

**IMPLEMENTASI ALGORITMA RABIN-KARP UNTUK MENDETEKSI
KEMIRIPAN KATA PADA KARYA TULIS AKADEMIK**

SKRIPSI



disusun oleh :

Fail Amir

17.11.1326

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2021**

**IMPLEMENTASI ALGORITMA RABIN-KARP UNTUK MENDETEKSI
KEMIRIPAN KATA PADA KARYA TULIS AKADEMIK**

SKRIPSI

untuk memenuhi sebagian persyaratan
mencapai gelar Sarjana
pada Program Studi Informatika



disusun oleh :

Fail Amir

17.11.1326

**PROGRAM SARJANA
PROGRAM STUDI SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2021**

PERSETUJUAN

SKRIPSI

IMPLEMENTASI ALGORITMA RABIN-KARP UNTUK MENDETEKSI KEMIRIPAN KATA PADA KARYA TULIS AKADEMIK

yang dipersiapkan dan disusun oleh

Fail Amir

17.11.1326

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 14 April 2021

Dosen Pembimbing,



Windha Mega PD, M.Kom.

NIK. 190302185

PENGESAHAN

SKRIPSI

**IMPLEMENTASI ALGORITMA CNN UNTUK DETEKSI
CITRA DIGITAL YANG TELAH DIMANIPULASI**

yang dipersiapkan dan disusun oleh

Fail Amir

17.11.1326

telah dipertahankan di depan Dewan Penguji
pada tanggal 21 April 2021

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Dwi Nurani, M.Kom

NIK. 190302236

Eli Pujastuti, M.Kom

NIK. 190302227

Windha Mega Pradnya, M.Kom

NIK. 190302148

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 21 April 2021

DEKAN UNIVERSITAS AMIKOM YOGYAKARTA

Hanif Al Fatta, M.Kom

NIK. 190302096

PERNYATAAN

Saya yang bertandatangan dibawah ini menyatakan bahwa, skripsi ini merupakan karya saya sendiri (ASLI), dan isi dalam skripsi ini tidak terdapat karya yang pernah diajukan oleh orang lain untuk memperoleh gelar akademis di suatu institusi pendidikan tinggi manapun, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis dan/atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Segala sesuatu yang terkait dengan naskah dan karya yang telah dibuat adalah menjadi tanggungjawab saya pribadi.

Yogyakarta, 15 Juni 2021



Fail Amir

NIM. 17.11.1326

MOTTO

”Isy Kariman Awmut Syahidan”



PERSEMBAHAN

Alhamdulillah penulis panjatkan puji-syukur kepada Allah SWT atas segala rahmat, taufiq, serta hidayah-Nya, sehingga diberi kesempatan untuk dapat menyelesaikan skripsi ini dengan sebaik-baiknya dengan segala kekurangan penulis. Segala syukur penulis ucapkan kepada-Mu karena telah menghadirkan mereka yang memberikan semangat dan doa disaat menjalani proses pembuatan skripsi ini. Dengan segala kerendahan hati penulis persembahkan skripsi ini kepada :

1. Kedua Orang Tua, Bapak Jumanto Dan Ibu Sarningsih yang selalu mendoakan, memberi semangat serta motivasi supaya dapat menyelesaikan skripsi ini dengan lancar serta bermanfaat bagi semua.
2. Saudariku Fatimah yang tidak berhenti untuk memberikan semangat.
3. Ibu Windha Mega, M.Kom. selaku dosen pembimbing dalam skripsi ini yang tidak lelah untuk tetap membimbing dan mengingatkan penulis dari awal hingga akhir proses pembuatan skripsi.
4. Dosen-dosen Universitas Amikom Yogyakarta yang telah memberikan banyak ilmu baik ilmu akademik maupun ilmu non-akademik selama kuliah.
5. Keluarga besar kelas 17-S1IF-06 yang telah bersama-sama menemani selama kuliah. Semoga silaturahmi kita tetap terjaga.
6. Serta orang-orang yang selalu membantu peneliti dalam mengerjakan skripsi yang tidak bisa disebut namanya satu-persatu.

Penulis ucapkan terima kasih yang sebesar-besarnya untuk kalian semua. Sukses untuk kalian semua, semoga Allah SWT memberikan rahmat dan hidayah-Nya kepada kita semua. Dan semoga skripsi ini dapat bermanfaat dan berguna untuk kemajuan ilmu pengetahuan kedepannya.

KATA PENGANTAR

Alhamdulillah penulis panjatkan puji syukur kepada Allah SWT atas segala rahmat, taufiq, serta hidayah-Nya kepada penulis sehingga dapat menyelesaikan skripsi yang berjudul “Implementasi Algoritma Rabin-Karp untuk Mendeteksi Kemiripan Kata Pada Karya Tulis Akademik”.

Selama proses pengerjaan skripsi ini penulis menyadari bahwa dalam proses penulisan skripsi ini banyak mengalami kendala, namun berkat bantuan, bimbingan, kerjasama dari berbagai pihak dan berkah dari Allah SWT sehingga kendala-kendala tersebut bisa diatasi. Selanjutnya ucapa terima kasih penulis sampaikan kepada :

1. Bapak Prof. Dr. M. Suyanto, M.M selaku Rektor Universitas Amikom Yogyakarta.
2. Ibu Windha Mega, M.Kom. selaku dosen pembimbing yang telah memberikan banyak masukan yang membantu membimbing dalam menyelesaikan skripsi ini.
3. Bapak Hanif Al Fatta, M.Kom selaku Dekan Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.
4. Bapak Sudarmawan, M.T selaku Kepala Prodi Informatika Universitas Amikom Yogyakarta.
5. Dosen Penguji (Bpk/Ibu Dosen Penguji, Bpk/Ibu Dosen Penguji, Bpk/Ibu Dosen Penguji, Bpk/Ibu Dosen Penguji) yang telah memberikan masukan terhadap penelitian ini.
6. Kedua orang tua dan keluarga yang selalu memberikan doa, dukungan dan semangat.
7. Keluarga Besar 17-S1IF-06.
8. Serta semua pihak yang tidak bisa penulis sebutkan satu-persatu yang telah membantu dalam penyusunan skripsi ini.

Penulis Menyadari bahwa masih banyak terdapat kekurangan-kekurangan dalam mengerjakan skripsi ini, sehingga penulis mengharapkan adanya saran dan kritik yang membangun demi kesempurnaan skripsi ini.

Yogyakarta, 10 November 2021

Fail Amir
NIM. 17.11.1326

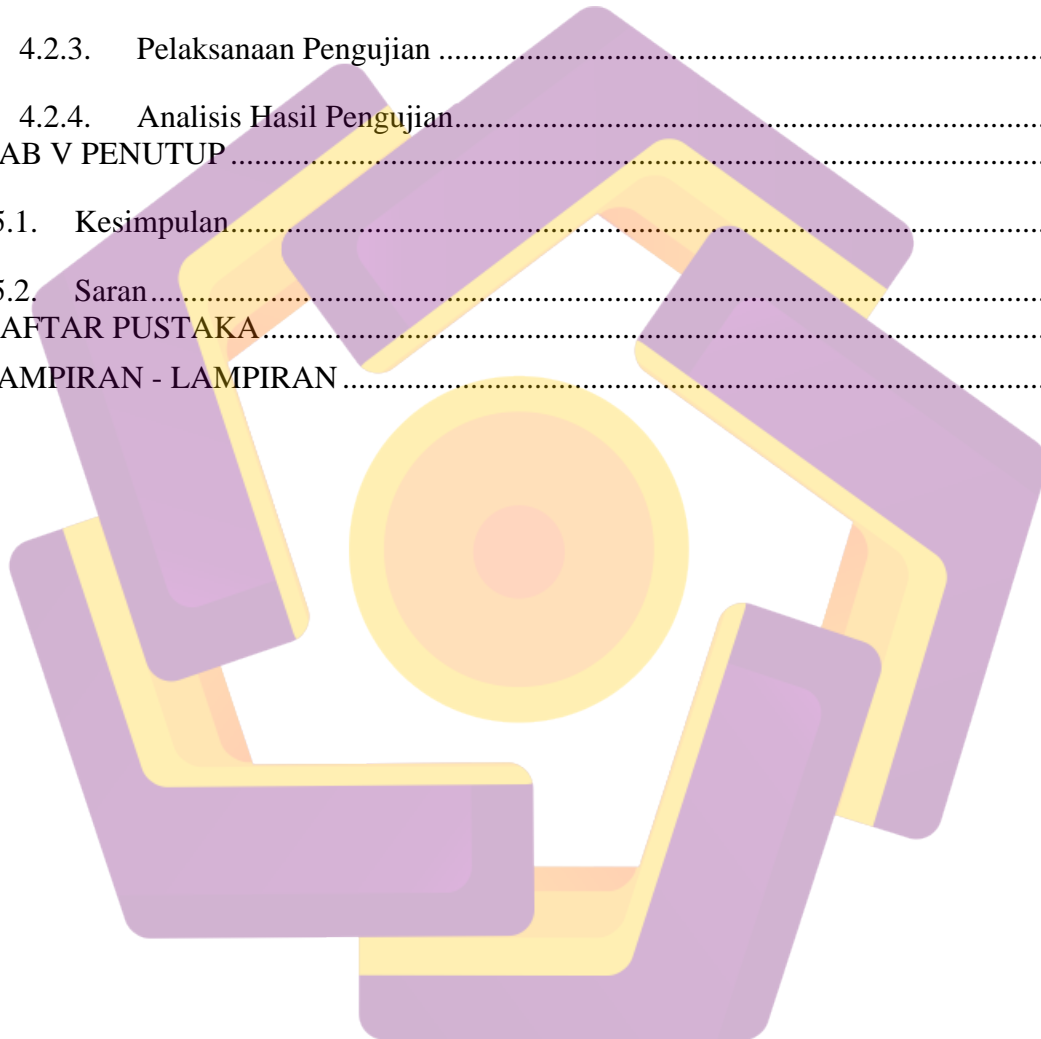
DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERNYATAAN KEASLIAN SKRIPSI.....	ii
HALAMAN PENGESAHAN	iii
HALAMAN PENGESAHAN	iv
ABSTRAK.....	v
ABSTRACT	vi
KATA PENGANTAR	vii
DAFTAR ISI	viii
DAFTAR GAMBAR.....	xii
DAFTAR TABEL	xiv
DAFTAR KODE	xv
DAFTAR LAMPIRAN	xvi
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	2
1.3. Tujuan dan Manfaat.....	2
1.4. Ruang Lingkup	3
1.5. Sistematika Penulisan.....	3
BAB II DASAR TEORI	5
2.1. Plagiarisme	5
2.1.1. Bentuk Plagiarisme.....	6
2.1.2. Metode Pendeteksi Plagiarisme.....	7
2.2. Metode Pencocokkan String.....	8
2.2.1. Algoritma pada Metode Pencocokkan <i>string</i>	9
2.2.1. Pencocokkan String dengan Algoritma Rabin-Karp	10
2.3. Rolling <i>Hash</i>	14

2.4.	Tahap <i>Text</i> Preprocessing.....	16
2.5.	Persentase Kemiripan	19
2.6.	Parsing K-Gram.....	18
2.7.	Tahap Pengembangan Sistem.....	19
2.8.	Data Flow Diagram (DFD).....	22
2.9.	Entity Relation Diagram (ERD)	22
2.10.	Flowchart.....	23
2.11.	MySQL.....	24
2.12.	PHP.....	25
2.13.	Pengukuran dan Kesalahan	25
BAB III SPESIFIKASI, ANALISIS, DAN PERANCANGAN.....		27
3.1.	Gambaran Umum Sistem	27
3.2.	Spesifikasi dan Analisis Kebutuhan Sistem	34
3.2.1.	Spesifikasi Sistem.....	34
3.2.2.	Permodelan Data.....	34
3.2.2.	Permodelan Fungsional	35
3.2.2.1.	Data Context Diagram	35
3.2.2.2.	DFD level 1	36
3.2.2.3.	DFD level 2 Sub Proses Preprocessing	38
3.3.	Desain Sistem	39
3.3.1.	Perancangan Proses Sistem Deteksi Kemiripan Kata.....	39
3.3.1.1.	Proses Upload dan Ekstraksi	41
3.3.1.2.	Proses <i>Text</i> Preprocessing	41
3.3.1.2.1.	Proses <i>Filtering</i>	42
3.3.1.2.2.	Proses Whitespace Insensitivity	43
3.3.1.2.3.	Proses Parsing 5-gram	43
3.3.1.3.	Proses Rolling <i>Hash</i>	44

3.3.1.4.	Proses Pencocokkan String dengan Algoritma Rabin-Karp	47
3.3.1.5.	Proses Perhitungan Kemiripan	48
3.3.1.6.	Proses Pencocokkan Kalimat Sama	48
3.3.2.	Perancangan Antarmuka Sistem	49
3.3.2.1.	Antarmuka Halaman Depan	50
3.3.2.2.	Antarmuka Upload Dokumen	50
3.3.2.3.	Antarmuka Pemrosesan Dokumen	51
3.3.2.4.	Antarmuka Hasil Pengujian	51
3.3.2.5.	Antarmuka Daftar Dokumen yang Diuji	52
3.3.2.6.	Antarmuka Stopwords	52
3.3.2.7.	Antarmuka Bantuan	53
BAB IV IMPLEMENTASI DAN PENGUJIAN		54
4.1.	Implementasi	54
4.1.1.	Spesifikasi Perangkat	54
4.1.2.	Implementasi Basis Data	54
4.1.3.	Implementasi Fungsi	55
4.1.3.1.	Fungsi Upload dan Ekstraksi	55
4.1.3.2.	Fungsi <i>Text Preprocessing</i>	57
4.1.3.3.	Fungsi <i>Rolling Hash</i>	58
4.1.3.4.	Fungsi Pencocokkan String	59
4.1.3.5.	Fungsi Perhitungan Kemiripan	59
4.1.3.6.	Fungsi Pencocokkan Kalimat Sama	60
4.1.4.	Implementasi Antarmuka	61
4.1.4.1.	Implementasi Antarmuka Halaman Depan	61
4.1.4.2.	Implementasi Antarmuka Upload Dokumen	61
4.1.4.3.	Implementasi Antarmuka Pemrosesan Dokumen	62
4.1.4.4.	Implementasi Antarmuka Hasil Pengujian	62

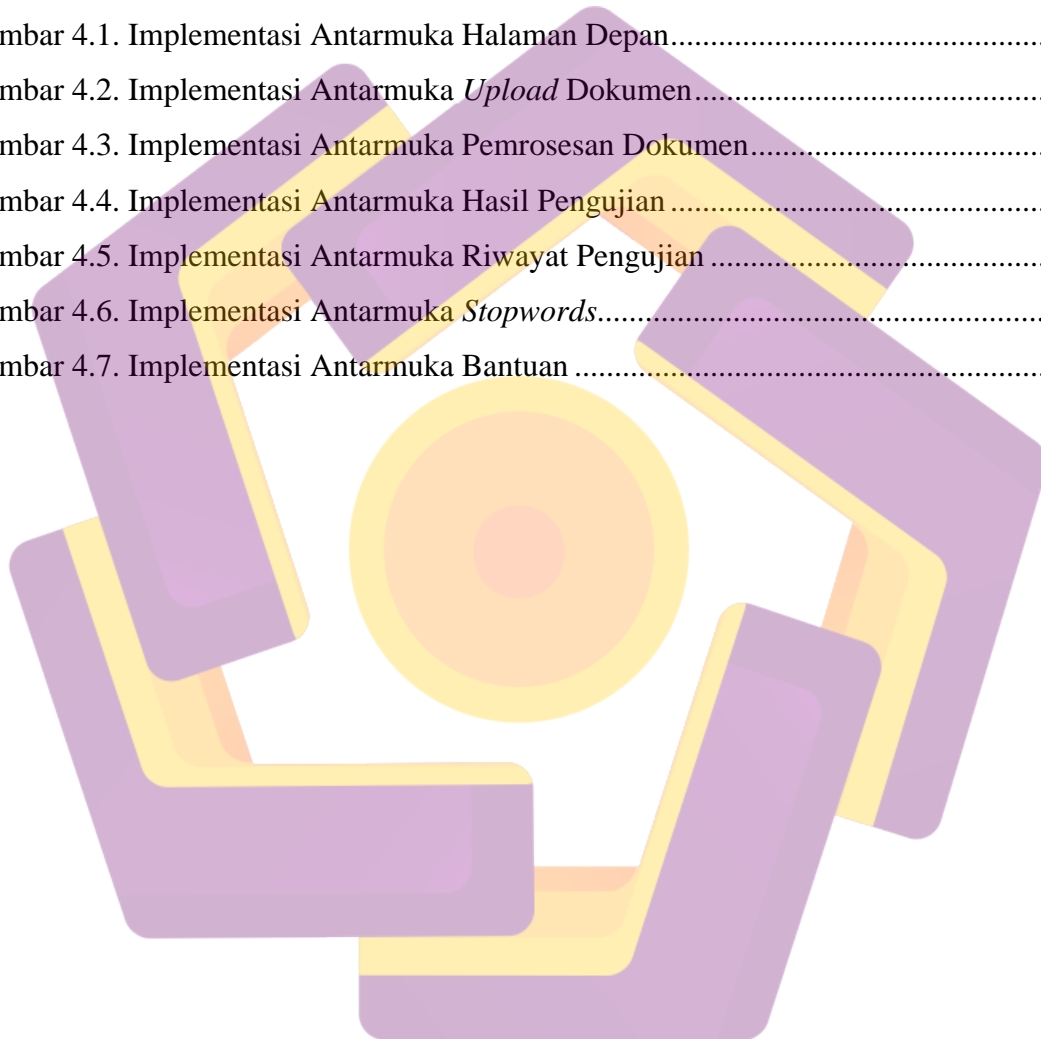
4.1.4.5.	Implementasi Antarmuka Riwayat Pengujian.....	63
4.1.4.6.	Implementasi Antarmuka Stopwords	63
4.1.4.7.	Implementasi Antarmuka Bantuan.....	64
4.2.	Pengujian	65
4.2.1.	Lingkungan Pengujian.....	65
4.2.2.	Rencana Pengujian	65
4.2.3.	Pelaksanaan Pengujian	67
4.2.4.	Analisis Hasil Pengujian.....	71
BAB V PENUTUP		72
5.1.	Kesimpulan.....	72
5.2.	Saran.....	72
DAFTAR PUSTAKA		73
LAMPIRAN - LAMPIRAN		75



DAFTAR GAMBAR

Gambar 2.1. Skema Metode Pendeteksi Plagiarisme	7
Gambar 2.2. Contoh <i>String Matching</i>	10
Gambar 2.3. <i>Fingerprint</i> awal	11
Gambar 2.4. Menggeser <i>fingerprint</i>	11
Gambar 2.5. Perbandingan kedua	12
Gambar 2.6. Perbandingan ketiga	12
Gambar 2.7. Perbandingan keempat (nilai <i>hash</i> sama)	12
Gambar 2.8. Perbandingan kelima (<i>string</i> ditemukan)	13
Gambar 2.9. Contoh <i>Whitespace Insensitivity text</i>	16
Gambar 2.10. Contoh <i>Parsing text</i>	16
Gambar 2.11. Contoh <i>Filtering text</i>	17
Gambar 2.12. Contoh <i>Stemming text</i>	17
Gambar 2.13. <i>Parsing 5-gram</i>	18
Gambar 2.14. Model <i>Waterfall</i>	20
Gambar 3.1 Gambaran Umum Sistem Deteksi Kemiripan Kata	27
Gambar 3.2 ERD Sistem Deteksi Kemiripan Kata	35
Gambar 3.3 <i>Data Context Diagram</i>	35
Gambar 3.4. <i>Data Flow Diagram</i> level 1	36
Gambar 3.5. DFD level 2 Proses <i>Preprocessing</i>	38
Gambar 3.6. <i>Flowchart</i> Sistem Deteksi Kemiripan pada Dua Dokumen	40
Gambar 3.7 <i>Flowchart</i> Proses <i>Upload</i> dan <i>Ekstraksi</i>	41
Gambar 3.8. <i>Flowchart</i> Proses <i>Preprocessing</i>	42
Gambar 3.9. <i>Flowchart</i> Proses <i>Filtering</i>	42
Gambar 3.10. <i>Flowchart</i> Proses <i>Whitespace Insensitivity</i>	43
Gambar 3.11. <i>Flowchart</i> Proses <i>Parsing 5-gram</i>	44
Gambar 3.12. <i>Flowchart</i> Proses <i>Hashing</i>	45
Gambar 3.13. <i>Flowchart</i> Proses <i>Rolling Hash</i>	45
Gambar 3.14. <i>Flowchart</i> Proses Pergeseran <i>Hash</i>	46
Gambar 3.15 <i>Flowchart</i> Perancangan Proses Pencocokkan String Algoritma <i>Rabin-Karp</i>	47
Gambar 3.16. <i>Flowchart</i> Proses Hitung Kemiripan	48
Gambar 3.17. <i>Flowchart</i> Proses Pencocokkan Kalimat Sama	49

Gambar 3.18. Desain Antarmuka Halaman Depan	50
Gambar 3.19. Desain Antarmuka <i>Upload</i> Dokumen	50
Gambar 3.20. Desain Antarmuka Pemrosesan Dokumen	51
Gambar 3.21. Desain Antarmuka Hasil Pengujian.....	52
Gambar 3.22. Desain Antarmuka Daftar Dokumen yang Diuji	52
Gambar 3.23. Desain Antarmuka <i>Stopwords</i>	53
Gambar 3.24. Desain Antarmuka Bantuan.....	53
Gambar 4.1. Implementasi Antarmuka Halaman Depan.....	61
Gambar 4.2. Implementasi Antarmuka <i>Upload</i> Dokumen.....	62
Gambar 4.3. Implementasi Antarmuka Pemrosesan Dokumen.....	62
Gambar 4.4. Implementasi Antarmuka Hasil Pengujian	63
Gambar 4.5. Implementasi Antarmuka Riwayat Pengujian	63
Gambar 4.6. Implementasi Antarmuka <i>Stopwords</i>	64
Gambar 4.7. Implementasi Antarmuka Bantuan	64

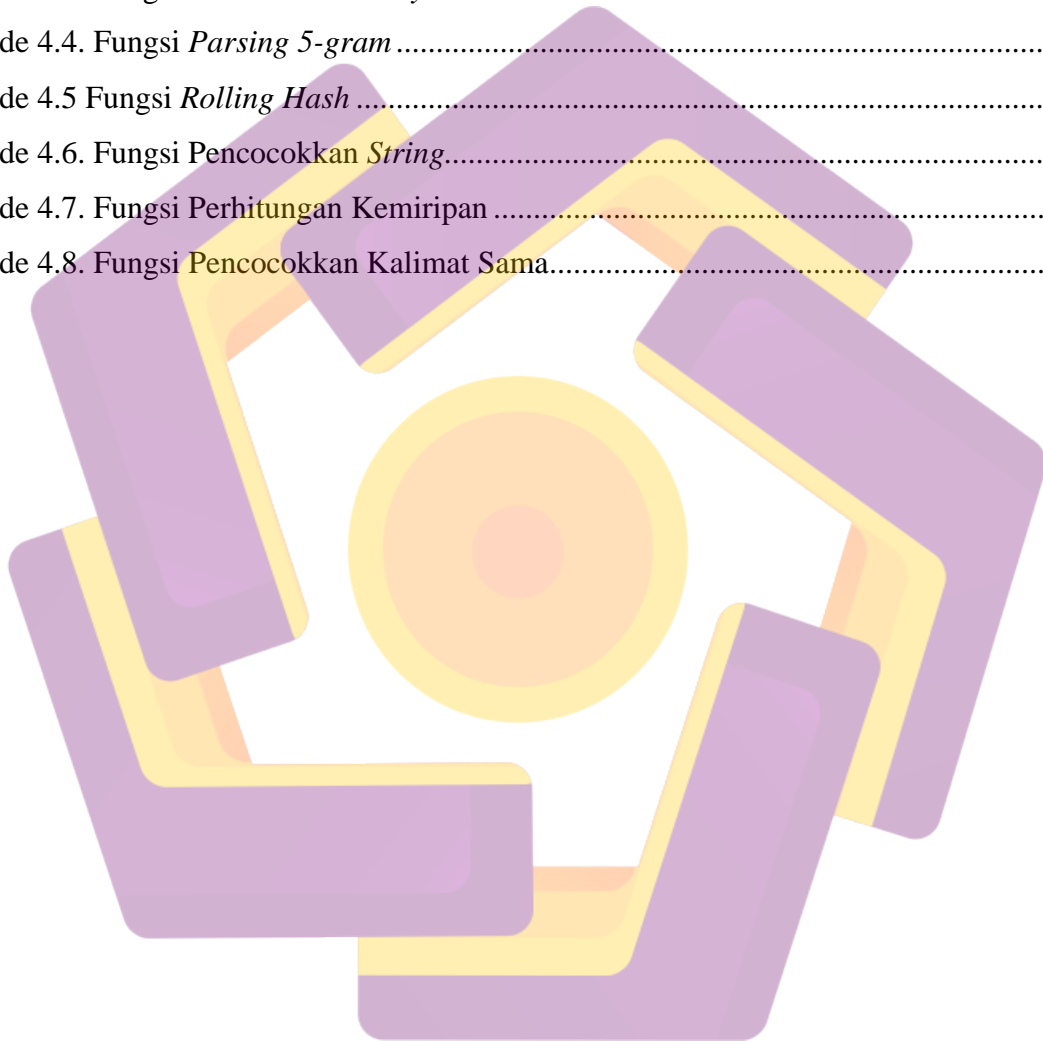


DAFTAR TABEL

Tabel 2.1 Simbol-simbol pada DFD	22
Tabel 2.2 Tabel Notasi <i>Entity Relation Diagram</i> (ERD)	23
Tabel 2.3 Simbol-simbol pada <i>Flowchart</i>	24
Tabel 3.1. Contoh Dokumen.....	30
Tabel 3.2. Dokumen hasil <i>filtering</i>	30
Tabel 3.3 Dokumen hasil <i>whitespace insensitivity</i>	30
Tabel 3.4 <i>Term</i> hasil <i>Parsing 5-gram</i> Dokumen Asli	31
Tabel 3.5 <i>Term</i> hasil <i>Parsing 5-gram</i> Dokumen Uji.....	31
Tabel 3.6. Nilai ASCII untuk kata “algor” dan “lgori”	31
Tabel 3.7 <i>Term</i> hasil <i>Rolling Hash</i> Dokumen Asli	32
Tabel 3.8. <i>Term</i> hasil <i>Rolling Hash</i> Dokumen Uji.....	33
Tabel 3.9 Hasil Pencocokkan Kalimat Sama.....	34
Tabel 3.10. Spesifikasi Kebutuhan Fungsional	34
Tabel 4.1. Daftar Tabel Sistem Deteksi Kemiripan.....	55
Tabel 4.2. Struktur Tabel DOKUMEN	55
Tabel 4.3. Struktur Tabel SIMILARITY	55
Tabel 4.4. Struktur Tabel STOPWORDS.....	55
Tabel 4.5 Rencana Pengujian Fungsional Sistem Deteksi Kemiripan	66
Tabel 4.6. Data Pengujian Validitas Sistem Deteksi Kemiripan	66
Tabel 4.7. Hasil dan Evaluasi Pengujian Sistem Deteksi Kemiripan.....	68
Tabel 4.8 Hasil Pengujian Validitas Sistem Deteksi Kemiripan	70

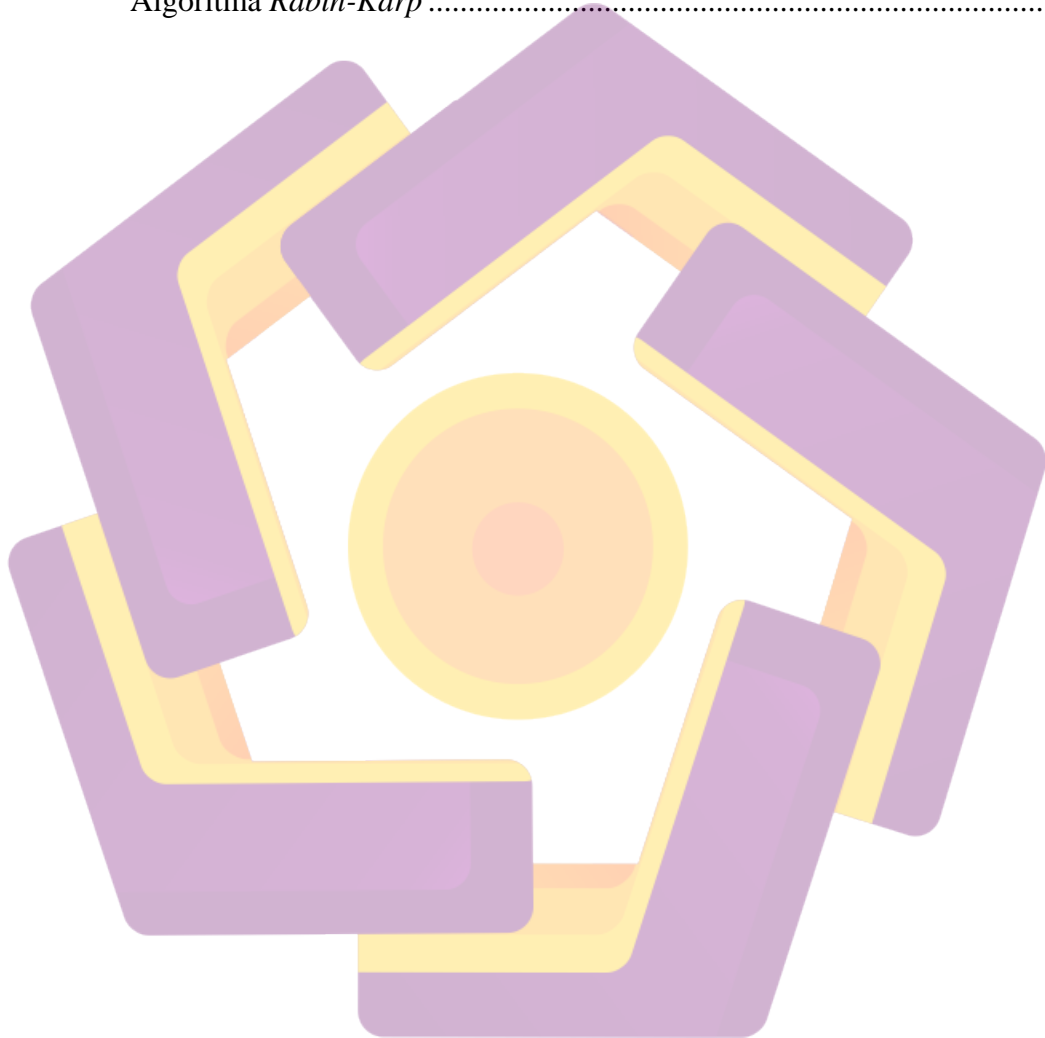
DAFTAR KODE

Kode 2.1. <i>Pseudocode</i> Algoritma <i>Rabin-Karp</i>	13
Kode 4.1. Fungsi <i>Upload</i> dan Ekstraksi.....	56
Kode 4.2. Fungsi <i>Filtering</i>	57
Kode 4.3. Fungsi <i>White Insensitivity</i>	57
Kode 4.4. Fungsi <i>Parsing 5-gram</i>	58
Kode 4.5 Fungsi <i>Rolling Hash</i>	59
Kode 4.6. Fungsi Pencocokkan <i>String</i>	59
Kode 4.7. Fungsi Perhitungan Kemiripan	60
Kode 4.8. Fungsi Pencocokkan Kalimat Sama.....	60



DAFTAR LAMPIRAN

Lampiran 1. Tabel Daftar <i>Stopwords</i>	76
Lampiran 2. Daftar Tanda Baca	77
Lampiran 3. Flowchart Sistem Deteksi Kemiripan Kata Pada Dua Dokumen menggunakan Algoritma <i>Rabin-Karp</i>	78



INTISARI

Munculnya berbagai kasus penjiplakan yang terjadi di beberapa perguruan tinggi Indonesia telah menjadi masalah tersendiri untuk dipecahkan. Dengan pencegahan dan pendeteksian dini bisa mengurangi kegiatan plagiasi dokumen. Deteksi secara manual sangat sulit jika dilakukan, jadi diperlukan sistem komputerisasi pendeteksi plagiasi secara cepat dan akurat untuk mendeteksi dan mengukur tingkat plagiasi dua dokumen.

Metode yang digunakan untuk membangun sistem adalah metode String Matching dengan menggunakan algoritma Rabin-Karp. Algoritma tersebut mencocokkan pecahan rangkaian kata 5-gram yang telah diubah menjadi nilai *hash*. Nilai *hash* yang sama menghasilkan persentase kemiripan kata.

Pencocokan kalimat sama digunakan untuk mengindikasikan keberadaan kalimat sama. Data uji pada penelitian ini menggunakan 10 dokumen publikasi dari berbagai sumber yang memiliki keterkaitan tema. Pengujian sistem sebanyak 20 dokumen percobaan menghasilkan tingkat keakuratan sebesar 93%.

Kata Kunci: *Plagiarisme, Algoritma Rabin-Karp, String Matching, Multi Pattern.*

ABSTRACT

The emergence of various plagiarism cases that have occurred in several universities has become a problem that cannot solve the problem. With prevention and detection, it can reduce plagiarism of documents. Manual detection is very difficult to do, so a computerized detection system is needed to quickly and accurately monitor and measure the plagiarism level of two documents.

The method used to build the system is the String Matching method using the Rabin-Karp algorithm. The algorithm matches the fraction of a 5-gram word string that has been converted into a hash value. The same hash value results in a similarity proportion of words.

Sentence match is used to indicate the existence of the same sentence. Test data using 10 published documents from various sources that have related themes. System testing of 20 experimental documents resulted in an accuracy rate of 93%.

Keywords: Plagiarism, Rabin-Karp Algorithm, String Matching, Multi Pattern.