

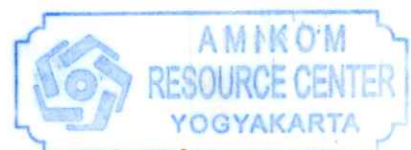
**IMPLEMENTASI ALGORITMA RABIN-KARP UNTUK
PENDETEKSI PLAGIARISME PADA DOKUMEN
TUGAS MAHASISWA**

SKRIPSI



disusun oleh
Asvarizal Filcha
15.11.9292

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2019**



**IMPLEMENTASI ALGORITMA RABIN-KARP UNTUK
PENDETEKSI PLAGIARISME PADA DOKUMEN
TUGAS MAHASISWA**

SKRIPSI

untuk memenuhi sebagian persyaratan
mencapai gelar Sarjana
pada Program Studi Informatika



disusun oleh
Asvarizal Filcha
15.11.9292

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2019**

PERSETUJUAN

SKRIPSI

**IMPLEMENTASI ALGORITMA RABIN-KARP UNTUK
PENDETEKSI PLAGIARISME PADA DOKUMEN
TUGAS MAHASISWA**

yang dipersiapkan dan disusun oleh

Asvarizal Filcha

15.11.9292

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 26 September 2018

Dosen Pembimbing,



Mardhiya Havaty, S.T, M.Kom

NIK. 190302108

PENGESAHAN

SKRIPSI

IMPLEMENTASI ALGORITMA RABIN-KARP UNTUK PENDETEKSI PLAGIARISME PADA DOKUMEN TUGAS MAHASISWA

yang dipersiapkan dan disusun oleh
Asvarizal Filcha

15.11.9292

telah dipertahankan di depan Dewan Penguji
pada tanggal 21 Februari 2019

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Ainul Yaqin, M.Kom
NIK. 190302255



Dina Maulina, M.Kom
NIK. 190302250




Mardhiya Hayaty, S.T, M.Kom
NIK. 190302108



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 28 Februari 2019

DEKAN FAKULTAS ILMU KOMPUTER




Krisnawati, S.Si, M.T
NIK. 190302038

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 tanggung jawab saya pribadi.

Yogyakarta, 28 Februari 2019



Asvarizal Filcha

15.11.9292

PERSEMBAHAN

Alhamdulillah segala puji syukur atas kehadiran Allah *Subhanahu wa Ta'ala* yang telah memberikan limpahan rahmat dan karunia-Nya sehingga penelitian ini dapat dilakukan dan diselesaikan dengan sebaik-baiknya. Saya juga ucapkan terimakasih untuk dukungan dan bantuan semua pihak yang membantu selesainya penelitian ini.

Skripsi ini saya persembahkan untuk kampus saya, Universitas AMIKOM Yogyakarta.

Untuk kedua orang tua saya yang telah susah payah mendukung saya tanpa kenal lelah.

Untuk ibu Mardhiya, yang sudah dengan sabar membimbing saya. Saya ucapkan terimakasih sedalam-dalamnya.

Untuk para pembaca, saya ucapkan terima kasih sebesar-besarnya. Semoga menjadi amal jariyah atas ilmu yang saya bagikan.

Dan untuk semua pihak yang telah banyak membantu dalam penyusunan skripsi ini yang tidak dapat saya sebutkan satu persatu.

MOTTO

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

”Boleh jadi kamu membenci sesuatu, padahal ia amat baik bagimu, dan boleh jadi (pula) kamu menyukai sesuatu, padahal ia amat buruk bagimu; Allah mengetahui, sedang kamu tidak mengetahui”

- Al-Qur'an Surah Al-Baqarah (2: 216)

”Siapa yang menempuh jalan untuk mencari ilmu, maka Allah akan mudahkan baginya jalan menuju surga”

- HR. Muslim, no. 2699

”If you can't make it good, at least make it look good.”

- Bill Gates

”Talk is cheap. Show me the code.”

- Linus Torvalds

”I actually hate programming, but I love solving problems”

- Rasmus Lerdorf

KATA PENGANTAR

Puji syukur penulis haturkan atas kehadiran Allah *Subhanahu wa Ta'ala* yang telah memberikan rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi dengan judul “Implementasi Algoritma Rabin-Karp Untuk Pendeteksi Plagiarisme Pada Dokumen Tugas Mahasiswa” dengan sebaik-baiknya. Tidak lupa shalawat serta salam penulis haturkan kepada junjungan Nabi Muhammad *Shallallahu 'alaihi wa Sallam* beserta keluarga dan para sahabatnya.

Dengan selesainya skripsi ini, penulis mengucapkan banyak terimakasih kepada:

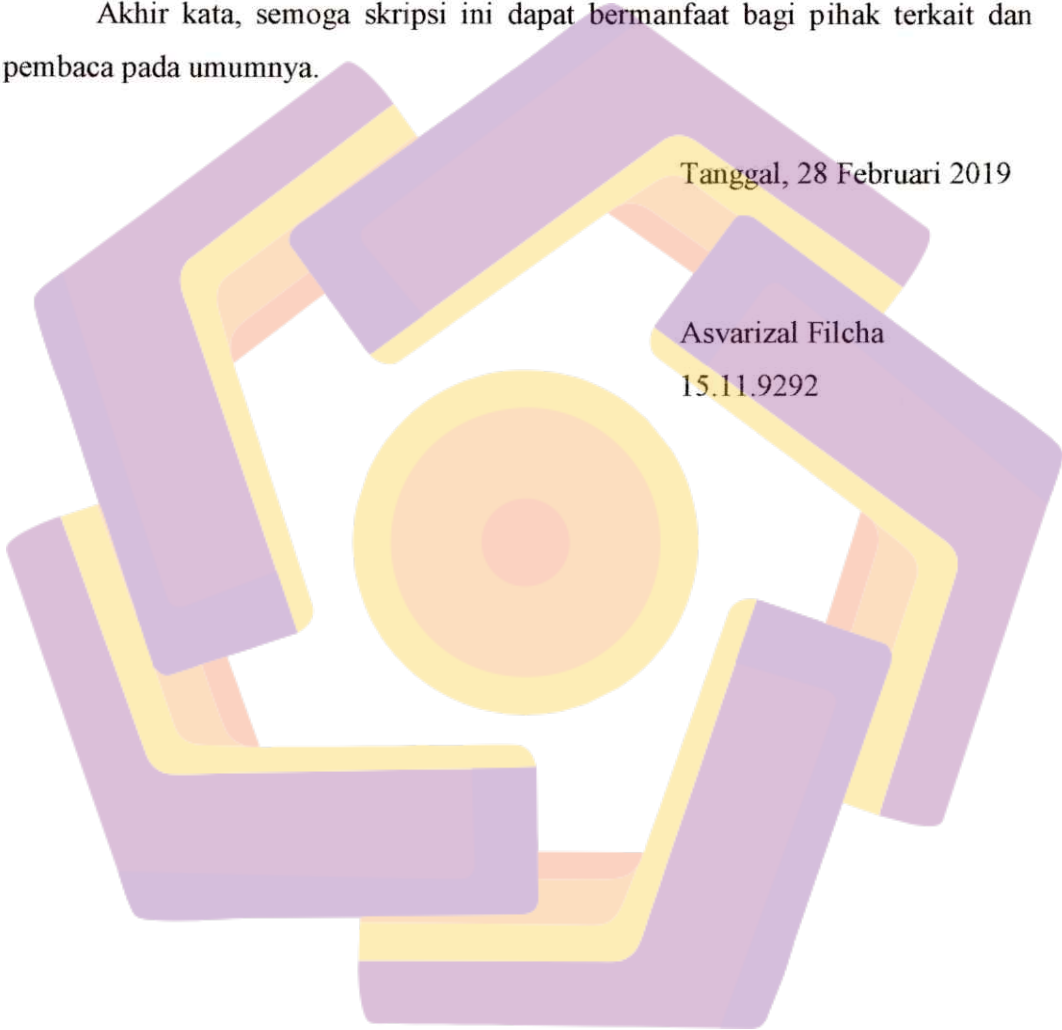
1. Bapak Prof. Dr. M. Suyanto selaku Rektor Universitas AMIKOM Yogyakarta
2. Ibu Krisnawati, S.Si, M.T. selaku Dekan Fakultas Ilmu Komputer.
3. Bapak Sudarmawan, S.T, M.T. selaku Ketua Program Studi S1 Informatika.
4. Ibu Mardhiya Hayaty, S.T, M.Kom. selaku pembimbing yang telah dengan sabar membimbing dan banyak memberikan kontribusi bagi penulis dalam pembuatan skripsi ini.
5. Para Dosen dan Staff Universitas AMIKOM Yogyakarta yang telah banyak memberi bantuan informasi, pengalaman dan kontribusi lainnya selama penulis berkuliah di kampus ini hingga terselesaikannya skripsi.
6. Orang tua dan keluarga penulis yang telah memberikan dukungan terbaiknya selama berkuliah.
7. PT. Durio Indigo dan rekan kerja yang telah memberi dukungan dalam bentuk materil maupun moril.
8. Makhluk-makhluk kontrakan yang sudah menemani dan bersenang-senang bersama penulis selama berkuliah.
9. Ibu Anna Baita, M.Kom. selaku dosen yang membuat saya semangat saat kuliah.

Penulis menyadari dengan betul bahawa dalam pembuatan skripsi ini masih banyak sekali kekurangan dan kelemahan. Oleh karena itu penulis berharap untuk semua pihak yang telah membaca dan memahami penelitian ini untuk dapat menyampaikan kritik dan saran yang membangun sehingga dapat menambah kesempurnaan skripsi ini.

Akhir kata, semoga skripsi ini dapat bermanfaat bagi pihak terkait dan pembaca pada umumnya.

Tanggal, 28 Februari 2019

Asvarizal Filcha
15.11.9292



DAFTAR ISI

COVER	i
PERSETUJUAN	iii
PENGESAHAN	iv
PERNYATAAN	v
PERSEMBAHAN	vi
MOTTO	vii
KATA PENGANTAR	viii
DAFTAR ISI	x
DAFTAR TABEL	xiii
DAFTAR GAMBAR	xiv
INTISARI	xvi
ABSTRACT	xvii
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	3
1.6 Metode Penelitian	4
1.7 Sistematika Penulisan	5
BAB II LANDASAN TEORI	7
2.1 Tinjauan Pustaka	7
2.2 Landasan Teori	9
2.2.1 Plagiarisme	9

2.2.2	Metode Pencocokan String	10
2.2.3	Text Preprocessing	11
2.2.4	Algoritma Rabin-Karp	13
2.2.5	Dice Coefficient Similarity	16
2.2.6	Confusion Matrix	16
2.2.7	OOAD (Object-Oriented Analysis and Design)	17
2.2.8	UML (Unified Modeling Language)	17
2.2.9	PHP	22
2.2.10	MySQL	22
BAB III ANALISIS DAN PERANCANGAN		23
3.1	Analisis Masalah	23
3.2	Analisis Sistem	24
3.2.1	Analisis Kelayakan Sistem	24
3.2.1.1	Kelayakan Teknis	24
3.2.1.2	Kelayakan Hukum	24
3.2.1.3	Kelayakan Operasional	25
3.2.2	Analisis Kebutuhan Sistem	25
3.2.2.1	Analisis Kebutuhan Fungsional	25
3.2.2.2	Analisis Kebutuhan Non Fungsional	26
3.3	Gambaran Sistem	26
3.4	Perancangan Sistem	29
3.4.1	Perancangan UML	30
3.4.1.1	Use Case Diagram	30
3.4.1.2	Acitivity Diagram	34
3.4.1.3	Sequence Diagram	36
3.4.2	Perancangan Antarmuka	37
BAB IV IMPLEMENTASI DAN PENGUJIAN		39
4.1	Implementasi Sistem	39
4.1.1	Spesifikasi Perangkat	39
4.1.2	Implementasi Basis Data	40



4.1.3 Implementasi Fungsi.....	41
4.1.3.1 Fungsi Unggah Dan Konversi Dokumen.....	41
4.1.3.2 Fungsi Text Preprocessing.....	47
4.1.3.3 Fungsi Parsing K-Gram	57
4.1.3.4 Fungsi Rolling Hash	59
4.1.3.5 Fungsi Fingerprint.....	61
4.1.3.6 Fungsi Perhitungan Similarity	63
4.1.3.7 Fungsi Simpan Similarity	64
4.1.4 Implementasi Antarmuka.....	64
4.2 Pengujian Sistem.....	66
4.3 Pengujian Akurasi.....	67
BAB V PENUTUP	70
5.1 Kesimpulan	70
5.2 Saran	70
DAFTAR PUSTAKA.....	71
LAMPIRAN.....	73

DAFTAR TABEL

Tabel 2.1 Perbedaan Penelitian.....	8
Tabel 2.2 Contoh K-Gram	14
Tabel 2.3 Contoh <i>Rolling Hash</i>	15
Tabel 2.4 Simbol <i>Use Case Diagram</i>	18
Tabel 2.5 Simbol <i>Acitivity Diagram</i>	20
Tabel 2.6 Simbol <i>Sequence Diagram</i>	21
Tabel 3.1 Dokumen Teks.....	27
Tabel 3.2 <i>Text Preprocessing</i>	27
Tabel 3.3 Hasil <i>Text Preprocessing</i>	28
Tabel 3.4 Proses Algoritma <i>Rabin-Karp</i>	28
Tabel 3.5 Penjelasan <i>Use Case</i> mendeteksi kemiripan tugas mahasiswa.....	31
Tabel 3.6 Penjelasan <i>Use Case Text Preprocessing</i>	32
Tabel 3.7 Penjelasan <i>Use Case</i> algoritma <i>Rabin-Karp</i>	33
Tabel 3.8 Penjelasan <i>Use Case</i> menampilkan <i>similarity</i>	34
Tabel 4.1 <i>Black box testing</i> sistem.....	66
Tabel 4.2 Klasifikasi data uji	67
Tabel 4.3 Hasil <i>confusion matrix</i>	69

DAFTAR GAMBAR

Gambar 2.1 Contoh <i>Case Folding</i>	11
Gambar 2.2 Contoh <i>Tokenizing</i>	12
Gambar 2.3 Contoh <i>Punctuation Removal</i>	12
Gambar 2.4 Contoh <i>Stopword Removal</i>	13
Gambar 2.5 Contoh <i>Stemming</i>	13
Gambar 2.6 <i>Confusion Matrix</i>	16
Gambar 3.1 <i>Use Case</i> mendeteksi kemiripan tugas mahasiswa	30
Gambar 3.2 <i>Activity Diagram</i> sistem deteksi plagiarisme.....	35
Gambar 3.3 <i>Sequence Diagram</i> sistem deteksi plagiarisme.....	36
Gambar 3.4 Rancangan halaman unggah berkas	37
Gambar 3.5 Rancangan halaman hasil <i>similarity</i> tugas mahasiswa	38
Gambar 4.1 Perintah SQL tabel hasil <i>similarity</i> tugas mahasiswa	40
Gambar 4.2 Struktur tabel hasil <i>similarity</i> tugas mahasiswa.....	41
Gambar 4.3 Struktur tabel hasil <i>similarity</i> tugas mahasiswa.....	42
Gambar 4.4 Dokumen tugas 15.11.9330	43
Gambar 4.5 Fungsi Unggah Dan Konversi Dokumen.....	45
Gambar 4.6 Hasil konversi dokumen tugas 15.11.9324	46
Gambar 4.7 Hasil konversi dokumen tugas 15.11.9330	46
Gambar 4.8 Fungsi <i>Case Folding</i>	47
Gambar 4.9 Hasil <i>case folding</i> dokumen tugas 15.11.9324	48
Gambar 4.10 Hasil <i>case folding</i> dokumen tugas 15.11.9330	48
Gambar 4.11 Fungsi <i>Tokenizing</i>	49
Gambar 4.12 Hasil <i>tokenizing</i> dokumen tugas 15.11.9324	50
Gambar 4.13 Hasil <i>tokenizing</i> dokumen tugas 15.11.9330	50
Gambar 4.14 Fungsi <i>Punctuation Removal</i> dan <i>Stopword Removal</i>	51
Gambar 4.15 Hasil <i>punctuation removal</i> dan <i>stopword removal</i> dokumen tugas 15.11.9324.....	52
Gambar 4.16 Hasil <i>punctuation removal</i> dan <i>stopword removal</i> dokumen tugas 15.11.9330.....	52

Gambar 4.17 Fungsi <i>Stemming</i>	53
Gambar 4.18 Hasil <i>stemming</i> dokumen tugas 15.11.9324.....	54
Gambar 4.19 Hasil <i>stemming</i> dokumen tugas 15.11.9330.....	54
Gambar 4.20 Fungsi <i>Text Preprocessing</i>	56
Gambar 4.21 Hasil <i>text preprocessing</i> dokumen tugas 15.11.9324	56
Gambar 4.22 Hasil <i>text preprocessing</i> dokumen tugas 15.11.9330	56
Gambar 4.23 Fungsi <i>Parsing K-Gram</i>	57
Gambar 4.24 Hasil <i>parsing k-gram</i> dokumen tugas 15.11.9324	58
Gambar 4.25 Hasil <i>parsing k-gram</i> dokumen tugas 15.11.9330	58
Gambar 4.26 Fungsi <i>Rolling Hash</i>	59
Gambar 4.27 Hasil <i>rolling hash</i> dokumen tugas 15.11.9324	60
Gambar 4.28 Hasil <i>rolling hash</i> dokumen tugas 15.11.9330	60
Gambar 4.29 Fungsi <i>Fingerprint</i>	61
Gambar 4.30 Hasil <i>fingerprint</i> dokumen tugas 15.11.9324.....	62
Gambar 4.31 Hasil <i>fingerprint</i> dokumen tugas 15.11.9330.....	62
Gambar 4.32 Hasil <i>fingerprint</i> gabungan dokumen tugas 15.11.9324 dan dokumen tugas 15.11.9330	63
Gambar 4.33 Fungsi Perhitungan <i>Similarity</i>	63
Gambar 4.34 Fungsi Simpan <i>Similarity</i>	64
Gambar 4.35 Hasil Fungsi Simpan <i>Similarity</i>	64
Gambar 4.36 Implementasi halaman unggah berkas	65
Gambar 4.37 Implementasi halaman hasil <i>similarity</i> tugas	66

INTISARI

Perkembangan pada dunia teknologi informasi mengakibatkan perguruan tinggi mengurangi penggunaan kertas sehingga banyak tugas mahasiswa yang dikumpulkan dalam bentuk digital. Penggunaan digital menyebabkan semakin mudahnya mahasiswa untuk melakukan plagiarisme. Sehingga diperlukan sebuah sistem untuk melakukan pemeriksaan plagiarisme pada dokumen tugas antar mahasiswa dengan cepat dan tepat. Metode yang dapat digunakan adalah menggunakan algoritma Rabin-Karp. Algoritma Rabin-Karp memiliki keunggulan pencarian string dengan pola yang panjang.

Algoritma Rabin-karp dalam sistem ini memiliki langkah - langkah *text preprocessing* yang terdiri *case folding*, *tokenizing*, *punctuation removal*, *stopword removal* dan *stemming*. Hasil dari *text preprocessing* inilah yang akan di proses menggunakan algoritma Rabin-karp. Hasil dari metode ini adalah nilai kemiripan dari tugas - tugas mahasiswa yang dihitung menggunakan *dice coefficient*.

Perhitungan akurasi dengan melakukan 20 perbandingan antara sistem pendeteksi plagiarisme dan *software* Plagiarisme Checker X menggunakan *confusion matrix* menghasilkan tingkat keakuratan sebesar 90%.

Kata Kunci: Rabin-Karp, plagiarisme, *text preprocessing*, *dice coefficient*

ABSTRACT

Information and technology development causes universities reduce paper usage so that the student's assignments can be collected in digital form. The digital form usage causes students can easily plagiarism the assignments. So, it is needed a system to check the plagiarism on assignment documents among students quickly and accurate. The method that can be used is Rabin-Karp algorithm. Rabin-Karp algorithm has excellence in searching strings with long pattern.

Rabin-Karp algorithm in this system has text preprocessing steps that consist of case folding, tokenizing, punctuation removal, stopword removal and stemming. The result from the text preprocessing will be processed using Rabin-Karp algorithm. The outcome of this method is the similiarity percentage of student's assignments calculated using dice coefficient.

The accuracy calculation by doing 20 comparisons between plagiarism checker system and Plagiarism Checker X software using confusion matrix is 90%.

Keyword: *Rabin-Karp, plagiarism, text preprocessing, dice coefficient*