyber.co.id/Capitalization/case-folding>. (2009, July 4). Stanford NLP Group. Retrieved December 16, 2021, from <https://nlp.stanford.edu/IR-book/html/htmledition/capitalizationcase-folding-1.html>
- Dewi, C. K. (2021, January 6). *Ciri dan Jenis Frasa - Materi Bahasa Indonesia Kelas 10*. Zenius Education. Retrieved December 16, 2021, from <https://www.zenius.net/blog/materi-bahasa-jenis-jenis-frasa>
- Gaikwad, D. K., & Mahender, C. (2016). A Review Paper on Text Summarization.
- Gotami, N. S. W., Indrianti, & Dewi, R. K. (2018). Peringkasan Teks Otomatis Secara Ekstraktif Pada Artikel Berita Kesehatan Berbahasa Indonesia Dengan Menggunakan Metode Latent Semantic Analysis. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komunikasi*, 2(9), 2821-2828. <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/2430>
- Grubnyak, A. (2019, Desember 18). *Multi-Class Metrics Made Simple, Part III: the Kappa Score (aka Cohen's Kappa Coefficient)*. Towards Data Science. <https://towardsdatascience.com/multi-class-metrics-made-simple-the-kappa-score-aka-cohens-kappa-coefficient-bdea137af09c>
- Kanani, B. (2020, March 12). *Text Preprocessing: Removal of Punctuations*. Machine Learning Tutorials. Retrieved December 16, 2021, from <https://studymachinelearning.com/text-preprocessing-removal-of-punctuations/>
- Kamath, V. (2014, May). The DF-ICF Algorithm- Modified TF-IDF. *International Journal of Computer Applications*, 93((13)), 28-30.
- Khanna, C. (2021, February 10). *Text preprocessing: Stop words removal / Chetna*. Towards Data Science. Retrieved December 16, 2021, from

- <https://towardsdatascience.com/text-pre-processing-stop-words-removal-using-different-libraries-f20bac19929a>
- Kramov, A. (2020, October 22). *Method of noun phrase detection in Ukrainian texts*. arXiv. <https://arxiv.org/abs/2010.11548>
- Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to information retrieval* (H. Schütze & P. Raghavan, Eds.). Cambridge University Press.
- Pratiwi, M. W. (2021, maret 22). *Text Summarization terjemahan hadits Berbahasa Indonesia menggunakan penggabungan Metode Fuzzy Logic Scoring dan Latent Semantic Analysis(LSA)*. Digital Library UIN Sunan Kalijaga Bandung. https://digilib.uinsgd.ac.id/37977/4/4_bab1.pdf
- Putranto, H. A., Setyawati, O., & Wijono. (2016). Pengaruh Phrase Detection dengan POS-Tagger terhadap Akurasi Klasifikasi Sentimen menggunakan SVM. *JNTETI (Jurnal Nasional Teknik Elektro dan Teknologi Informasi)*, 5(4), 252-259. <http://dx.doi.org/10.22146/jnteti.v5i4.271>
- Rozi, I. F., Batubulan, K. S., & Rusbandi, M. (2021). Otomatisasi Peringkasan Teks Pada Dokumen Hukum Menggunakan Metode Latent Semantic Analysis. *Jurnal Informatika Polinema*, 7(3). <https://doi.org/10.33795/jip.v7i3.515>
- Singh, S. (2019, July 18). *What is Tokenization | Methods to Perform Tokenization*. Analytics Vidhya. Retrieved December 16, 2021, from <https://www.analyticsvidhya.com/blog/2019/07/how-get-started-nlp-6-unique-ways-perform-tokenization/>
- Suhartono, D. (2015, Agustus 3). *Penggunaan Latent Semantic Analysis (LSA) dalam Pemrosesan Teks*. socs.binus.ac.id. <https://socs.binus.ac.id/2015/08/03/penggunaan-latent-semantic-analysis-lsa-dalam-pemrosesan-teks/>
- What is data labeling?* (2019, Febuari 28). Amazon AWS. Retrieved December 16, 2021, from <https://aws.amazon.com/sagemaker/data-labeling/what-is-data-labeling/>
- Widmann, M. (2016). *Kappa Cohen: Apa itu, kapan menggunakannya, dan bagaimana menghindari jebakannya*. ichi.pro. Retrieved December 16, 2021, from <https://ichi.pro/id/kappa-cohen-apa-itu-kapan->

menggunakannya-dan-bagaimana-menghindari-jebakannya-
250844470979516

- Winanta, F., & Rainarli, E. (2016). IMPLEMENTASI CROSS METHOD LATENT SEMANTIC ANALYSIS UNTUK MERINGKAS DOKUMEN BERITA BERBAHASA INDONESIA. *Techno.COM*, 15(4), 266-277.
- Yulyardo, Rahardian, O. P., & Sujono, M. (2018, Desember 26). *Peringkas Teks Otomatis*. mti.binus.ac.id. <https://mti.binus.ac.id/2018/12/26/peringkas-teks-otomatis-automatic-text-summarization/>
- Yunus, M. (2020, April 26). *Basic Text Preprocessing menggunakan NLTK*. medium.com. <https://yunusmuhammad007.medium.com/basic-text-preprocessing-menggunakan-nltk-86ba3e65a1dc>
- Zamzam, M. A., Crysdiان, C., & Holle, K. F. H. (2020). Sistem Automatic Text Summarization Menggunakan Algoritma Textrank. *Matics Jurnal Komputer dan Teknologi Informasi*, 12(2), 111-116. DOI: 10.18860/mat.v12i2.8372

