

**DETEKSI PLAGIARISME DOKUMEN TEKS
MENGUNAKAN ALGORITMA RABIN-KARP
DENGAN *SYNONYM RECOGNITION***

SKRIPSI

Untuk memenuhi sebagian persyaratan
untuk mencapai gelar Sarjana Komputer



Disusun Oleh:
SANDY DEWANTO
0610963055

**KEMENTERIAN PENDIDIKAN DAN KEBUDAYAAN
UNIVERSITAS BRAWIJAYA
PROGRAM TEKNOLOGI INFORMASI DAN ILMU KOMPUTER
MALANG
2013**

DAFTAR ISI

	Halaman
SAMPUL	i
LEMBAR PERSETUJUAN	i
LEMBAR PENGESAHAN	ii
PERNYATAAN ORISINALITAS	iii
KATA PENGANTAR	iv
ABSTRAK	vi
ABSTRACT	vii
DAFTAR ISI	viii
DAFTAR GAMBAR	xi
DAFTAR TABEL	xiii
BAB I PENDAHULUAN	1
1.1 LatarBelakang	1
1.2 RumusanMasalah	3
1.3 BatasanMasalah.....	3
1.4 Tujuan.....	4
1.5 Manfaat.....	4
1.6 SistematikaPenulisan.....	4
BAB II KAJIAN PUSTAKA DAN DASAR TEORI	6
2.1 Kajian Pustaka.....	6
2.2 Plagiarisme	7
2.2.1 Definisi	7
2.2.2 DeteksiPlagiarisme	8
2.3 Pencocokan String (<i>String Matching</i>)	10
2.3.1 Definisi	10
2.3.2 AlgoritmaPencocokanString (<i>StringMatching Algorithm</i>).....	11
2.4Algoritma Rabin-Karp.....	12
2.4.1Hashing.....	14
2.4.2Rolling Hash	14
2.4.3PrinsipKerjaAlgoritma Rabin-Karp.....	15
2.4.4Pencocokan String BerpolaBanyak (<i>Multiple Pattern Search</i>).....	19



2.4.5	Penilaian Kecocokan	20
2.5	<i>Text Pre-processing</i>	22
2.5.1	Ekstraksi Dokumen	22
2.5.1.1	<i>Case Folding</i> dan <i>Tokenizing</i>	23
2.5.1.2	<i>Filtering</i>	24
2.6	<i>Synonym Recognition</i>	24
BAB III METODE PENELITIAN DAN PERANCANGAN		25
3.1	Data Penelitian	25
3.2	Deskripsi Umum Sistem	26
3.3	Perancangan Proses	27
3.3.1	Perancangan <i>Pre-processing</i>	29
3.3.1.1	Perancangan <i>Case Folding</i> dan <i>Tokenizing</i>	30
3.3.1.2	Perancangan <i>Filtering</i>	30
3.3.2	Perancangan Proses <i>Parsing K-gram</i>	32
3.3.3	Perancangan Proses <i>Hashing</i>	34
3.3.4	Perancangan Proses Pencocokan <i>String</i>	34
3.3.5	Perancangan Proses <i>Synonym Recognition</i>	36
3.4	Perhitungan Manual	37
3.5	Rancangan Uji Coba	41
3.5.1	Bahan Pengujian	41
3.5.2	Tujuan Pengujian	41
3.5.3	Skenario Pengujian	41
3.5.4	Pengukuran Nilai Kesamaan	42
3.5.5	Pengukuran Waktu Proses	42
3.5.6	Pengukuran Presentase Kesalahan (<i>Error</i>)	42
3.5.7	Perancangan Dokumen Asli dan Dokumen Uji	42
3.6	Perancangan Antarmuka	43
3.6.1	Perancangan Input Sistem	43
3.6.2	Perancangan <i>Protoype</i> Sistem	43
BAB IV IMPLEMENTASI		45
4.1	Lingkungan Perangkat Keras	45
4.2	Lingkungan Perangkat Lunak	45

4.3 ImplementasiAntarmuka	45
4.4 Implementasi Proses.....	47
4.4.1 Implementasi Proses Input Dokumen	47
4.4.2 Implementasi <i>Pre-processing</i>	49
4.4.3Implementasi <i>Parsing k-gram</i>	49
4.4.4Implementasi <i>Hashing</i>	50
4.4.5Implementasi Pencocokan <i>String</i>	51
4.4.6Implementasi <i>Synonym Recognition</i>	51
4.4.7Implementasi HitungNilaiKesamaan.....	53
4.5 ImplementasiAplikasiSistem.....	53
BAB V PENGUJIAN DAN ANALISIS	58
5.1 PengujianSistem	58
5.2 DataPengujian	58
5.3 HasilPercobaanSistem.....	58
5.3.1HasilUjiCobak-gram.....	58
5.3.2HasilUji Coba Presentase Kesalahan (<i>Error</i>).....	61
5.4AnalisaHasil	64
5.4.1Analisa Hasil Parameter <i>k-gram</i>	64
5.4.2Analisa Hasil <i>Synonym Recognition</i>	66
BAB VI PENUTUP	69
6.1 Kesimpulan.....	69
6.2 Saran	69
DAFTAR PUSTAKA	70
LAMPIRAN	72

DAFTAR GAMBAR

	Halaman
Gambar 2.1 Definisi plagiarisme	7
Gambar 2.2 Operasi-operasi teks yang bisa mengarah pada plagiarisme	7
Gambar 2.3 Perbandingan pertama	16
Gambar 2.4 Pergeseran dan <i>updatesubstring</i>	16
Gambar 2.5 Perbandingan kedua	16
Gambar 2.6 Perbandingan keempat. <i>Spurious hit</i>	17
Gambar 2.7 Perbandingan kelima. <i>Successful hit</i>	17
Gambar 2.8 Algoritma Rabin-Karp	18
Gambar 2.9 Algoritma Rabin-Karp untuk pencarian pola banyak	20
Gambar 2.10 <i>Pre-processing</i>	23
Gambar 2.11 <i>Case folding</i> dan <i>tokenizing</i>	23
Gambar 2.12 <i>Filtering</i>	24
Gambar 3.1 Skema aliran data	26
Gambar 3.2 <i>Flowchart</i> sistem secara global	28
Gambar 3.3 <i>Flowchart</i> fase <i>pre-processing</i>	29
Gambar 3.4 <i>Flowchart</i> <i>case folding</i> dan <i>tokenizing</i>	30
Gambar 3.5 <i>Flowchart</i> <i>filtering</i>	31
Gambar 3.6 <i>Flowchart</i> proses <i>parsing-k-gram</i>	32
Gambar 3.7 <i>Flowchart</i> <i>hashing</i>	33
Gambar 3.8 Blok <i>Flowchart</i> pencocokan <i>string</i>	35
Gambar 3.9 <i>Flowchart</i> <i>synonym recognition</i>	36
Gambar 3.10 Desain antarmuka sistem	44
Gambar 4.1 Tampilan antarmuka sistem	46
Gambar 4.2 <i>Source code</i> proses baca <i>file</i> dokumen	48
Gambar 4.3 <i>Source code</i> <i>pre-processing</i>	49
Gambar 4.4 <i>Source code</i> <i>parsing k-gram</i>	50
Gambar 4.5 <i>Source code</i> <i>hashing</i>	50
Gambar 4.6 <i>Source code</i> pencocokan <i>string</i>	51
Gambar 4.7 <i>Source code</i> hitung nilai kesamaan	52

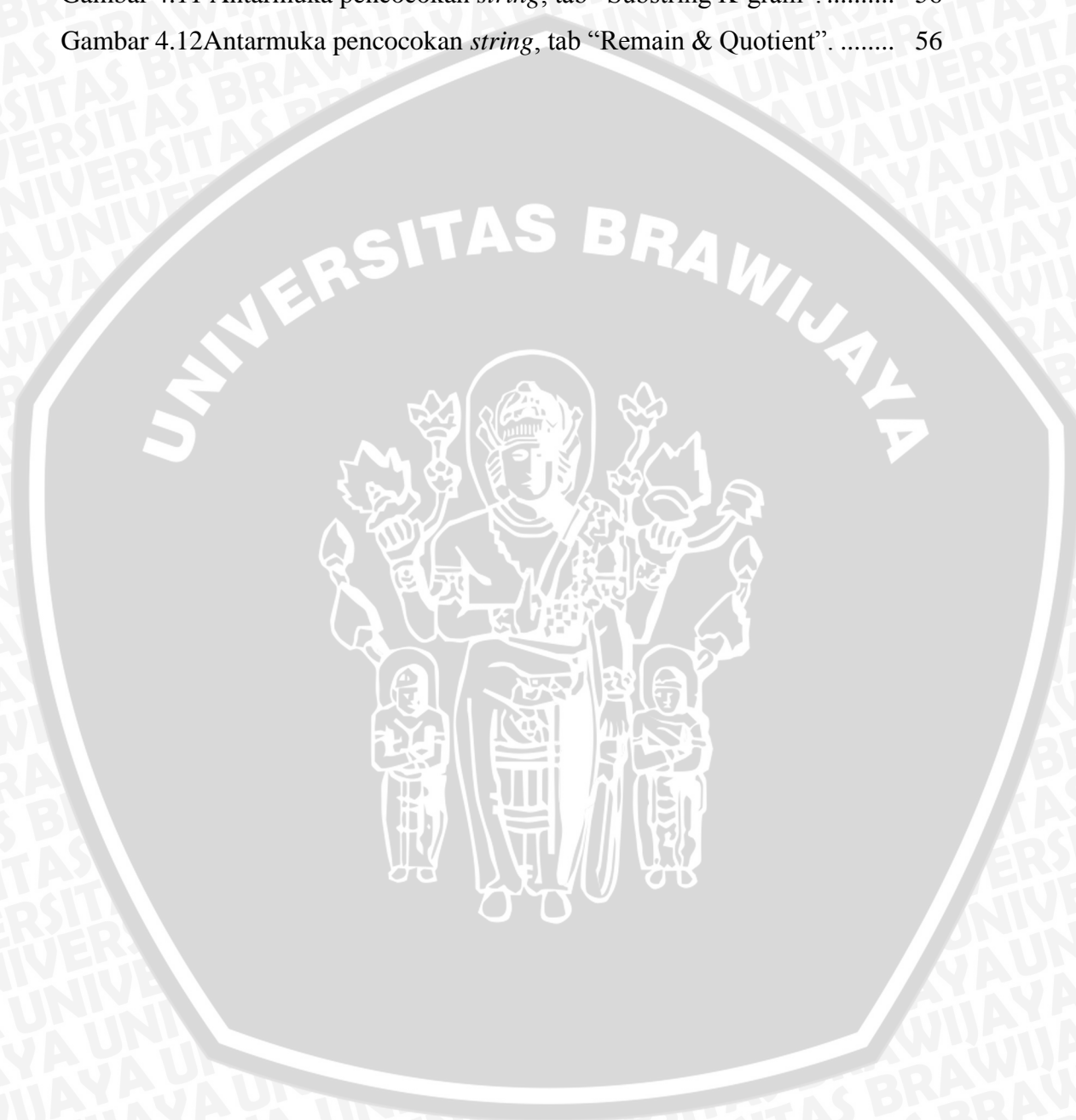
Gambar 4.8 Antarmuka penginputan dokumen 53

Gambar 4.9 Antarmuka *pre-processing* 54

Gambar 4.10 Antarmuka pencocokan *string*, tab “Hasil Filtering” 55

Gambar 4.11 Antarmuka pencocokan *string*, tab “Substring K-gram” 56

Gambar 4.12 Antarmuka pencocokan *string*, tab “Remain & Quotient” 56



DAFTAR TABEL

	Halaman
Tabel 2.1 Perbandingan kompleksitas waktu algoritma pencocokan <i>string</i>	12
Tabel 2.2 Kesamaan antar dua <i>string</i> dari kata <i>photography</i> , <i>photographic</i> , dan <i>phonetic</i>	21
Tabel 3.1 Pencocokan <i>string</i>	39
Tabel 3.2 Pencocokan <i>string</i> setelah <i>synonym recognition</i>	40
Tabel 5.1 Data uji dokumen teks.....	58
Tabel 5.2 Tabel Uji <i>k-gram</i> dengan modulo 79	59
Tabel 5.3 Tabel Uji <i>k-gram</i> dengan modulo 83	59
Tabel 5.4 Tabel Uji <i>k-gram</i> dengan modulo 89	59
Tabel 5.5 Tabel Uji <i>k-gram</i> dengan modulo 97	59
Tabel 5.6 Tabel Uji <i>k-gram</i> dengan modulo 101	59
Tabel 5.7 Tabel Uji <i>k-gram</i> dengan modulo 103	60
Tabel 5.8 Tabel Uji <i>k-gram</i> dengan modulo 127	60
Tabel 5.9 Tabel Uji <i>k-gram</i> dengan modulo 151	60
Tabel 5.10 Tabel Uji <i>k-gram</i> dengan modulo 173	60
Tabel 5.11 Tabel Uji <i>k-gram</i> dengan modulo 251	60
Tabel 5.12 Uji coba <i>synonym recognition</i>	64

