



### **Hak cipta dan penggunaan kembali:**

Lisensi ini mengizinkan setiap orang untuk menggubah, memperbaiki, dan membuat ciptaan turunan bukan untuk kepentingan komersial, selama anda mencantumkan nama penulis dan melisensikan ciptaan turunan dengan syarat yang serupa dengan ciptaan asli.

### **Copyright and reuse:**

This license lets you remix, tweak, and build upon work non-commercially, as long as you credit the origin creator and license it on your new creations under the identical terms.

**IMPLEMENTASI ALGORITMA RABIN-KARP DALAM  
APLIKASI PENDETEKSI PLAGIARISME PADA  
DOKUMEN TEKS BERBAHASA INDONESIA**

Nama : Kevin Leonardi  
NIM : 09110110006  
Program Studi : Teknik Informatika  
Fakultas : Teknologi Informasi dan Komunikasi

**SKRIPSI**

**Diajukan sebagai salah satu syarat untuk memperoleh gelar  
Sarjana Komputer (S.Kom.)**

**UNIVERSITAS MULTIMEDIA NUSANTARA  
TANGERANG  
2013**

**HALAMAN PENGESAHAN SKRIPSI**

**IMPLEMENTASI ALGORITMA RABIN-KARP DALAM APLIKASI  
PENDETEKSI PLAGIARISME PADA DOKUMEN TEKS  
BERBAHASA INDONESIA**

Oleh

Nama : Kevin Leonardi

NIM : 09110110006

Program Studi : Teknik Informatika

Fakultas : Teknologi Informasi dan Komunikasi

Tangerang, 18 Juli 2013

Mengetahui,

Ketua Sidang

Dosen Penguji

Maria Irmina Prasetyowati, S.Kom.,  
M.T.

Adhi Kusnadi, M.Si.

Dosen Pembimbing

Ketua Program Studi  
Teknik Informatika

Siauw Yohanes Darmawan, S.Kom.,  
M.Sc.

Maria Irmina Prasetyowati, S.Kom.,  
M.T.

## **PERNYATAAN TIDAK MELAKUKAN PLAGIAT**

Dengan ini saya:

Nama : Kevin Leonardi

NIM : 09110110006

Program Studi : Teknik Informatika

Fakultas : Teknologi Informasi dan Komunikasi

menyatakan bahwa skripsi yang berjudul “Implementasi Algoritma Rabin-Karp Dalam Aplikasi Pendeteksi Plagiarisme Pada Dokumen Teks Berbahasa Indonesia” adalah karya ilmiah saya sendiri, bukan plagiat dari karya ilmiah yang ditulis oleh orang lain atau lembaga lain. Semua karya ilmiah orang atau lembaga lain yang dirujuk dalam skripsi ini telah disebutkan sumbernya serta dicantumkan dalam daftar pustaka.

Tangerang, 18 Juli 2013

Kevin Leonardi

# IMPLEMENTASI ALGORITMA RABIN-KARP DALAM APLIKASI PENDETEKSI PLAGIARISME PADA DOKUMEN TEKS BERBAHASA INDONESIA

## ABSTRAKSI

Internet kini menjadi sumber informasi yang umum dipakai. Paradigma ini menimbulkan kecenderungan untuk menggandakan dokumen teks dari internet dan mengakuinya sebagai karya sendiri. Inilah yang disebut dengan tindakan plagiarisme. Salah satu upaya untuk mencegah tindakan plagiarisme adalah membuat aplikasi pendeteksi plagiarisme pada dokumen teks, khususnya yang berbahasa Indonesia.

Skripsi ini mengimplementasikan aplikasi untuk mendeteksi plagiarisme menggunakan algoritma Rabin-Karp. Algoritma pencocokan *string* ini dipakai karena mampu melakukan pencocokan banyak pola. Pencocokan dilakukan untuk mengetahui pola mana saja yang sama dari kedua dokumen sehingga persentase kemiripan antar dokumen dapat dikalkulasi. Selain itu, terdapat algoritma Rabin-Karp modifikasi. Waktu proses yang dibutuhkan algoritma Rabin-Karp modifikasi untuk mendeteksi lebih singkat dibandingkan algoritma Rabin-Karp tanpa modifikasi. Selain itu, persentase kemiripanyang dihasilkan pun tidak berbeda jauh.

**Kata kunci** : plagiarisme, dokumen teks, Bahasa Indonesia, pencocokan *string*, algoritma Rabin-Karp, algoritma Rabin-Karp modifikasi

# UMMN

# IMPLEMENTATION OF RABIN-KARP ALGORITHM ON INDONESIAN TEXT DOCUMENT PLAGIARISM DETECTION APPLICATION

## ABSTRACT

Nowadays, internet is a common source of information. This paradigm causes a tendency to duplicate text documents from the Internet and claim it as his/her own works. This action is called plagiarism. One effort to prevent plagiarism is to make an application that detects plagiarism in text documents, especially in Indonesian language.

This thesis implements a plagiarism detection application using Rabin-Karp algorithm. This string matching algorithm is used because it is capable to match multiple patterns. Pattern matching is performed to determine which same patterns are in two documents so that the percentage of similarity between those documents can be calculated. In addition, there is a modified Rabin-Karp algorithm. Modified Rabin-Karp algorithm requires shorter detection time than Rabin-Karp algorithm without modification. Beside that, the resulting percentage of similarity does not differ much.

**Keywords:** plagiarism, text document, Indonesian language, string matching, Rabin-Karp algorithm, modified Rabin-Karp algorithm

UMMN

## HALAMAN PERSEMBAHAN

*“Initium sapientiae timor Domini.”*

Kupersembahkan karya ini untuk Ayah dan Ibu tercinta, yang senantiasanya ada saat suka maupun duka, selalu setia mendampingi dan memanjatkan doa kepada putra keduanya. Terima kasih untuk semuanya.

UMMN

## KATA PENGANTAR

Komputer tidak bisa beroperasi tanpa sistem operasi. Begitu juga penulis. Anugerah dari Tuhan Yang Maha Esa membuahkan inspirasi kepada penulis sehingga mampu menyelesaikan skripsi yang berjudul “Implementasi Algoritma Rabin-Karp Dalam Aplikasi Pendeteksi Plagiarisme Pada Dokumen Teks Berbahasa Indonesia”. Tanpa berkat-Nya, penulis bukan apa-apa.

Penyelesaian skripsi ini juga tidak lepas dari bantuan berbagai pihak. Oleh karena itu, pada kesempatan kali ini penulis ingin mengucapkan terima kasih kepada

1. Dr. Ninok Leksono, selaku Rektor Universitas Multimedia Nusantara.
2. Maria Irmina Prasetyowati, S.Kom., M.T., selaku Ketua Program Studi Teknik Informatika Universitas Multimedia Nusantara.
3. Siauw Yohanes Darmawan, S.Kom., M.Sc., selaku Dosen Pembimbing yang tak kenal lelah dalam membimbing dan memberi masukan-masukan yang bersifat membangun.
4. Seluruh bapak dan ibu dosen yang telah mendidik dan mengamalkan ilmunya selama delapan semester kepada penulis.
5. Sahabat-sahabat tercinta dari Teknik Informatika 2009 Universitas Multimedia Nusantara, khususnya Eddy Wijaya, Edward Kurnia, Evan Virnando, Herry Gunawan, Jansen, Kevin Adisaputra Bachtiar, Kevin Malviyanto, Niko Jenius, Ronald, dan Winston yang sudah



memberikan pengalaman berharga yang tak terlupakan serta dukungan yang tak pernah habis kepada penulis.

6. Para sahabat terkasih, Albert Hadi, Brian Alvin, David Chandra, Jean Gabrielle Dufresse, Jonathan Ferro, Joshua Thaddeus Kurniawan, Mario Marcello, Riski Lasma Dwianto Putra, Robby, dan Vincent Halim, yang memberikan inspirasi, motivasi, dan juga canda tawa serta sukacita sehingga penulis tetap semangat dalam menulis skripsi.
7. Keluarga tercinta penulis yang menjadi sumber semangat dan inspirasi selama pengerjaan skripsi.
8. Pihak lain yang telah membantu dalam penyelesaian skripsi ini yang tidak bisa penulis sebutkan.

Tak ada yang sempurna selain Sang Pencipta, demikian pula skripsi ini yang masih memiliki banyak kekurangan. Penulis sangat terbuka untuk menerima saran maupun kritik dari para pembaca demi menyempurnakan skripsi ini.

Akhir kata, penulis mengucapkan terima kasih kepada siapa saja yang telah meluangkan sedikit waktunya untuk membaca skripsi ini. Harapan terbesar penulis adalah skripsi ini dapat menambah wawasan bagi para pembacanya dan digunakan untuk membantu menekan tindakan plagiarisme.

Tangerang, 18 Juli 2013

Penulis

## DAFTAR ISI

HALAMAN PENGESAHAN SKRIPSI.....	ii
PERNYATAAN TIDAK MELAKUKAN PLAGIAT .....	iii
ABSTRAKSI.....	iv
<i>ABSTRACT</i> .....	v
HALAMAN PERSEMBAHAN.....	vi
KATA PENGANTAR .....	vii
DAFTAR ISI.....	ix
DAFTAR GAMBAR.....	xii
DAFTAR TABEL .....	xvi
DAFTAR RUMUS .....	xvii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	3
1.3 Batasan Masalah .....	3
1.4 Tujuan Penelitian .....	4
1.5 Manfaat Penelitian .....	4
1.6 Sistematika Penulisan.....	4
BAB II TINJAUAN PUSTAKA.....	6
2.1 Plagiarisme .....	6
2.2 Dokumen Elektronik .....	8
2.3 Bahasa Indonesia.....	10
2.4 Pencocokan <i>String</i> .....	11

2.5	<i>Text Mining</i> .....	12
2.6	Algoritma Rabin-Karp.....	15
2.6.1	Definisi Algoritma Rabin-Karp.....	15
2.6.2	Konsep Algoritma Rabin-Karp .....	16
2.6.3	<i>Multiple Pattern Search</i> .....	17
2.6.4	Skenario Pencarian Algoritma Rabin-Karp .....	18
2.6.5	Algoritma Rabin-Karp Modifikasi .....	19
2.7	<i>Hashing</i> .....	20
2.8	<i>N-grams</i> .....	23
2.9	Nilai <i>Similarity</i> .....	24
<b>BAB III ANALISIS DAN PERANCANGAN APLIKASI</b> .....		27
3.1	Analisis Permasalahan.....	27
3.2	Usulan Solusi Permasalahan.....	27
3.3	Spesifikasi Umum Aplikasi .....	29
3.4	Fitur Lainnya.....	42
3.5	Desain Antarmuka Aplikasi.....	43
<b>BAB IV IMPLEMENTASI DAN UJI COBA APLIKASI</b> .....		48
4.1	Implementasi Sistem .....	48
4.1.1	Lingkup Implementasi .....	48
4.1.2	Implementasi Antarmuka.....	49
4.1.3	Implementasi Proses .....	53
4.2	Uji Coba.....	71
4.2.1	Data dan Proses Pengujian.....	71

4.2.2 Hasil Pengujian .....	74
BAB V SIMPULAN DAN SARAN.....	90
5.1 Simpulan.....	90
5.2 Saran.....	91
DAFTAR PUSTAKA.....	93
LAMPIRAN.....	100



## DAFTAR GAMBAR

Gambar 2.1	Tahap <i>Pre-processing</i> .....	12
Gambar 2.2	Tahap <i>Case Folding</i> dan <i>Tokenizing</i> .....	13
Gambar 2.3	Tahap <i>Filtering</i> .....	14
Gambar 2.4	Tahap <i>Stemming</i> .....	15
Gambar 2.5	<i>Pseudocode</i> Algoritma Rabin-Karp.....	16
Gambar 2.6	Ilustrasi Algoritma Rabin-Karp.....	17
Gambar 2.7	<i>Pseudocode Multiple Pattern Search</i> Algoritma Rabin-Karp .....	18
Gambar 2.8	<i>Successful Hit</i> .....	18
Gambar 2.9	<i>Spurious Hit</i> .....	19
Gambar 2.10	<i>Unsuccessful Hit</i> .....	19
Gambar 2.11	Penempatan Data ke <i>Hash Table</i> .....	21
Gambar 3.1	<i>Flowchart</i> Umum Aplikasi .....	28
Gambar 3.2	<i>Flowchart</i> Mendeteksi Plagiarisme .....	30
Gambar 3.3	<i>Flowchart</i> Tahap <i>Pre-processing</i> .....	31
Gambar 3.4	<i>Flowchart</i> Proses <i>Case Folding</i> .....	32
Gambar 3.5	<i>Flowchart</i> Proses <i>Tokenizing</i> .....	33
Gambar 3.6	<i>Flowchart</i> Proses <i>Filtering</i> .....	34
Gambar 3.7	<i>Flowchart String Matching</i> Menggunakan Algoritma Rabin-Karp.....	35
Gambar 3.8	<i>Flowchart</i> Proses Pembentukan <i>N-grams</i> .....	36
Gambar 3.9	<i>Flowchart</i> Algoritma Rabin-Karp .....	37
Gambar 3.10	<i>Flowchart Hashing</i> .....	39

Gambar 3.11	<i>Flowchart</i> Algoritma Rabin-Karp yang Dimodifikasi .....	40
Gambar 3.12	<i>Flowchart</i> Perhitungan Nilai <i>Similarity</i> .....	41
Gambar 3.13	<i>Flowchart</i> Modul Penambahan Kata .....	42
Gambar 3.14	Desain Antarmuka Aplikasi Pendeteksi Plagiarisme .....	43
Gambar 3.15	Desain Antarmuka <i>Output Window</i> .....	44
Gambar 3.16	Desain Antarmuka <i>Tab</i> “Tambah Kata” .....	45
Gambar 3.17	Desain Antarmuka <i>Tab</i> “Cara Menggunakan” .....	46
Gambar 3.18	Desain Antarmuka <i>Tab</i> “Tentang” .....	47
Gambar 4.1	Tampilan Aplikasi Pendeteksi Plagiarisme.....	49
Gambar 4.2	Tampilan <i>Window Output</i> .....	50
Gambar 4.3	Tampilan <i>Tab</i> “Tambah Kata” .....	51
Gambar 4.4	Tampilan <i>Tab</i> “Cara Menggunakan” .....	52
Gambar 4.5	Tampilan <i>Tab</i> “Tentang” .....	53
Gambar 4.6	Kode Pembentukan Daftar Kata.....	54
Gambar 4.7	Kode Pemilihan Dokumen .....	55
Gambar 4.8	<i>Open File Dialog</i> untuk Dokumen Asli.....	56
Gambar 4.9	<i>Open File Dialog</i> untuk Dokumen Uji .....	56
Gambar 4.10	Kode Fungsi “FileSedangDibuka” .....	57
Gambar 4.11	Direktori Dokumen Ditampilkan.....	58
Gambar 4.12	Kode Ekstraksi Teks .....	58
Gambar 4.13	Kode Tahap <i>Pre-processing</i> Teks (1).....	59
Gambar 4.14	Kode Tahap <i>Pre-processing</i> Teks (2).....	60
Gambar 4.15	Contoh Hasil <i>Pre-processing</i> .....	60

Gambar 4.16	<i>Class</i> NGram.....	61
Gambar 4.17	Kode Pembentukan <i>N-grams</i> .....	62
Gambar 4.18	Contoh Pembentukan <i>N-grams</i> .....	63
Gambar 4.19	Kode Algoritma Rabin-Karp.....	64
Gambar 4.20	Kode Fungsi “Hash”.....	65
Gambar 4.21	Contoh Perhitungan Nilai <i>Hash</i> .....	65
Gambar 4.22	Kode Algoritma Rabin-Karp Modifikasi .....	66
Gambar 4.23	Kode Perhitungan Nilai <i>Similarity</i> .....	67
Gambar 4.24	<i>Class</i> HasilProses .....	68
Gambar 4.25	Tampilan Ditemukannya Kata “Kamu” .....	69
Gambar 4.26	Tampilan <i>Message Box</i> Ketika Menambahkan Kata “skripsi” .....	70
Gambar 4.27	Kode Penambahan Kata Baru .....	70
Gambar 4.28	Grafik Waktu Proses Algoritma Rabin-Karp dan yang Dimodifikasi pada Dokumen A.....	76
Gambar 4.29	Grafik Waktu Proses Algoritma Rabin-Karp dan yang Dimodifikasi pada Dokumen B .....	76
Gambar 4.30	Grafik Nilai <i>Similarity</i> Algoritma Rabin-Karp dan yang Dimodifikasi pada Dokumen A.....	77
Gambar 4.31	Grafik Nilai <i>Similarity</i> Algoritma Rabin-Karp dan yang Dimodifikasi pada Dokumen B .....	78
Gambar 4.32	Grafik Waktu Proses Rabin-Karp Modifikasi Dokumen A.....	82
Gambar 4.33	Grafik Waktu Proses Rabin-Karp Modifikasi Dokumen B .....	83
Gambar 4.34	Grafik Waktu Proses Rabin-Karp Modifikasi Dokumen C .....	84

Gambar 4.35	Grafik Waktu Proses Rabin-Karp Modifikasi Dokumen D.....	84
Gambar 4.36	Grafik Nilai <i>Similarity</i> Rabin-Karp Modifikasi Dokumen A .....	85
Gambar 4.37	Grafik Nilai <i>Similarity</i> Rabin-Karp Modifikasi Dokumen B.....	86
Gambar 4.38	Grafik Nilai <i>Similarity</i> Rabin-Karp Modifikasi Dokumen C .....	86
Gambar 4.39	Grafik Nilai <i>Similarity</i> Rabin-Karp Modifikasi Dokumen D .....	87
Gambar 4.40	Tampilan Plagiarism Checker X .....	88





## DAFTAR TABEL

Tabel 2.1	<i>String Hashing</i> .....	22
Tabel 2.2	Nama <i>N-Grams</i> Berdasarkan Ukuran.....	24
Tabel 2.3	Contoh Perhitungan Nilai <i>Similarity</i> .....	25
Tabel 4.1	Hasil Pengujian dengan Algoritma Rabin-Karp dan Rabin-Karp Modifikasi .....	75
Tabel 4.2	Hasil Pengujian dengan Algoritma Rabin-Karp Modifikasi.....	79
Tabel 4.3	Hasil Perbedaan Nilai <i>Similarity</i> .....	89

UMMN

## DAFTAR RUMUS

Rumus 2.1	Kriteria Algoritma Rabin-Karp Modifikasi .....	20
Rumus 2.2	<i>Hash Function</i> .....	22
Rumus 2.3	<i>Dice's Coefficient</i> .....	25



UMN