

**PENERAPAN ALGORITMA COSINE SIMILARITY DAN
OPTIMASI STEMMER LIBRARY SASTRAWI DALAM
MENGHITUNG PERSENTASE KEMIRIPAN JUDUL DAN
ABSTRAK SKRIPSI (STUDI: REPOSITORY ATMALUHUR
PANGKALPINANG)**

SKRIPSI

Diajukan Untuk Melengkapi Salah Satu Syarat

Memperoleh Gelar Sarjana Komputer



**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNOLOGI INFORMASI
INSTITUT SAINS DAN BISNIS ATMA LUHUR
PANGKALPINANG**

2020

LEMBAR PERNYATAAN

Yang bertanda tangan di bawah ini :

NIM : 1611500121

Nama : Nurfajar Alasyari

Judul Skripsi : PENERAPAN ALGORITMA COSINE SIMILARITY DAN OPTIMASI STEMMER LIBRARY SASTRAWI DALAM MENGHITUNG PERSENTASE KEMIRIPAN JUDUL DAN ABSTRAK SKRIPSI (STUDI: REPOSITORY ATMALUHUR PANGKALPINANG)

Menyatakan bahwa Laporan Tugas Akhir saya adalah hasil karya sendiri dan bukan plagiat. Apabila ternyata ditemukan didalam laporan Tugas Akhir saya terdapat unsur plagiat, maka saya siap untuk mendapatkan sanksi akademik yang terkait dengan hal tersebut.

Pangkalpinang, 1 Juli 2020



Nurfajar Alasyari
Nurfajar Alasyari
NIM 1611500121

LEMBAR PENGESAHAN SKRIPSI

PENERAPAN ALGORITMA COSINE SIMILARITY DAN OPTIMASI STEMMER
LIBRARY SASTRAWI DALAM MENGHITUNG PERSENTASE KEMIRIPAN
JUDUL DAN ABSTRAK SKRIPSI (STUDI: REPOSITORY ATMALUHUR
PANGKALPINANG)

Yang dipersiapkan dan disusun oleh

Nurfajar Alasyari

1611500121

Telah dipertahankan di depan Dewan Penguji
Pada tanggal 21 Juli 2020

Anggota Penguji



Lukas Tommy, M.Kom
NIDN. 0215099201

Dosen Pembimbing



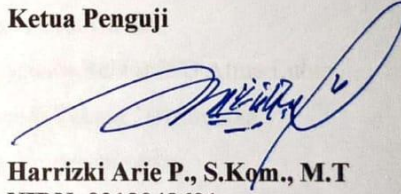
Benny Wijaya, S.T., M.Kom
NIDN. 0202097902

Kaprodi Teknik Informatika



Candra Kirana, M.Kom
NIDN. 0228108501

Ketua Penguji



Harrizki Arie P., S.Kom., M.T
NIDN. 0213048601

Skripsi ini telah diterima dan sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 16 Agustus 2020

**DEKAN FAKULTAS TEKNOLOGI INFORMASI
INSTITUT SAINS DAN BISNIS ATMA LUHUR**



Ellya Helmiud, M.Kom
NIDN. 0201027901

KATA PENGANTAR

Puji syukur Alhamdulillah kehadiran Allah SWT yang telah melimpahkan segala rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan skripsi yang berjudul **“Penerapan Algoritma Cosine Similarity Dan Optimasi Stemmer Library Sastrawi Dalam Menghitung Persentase Kemiripan Judul Dan Abstrak Skripsi (Studi Kasus: Repository Atma Luhur Pangkalpinang)”**

Skripsi ini mengambil topik *Mobile Android* dengan masalah penelitian yaitu pendaftaran peserta diklat teknis dan fungsional yang selanjutnya dikembangkan melalui aplikasi berbasis *Mobile Android*.

Penulis menyadari bahwa skripsi ini takkan terwujud tanpa bantuan, bimbingan, dan dorongan dari berbagai pihak. Dengan segala kerendahan hati, penulis menyampaikan ucapan terima kasih kepada :

1. Allah SWT yang telah menciptakan dan memberikan kehidupan di dunia
2. Bapak dan Ibu tercinta yang telah mendukung penulis baik spirit maupun materi.
3. Bapak Drs. Djaetun Hs yang telah mendirikan Atma Luhur .
4. Bapak Dr. Husni Teja Sukmana, S.T., M.Sc, selaku Rektor ISB Atma Luhur.
5. Bapak Chandra Kirana, M. Kom Selaku Kaprodi Teknik Informatika.
6. Bapak Benny Wijaya, S.T, M.Kom selaku dosen pembimbing.
7. Saudara dan sahabat-sahabatku terutama kawan-kawan angkatan 2020 yang telah memberikan dukungan moral untuk terus menyelesaikan skripsi ini.

Diharapkan kiranya skripsi ini dapat bermanfaat bagi mereka yang nantinya akan menulis Laporan penelitian dengan topik yang sama.

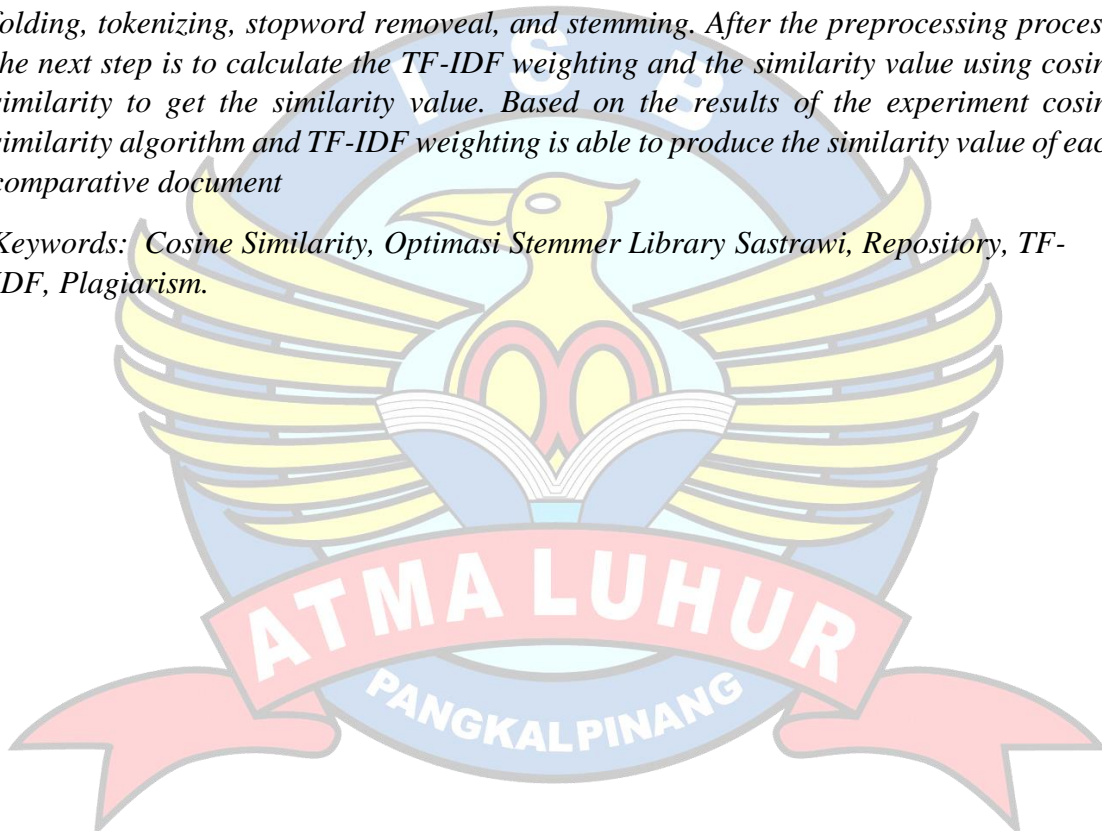
Pangkalpinang, Juli 2020

Penulis

ABSTRACT

Plagiarism is the act of taking part or all of a person's ideas in the form of documents or texts without including the source of information retrieval. This study aims to detect the similarity of text documents using cosine similarity algorithm and TF-IDF weighting so that it can be used to determine the value of plagiarism. The document used for this text comparison is the title of the thesis. The results of the research are that when stemming the similarity value is 10% higher on average than stemming is not done. This study produces a similarity value above 50% for documents with a high degree of similarity. Whereas for documents with a low level of similarity or non-plagiarism, the similarity value is below 40%. With the method used in preprocessing which consists of case folding, tokenizing, stopword removal, and stemming. After the preprocessing process, the next step is to calculate the TF-IDF weighting and the similarity value using cosine similarity to get the similarity value. Based on the results of the experiment cosine similarity algorithm and TF-IDF weighting is able to produce the similarity value of each comparative document

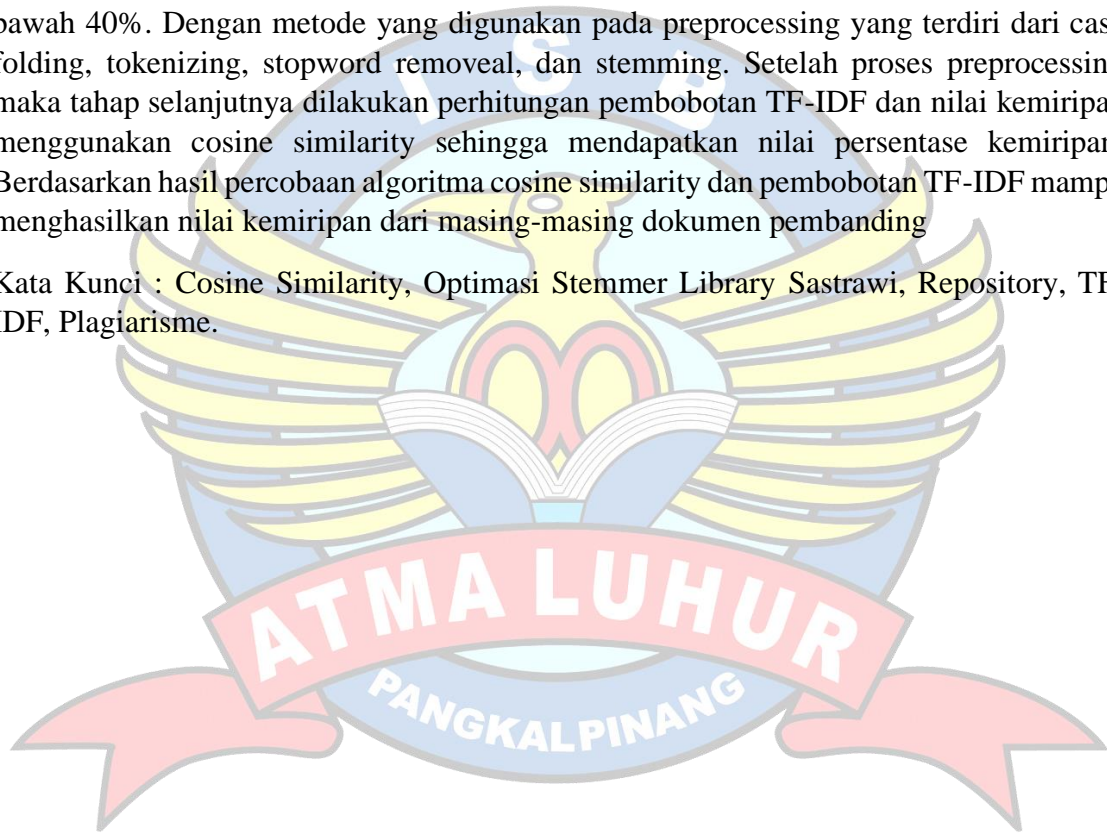
Keywords: Cosine Similarity, Optimasi Stemmer Library Sastrawi, Repository, TF-IDF, Plagiarism.



ABSTRAK

Plagiarisme merupakan tindakan mengambil sebagian atau seluruh ide seseorang berupa dokumen maupun teks tanpa mencantumkan sumber pengambilan informasi. Penelitian ini bertujuan untuk mendeteksi kemiripan dokumen teks menggunakan algoritma cosine similarity dan pembobotan TF-IDF sehingga dapat digunakan untuk menentukan nilai plagiarisme. Dokumen yang digunakan untuk perbandingan teks ini adalah judul skripsi. Hasil penelitian yaitu saat dilakukan stemming nilai kemiripan lebih tinggi rata-rata 10% daripada tidak dilakukan proses stemming. Penelitian ini menghasilkan nilai similaritas di atas 50% untuk dokumen yang tingkat kemiripannya tinggi. Sedangkan untuk dokumen dengan tingkat kemiripan rendah atau tidak berplagiat menghasilkan nilai similarity di bawah 40%. Dengan metode yang digunakan pada preprocessing yang terdiri dari case folding, tokenizing, stopword removal, dan stemming. Setelah proses preprocessing maka tahap selanjutnya dilakukan perhitungan pembobotan TF-IDF dan nilai kemiripan menggunakan cosine similarity sehingga mendapatkan nilai persentase kemiripan. Berdasarkan hasil percobaan algoritma cosine similarity dan pembobotan TF-IDF mampu menghasilkan nilai kemiripan dari masing-masing dokumen perbandingan

Kata Kunci : Cosine Similarity, Optimasi Stemmer Library Sastrawi, Repository, TF-IDF, Plagiarisme.



DAFTAR ISI

	Halaman
LEMBAR PERNYATAAN	i
LEMBAR PENGESAHAN SKRIPSI	ii
KATA PENGANTAR	ii
ABSTRACT	v
ABSTRAK	vi
DAFTAR ISI	vii
DAFTAR GAMBAR	ix
DAFTAR TABEL	x
DAFTAR SIMBOL	xi
BAB I PENDAHULUAN	
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Tujuan dan Manfaat	3
1.3.1 Tujuan	3
1.3.2 Manfaat	3
1.4 Batasan Masalah.....	3
1.5 Metodologi Penelitian	4
1.6 Sistematika Penulisan.....	4
BAB II LANDASAN TEORI	
2.1 Model Pengembangan Prototype	6
2.2 Revolusi Industri	7
2.3 Metode <i>Cosine Similarity</i>	7
2.4 <i>Natural Language Processing (NLP)</i>	8
2.5 Penelitian Terdahulu.....	9

BAB III METODOLOGI PENELITIAN

3.1	Motodologi Penelitian.....	15
3.1	Model Penelitian	15
3.1	Perancangan Sistem.....	16
3.1	Text Preprocessing	17
3.1	Algoritma cosine similarity	20
3.1	Implementasi	21

BAB IV HASIL DAN PEMBAHASAN

4.1	Profil Instansi	23
4.1.1	Sejarah Instansi	24
4.1.2	Visi Dan Misi Instansi.....	24
4.1.3	Struktur Organisasi.....	24
4.2	Analisis Masalah	25
4.2.1	Analisis Kebutuhan.....	26
4.2.2	Analisis Sistem Berjalan	27
4.3	Rancangan Sistem.....	29
4.3.1	Proses Sistem Usulan	29
4.3.2	Class Diagram Sistem Usulan.....	39
4.3.3	Rancangan Layar.....	40
4.4	Implementasi.....	42
4.4.1	Pengujian <i>Cosine similarity manual</i>	42
4.4.2	Hasil pengujian	47
4.4.3	Tampilan Layar	58

BAB V PENUTUP

4.1	Kesimpulan	59
4.2	Saran.....	59

DAFTAR PUSTAKA.....	59
----------------------------	-----------

LAMPIRAN	59
-----------------------	-----------

DAFTAR GAMBAR

	Halaman
Gambar 3 1 <i>Text procesing</i>	29
Gambar 3. 2 <i>Proses tokenizing</i>	29
Gambar 3. 3 <i>Proses fltering</i>	29
Gambar 3. 4 <i>Poses stemming</i>	29
Gambar 4. 1 <i>Struktur Organisasi</i>	29
Gambar 4. 2 <i>Activity Diagram Berjalan mahasiswa</i>	34
Gambar 4. 2 <i>Activity Diagram Berjalan admin</i>	34
Gambar 4. 3 <i>Usecase Diagram mahasiswa</i>	35
Gambar 4. 4 <i>Usecase Diagram Admin</i>	36
Gambar 4. 5 <i>Sequence Diagram Login Admin</i>	37
Gambar 4. 7 <i>Sequence Diagram input judul dan abstrak</i>	39
Gambar 4. 8 <i>Sequence Diagram cosine proses tahap 1</i>	40
Gambar 4. 9 <i>Sequence Diagram cosine proses tahap 2</i>	41
Gambar 4. 10 <i>Sequence Diagram cosine proses tahap 3</i>	41
Gambar 4. 23 <i>class diagram</i>	52
Gambar 4. 24 <i>Rancangan layar login</i>	53
Gambar 4. 29 <i>Rancangan layar utama user</i>	56
Gambar 4. 29 <i>Rancangan layar utama Admin</i>	56
Gambar 4. 32 <i>Tampilan layar hasil pengujian d1</i>	58
Gambar 4. 32 <i>Tampilan layar hasil pengujian d2</i>	58
Gambar 4. 33 <i>Tampilan layar hasil pengujian d3</i>	58
Gambar 4. 34 <i>Tampilan layar login admin</i>	59
Gambar 4. 35 <i>Tampilan layar halaman utama user</i>	59
Gambar 4. 36 <i>Tampilan layar halaman utama admin</i>	60




DAFTAR TABEL

	Halaman
Tabel 2. 1 Ringkasan Penelitian Terdahulu	14
Tabel 4. 1 Analisis Kebutuhan.....	24
Tabel 4. 1 Usecase mahasiswa.....	30
Tabel 4. 1 usecase admin	31
Tabel 4. 1 Kebutuhan perangkat	43
Tabel 4. 1 Tabel Sampel Query	43
Tabel 4. 1 Tabel Perhitungan TF dan IDF	44
Tabel 4. 1 Tabel Perhitungan TF/IDF	45
Tabel 4. 1 Tabel Hasil Perkalian Skalar Tiap Dn terhadap Q.....	46
Tabel 4. 1 Tabel Hasil Perkalian Vektor.....	46
Tabel 4. 1 T Tabel Hasil cosine	46
Tabel 4. 5 Deskripsi <i>Fungsi dan Pengujian</i>	49
Tabel 4. 6 Identifikasi dan Perencanaan	49
Tabel 4. 7 Kelas Uji login.....	49
Tabel 4. 8 Kelas Uji mahasiswa.....	50
Tabel 4. 9 Kelas Uji admin	50

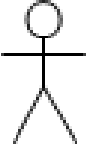




DAFTAR SIMBOL






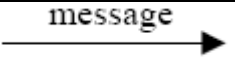
Activity Diagram

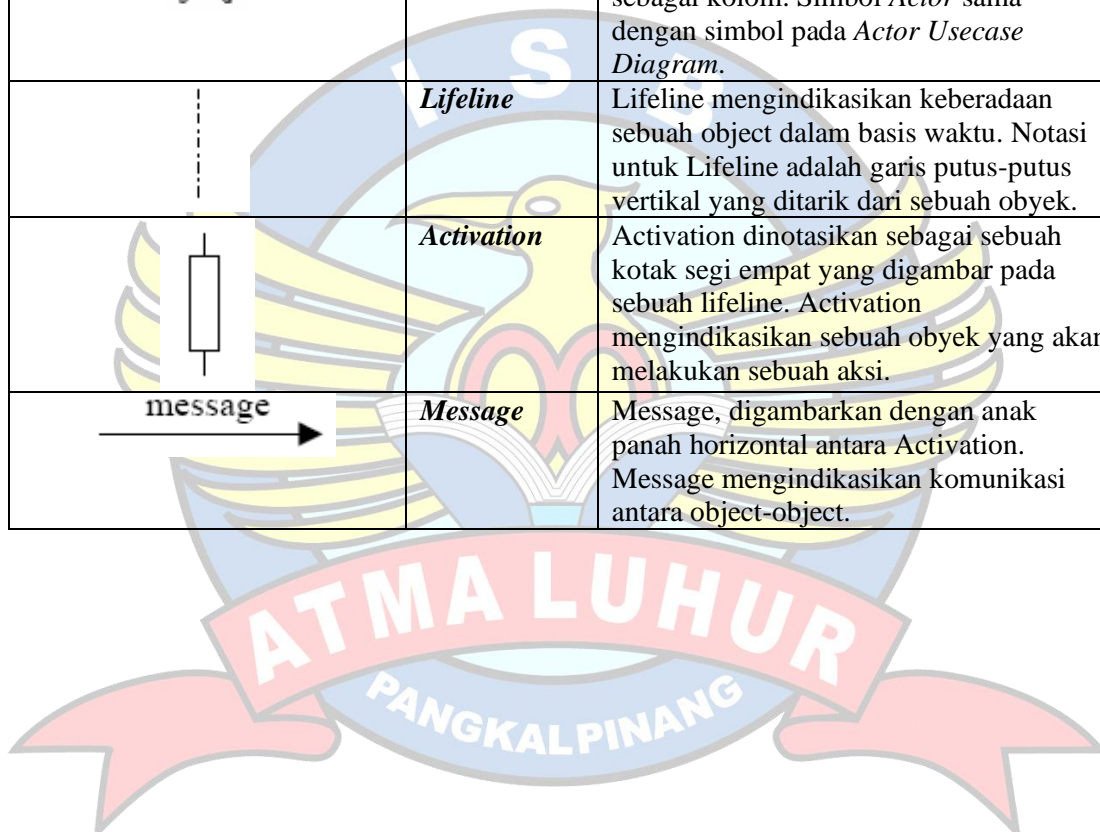
Simbol	Nama	Keterangan
	Initial Node	Titik Awal
	Activity Final	Titik Akhir
	Action	State dari sistem yang mencerminkan eksekusi


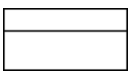

Usecase Diagram

Simbol Use Case Diagram	
	<p>Aktor</p> <p>Menggambarkan orang atau sistem yang menyediakan atau menerima informasi dari sistem yang dibuat atau bisa disebut dengan pengguna aplikasi</p>
	<p>Association</p> <p>Menggambarkan hubungan aktor dengan <i>use case</i></p>
	<p>Use Case</p> <p>Menggambarkan fungsionalitas dari suatu sistem sehingga pengguna sistem paham dan mengerti kegunaan sistem yang akan dibangun.</p>

Sequence Diagram

Simbol	Nama	Keterangan
	Boundary Class	Menggambarkan sebuah penggambaran dari form.
	Control Class	Menggambarkan penghubung antara boundary dengan tabel.
	Actor	<i>Actor</i> juga dapat berkomunikasi dengan object, maka <i>Actor</i> juga dapat diurutkan sebagai kolom. Simbol <i>Actor</i> sama dengan simbol pada <i>Actor Usecase Diagram</i> .
	Lifeline	Lifeline mengindikasikan keberadaan sebuah object dalam basis waktu. Notasi untuk Lifeline adalah garis putus-putus vertikal yang ditarik dari sebuah obyek.
	Activation	Activation dinotasikan sebagai sebuah kotak segi empat yang digambar pada sebuah lifeline. Activation mengindikasikan sebuah obyek yang akan melakukan sebuah aksi.
	Message	Message, digambarkan dengan anak panah horizontal antara Activation. Message mengindikasikan komunikasi antara object-object.



Class Diagram	
	<p>Pesan Tipe Send</p> <p>Menggambarkan suatu <i>object</i> mengirim data masuk</p>
	<p>Class</p> <p>Himpunan dari objek-objek yang berbagi atribut serta operasi yang sama.</p>
	<p>Association</p> <p>Apa yang menghubungkan antara objek satu dengan objek lainnya.</p>

