

**PERANCANGAN SISTEM DETEKSI PLAGIAT PADA DOKUMEN TEKS
DENGAN KONSEP SIMILARITY MENGGUNAKAN
ALGORITMA RABIN KARP**

SKRIPSI



disusun oleh :

Salmuasih

08.11.2281

**JURUSAN TEKNIK INFORMATIKA
SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER
AMIKOM
YOGYAKARTA
2013**

**PERANCANGAN SISTEM DETEKSI PLAGIAT PADA DOKUMEN TEKS
DENGAN KONSEP SIMILARITY MENGGUNAKAN
ALGORITMA RABIN KARP**

SKRIPSI

untuk memenuhi sebagian persyaratan
mencapai derajat Sarjana T1
pada jurusan Teknik Informatika



disusun oleh :

Salmuasih

08.11.2281

**JURUSAN TEKNIK INFORMATIKA
SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER
AMIKOM
YOGYAKARTA
2013**

PERSETUJUAN

SKRIPSI

**PERANCANGAN SISTEM DETEKSI PLAGIAT PADA DOKUMEN TEKS
DENGAN KONSEP SIMILARITY MENGGUNAKAN
ALGORITMA RABIN KARP**

yang dipersiapkan dan disusun oleh

Salmuasih

08.11.2281

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 25 Januari 2013

Dosen Pembimbing,


Andi Sunvoto, M.Kom.

NIK. 190302052

PENGESAHAN

SKRIPSI

**PERANCANGAN SISTEM DETEKSI PLAGIAT PADA DOKUMEN TEKS
DENGAN KONSEP SIMILARITY MENGGUNAKAN
ALGORITMA RABIN KARP**

yang dipersiapkan dan disusun oleh

Salmuasih

08.11.2281

telah dipertahankan di depan Dewan Penguji
pada tanggal 21 Februari 2013

Susunan Dewan Penguji

Nama Penguji

Andi Sunyoto, M.Kom.

NIK. 190302052

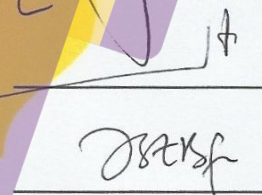
Dr. Kusrini, M.Kom.

NIK. 190302106

Yuli Astuti, M.Kom.

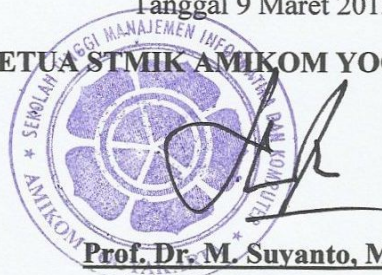
NIK. 190302146

Tanda Tangan



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh Sarjana Komputer
Tanggal 9 Maret 2013

KETUA STMIK AMIKOM YOGYAKARTA



Prof. Dr. M. Suyanto, M. M.

NIK. 190302001

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, 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.



HALAMAN MOTTO

Man jadda wa jadda

Sesungguhnya bersama kesulitan itu ada kemudahan

QS. Al Insyirah : 6

Saat sebuah urusan telah terselesaikan mulailah untuk mengerjakan urusan yang lain dengan sungguh-sungguh

QS. Al Insyirah : 8

Sebaik-baik kalian adalah yang paling bermanfaat bagi orang lain

HR. Tirmidzi

"...sesungguhnya yang benar-benar takut kepada Allah, adalah hamba-hamba-Nya yang mempunyai ilmu.."

QS. Fathir : 28

HALAMAN PERSEMBAHAN

Skripsi ini kupersembahkan dengan rasa syukur kepada saksi dan penolong seumur hidupku, Allah SWT

Untuk Rama, Biyung dan Jibal *jazakumullah khairan katsiir* atas pengertian dan harapan kalian, semoga Allah kuatkan kami untuk selalu berbakti..

Untuk mba Dyah Fajar Nur Rohmah *jazakillah* atas bimbingan, ilmu, alur berfikir, kerja keras dan pengorbanannya..

Jazakillah untuk mba Nevee yang selalu setia 'ngalor ngidul' bareng ngurus skripsi..

Wida, *jazakillah* editingnya. Untuk mba eRlinD, Istiqomah, Pi-Shop, Ka' Maning, Ka' Pal, Ka' aDew, Eyna, Arumy, Ragil *jazakumullah* atas *ukhuwah* dan *support* nya..

Jazakumullah untuk murobbi dan teman-teman lingkaranku, semoga kita tetap *istiqomah*..

Tuan rumah di Tiga Piranha : Ammah Eiya, mba Fajar 'DyahSaga', Pi-Shop, *jazakumullah* atas kehangatan kalian..

Mba-mba dan teman-teman di Baren lante 4 : mba ella Ndul, mba Restu, mba Nina, Tri, Mput, Esa, mba Rina, mba Nin-chan, Sasha, mba Anas, Dilla, Nyla *jazakumullah* untuk kebersamaan kita dan *support* nya..

Tidak lupa untuk teman-teman seperjuangan di Jashtis, kalian luar biasa..

KATA PENGANTAR

Bismillaahirrahmaanirrahiim

Alhamdulillah, syukur kehadiran Allah SWT yang selalu melimpahkan segala rahmat, nikmat, dan petunjuk-Nya sehingga skripsi ini akhirnya dapat terselesaikan. Sholawat teriring salam penulis persembahkan kepada manusia paling berpengaruh nomor satu di dunia, Rosulullah saw, yang ajarannya tetap murni dan diamalkan sampai detik ini.

Skripsi ini disusun untuk memenuhi salah satu persyaratan kelulusan di STMIK AMIKOM Yogyakarta. Mengangkat judul “Perancangan Sistem Deteksi Plagiat pada Dokumen Teks dengan Konsep Similarity Menggunakan Algoritma Rabin Karp”, skripsi ini dimaksudkan agar mahasiswa dapat menerapkan analisis dan pengetahuan lain terkait bidang *study*.

Banyak pihak yang telah mendukung terselesaikannya skripsi ini, sehingga pada kesempatan ini penulis mengucapkan banyak terima kasih kepada :

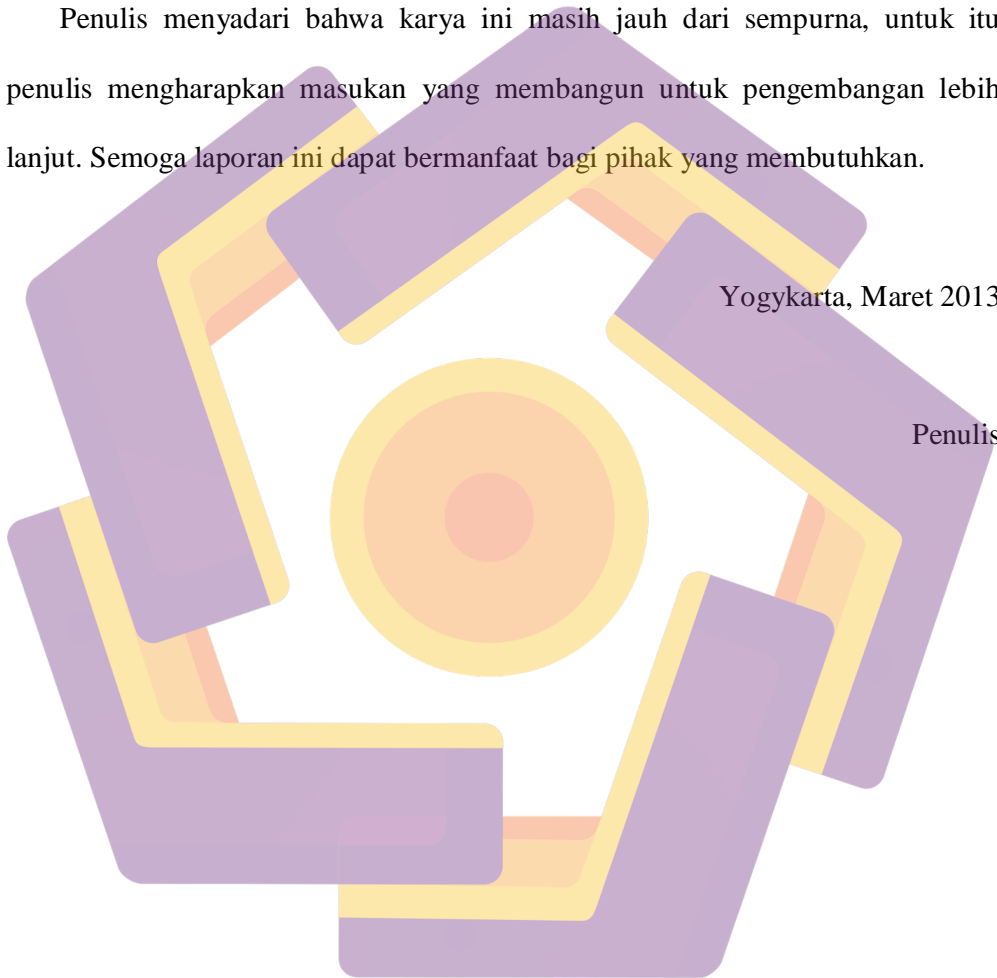
3. Bapak Prof. DR. Mohammad Suyanto, MM selaku ketua STMIK AMIKOM Yogyakarta
4. Bapak Andi Sunyoto, M. Kom. Selaku dosen pembimbing yang selalu memberikan waktu, bimbingan dan ilmu yang bermanfaat.
5. Dosen-dosen STMIK AMIKOM Yogyakarta yang telah memberikan ilmu-ilmu yang bermanfaat.
6. Kedua orangtua yang selalu menuntun dan memberikan kepercayaan kepada penulis sampai saat ini
7. Keluarga besar UKI Jashtis

8. Keluarga besar Asrama Pocut Baren
9. Keluarga besar S1 Teknik Informatika 2008 kelas F
10. Semua pihak yang telah mendukung kelancaran penyusunan skripsi ini yang tidak dapat dituliskan satu persatu.

Penulis menyadari bahwa karya ini masih jauh dari sempurna, untuk itu penulis mengharapkan masukan yang membangun untuk pengembangan lebih lanjut. Semoga laporan ini dapat bermanfaat bagi pihak yang membutuhkan.

Yogyakarta, Maret 2013

Penulis



DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN	Error! Bookmark not defined.
HALAMAN PENGESAHAN	Error! Bookmark not defined.
HALAMAN PERNYATAAN	iv
HALAMAN MOTTO.....	v
HALAMAN PERSEMBAHAN.....	vi
KATA PENGANTAR	vii
DAFTAR ISI.....	ix
DAFTAR TABEL	xi
DAFTAR GAMBAR.....	xii
DAFTAR LAMPIRAN.....	xiv
INTISARI.....	xv
<i>ABSTRACT</i>	xvi
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah.....	3
1.4 Tujuan Penelitian.....	3
1.5 Manfaat Penelitian.....	4
1.6 Metodologi Penelitian.....	4
1.7 Sistematika Penulisan	5
BAB II LANDASAN TEORI	7
2.1 Plagiarisme.....	7
2.2 Pendekatan Deteksi Plagiat	9
2.3 Similarity Dokumen	10
2.4 Text Mining.....	11
2.5 Ekstraksi Dokumen Teks Mining.....	12
2.6 Indonesian Porter Stemmer.....	14
2.7 String Matching	17

2.8	Algoritma Rabin Karp	18
2.9	K-Gram	19
2.10	Hashing	19
2.11	Prinsip kerja algoritma Rabin Karp	20
2.12	Peningkatan Performa Algoritma Rabin Karp	23
2.13	Pengukuran Nilai Similarity.....	27
BAB III ANALISIS DAN PERANCANGAN SISTEM.....		29
3.1	Analisis Sistem	29
3.1.1	Analisis Masalah	30
3.1.2	Analisis Kebutuhan Sistem	30
3.2	Perancangan Sistem	32
3.2.1	Perancangan Proses	34
3.2.2	Perancangan basisdata	40
3.2.3	Perancangan interface.....	43
3.4	Perhitungan Manual.....	46
BAB IV IMPLEMENTASI DAN PEMBAHASAN.....		49
4.1	Implementasi	49
4.1.1	Implementasi database.....	49
4.1.2	Implementasi interface.....	52
4.2	Pengujian Sistem	62
4.2.1	Hasil uji preprocessing dan string matching	62
4.2.2	Hasil uji berdasarkan banyaknya konten file	64
4.2.3	Hasil uji tanpa menggunakan stemming	64
4.2.4	Hasil uji tanpa menggunakan preprocessing.....	64
4.2.5	Hasil uji modulo pada algoritma Rabin-Karp	65
4.2.6	Hasil uji k-gram pada algoritma Rabin-Karp.....	65
BAB V PENUTUP		66
5.1	Kesimpulan	66
5.2	Saran	67
DAFTAR PUSTAKA.....		68

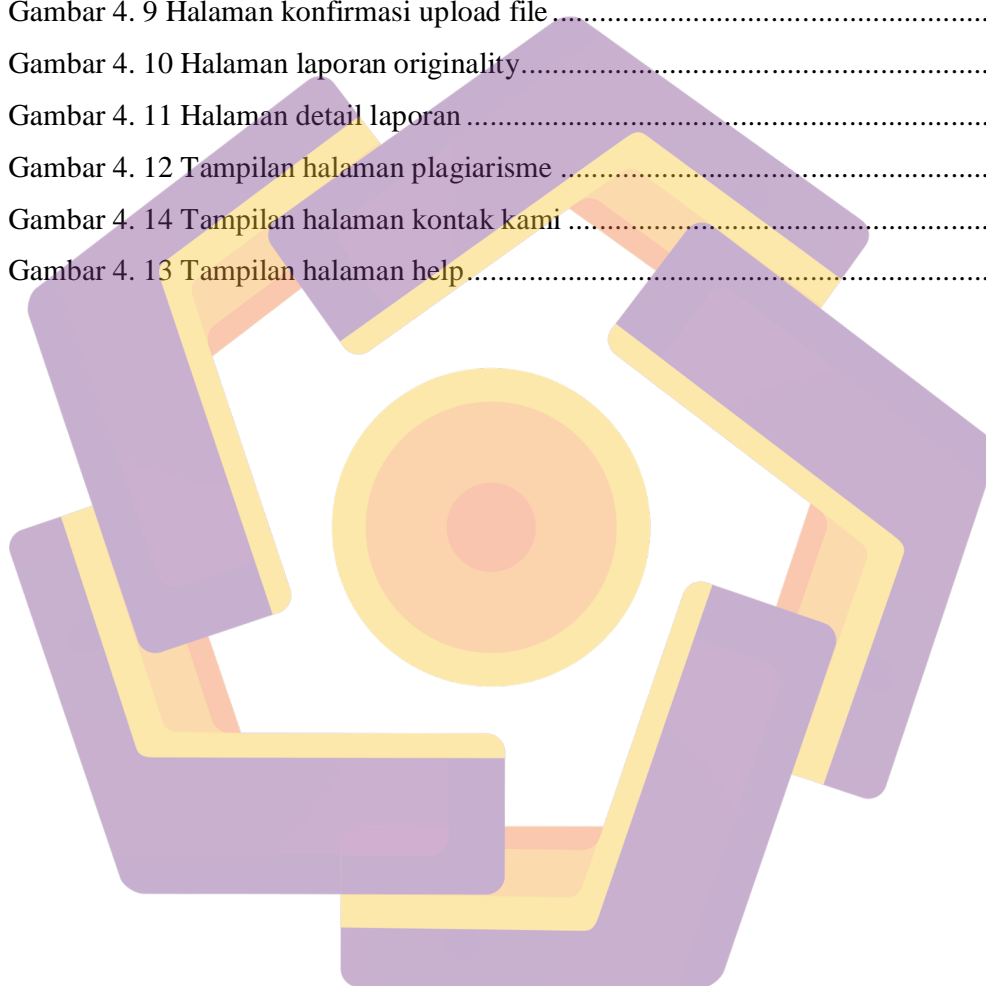
DAFTAR TABEL

Tabel 2. 1 Kelompok rule pertama: <i>inflectional particles</i>	16
Tabel 2. 2 Kelompok rule kedua: <i>inflectional possessive pronouns</i>	16
Tabel 2. 3 Kelompok rule ketiga: <i>first order of derivational prefixes</i>	16
Tabel 2. 4 Kelompok rule keempat: <i>second order of derivational prefixes</i>	16
Tabel 2. 5 Kelompok rule kelima: <i>derivational suffixes</i>	17
Tabel 3. 1 Tabel Use case	35
Tabel 3. 2 Rancangan tabel file	41
Tabel 3. 3 Rancangan tabel stopword.....	41
Tabel 3. 4 Rancangan tabel ktdasar	41
Tabel 3. 5 Rancangan tabel hasil_uji.....	41
Tabel 3. 6 Rancangan tabel detail_similarity.....	42
Tabel 3. 7 Rancangan tabel indexing.....	42
Tabel 3. 8 Hasil parsing k-gram	46
Tabel 3. 9 Hasil penghitungan <i>modulo</i> dan <i>remainder</i> (hasil bagi)	48
Tabel 4. 1 Hasil <i>preprocessing file</i>	62
Tabel 4. 4 Informasi <i>file</i> uji.....	64
Tabel 4. 5 Hasil uji berdasarkan banyaknya konten <i>file</i>	64
Tabel 4. 6 Hasil uji tanpa menggunakan <i>stemming</i>	64
Tabel 4. 7 Hasil uji tanpa menggunakan <i>preprocessing</i>	64
Tabel 4. 8 Informasi <i>file</i> untuk uji <i>modulo</i> dan <i>k-gram</i>	65
Tabel 4. 9 Hasil uji <i>modulo</i>	65
Tabel 4. 10 Hasil uji <i>k-gram</i>	65

DAFTAR GAMBAR

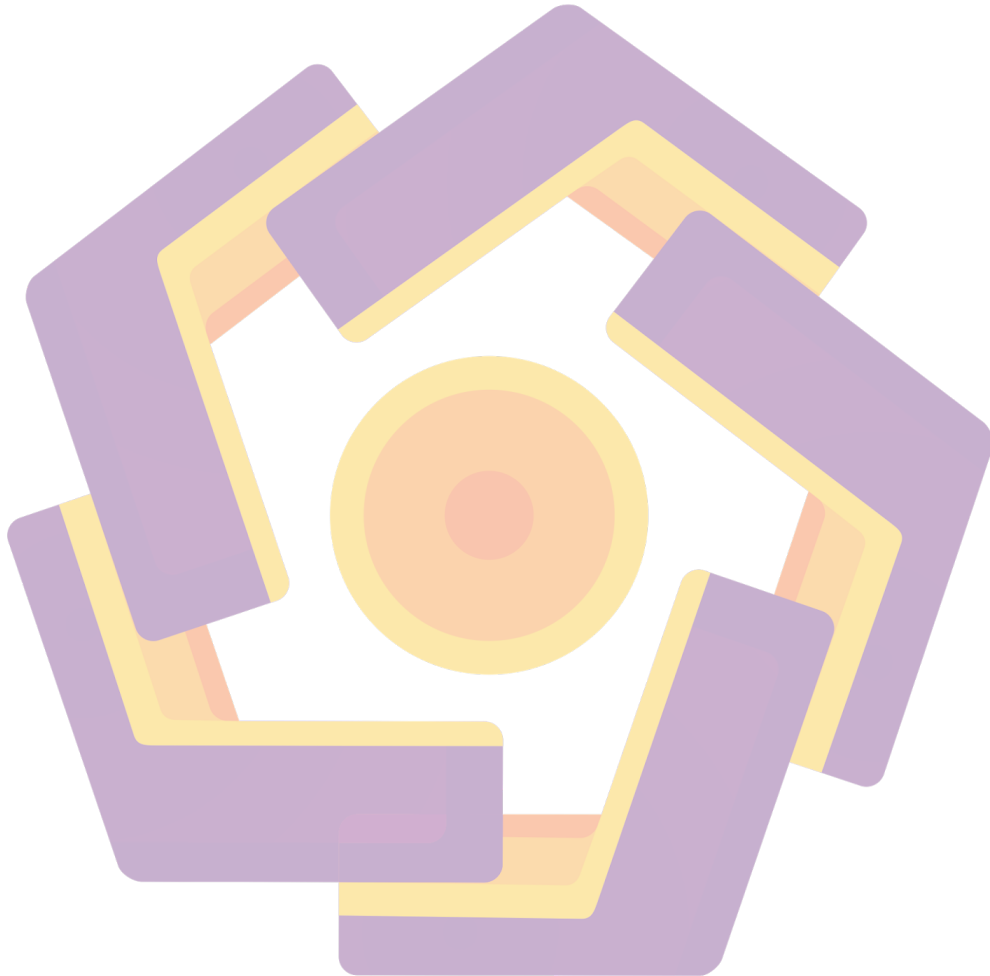
Gambar 2. 1 Klasifikasi Plagiat.....	8
Gambar 2. 2 Tahap <i>preprocessing</i> dokumen	12
Gambar 2. 3 Tahap <i>Tokenizing</i>	13
Gambar 2. 4 Tahap <i>Filtering</i>	13
Gambar 2. 5 Tahap <i>Stemming</i>	14
Gambar 2. 6 Desain Porter Stemmer	15
Gambar 2. 7 Algoritma Rabin Karp	23
Gambar 2. 8 Pengecekan tiga karakter pertama.....	25
Gambar 2. 9 Pengecekan tiga karakter berikutnya.....	25
Gambar 2. 10 Perbandingan <i>pattern</i> “cab” dengan <i>substring</i> “abb”	26
Gambar 2. 11 Perbandingan <i>pattern</i> dan <i>substring</i> dengan nilai <i>hash</i> sama	26
Gambar 2. 12 <i>String</i> ditemukan	26
Gambar 3. 1 Skema aliran data pada sistem	32
Gambar 3. 2 Arsitektur sistem.....	34
Gambar 3. 3 <i>Use case diagram</i>	35
Gambar 3. 4 <i>Activity diagram</i>	37
Gambar 3. 5 <i>Class diagram</i>	38
Gambar 3. 6 <i>Sequence diagram</i>	39
Gambar 3.7 <i>Entity Relational Diagram</i>	40
Gambar 3. 8 Relasi Tabel.....	42
Gambar 3. 9 Halaman Home.....	43
Gambar 3. 10 Halaman Upload File	43
Gambar 3. 11 Halaman konfirmasi upload file	44
Gambar 3. 12 Halaman laporan originality.....	44
Gambar 3. 13 Halaman Detail Similarity	45
Gambar 3. 14 Halaman Help.....	45
Gambar 4. 1 Tabel file	49
Gambar 4. 2 Tabel file_uji	50
Gambar 4. 3 Tabel ktdasar	50

Gambar 4. 4 Tabel ktdasar	51
Gambar 4. 5 Tabel hasil_uji	51
Gambar 4. 6 Halaman utama (Home).....	52
Gambar 4. 7 Halaman upload file.....	53
Gambar 4. 8 Validasi input file “false”.....	54
Gambar 4. 9 Halaman konfirmasi upload file.....	54
Gambar 4. 10 Halaman laporan originality.....	60
Gambar 4. 11 Halaman detail laporan	60
Gambar 4. 12 Tampilan halaman plagiarisme	61
Gambar 4. 14 Tampilan halaman kontak kami	62
Gambar 4. 13 Tampilan halaman help.....	61



DAFTAR LAMPIRAN

Lampiran I : Daftar Stopword	71
Lampiran II : Pencocokan Manual	73
Lampiran III : Dokumen Uji	78



INTISARI

Masalah plagiarisme menjadi semakin serius seiring pesatnya perkembangan teknologi. Berbagai kalangan dapat dengan mudah mengakses konten *digital* seperti gambar, *video*, mp3, dokumen teks, maupun *source code* yang biasanya menjadi target plagiat. Ironisnya tindak plagiat paling banyak terjadi dikalangan akademisi seperti mahasiswa dan pelajar. Dalam hal ini yang paling sering dilakukan adalah *copy-paste-edit* dokumen teks.

Plagiarisme merupakan tindak kriminal, karena mengakui hak milik orang sebagai karya pribadi. Untuk mengantisipasinya diperlukan alat deteksi yang dapat menganalisis tindak plagiat yang dilakukan. Alat deteksi ini dibangun dengan strategi pendekatan algoritma *Rabin-Karp*, yang merupakan algoritma *multiple pattern search*.

Algoritma ini adalah salah satu yang terbaik dalam menganalisis kecocokan antar dokumen. Teknik yang digunakan algoritma ini adalah teknik *hashing*. Pencarian akan lebih efisien dengan teknik ini karena hanya akan membandingkan beberapa *digit* angka.

Kata-kunci : Plagiarisme, *String matching*, *Pattern matching*, *Rabin-Karp*, *Hashing*, *Similarity*

ABSTRACT

Plagiarism is becoming increasingly serious problem as the rapid development of technology. Various people can easily access digital content such as images, video, mp3, text documents, and source code that is usually the target of plagiarism. Ironically acts of plagiarism is most prevalent among academics as students and learners. In this case most often done is copy-paste-edit text documents.

Plagiarism is a crime, because it recognizes people's property as a personal masterpiece. As anticipated, detection tools needed to analyze the acts committed plagiarism. This detection tool is built with a strategic approach Rabin-Karp algorithm, which is a multiple-pattern search algorithm.

This algorithm is one of the best in analyzing compatibility between documents. It uses a hashing technique. The search will be more efficient with this technique because it will compare multiple digit numbers.

Keywords: *Plagiarism, String matching, Pattern Matching, Rabin-Karp, Hashing, Similarity*

