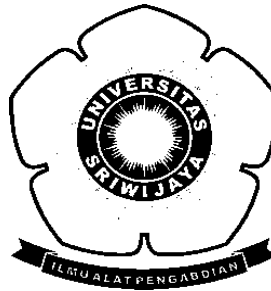


**DETEKSI KEMIRIPAN TEKS BERBAHASA INDONESIA
MENGUNAKAN ALGORITMA RATCLIFF/OBERSHELP**

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



Oleh :

Nurul Izzah

NIM : 09021381823127

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

LEMBAR PENGESAHAN SKRIPSI

DETEKSI KEMIRIPAN TEKS BERBAHASA INDONESIA MENGUNAKAN ALGORITMA *RATCLIFF/OBERSHELP*

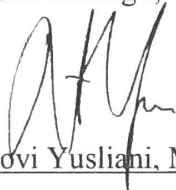
Oleh :

Nurul Izzah

NIM : 09021381823127

Palembang, 29 Desember 2021

Pembimbing I,



Novi Yusliani, M.T

NIP. 198211082012122001

Pembimbing II,

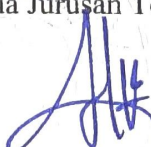


Desty Rochah, M.T.

NIP. 198912212020122011

Mengetahui,

Ketua Jurusan Teknik Informatika



Alvi Syahrini Utami, M.Kom

NIP. 197812222006042003

TANDA LULUS UJIAN SIDANG SKRIPSI

Pada hari Rabu tanggal 29 Desember 2021 telah dilaksanakan ujian sidang skripsi oleh
Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya

Nama : Nurul Izzah

NIM : 09021381823127

Judul : DETEKSI KEMIRIPAN TEKS BERBAHASA INDONESIA
MENGUNAKAN ALGORITMA *RATCLIFF/OBERSHELP*

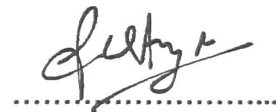
1. Pembimbing I

Novi Yuslinai, M.T.
NIP. 198211082012122001



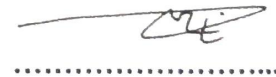
2. Pembimbing II

Desty Rodiah, M.T.
NIP. 198912212020122011



3. Penguji I

Osvari Arsalan, M.T.
NIP. 198806282018031001

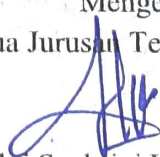


4. Penguji II

Kanda Januar Miraswan, M.T.
NIP. 199001092019032016



Mengetahui,
Ketua Jurusan Teknik Informatika



Alvi Syahrini Utami, M.Kom
NIP. 197812222006042003

HALAMAN PERNYATAAN BEBAS PLAGIAT

Yang bertanda tangan di bawah ini :

Nama : Nurul Izzah

NIM : 09021381823127

Program Studi : Teknik Informatika

Judul Skripsi : Deteksi Kemiripan Teks Berbahasa Indonesia

Menggunakan Algoritma *Ratcliff/Obershelp*

Hasil Pengecekan Software *iThenticate/Turnitin* : 3%

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 berlaku.

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



Palembang, 3 Desember 2021



Nurul Izzah
NIM. 09021381823127

Motto :

- ☐ Allah tidak akan membebani seseorang melainkan sesuai dengan kesanggupannya, nikmati prosesmu saat ini karena Allah tidak pernah meninggalkan hamba-Nya yang beriman.
- ☐ Belajarlah dari sebuah kegagalan agar menjadi pribadi yang lebih baik.
- ☐ Jangan pernah takut untuk mencoba, karena kita tidak pernah tahu bagaimana cara Allah memberikan nikmat-Nya untuk kita.
- ☐ Kebiasaan kecil yang baik akan menghasilkan kualitas diri yang baik pula.
- ☐ Setiap insan adalah seniman pada bidangnya dan akan berhasil pada waktunya.

Kupersembahkan karya tulis ini kepada :

- ☐ Allah SWT atas segala nikmat dan rahmat-nya laporan ini dapat terselesaikan dengan baik.
- ☐ Abi dan Umi tercinta yang selalu berusaha memberikan yang terbaik dan mendukung setiap langkah putrinya.
- ☐ Teman-teman kelas TIBIL C yang selalu mendukung dan menyemangati.
- ☐ Gege dan Sekar yang selalu membantu, mendukung, memotivasi, dan menyemangati.

ABSTRACT

Acts of plagiarism in written works are often encountered in everyday life. More and more texts circulating on the internet result in increased acts of plagiarism without citations due to a lack of creativity in the community. Prevention of plagiarism can be done by creating a software that is able to detect text similarities. In this study, the Ratcliff/Obershelp algorithm is used to detect text similarity. The steps taken are in the form of combining strings from 2 pieces of text to get the total character length of the text (sequence (string) matching). Then the stage (sub-sequence) is carried out to find the same word in the 2 texts and calculate the character length. The calculation of the value and percentage of similarity strings is carried out to classify the types of plagiarism that exist in the text. Classification of types of plagiarism is divided into 5 types and categorized based on the results of the percentage similarity string. The data used in this study is secondary data in the form of 20 Indonesian news texts with different website sources and divided into 5 topics. This data is obtained by the documentation method and stored in a file with .txt format. This text similarity detection software can generate percentage values and classifications of text similarity types. Software testing is carried out based on 3 scenarios, each of which will detect the similarity of the text. This test produces an average error percentage value of 0% in scenario 1, 8.33% in scenario 2, and 14.65% in scenario 3. Based on the test results of the 3 scenarios, the average error percentage of the text similarity detection software is obtained. in Indonesian using the Ratcliff/Obershelp algorithm of 7.66%.

*Keywords: Indonesian Text, Internet, Plagiarism, Ratcliff/Obershelp
Algorithm, Text Similarity Detection*

ABSTRAK

Tindakan plagiarisme pada karya tulis sering sekali dijumpai dalam kehidupan sehari – hari. Semakin banyak teks yang beredar di internet mengakibatkan meningkatnya tindakan plagiarisme tanpa melakukan sitasi dikarenakan kurangnya kreativitas pada masyarakat. Pencegahan terhadap plagiarisme dapat dilakukan dengan membuat suatu perangkat lunak yang mampu mendeteksi kemiripan teks. Pada penelitian ini, algoritma *Ratcliff/Obershelp* digunakan untuk mendeteksi kemiripan teks. Tahapan yang dilakukan berupa penggabungan *string* dari 2 buah teks untuk mendapatkan panjang karakter total teks (*sequence (string) matching*). Kemudian tahap (*sub-sequence*) dilakukan untuk mencari kata yang sama pada 2 teks tersebut dan menghitung panjang karakternya. Perhitungan nilai dan persentase *similarity string* dilakukan untuk mengklasifikasikan jenis plagiarisme yang ada pada teks. Klasifikasi jenis plagiarisme dibagi menjadi 5 jenis dan dikategorikan berdasarkan hasil dari persentase *similarity string*. Data yang digunakan pada penelitian ini merupakan data sekunder berupa 20 teks berita berbahasa Indonesia dengan sumber website yang berbeda dan terbagi atas 5 topik. Data ini diperoleh dengan metode dokumentasi dan disimpan dalam file dengan format *.txt*. Perangkat lunak deteksi kemiripan teks ini dapat menghasilkan nilai persentase dan klasifikasi jenis kemiripan teksnya. Pengujian perangkat lunak dilakukan berdasarkan 3 skenario yang masing – masing akan mendeteksi kemiripan teks. Pengujian ini menghasilkan nilai persentase kesalahan rata – rata 0% pada skenario 1, 8,33% pada skenario 2, dan 14,65 % pada skenario 3. Berdasarkan hasil pengujian 3 skenario tersebut, diperoleh persentase kesalahan rata – rata perangkat lunak deteksi kemiripan teks berbahasa Indonesia menggunakan algoritma *Ratcliff/Obershelp* sebesar 7,66%.

Kata Kunci : Algoritma *Ratcliff/Obershelp*, Internet, Plagiarisme, Deteksi

Kemiripan Teks, Teks Berbahasa Indonesia

KATA PENGANTAR

Alhamdulillahirabbiláalamiin. Segala puji dan syukur penyusun haturkan kehadiran Allah SWT atas rahmat dan karunia-Nya penyusunan skripsi yang berjudul “**Sistem Deteksi Kemiripan Teks Berbahasa Indonesia Menggunakan Algoritma Ratcliff/Obershelp**” dapat diselesaikan dengan baik dan tepat waktu sesuai dengan waktu yang telah ditentukan. Skripsi ini dimaksudkan untuk memperoleh gelar sarjana strata satu (S1) Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya

Untuk selanjutnya penyusun mengucapkan banyak terima kasih kepada pihak-pihak yang telah membantu dalam penyelesaian skripsi ini, yaitu:

1. Dekan Fakultas Ilmu Komputer Universitas Sriwijaya, Bapak Jaidan Jauhari, SPd., M.T.
2. Ketua Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya, Ibu Alvi Syahrini Utami, M.Kom.
3. Kedua orang tua, Abi dan Umi tercinta yang selalu memberikan dukungan, motivasi, dan semangat kepada penyusun.
4. Dosen Pembimbing Akademik, Ibu Yunita, S.Si., M.CS. yang telah meluangkan waktu, tenaga, dan pikiran untuk membimbing, mengarahkan, serta memotivasi penyusun dalam menyelesaikan skripsi ini.
5. Dosen Pembimbing 1, Ibu Novi Yusliani, S.Kom., M.T. yang telah meluangkan waktu, tenaga, dan pikiran untuk membimbing, mengarahkan, serta memotivasi penyusun dalam menyelesaikan skripsi ini.
6. Dosen Pembimbing 2, Ibu Desty Rodiah, M.T. yang telah meluangkan waktu, tenaga, dan pikiran untuk membimbing, mengarahkan, serta memotivasi penyusun dalam menyelesaikan skripsi ini.
7. Seluruh dosen Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya yang telah memberikan ilmu pengetahuan kepada penyusun.

8. Rekan – rekan mahasiswa kelas TIBIL C Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya yang selalu memberikan dukungan dan semangat kepada penyusun.
9. Semua pihak yang telah membantu, yang tidak dapat disebutkan satu persatu.

Penyusun menyadari bahwa skripsi ini belumlah sempurna dan masih terdapat banyak kekurangan dan kekeliruan. Untuk itu, penulis memohon maaf serta mengharapkan dan menerima segala kritik dan saran yang sifatnya membangun serta bermanfaat bagi pengembangan ilmu pengetahuan, khususnya pada bidang Teknik Informatika.

Palembang, November 2021

Penyusun,

DAFTAR ISI

	Halaman
JUDUL	i
HALAMAN PENGESAHAN.....	ii
HALAMAN PERSETUJUAN KOMISI PENGUJI	iii
HALAMAN PERNYATAAN BEBAS PLAGIAT	iv
HALAMAN MOTTO DAN PERSEMBAHAN.....	v
ABSTRACT.....	vi
ABSTRAKSI.....	vii
KATA PENGANTAR	viii
DAFTAR ISI.....	x
DAFTAR TABEL.....	xiv
DAFTAR GAMBAR	xvii
DAFTAR ALGORITMA.....	xix
DAFTAR LAMPIRAN	xx

BAB I PENDAHULUAN

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-4
1.6	Batasan Masalah.....	I-5
1.7	Sistematika Penulisan.....	I-5
1.8	Kesimpulan.....	I-7

BAB II STUDI PUSTAKA

2.1	Pendahuluan.....	II-1
2.2	Landasan Teori.....	II-1

2.2.1	Plagiarisme	II-1
2.2.2	Deteksi Kemiripan Teks.....	II-2
2.2.3	<i>Pre-Processing</i> Teks	II-3
2.2.4	Algoritma <i>Ratcliff/Obershelp</i> (RO).....	II-8
2.2.5	Klasifikasi Jenis Plagiarisme.....	II-11
2.2.6	<i>Cosine Similarity</i>	II-12
2.2.7	Pengukuran dan Kesalahan	II-13
2.2.8	Metode <i>Rational Unified Process</i> (RUP).....	II-13
2.2.9	Bahasa Pemrograman <i>Python</i>	II-15
2.3	Penelitian Terkait	II-16
2.4	Kesimpulan	II-17

BAB III METODELOGI PENELITIAN

3.1	Pendahuluan.....	III-1
3.2	Pengumpulan Data	III-1
3.2.1	Jenis dan Sumber Data	III-1
3.2.2	Metode Pengumpulan Data	III-1
3.3	Tahapan Penelitian.....	III-2
3.3.1	Menentukan Kerangka Kerja Penelitian	III-2
3.3.2	Menentukan Kriteria Pengujian	III-5
3.3.3	Menentukan Format Data Pengujian.....	III-5
3.3.4	Menentukan Alat Bantu Penelitian	III-6

3.3.5	Melakukan Pengujian Penelitian.....	III-6
3.3.6	Melakukan Analisis Hasil Pengujian dan Membuat Kesimpulan Penelitian.....	III-7
3.4	Metode Pengembangan Perangkat Lunak.....	III-8
3.4.1	Tahap <i>Inception</i>	III-8
3.4.2	Tahap <i>Elaboration</i>	III-8
3.4.3	Tahap <i>Construction</i>	III-9
3.4.4	Tahap <i>Transition</i>	III-9
3.4.5	Manajemen Proyek Penelitian.....	III-9
3.5	Kesimpulan	III-12
 BAB IV PENGEMBANGAN PERANGKAT LUNAK		
4.1	Pendahuluan	IV-1
4.2	<i>Rational Unified Process (RUP)</i>	IV-1
4.2.1	Tahap <i>Inception</i>	IV-1
4.2.2	Tahap <i>Elaboration</i>	IV-4
4.2.3	Tahap <i>Construction</i>	IV-12
4.2.4	Tahap <i>transition</i>	IV-17
4.3	Kesimpulan.....	IV-22
 BAB V HASIL DAN ANALISIS PENELITIAN		
5.1	Pendahuluan	V-1
5.2	Data Hasil Percobaan/Penelitian	V-1

5.2.1	Konfigurasi Percobaan	V-I
5.2.2	Data Hasil Konfigurasi I	V-2
5.2.3	Data Hasil Konfigurasi II	V-6
5.2.4	Data Hasil Konfigurasi III.....	V-10
5.3	Analisis Hasil Penelitian	V-20
5.4	Kesimpulan.....	V-22

BAB VI KESIMPULAN DAN SARAN

6.1	Kesimpulan.....	VI-1
6.2	Saran.....	VI-1

DAFTAR PUSTAKA	xxi
----------------------	-----

LAMPIRAN

DAFTAR TABEL

	Halaman
II-1 Contoh <i>Case folding Text</i>	II-5
II-2 Contoh <i>Punctuation Text</i>	II-5
II-3 Aturan Algoritma Nazief dan Adriani <i>Stemming</i>	II-6
II-4 Contoh <i>Stemming Text</i>	II-7
II-5 Contoh <i>Tokenization Text</i>	II-7
II-6 Contoh <i>Whiitespace Removal Text</i>	II-8
II-7 Contoh Perhitungan <i>Sequence (String) Matching</i>	II-9
II-8 Contoh Perhitungan <i>Sub-Sequence (Sub-String)</i>	II-10
II-9 Contoh perhitungan Nilai <i>Similarity String</i> dan Persentasenya	II-11
II-10 Klasifikasi Jenis Kemiripan Teks.....	II-11
II-11 Contoh Perhitungan <i>Cosine Similarity</i>	II-12
III-1 Format Tabel Hasil Pengujian Skenario	III-6
III-2 Alat Bantu Penelitian	III-7
III-3 Format Tabel Hasil Perhitungan Kesalahan	III-7
III-4 Rencana Pengembangan Perangkat Lunak	III-10
IV-1 Kebutuhan Fungsional Perangkat Lunak.....	IV-1
IV-2 Kebutuhan Non-Fungsional Perangkat Lunak	IV-2
IV-3 Hasil Validasi <i>Use Case Diagram</i> Perangkat Lunak.....	IV-3
IV-4 Alat Bantu Penelitian.....	IV-4

IV-5	Skenario <i>Use Case</i> Masukkan Data Teks.....	IV-5
IV-6	Skenario <i>Use Case</i> Perhitungan dan Penyimpanan Hasil Deteksi Kemiripan Teks.....	IV-5
IV-7	Skenario <i>Use Case</i> Hapus Data Teks	IV-6
IV-8	Skenario <i>Use Case</i> Lihat Data Hasil Deteksi Kemiripan Teks	IV-6
IV-9	Implementasi Kelas dalam Kode Program	IV-12
IV-10	Alat Pengujian Perangkat Lunak	IV-17
IV-11	Skenario Pengujian <i>Use Case</i> Masukkan Data Teks.....	IV-18
IV-12	Skenario Pengujian <i>Use Case</i> Perhitungan dan Penyimpanan Hasil Deteksi Kemiripan Teks	IV-18
IV-13	Skenario Pengujian <i>Use Case</i> Hapus Data Teks	IV-19
IV-14	Skenario Pengujian <i>Use Case</i> Lihat Data Hasil Deteksi Kemiripan Teks	IV-19
IV-15	Hasil Pengujian <i>Use Case</i> Masukkan Data Teks	IV-19
IV-16	Hasil Pengujian <i>Use Case</i> Perhitungan dan Penyimpanan Hasil Deteksi Kemiripan Teks	IV-20
IV-17	Hasil Pengujian <i>Use Case</i> Hapus Data Teks.....	IV-20
IV-18	Hasil Pengujian <i>Use Case</i> Lihat Data Hasil Deteksi Kemiripann Teks	IV-21
V-1	Hasil Perhitungan Algoritma <i>Ratcliff/Obershelp</i> Skenario 1.....	V-2
V-2	Data Hasil Pengujian pada Skenario 1.....	V-3
V-3	Hasil Perhitungan Algoritma <i>Ratcliff/Obershelp</i> Skenario 2.....	V-6
V-4	Data Hasil Pengujian pada Skenario 2.....	V-7

V-5	Hasil Perhitungan Algoritma <i>Ratcliff/Obershelp</i> Skenario 3.....	V-10
V-6	Data Hasil Pengujian pada Skenario 3.....	V-14
V-7	Data Hasil Analisis	V-20

DAFTAR GAMBAR

	Halaman
II-1 Alur Deteksi Plagiarisme Teks.....	II-3
II-2 Tahapan <i>Pre-Processing</i> Teks	II-4
II-3 Tahapan Metode <i>Rational Unified Process</i> (RUP).....	II-14
III-1 Contoh Hasil Penyalinan Data	III-2
III-2 Alur Tahapan penelitian.....	III-2
III-3 Diagram Alur proses Umum Perangkat Lunak.....	III-3
IV-1 <i>Use Case Diagram</i> Perangkat Lunak	IV-3
IV-2 <i>Activity Diagram</i> Masukkan Data Teks	IV-7
IV-3 <i>Activity Diagram</i> Perhitungan dan Penyimpanan Hasil Deteksi Kemiripan teks	IV-8
IV-4 <i>Activity Diagram</i> Hapus Data Teks.....	IV-8
IV-5 <i>Activity Diagram</i> Lihat Data Hasil Deteksi Kemiripan Teks.....	IV-9
IV-6 <i>Sequence Diagram</i> Masukkan Data Teks.....	IV-9
IV-7 <i>Sequence Diagram</i> Perhitungan dan Penyimpanan Hasil Deteksi Kemiripan Teks.....	IV-10
IV-8 <i>Sequence Diagram</i> Hapus Data Teks	IV-10
IV-9 <i>Sequence Diagram</i> Lihat Data Hasil Deteksi Kemiripan teks ...	IV-11
IV-10 <i>Class Diagram</i> Perangkat Lunak.....	IV-11
IV-11 Rancangan Antarmuka Perangkat Lunak Halaman Main_Page	IV-12

IV-12	Implementasi Antarmuka Perangkat Lunak Halaman	
	Main_Page.....	IV-16
V-1	Grafik Perbandingan Persentase dan Keakuratan	
	Skenario 1	V-5
V-2	Grafik Hasil Perhitungan Prediksi dan Hasil Persentase	
	Skenario 1	V-5
V-3	Grafik Perbandingan Persentase dan Keakuratan	
	Skenario 2	V-9
V-4	Grafik Hasil Perhitungan Prediksi dan Hasil Persentase	
	Skenario 2	V-10
V-5	Grafik Perbandingan Persentase dan Keakuratan	
	Skenario 3	V-19
V-6	Grafik Hasil Perhitungan Prediksi dan Hasil Persentase	
	Skenario 3	V-20
V-7	Grafik Persentase Kesalahan Data Analisis.....	V-22

DAFTAR ALGORITMA

	Halaman
IV-1 Kode Program Unrtuk Proses <i>Pre-Processing Text</i> , Perhitungan Algoritma <i>Ratcliff/Obershelp</i> , Dan Klasifikasi Jenis Plagiarisme	IV-16

DAFTAR LAMPIRAN

- Lampiran 1. Data Teks Berita Berbaha Indonesia
- Lampiran 2. Kode Program Fungsi Pada Kelas SampleApp
- Lampiran 3. Kode Program Fungsi Pada Kelas Input_Paage
- Lampiran 4. Kode Program Fungsi Pada Kelas View_Page

BAB I

PENDAHULUAN

1.1 Pendahuluan

Pada bab pendahuluan akan dibahas mengenai latar belakang masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, perangkat lunak ketika penulisan dan kesimpulan. Bab ini akan menjelaskan mengenai gambaran umum dari keseluruhan kegiatan penelitian.

1.2 Latar Belakang Masalah

Plagiarisme merupakan suatu tindakan salin dan tempel dari produk intelektual orang lain yang disalahgunakan tanpa menyebutkan nama penulis, penemu, dan penggagas orisinal (Syadiqi, 2019). Tindakan plagiarisme diatur dalam UU nomor 19 tahun 2002 mengenai hak cipta. Jenis plagiarisme dapat berupa kata, ide, sumber, dan kepengarangan (Shadiqi, 2019) yang mengakibatkan berkurangnya apresiasi terhadap hasil karya dan kreativitas dari si pencipta aslinya. Dalam hal ini, karya dapat berupa karya tulis, desain, gambar, video, maupun kode program.

Kemajuan teknologi membuat teks berbahasa Indonesia dapat diakses dengan mudah di berbagai media, salah satunya adalah internet. Namun, semakin banyak teks berbahasa Indonesia yang beredar di internet mengakibatkan meningkatnya tindakan plagiarisme tanpa adanya sitasi, ditambah lagi dengan

adanya fasilitas salin-tempel pada teknologi saat ini yang memberi kemudahan dalam melakukan plagiarisme. Hal ini memicu para pelaku plagiarisme untuk bertindak curang sehingga menurunnya kreativitas dalam menciptakan teks berbahasa Indonesia yang original.

Untuk meminimalisir hal tersebut, maka diperlukan suatu upaya pencegahan terhadap plagiarisme teks berbahasa Indonesia. Pencegahan dengan mendeteksi kemiripan teks merupakan salah satu cara yang paling efektif untuk meminimalisirnya (Sulistyaningsih, 2017). Dengan mendeteksi kemiripan teks, maka dapat dilakukan pencarian pada bagian yang diduga terdapat plagiarisme atau kesamaan pada teks. Namun, pendeteksian sangat sulit dilakukan secara manual, sehingga diperlukan suatu perangkat lunak yang mampu mendeteksi kemiripan teks. Perangkat lunak deteksi kemiripan teks dibangun dengan berbagai algoritma, salah satunya adalah algoritma *Ratcliff/Obershelp*.

Algoritma *Ratcliff/Obershelp* merupakan algoritma yang dapat melakukan proses identifikasi kemiripan *string* yang akurat dan lebih baik dengan jumlah *string* yang sangat banyak. Algoritma ini akan melakukan pendeteksian kemiripan *string* pada dua buah teks. Pada algoritma ini akan dilakukan proses pencarian total karakter pada teks (*sequence (string) matching*), pencarian kata yang sama dan perhitungan total karakternya (*sub-string*), serta melakukan perhitungan nilai dan persentasi kemiripan *string* (Ilyankou, 2014).

Pada penelitian sebelumnya (Joane, 2017) dilakukan penelitian terkait implementasi aplikasi deteksi kemiripan dokumen teks menggunakan algoritma

Ratcliff/Obershelp dengan metode *Data Flow Diagram* (DFD). Penelitian ini menghasilkan perangkat lunak dengan persentase nilai yang hampir mendekati nilai persentase yang telah diprediksi sebelumnya. Namun, perangkat lunak ini menghasilkan nilai yang kurang akurat karena kurangnya proses *stemming* pada tahapan *pre-processing text*, sehingga mempengaruhi hasil dari proses pendeteksian kemiripan teks menggunakan algoritma *Ratcliff/Obershelp*.

Yusuf, dkk (2019) telah meneliti mengenai perbandingan algoritma *Ratcliff/Obershelp* dan *Rabin Karp* dalam menghitung kemiripan teks berbahasa Indonesia. Penelitian ini menghasilkan nilai persentase, perhitungan nilai similarity, waktu eksekusi, dan nilai konsistensi terhadap dua algoritma tersebut. Dalam perbandingannya, penelitian ini membuktikan bahwa algoritma *Ratcliff/Obershelp* lebih baik digunakan dalam mendeteksi kemiripan teks dengan nilai kurang lebih 3% lebih tinggi dibandingkan dengan algoritma *Rabin Karp*.

Ilyankou (2014) telah meneliti mengenai perbandingan algoritma *Jaro-Winkler* dan *Ratcliff Obershelp* dalam mengecek ejaan. Penelitian ini diujikan dalam *database FreeBSD* dan *database Mieliestroke* dan menghasilkan nilai skor. Algoritma *Ratcliff/Obershelp* dapat mendeteksi kata yang salah ketik baik itu di awal maupun di akhir *string*, sehingga membuktikan bahwa algoritma ini lebih akurat dan efisien dengan nilai 4,0% - 18,6% lebih tinggi dibandingkan algoritma *Jaro-Winkler*.

Berdasarkan latar belakang tersebut, maka penelitian dilakukan untuk menghasilkan sebuah perangkat lunak yang dapat mendeteksi kemiripan berbahasa

Indonesia menggunakan algoritma *Ratcliff/Obershelp* dan proses *stemming* dengan algoritma Nazief dan Adriani pada tahapan *pre-processing text*.

1.3 Rumusan Masalah

Berdasarkan latar belakang di atas, adapun rumusan masalah yang akan diselesaikan dalam penelitian ini adalah:

1. Bagaimana membangun perangkat lunak yang dapat mendeteksi kemiripan teks berbahasa Indonesia menggunakan algoritma *Ratcliff/Obershelp*?
2. Berapakah persentase kesalahan perangkat lunak deteksi plagiarisme pada teks berbahasa Indonesia dengan algoritma *Ratcliff/Obershelp* ?

1.4 Tujuan Penelitian

Penelitian ini memiliki tujuan sebagai berikut:

1. Menghasilkan perangkat lunak deteksi plagiarisme pada teks berbahasa Indonesia menggunakan algoritma *Ratcliff/Obershelp* .
2. Mengetahui persentase kesalahan perangkat lunak deteksi kemiripan teks pada teks berbahasa Indonesia dengan algoritma *Ratcliff/Obershelp*.

1.5 Manfaat Penelitian

Penelitian ini diharapkan memiliki manfaat sebagai berikut:

1. Mendapatkan hasil persentase kemiripan teks pada teks pada teks berbahasa Indonesia menggunakan algoritma *Ratcliff/Obershelp*.
2. Dapat digunakan sebagai referensi dalam membangun perangkat lunak deteksi kemiripan teks menggunakan algoritma *Ratcliff/Obershelp*.

1.6 Batasan Masalah

Pada penelitian ini terdapat batasan – batasan sebagai berikut:

1. Data yang digunakan dalam penelitian berupa teks berita berbahasa Indonesia yang didapat dari internet dengan sumber 20 website berbeda yang membahas mengenai topik tertentu, yaitu jabar.inews.id, news.okezone.com, today.line.me/id, cirebon.inews.id, regional.kompas.com, video.tribunnews.com, cirebon.tribunnews.com, zonamahasiswa.id, cnn.indonesia.com, news.detik.com, regional.kontan.co.id, covid19.go.id, bbc.com, kabar24.bisnis.com, nasional.kompas.com, suara.com, dan republika.co.id.
2. Proses *stemming* pada tahapan *pre-processing text* akan dilakukan dengan menggunakan algoritma Nazief dan Andriani.
3. Nilai prediksi persentase didapatkan dengan menggunakan *cosine similarity* untuk dibandingkan dengan hasil persentase algoritma *Ratcliff/Obershelp*.

1.7 Perangkat lunak ketika Penulisan

Adapun perangkat lunak ketika penulisan skripsi ini adalah sebagai berikut :

BAB I PENDAHULUAN

Pada bab pendahuluan akan dibahas mengenai latar belakang masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, dan perangkat lunak pada penulisan. Bab ini akan menjelaskan mengenai gambaran umum dari keseluruhan kegiatan penelitian.

BAB II STUDI PUSTAKA

Pada bab studi pustaka akan membahas penelitian terkait dan dasar teori dalam menyelesaikan masalah – masalah pada penelitian. Studi pustaka bertujuan sebagai landasan dalam pemecahan masalah pembuatan perangkat lunak.

BAB III METODELOGI PENELITIAN

Pada bab metodologi penelitian akan membahas mengenai pengumpulan data, tahapan penelitian, serta metode pengembangan perangkat lunak. Rencana tahapan penelitian akan dijelaskan secara rinci dengan mengacu pada kerangka kerja dan perancangan proyek perangkat lunak agar tercapainya tujuan penelitian.

BAB IV PENGEMBANGAN PERANGKAT LUNAK

Pada bab pengembangan perangkat lunak akan membahas mengenai

metode *Rational Unifed Process* (RUP) dan tahapannya yang akan digunakan dalam mengembangkan perangkat lunak.

BAB V HASIL DAN ANALISIS

Pada bab hasil dan analisis akan membahas mengenai data hasil percobaan atau penelitian dan analisis hasil penelitian yang telah dilakukan.

BAB VI KESIMPULAN DAN SARAN

Pada bab kesimpulan dan saran akan membahas mengenai kesimpulan dan saran yang diperoleh dari penelitian yang telah dilakukan.

1.8 Kesimpulan

Pada bab pendahuluan ini dapat disimpulkan bahwa penelitian dilakukan untuk membangun sebuah perangkat lunak deteksi plagiarisme pada teks berbahasa Indonesianesia menggunakan algoritma *Ratcliff/Obershelp* dan *stemming* dengan algoritma Nazief dan Adriani pada tahapan *pre-processing text*.

DAFTAR PUSTAKA

- Adriani, M., Nazief, B., et al. 2007. *Stemming Indonesian : A Confix-Stripping Approach*. *ACM Transaction on Asian Language Information Processing*, Vol.6, No.4, Article 13.
- Agustina, S., et al. 2008. *Anti Plagiarism Application with Algoritma Karp-Rabin at Thesis in Gunadarma University*. Universitas Gunadarma, Depok.
- Bird, S., et al. 2009. *Natural Language Processing with Python*. O'Reilly Media, Inc. : 1005 Gravenstein Highway North, Sebastopol, CA.
- Firdaus, et al. 2019. Implementasi *Cosine Similarity* Untuk Peningkatan Akurasi Pengukuran Kesamaan Dokumen Pada Klasifikasi Dokumen Berita Dengan K Nearest Neighbour. *Jurnal Teknologi Informasi dan Komunikasi*, Volume 9, Nomor 1:69-74.
- Ilyankou, I. 2014. *Comparison of Jaro-Winkler and Ratcliff/Obershelp Algorithms in Spell Check*. *IB Extended Essay Computer Science*:11-13.
- Joane, Y.L., et al. 2017. Aplikasi Deteksi Kemiripan Dokumen Teks Menggunakan Algoritma *Ratcliff/Obershelp*. *E-Journal Teknik Informatika*, Vol 11, No.1.
- Jurafsky, D., et al. 2009. *Speech and Language Processing*. Prentice Hall, Englewood Cliffs : New Jersey.
- Kroll, P., et al. 2003. *The Rational Unified Process Made Easy : a practitioner's guide to the RUP*. Person Education : Canada.

- Shadiqi, M.A. 2019. Memahami dan Mencegah Perilaku Plagiarisme dalam Menulis Karya Ilmiah. Buletin Psikologi, Vol 22, No 1:30:42.
- Sulistyaningsih, I. 2017. Plagiarisme, Upaya Pencegahan, Penanggulangan, dan Solusinya. Jurnal Pustaka Ilmiah, Vol 3, No 1.
- Tsuneo, F., et al. 2011. Bab 1 : Pengukuran dan Kesalahan. Instrumentasi Elektronik dan Pengukuran:5-6.
- Yusuf, B., et al. 2019. Perbandingan Algoritma *Rabin-Karp* dan *Ratcliff/Obershelp* untuk Menghitung Kesamaan Teks dalam Bahasa Indonesia. Seminar Nasional APTIKOM (SEMNASTIK):61-69.