

**PENENTUAN TOPIK UTAMA TEKS BAHASA
INDONESIA MENGGUNAKAN *COSINE SIMILARITY*
DAN *POS TAGGER***

Diajukan Sebagai Syarat untuk Menyelesaikan Pendidikan
Program Studi Strata-1 pada Jurusan Teknik Informatika
Fakultas Ilmu Komputer Universitas Sriwijaya



Oleh :

AREGA GERI FERNANDO
NIM : 09021281320003

**Jurusan Teknik Informatika
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA
2021**

LEMBAR PENGESAHAN SKRIPSI

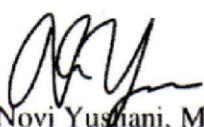
PENENTUAN TOPIK UTAMA TEKS BAHASA INDONESIA MENGGUNAKAN COSINE SIMILARITY DAN POS TAGGER

OLEH:

AREGA GERI FERNANDO

09021281320003

Pembimbing I,


Novi Yushani, M.T
NIP. 198211082012122001

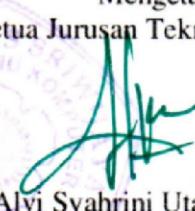
Indralaya, 22 Januari 2021

Pembimbing II,


Desty Kodiah, M.T
NIP. 1671016112890005

Mengetahui,

Ketua Jurusan Teknik Informatika


Alvi Syahrini Utami, M.Kom.
NIP. 197812222006042003

TANDA LULUS UJIAN SIDANG TUGAS AKHIR

Pada hari rabu, tanggal 30 Desember 2020 telah dilaksanakan Ujian Sidang Tugas Akhir oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Arega Geri Fernando

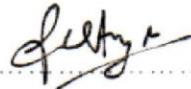
NIM : 09021281320003

Judul : Penentuan Topik Utama Teks Bahasa Indonesia Menggunakan *Cosine Similarity* dan *POS Tagger*

1. Pembimbing I
Novi Yusliani, M.T
NIP. 198211082012122001



2. Pembimbing II
Desty Rodiah, M.T
NIP. 1671016112890005



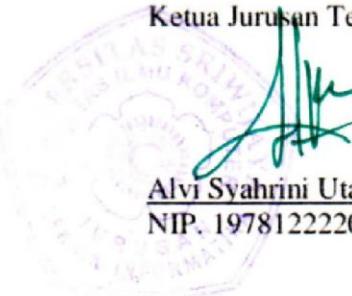
3. Penguji I
Samsuryadi, M.Kom., Ph.D
NIP. 197102041997021003



4. Penguji II
M Qurhanul Rizqie, S.Kom, M.T
NIP. 1671060312870008



Mengetahui,
Ketua Jurusan Teknik Informatika



Alvi Syahrini Utami, M.Kom.
NIP. 197812222006042003

HALAMAN PERNYATAAN

Yang bertanda tangan dibawah ini:

Nama : Arega Geri Fernando

NIM : 09021281320003

Jurusan : Teknik Informatika

Judul Skripsi : Penentuan Topik Utama Teks Bahasa Indonesia Menggunakan
Cosine Similarity dan *POS Tagger*

Hasil pengecekan *Software iThenticate/Turnitine* : 10%

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

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



Indralaya, 28 Januari 2021



Arega Geri Fernando

NIM. 09021281320003

MOTTO DAN PERSEMPAHAN

- Salah satu kunci bahagia dalam hidup ini adalah Bersyukur.
- Ketika kamu merasa lelah dalam menjalani hidup ini, ingatlah kedua orang tua mu yang lebih lelah dalam menghidupimu.

Kupersembahkan Hasil Karyaku ini Kepada :

- ❖ Allah SWT
- ❖ Kedua Orang Tuaku (Ibunda Sumiati dan Ayahanda Amanul Ihwan) yang Tidak Pernah Berhenti Memberikan Dukungan dan Doa untuk Anaknya
- ❖ Saudara Saudari ku tercinta (Silvia Susanti, Vonny Ayu Elintiya dan M Fadhil Arya Jaya)
- ❖ Keluarga dan Kerabat yang Selalu Mendukungku
- ❖ IF Reguler 2013
- ❖ Almamaterku

Determination of Main Topic of Indonesian Text Using *Cosine Similarity* and *Pos Tagger*

Arega Geri Fernando (09021281320003)

Informatics Engineering Department, Faculty of Computer Science

Sriwijaya University

ABSTRACT

Information is a very important thing to add insight and gain knowledge. Hence, there are many ways to obtain information. One of them is by reading but it takes a long time to determine the type of information in the text. Therefore, it takes a way to find out the type of information quickly, one of which is by building a system that can recognize the main topic of Indonesian text. This study applied *Cosine Similarity* and *POS Tagger*. *Cosine similarity* was used to compare the similarity of the text title to the content of each sentence and *POS Tagger* was used to recognize the types of words and noun phrases that would be recognized as the main topic. Besides, the data used in this study were 50 abstract texts of Indonesian scientific papers. The software to determine the main topic using *Cosine Similarity* and *POS Tagger* was successfully performed with an accuracy of 28%.

Keywords: *Topic Determination, Cosine Similarity, POS Tagger, Phrase Chunking*

**Penentuan Topik Utama Teks Bahasa Indonesia Menggunakan *Cosine Similarity*
dan *Pos Tagger***

Arega Geri Fernando (09021281320003)

Jurusan Teknik Informatika, Fakultas Ilmu Komputer

Universitas Sriwijaya

ABSTRAK

Informasi adalah suatu hal yang sangat penting sekali untuk menambah wawasan dan pengetahuan, ada banyak cara untuk mendapatkan informasi, salah satunya ialah dengan membaca, namun akan memakan waktu yang lama untuk menentukan jenis informasi dalam teks. Diperlukan cara untuk mengetahui jenis informasi dengan cepat, salah satunya dengan membangun sistem yang dapat mengenali topik utama dari sebuah teks Bahasa Indonesia. Penelitian ini menggunakan *cosine similarity* dan *POS Tagger*. *Cosine similarity* dipakai untuk membandingkan kemiripan judul teks dengan isi teks perkalamat, dan *POS Tagger* dipakai untuk mengenali jenis kata dan frasa benda yang kemudian akan dikenali sebagai topik utama. Data yang digunakan pada penelitian ini adalah 50 teks abstrak karya ilmiah Bahasa Indonesia. Perangkat lunak penentuan topik utama menggunakan *cosine similarity* dan *POS Tagger* berhasil dilakukan dengan akurasi sebesar 28%.

Kata kunci : Penentuan topik, *cosine similarity*, *POS Tagger*, *Phrase Chunking*

KATA PENGANTAR

Segala puji dan syukur saya persembahkan kepada Allah SWT atas berkah dan rahmat-Nya yang telah diberikan kepada Penulis sehingga dapat menyelesaikan Tugas Akhir ini dengan baik. Tugas Akhir yang berjudul “Penentuan Topik Utama Teks Bahasa Indonesia Menggunakan *Cosine Similarity* dan *POS Tagger*” disusun untuk memenuhi salah satu syarat dalam menyelesaikan pendidikan program Strata-1 pada Fakultas Ilmu Komputer Jurusan Teknik Informatika di Universitas Sriwijaya.

Pada kesempatan ini Penulis ingin menyampaikan ucapan terima kasih kepada pihak-pihak yang telah memberikan dukungan, bimbingan dan motivasi selama proses penelitian ini dilaksanakan. Secara khusus Penulis ingin mengucapkan terima kasih kepada :

1. Allah SWT atas segala berkah dan rahmat-Nya.
2. Keluarga tercinta, Ibu Sumiati dan Ayah Amanul Ihwan yang selalu menyemangati, memberikan dukungan moral dan materil, serta selalu menyebut namaku didalam setiap doanya. Saudaraku, Silvia Susanti, Vonny Ayu Elintiya dan M Fadhil Arya Jaya, serta keluarga besar beserta kerabat yang juga selalu memberikan semangat dan dukungan.
3. Bapak Jaidan Jauhari, M.T., selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya.
4. Ibu Alvi Syahrini Utami, M.Kom., selaku Ketua Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.
5. Ibu Novi Yusliani, M.T., selaku Dosen Pembimbing I dan Ibu Desty Rodiah, M.T., selaku Dosen Pembimbing II yang telah membimbing, mengarahkan dan memotivasi saya untuk menyelesaikan Tugas Akhir ini.
6. Bapak Samsuryadi, M.Kom., Ph.D., Bapak Muhammad Ali Buchari, S.Kom., M.T., dan Bapak M Qurhanul Rizqie, S.Kom, M.T selaku penguji Tugas Akhir yang telah memberikan nasihat dan saran yang membangun.

7. Bapak Ali Ibrahim, M.T. Selaku dosen pembimbing akademik yang selalu memberikan saran dan arahan selama masa perkuliahan.
8. Seluruh Dosen Jurusan Teknik Informatika nyang telah memberikan bekal ilmu selama masa perkuliahan.
9. Teman seperjuangan Teknik Informatika Angkatan 2013.
10. Seluruh Staf Administrasi dan Pegawai yang telah membantu dalam urusan administrasi.

Penulis menyadari masih banyak terdapat kekurangan dalam penyusunan Tugas Akhir yang disebabkan oleh keterbatasan pengetahuan dan pengalaman. Oleh karena itu, kritik dan saran membangun sangat diharapkan untuk menyempurnakan Tugas Akhir ini. Semoga Tugas Akhir ini dapat bermanfaat bagi kita semua.

Indralaya, 28 Januari 2021

Arega Geri Fernando

DAFTAR ISI

	Halaman
HALAMAN JUDUL	i
LEMBAR PENGESAHAN SKRIPSI	ii
TANDA LULUS UJIAN SIDANG TUGAS AKHIR	iii
HALAMAN PERNYATAAN	iv
MOTTO DAN PERSEMBAHAN	v
ABSTRACT	vi
ABSTRAK	vii
KATA PENGANTAR	viii
DAFTAR ISI	x
DAFTAR GAMBAR	xv
DAFTAR TABEL	xvi
DAFTAR LAMPIRAN.....	xviii
 BAB I PENDAHULUAN	
1.1 Pendahuluan	I-1
1.2 Latar Belakang	I-1
1.3 Rumusan Masalah	I-2
1.4 Tujuan Penelitian	I-3
1.5 Manfaat Penelitian	I-3
1.6 Batasan Masalah	I-3
1.7 Sistematika Penulisan.....	I-4
1.8 Kesimpulan.....	I-6

BAB II TINJAUAN PUSTAKA

2.1 Pendahuluan	II-1
2.2 Penelitian Terkait	II-1
2.3 Pemrosesan Bahasa Alami.....	II-3
2.3.1 <i>Preprocessing</i>	II-4
2.3.1.1 <i>Case Folding</i>	II-4
2.3.1.2 <i>Tokenizing</i>	II-5
2.3.1.3 <i>Stopword Removal</i>	II-5
2.3.2 Pengambilan Kalimat.....	II-6
2.3.2.1 <i>Term Frequency</i>	II-7
2.3.2.2 <i>Cosine Similarity</i>	II-7
2.3.3 Pelabelan Frasa.....	II-8
2.3.3.1 <i>Pos Tagging</i>	II-8
2.3.3.2 <i>Phrase Chunking</i>	II-12
2.3.4 Frekuensi Frasa	II-13
2.4 Metode Pengembangan Perangkat Lunak	II-14
2.4.1 <i>Rational Unified Process (RUP)</i>	II-14
2.5 Kesimpulan.....	II-17

BAB III METODOLOGI PENELITIAN

3.1 Pendahuluan	III-1
3.2 Unit Penelitian	III-1
3.3 Pengumpulan Data	III-1
3.3.1 Jenis Data.....	III-1
3.3.2 Sumber Data.....	III-2
3.3.3 Metode Pengumpulan Data	III-2
3.4 Langkah-Langkah Penelitian	III-2

3.4.1 Mengkaji Konsep Mengenai <i>Cosine Similarity</i> dan <i>POS Tagger</i>	III-3
3.4.2 Mengumpulkan Data Teks Abstrak.....	III-3
3.4.3 Melakukan Pengembangan Perangkat Lunak Penentuan Topik Utama Menggunakan Metode RUP (<i>Rational Unified Process</i>)	III-3
3.4.4 Menentukan Kerangka Kerja Perangkat Lunak	III-4
3.4.5 Menetapkan Kriteria Pengujian	III-6
3.4.6 Menetapkan Format Data Pengujian	III-7
3.4.7 Menentukan Alat Yang Digunakan Dalam Pelaksanaan Penelitian.....	III-7
3.4.8 Melakukan Pengujian Penelitian.....	III-8
3.4.9 Melakukan Analisa Hasil Pengujian dan Membuat Kesimpulan	III-8
3.5 Manajemen Proyek Penelitian.....	III-8

BAB IV PENGEMBANGAN PERANGKAT LUNAK

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
4.2.3.1 Analisis Perangkat Lunak	IV-4
4.2.3.1.1 Analisis Kebutuhan Perangkat Lunak	IV-4
4.2.3.1.2 Analisis Data	IV-5
4.2.3.1.2.1 Analisis Teks Masukan	IV-6
4.2.3.1.3 Analisis <i>Preprocessing</i>	IV-7
4.2.3.1.4 Analisis <i>Term Frequency</i>	IV-11
4.2.3.1.5 Analisis <i>Cosine Similarity</i>	IV-13

4.2.3.1.6 Analisis <i>POS Tagger</i>	IV-15
4.2.3.1.7 Analisis <i>Phrase Chunking</i>	IV-16
4.2.3.1.8 Analisis Perhitungan Frekuensi Frasa	IV-17
4.2.3.2 Desain Perangkat Lunak.....	IV-17
4.3 Fase Elaborasi.....	IV-25
4.3.1 Pemodelan Bisnis	IV-25
4.3.1.1 Perancangan Data	IV-25
4.3.1.2 Perancangan Antar Muka	IV-25
4.3.2 Kebutuhan Sistem.....	IV-27
4.3.3 Diagram Sequence.....	IV-27
4.4 Fase Kontruksi.....	IV-31
4.4.1 Kebutuhan Sistem.....	IV-31
4.4.2 Diagram Kelas.....	IV-31
4.4.3 Implementasi	IV-32
4.4.3.1 Implementasi Kelas.....	IV-32
4.4.3.2 Implementasi Antarmuka	IV-34
4.5 Fase transisi	IV-36
4.5.1 Pemodelan Bisnis	IV-36
4.5.2 Kebutuhan Sistem.....	IV-37
4.5.3 Rencana Pengujian	IV-37
4.5.3.1 Rencana Pengujian Use Case Memuat File	IV-37
4.5.3.2 Rencana Pengujian Use Case Menentukan Topik Utama	IV-38
4.5.4 Implementasi	IV-38
4.5.4.1 Pengujian Use Case Memuat File	IV-39
4.5.4.2 Pengujian Use Case Menentukan Topik Utama	IV-40
4.6 Kesimpulan.....	IV-41

BAB V ANALISIS PENELITIAN

5.1 Pendahuluan	V-1
5.2 Hasil Penelitian.....	V-1
5.2.1 Analisis Hasil Pengujian Metode 1	V-1
5.2.2 Analisis Hasil Pengujian Metode 2	V-3
5.3 Hasil Pengujian.....	V-4
5.4 Kesimpulan.....	V-4

BAB VI KESIMPULAN DAN SARAN

6.1 Pendahuluan	VI-1
6.2 Kesimpulan.....	VI-1
6.3 Saran	VI-2

DAFTAR PUSTAKA.....	xix
----------------------------	------------

DAFTAR GAMBAR

	Halaman
Gambar II-1. <i>Case Folding</i>	II-2
Gambar II-2. <i>Tokenizing</i>	II-5
Gambar II-3. Proses <i>Rational Unified Proses</i>	II-15
Gambar III-1. Kerangka Kerja Perangkat Lunak.....	III-4
Gambar IV-1 Diagram <i>Use Case</i>	IV-16
Gambar IV-2 <i>Analysis Class</i> Memuat File.....	IV-20
Gambar IV-3 <i>Analysis Class</i> Menentukan Topik Utama	IV-21
Gambar IV-4 Diagram Aktivitas Memuat File.....	IV-22
Gambar IV-5 Diagram Aktivitas Menentukan Topik Utama	IV-23
Gambar IV-6 Perancangan Antarmuka	IV-25
Gambar IV-7 Sequence Diagram Memuat File	IV-28
Gambar IV-8 Sequence Diagram Menentukan Topik Utama	IV-29
Gambar IV-9 Diagram Kelas	IV-30
Gambar IV-10 Antarmuka Halaman Menu Utama	IV-33
Gambar IV-11 Antarmuka Halaman Output Preview	IV-33

DAFTAR TABEL

	Halaman
TABEL II-1. Kelas Kata Bahasa Indonesia	II-8
TABEL II-2. Proses Generalisasi Kelas Kata.....	II-10
TABEL II-3. Frasa Dalam Bahasa Indonesia	II-12
TABEL III-1. Penjadwalan Penelitian dalam Bentuk <i>Work Breakdown Structure</i> (WBS)	III-4
TABEL IV-1. Kebutuhan Fungsional.....	IV-3
TABEL IV-2. Kebutuhan Non-Fungsional	IV-3
TABEL IV-3. Teks Masukan.....	IV-6
TABEL IV-4. Tahap <i>Case Folding</i> Judul Dokumen	IV-7
TABEL IV-5. Tahap <i>Case Folding</i> Isi Dokumen.....	IV-7
TABEL IV-6. Tahap Tokenizing Judul Dokumen.....	IV-8
TABEL IV-7. Tahap Tokenizing Isi Dokumen	IV-8
TABEL IV-8. Tahap <i>Stopword Removal</i> Judul Dokumen	IV-9
TABEL IV-9. Tahap <i>Stopword Removal</i> Isi Dokumen.....	IV-9
TABEL IV-10. Tahap <i>Term Frequency</i> Judul dan Isi Dokumen	IV-10
TABEL IV-11. Tahap <i>POS Tagger</i> K1 dan K3.....	IV-13
TABEL IV-12. Menunjukkan Hasil Dari <i>Phrase Chunking</i> Kalimat 1 dan Kalimat 2.....	IV-14
TABEL IV-13. Tabel Frasa Benda	IV-15

TABEL IV-14. Tabel Perhitungan Frekuensi Frasa Benda	IV-15
TABEL IV-15. Definisi Aktor	IV-17
TABEL IV-16. Definisi <i>Use Case</i>	IV-17
TABEL IV-17. Skenario <i>Use Case</i> Memuat File	IV-18
TABEL IV-18. Skenario <i>Use Case</i> Menentukan Topik Utama	IV-19
TABEL IV-19. Implementasi Kelas	IV-30
TABEL IV-20. Rencana Pengujian Use Case Memuat File.....	IV-34
TABEL IV-21. Rencana Pengujian Use Case Menentukan Topik Utama.....	IV-35
TABEL IV-22. Pengujian <i>Use Case</i> Memuat File.....	IV-36
TABEL IV-23. Pengujian <i>Use Case</i> Menentukan Topik utama	IV-37

DAFTAR LAMPIRAN

Halaman

Hasil Coding L1

BAB I

PENDAHULUAN

1.1 Pendahuluan

Bab ini menjelaskan secara detail tentang latar belakang, perumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, dan sistematika penulisan dokumen penelitian.

1.2 Latar Belakang Masalah

Informasi adalah suatu hal yang sangat penting sekali untuk menambah wawasan dan pengetahuan, ada banyak cara untuk mendapatkan informasi, salah satunya ialah dengan membaca, namun akan memakan waktu yang lama untuk menentukan jenis informasi dalam teks. Diperlukan cara untuk mengetahui jenis informasi dengan cepat, salah satunya dengan membangun sistem yang dapat mengenali topik utama dari sebuah teks Bahasa Indonesia.

Penggunaan pemrosesan bahasa alami dalam membangun sebuah perangkat lunak untuk menentukan topik utama teks bahasa Indonesia sudah pernah dilakukan (Pratama, 2016), penelitian telah dilakukan dengan menggunakan metode *Rule Reduction*. Penelitian tersebut berhasil dilakukan dengan rata-rata nilai akurasi sebesar 68% untuk 50 data teks dengan dua metode pengujian. Pada penelitian ini, penelitian dirasa belum maksimal karena hasil yang didapat tidak memiliki hubungan dengan judul dari teks dokumen yang diolah.

Salah satu metode yang dapat digunakan untuk mengukur kemiripan dokumen ialah *cosine similarity*. Metode ini pernah dilakukan oleh (Ariantini, Lumenta, & Jacobus, 2016) untuk mengukur kemiripan dokumen teks bahasa Indonesia, hasilnya pada sistem deteksi kemiripan dokumen sudah berhasil diterapkan dan sudah bisa digunakan untuk mengukur tingkat kemiripan dokumen. Dan *POS Tagger* pernah dilakukan untuk mengenali frasa setiap kata didalam suatu dokumen, sebelumnya pernah dilakukan penelitian oleh (Wicaksono & Purwarianti, 2010) dengan judul “*HMM Based Part-of-Speech Tagger for Bahasa Indonesia*”. Tingkat akurasi dari hasil penelitian ini sebesar 99,4% dalam mengenali jenis kata bahasa Indonesia dengan menggunakan metode *Hidden Markov Model*.

Oleh karena itu, berdasarkan penjelasan diatas, pada penelitian ini, peneliti akan mengembangkan penelitian sebelumnya tentang menentukan topik utama teks Bahasa Indonesia dengan menggunakan *cosine similarity* dan *POS Tagger*.

1.3 Rumusan Masalah

Rumusan masalah pada penelitian ini adalah bagaimana mengembangkan perangkat lunak untuk menentukan topik utama teks bahasa Indonesia. Dalam penelitian ini *cosine similarity* dipakai untuk membandingkan kemiripan judul teks dengan isi teks perkalimat, dan *POS Tagger* dipakai untuk mengenali frasa benda yang kemudian akan dikenali sebagai topik utama.

Untuk menyelesaikan permasalahan yang disebutkan di atas, maka dalam penelitian ini dibatasi dengan beberapa pertanyaan penelitian (*research question*) antara lain :

1. Bagaimana mengembangkan perangkat lunak untuk penentuan topik utama menggunakan *cosine similarity* dan *POS Tagger*.
2. Bagaimana tingkat akurasi *cosine similarity* dan *POS Tagger* untuk penentuan topik utama.

1.4 Tujuan Penelitian

Tujuan penelitian ini adalah sebagai berikut :

1. Membuat perangkat lunak untuk penentuan topik utama menggunakan *cosine similarity* dan *POS Tagger*.
2. Mengetahui tingkat akurasi dari penggunaan *cosine similarity* dan *POS Tagger* untuk menentukan topik utama.

1.5 Manfaat Penelitian

Manfaat penelitian ini adalah :

1. Memahami penggunaan *cosine similarity* dan *POS Tagger* untuk penentuan topik utama.
2. Sebagai alat bantu untuk mengetahui topik utama teks Bahasa Indonesia.

1.6 Batasan Masalah

Penelitian ini menggunakan beberapa batasan masalah di antaranya :

1. Data yang digunakan diperoleh dari perpustakaan Fakultas Ilmu Komputer Universitas Sriwijaya dan internet.
2. Jenis data yang digunakan dalam penelitian ini adalah teks abstrak karya ilmiah dalam Bahasa Indonesia.
3. Jenis data yang digunakan adalah dokumen teks dengan format (.txt).
4. Jenis data yang digunakan adalah 50 teks abstrak karya ilmiah dalam Bahasa Indonesia dengan panjang kata berada di rentang 100-200 kata.

1.7 Sistematika Penulisan

Sistematika penulisan tugas akhir ini mengikuti standar penulisan tugas akhir Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya, antara lain :

BAB I. PENDAHULUAN

Pada bab ini diuraikan mengenai latar belakang, perumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, dan sistematika penulisan.

BAB II. TINJAUAN PUSTAKA

Pada bab ini akan dibahas dasar-dasar teori yang digunakan dalam penelitian, seperti *cosine similarity* , *POS Tagger* dan penelitian lain yang relevan dengan penelitian yang sedang dikembangkan.

BAB III. METODOLOGI PENELITIAN

Pada bab ini akan dibahas mengenai tahapan yang akan dilaksanakan pada penelitian ini. Masing-masing rencana tahapan penelitian dideskripsikan dengan rinci dengan mengacu pada suatu kerangka kerja. Di akhir bab ini berisi perancangan manajemen proyek pada pelaksanaan penelitian.

BAB IV. PENGEMBANGAN PERANGKAT LUNAK

Pada bab ini akan dibahas mengenai perancangan dan implementasi perangkat lunak dengan metode pemrograman berorientasi objek berdasarkan panduan *rational unified process* yang di dalamnya terdapat fase insepsi, elaborasi, konstruksi, dan transisi.

BAB V. HASIL DAN ANALISIS PENELITIAN

Pada bab ini akan dibahas mengenai hasil dari pengembangan perangkat lunak dengan menggunakan *cosine similarity* dan *POS Tagger*. Pada akhir bab ini berisi analisis dari hasil yang telah didapatkan.

BAB VI. KESIMPULAN DAN SARAN

Pada bab ini akan menjelaskan kesimpulan dan saran berdasarkan hasil dan analisis penelitian.

1.8 Kesimpulan

Berdasarkan uraian di atas, maka dalam penelitian ini akan dikembangkan sebuah perangkat lunak untuk menentukan topik utama teks Bahasa Indonesia dengan menggunakan *cosine similarity* dan *POS Tagger*.

DAFTAR PUSTAKA

- Ariantini, D. A., Lumenta, A. S., & Jacobus, A. (2016). PENGUKURAN KEMIRIPAN DOKUMEN TEKS BAHASA INDONESIA MENGGUNAKAN METODE COSINE SIMILARITY. *E-Jurnal Teknik Informatika Volume 9, No 1*, 1.
- Fauzan, R., & Oranova Siahaan, D. (2014). POS TAGGING BAHASA INDONESIA DENGAN MAXIMUM ENTROPY MENGGUNAKAN STANFORD POSTAGGER. *Polhasains, Vol 2*, 26.
- Gunawan, & Lovina, G. (2006). QUESTION ANSWERING SYSTEM DAN PENERAPANNYA PADA ALKITAB. *JURNAL INFORMATIKA*, 1-9.
- Hasan, S. A. (2016). Rancang Bangun API Part of Speech Tagger untuk Bahasa Indonesia Menggunakan Hidden Markov Model.
- Herwin, & Andesa, K. (2019). Super Agent Chatbot “3S” Sebagai Media Informasi Menggunakan Metoda Natural Language Processing(NLP). *JURNAL TEKNOLOGI DAN OPEN SOURCE*, 53 - 64.
- Kannan, & Gurusamy, V. (2014). Preprocessing Techniques for Text Mining. *International Journal of Computer Science*, 2.
- Kruchten, P. (2004). *The Rational Unified Process: An Introduction*.
- Lelywiary, C. J., Widowati, S., & Muslim, K. L. (2019). Deteksi Pola Ambiguitas Struktural pada Spesifikasi Kebutuhan Perangkat Lunak menggunakan Pemrosesan Bahasa Alami. *Indonesia Journal on Computing*, 54.
- Novantara, P., & Pasruli, O. (2018). Implementasi Algoritma Jaro-Winkler Distance Untuk Sistem Pendekripsi Plagiarisme Pada Dokumen Skripsi. *Jurnal Buffer Informatika*.
- Pratama, M. R. (2016). Penentuan topik utama teks bahasa indonesia menggunakan metode Rule Reduction.
- Putriwana Pratama, R., Faisal, M., & Hanani, A. (2019). Deteksi Plagiarisme pada Artikel Jurnal Menggunakan Metode Cosine Simililarity. *SMARTICS Journal*, 23.
- Setiawan, A., Kurniawan, E., & Handiwidjodjo, W. (2013). IMPLEMENTASI STOP WORD REMOVAL UNTUK PEMBANGUNAN APPLIKASI ALKITAB BERBASIS WINDOWS 8. *Jurnal EKSIS*, 3.
- Wicaksono, A. f., & Purwarianti, A. (2010). HMM Based Part-of-Speech Tagger for Bahasa Indonesia. *On Proceedings of 4th International MALINDO (Malay and Indonesian Language) Workshop*, 1.

- Yamamoto, M., & Church, K. W. (2001). Using Suffix Arrays to Compute Term Frequency and Document Frequency for All Substrings in a Corpus. *Institute of Information Sciences and Electronics*, , 1.
- Ying, Q. (2012). APPLYING FREQUENCY ANDLOCATION INFORMATION TO KEYWORD EXTRACTION IN SINGLE DOCUMENT. *Proceedings of IEEE*.