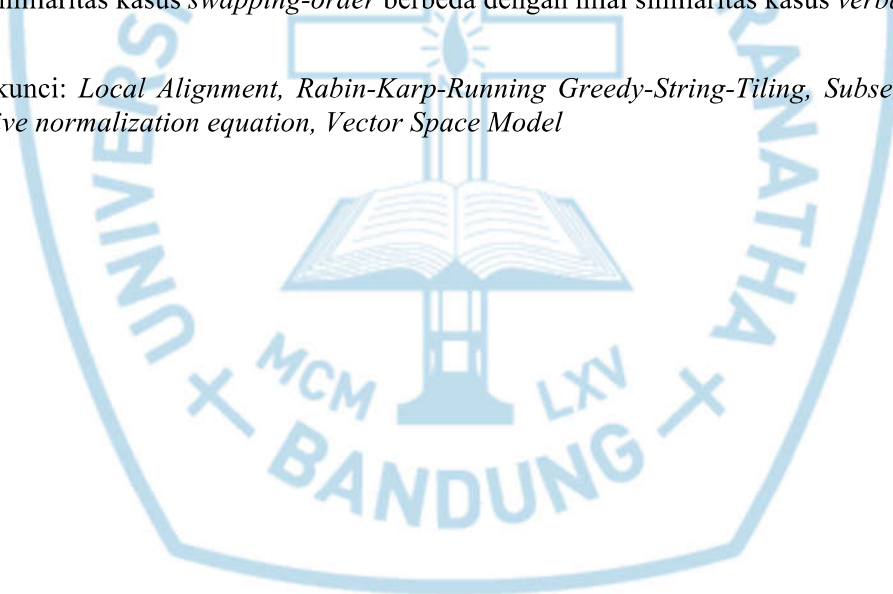


ABSTRAK

Plagiarisme merupakan tindakan ilegal dalam kegiatan akademis. Dibutuhkan suatu sistem yang dapat membantu manusia dalam melakukan pengecekan dugaan plagiarisme pada kode sumber. Akan tetapi, sistem pengecekan plagiarisme kode sumber yang ada hingga saat ini berjalan lambat, hal ini dikarenakan tingkat kompleksitas dari algoritma *string-matching* yang digunakan untuk melakukan pengecekan memiliki tingkat kompleksitas yang tinggi, yaitu $O(n^3)$. Selain permasalahan lambatnya waktu proses, sistem juga dirasa kurang sensitif dalam menilai plagiarisme bentuk *swapping-order*. Sehingga, pada penelitian ini, akan dibangun sistem yang lebih efisien dalam segi waktu proses dalam penerapan *string-matching-algorithm*, berupa algoritma *Rabin-Karp-Running Greedy-String-Tiling* dan *Local Alignment*. Dikatakan efisien, karena sistem yang dibangun akan menerapkan IR berbasis *Vector Space Model* sebagai *initial filtering* dengan tujuan untuk mengurangi kandidat pasangan kode sumber yang akan dicek nilai similaritas menggunakan persamaan dalam *string-matching*. Berbeda pula dengan sistem yang ada, pada penelitian ini akan diterapkan pula perhitungan similaritas yang sensitif terhadap jumlah *subsequences* terbentuk. Berdasarkan pengujian yang dilakukan, didapatkan hasil bahwa dalam segi waktu, sistem yang dibangun lebih unggul karena terdapat pengurangan waktu yang cukup signifikan dan penerapan *subsequence-sensitive normalization equation* yang ditawarkan terbukti lebih sensitif terhadap jumlah *subsequences* terbentuk sehingga nilai similaritas kasus *swapping-order* berbeda dengan nilai similaritas kasus *verbatim*.

Kata kunci: *Local Alignment*, *Rabin-Karp-Running Greedy-String-Tiling*, *Subsequence-sensitive normalization equation*, *Vector Space Model*



ABSTRACT

Plagiarism is illegal in academic activities. It takes a system that can help human beings in checking plagiarism allegations in the source code. However, the current code plagiarism checking system is running slowly, this is done at the level of the string-matching algorithm used to check which has a high degree of complexity, i.e. $O(n^3)$. In addition, the process is also considered less sensitive in assessing the swapping-order plagiarism. So, in this research, a more efficient system will be built in some process time in the application of string-matching-algorithm, namely Rabin-Karp-Running Greedy-String-Tiling algorithm and Local Alignment. It is said to be efficient, since the built system will apply IRs based on the Vector Space Model as initial screening to reduce the emission of code pairs to be checked for similarity values by using the formula in string-matching. In contrast to existing systems, in this study will also be made similarity calculations are sensitive to the number of subsequences that are formed. Based on the tests conducted, it is found that in terms of time, the building system is superior because there is a significant time and the effect of subsequence-sensitive normalization equation is more apparent to the number of events formed from cases equal to the value of similarity case verbatim.

Keywords: Local Alignment, Rabin-Karp-Running Greedy-String-Tiling, Subsequence-sensitive normalization equation, Vector Space Model



DAFTAR ISI

LEMBAR PENGESAHAN	i
PERNYATAAN ORISINALITAS LAPORAN PENELITIAN	ii
PERNYATAAN PUBLIKASI LAPORAN PENELITIAN	iii
PRAKATA.....	iv
ABSTRAK.....	vi
ABSTRACT.....	vii
DAFTAR ISI.....	viii
DAFTAR GAMBAR.....	xi
DAFTAR TABEL.....	xiii
DAFTAR KODE SUMBER.....	xiv
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	2
1.3 Tujuan Pembahasan	2
1.4 Ruang Lingkup.....	3
1.5 Sumber Data.....	3
1.6 Sistematika Penyajian	3
BAB 2 KAJIAN TEORI	5
2.1 Plagiarisme.....	5
2.2 Pendekatan Deteksi Plagiarisme Kode Sumber	7
2.3 Sistem Deteksi Plagiarisme Kode Sumber	8
2.4 Pemodelan <i>Vector Space Model</i>	9
2.5 Algoritma <i>Running-Karp-Rabin Greedy-String-Tiling</i>	10
2.6 Algoritma <i>Local Alignment</i>	11
2.7 ANTLR	12

BAB 3 ANALISIS DAN RANCANGAN SISTEM.....	14
3.1 Analisis	14
3.2 Gambaran Aplikasi	23
3.2.1 Persyaratan Antarmuka Eksternal.....	24
3.2.2 Antarmuka dengan Pengguna	24
3.2.3 Antarmuka Perangkat Keras	24
3.2.4 Antarmuka Perangkat Lunak	24
3.2.5 Fitur Produk Perangkat Lunak.....	25
3.2.5.1 Fitur Menghitung Nilai Similaritas.....	25
3.2.5.2 Fitur Menampilkan Warna Gradasi Pada Matriks	27
3.3 Pemodelan Perangkat Lunak.....	27
3.3.1 <i>Use Case</i> Aplikasi.....	27
3.3.2 <i>Use Case Scenario</i>	28
3.3.3 <i>Class Diagram</i>	29
3.3.4 Rancangan <i>User Interface</i>	30
BAB 4 IMPLEMENTASI	34
4.1 Implementasi Tampilan Antar Muka.....	34
4.1.1 Implementasi Tampilan Utama.....	34
4.1.2 Implementasi Tampilan <i>Source Code Preview</i>	37
4.1.3 Implementasi Tampilan <i>Token Match</i>	38
4.2 Implementasi Pemodelan VSM	38
4.3 Implementasi Algoritma RKRGS.....	41
4.4 Implementasi Algoritma <i>Local Alignment</i>	44
BAB 5 Pengujian	47
5.1 Skenario Pengujian	47
5.2 <i>Human Perspective</i>	47

5.2.1 Pengujian Dampak Penalti Terhadap <i>Human Perspective: Dataset</i> Terkontrol	47
5.2.2 Pengujian Dampak Penalti Terhadap <i>Human Perspective: Dataset</i> Empirik	50
5.2.3 Pengujian Aspek Pendeteksian Plagiarisme	60
5.3 Waktu Proses Algoritma	60
5.3.1.1 <i>Introductory Programming Plagiarism Dataset</i>	60
5.3.1.1.1 RKRGSST	61
5.3.1.1.2 <i>Local Alignment</i>	62
5.3.1.1.3 RKRGSST Vs. <i>Local Alignment</i>	63
5.3.1.2 <i>Course Dataset</i>	64
5.3.1.2.1 RKRGSST	64
5.3.1.2.2 <i>Local Alignment</i>	65
5.3.1.2.3 RKRGSST Vs. <i>Local Alignment</i>	66
5.4 Pengujian Aplikasi	67
5.4.1 <i>Black Box Testing</i>	67
5.4.2 Perbandingan Nilai Similaritas	68
5.4.2.1 Running-Karp-Rabin Greedy-String-Tiling dan Local Alignment	68
5.4.2.2 Running-Karp-Rabin Greedy-String-Tiling dan Vector Space Model	69
5.4.2.3 RKRGSST <i>Standard</i> dan <i>Penalty</i>	70
5.4.3 Fungsionalitas Aplikasi.....	70
BAB 6 Simpulan dan saran.....	74
6.1 Simpulan	74
6.2 Saran	74
DAFTAR PUSTAKA	75

DAFTAR GAMBAR

Gambar 2.1 Tampilan Antar Muka JPlag	8
Gambar 2.2 Tampilan Antar Muka P.L.A.G.U.E	9
Gambar 2.3 Cara Kerja ANTLR.....	13
Gambar 3.1 <i>Flowcharts</i> Cara Kerja Aplikasi	17
Gambar 3.2 <i>Flowchart</i> Langkah Penggunaan User	26
Gambar 3.3 <i>Use Case Aplikasi</i>	28
Gambar 3.4 <i>Class Diagram</i> Aplikasi.....	30
Gambar 3.5 Rancangan <i>User Interface</i> Tampilan Utama Aplikasi	32
Gambar 3.6 Rancangan <i>User Interface Source Code Preview</i>	32
Gambar 3.7 Rancangan <i>User Interface Token Preview</i>	33
Gambar 4.1 Tampilan Awal Aplikasi	35
Gambar 4.2 Tampilan Aplikasi Saat Menggunakan Pendekatan <i>String Matching</i>	36
Gambar 4.3 Tampilan Aplikasi Saat Menggunakan Pendekatan VSM.....	36
Gambar 4.4 Tampilan Aplikasi Saat Menggunakan Pendekatan VSM+ <i>String Matching</i>	37
Gambar 4.5 Tampilan Aplikasi bagian <i>Source Code Preview Dialog</i>	37
Gambar 4.6 Tampilan Aplikasi bagian <i>Token Match Dialog</i>	38
Gambar 5.1 <i>Averaged Negated Rank Based on Human Rank</i>	48
Gambar 5.2 Persebaran Similaritas <i>Dataset</i> Terkontrol	49
Gambar 5.3 Persentase Jumlah Kode Sumber yang Dipilih Responden Untuk Setiap Skenario	50
Gambar 5.4 Persebaran Similaritas <i>Dataset</i> Empirik.....	51
Gambar 5.5 Perbandingan Waktu Proses Algoritma <i>Introductory Programming Plagiarism Dataset RKRGS</i> T.....	61
Gambar 5.6 Perbandingan Waktu Proses Algoritma <i>Introductory Programming Plagiarism Dataset LA</i>	62
Gambar 5.7 Waktu Proses Algoritma <i>Introductory Programming Plagiarism Dataset Based on Project RKRGS</i> T dan <i>LA</i>	64
Gambar 5.8 Perbandingan Nilai Similaritas RKRGST dan <i>Local Alignment</i>	69
Gambar 5.9 Perbandingan Nilai Similaritas RKRGST dan VSM.....	69

Gambar 5.10 Perbandingan Nilai Similaritas RKRGS <i>Average</i> dan <i>Penalty</i>	70
Gambar 5.11 Ringkasan Tanggapan Terhadap Pernyataan Pertama	71
Gambar 5.12 Ringkasan Tanggapan Terhadap Pernyataan Kedua	71
Gambar 5.13 Ringkasan Tanggapan Terhadap Pernyataan Ketiga	72
Gambar 5.14 Ringkasan Tanggapan Terhadap Pernyataan Keempat	72
Gambar 5.15 Ringkasan Tanggapan Terhadap Pernyataan Kelima	73



DAFTAR TABEL

Tabel 2.1 Contoh <i>Natural-Language-based Plagiarism</i>	6
Tabel 3.1 Kumpulan Token Kode Sumber A	18
Tabel 3.2 Kumpulan Token Kode Sumber B.....	19
Tabel 3.3 Kumpulan Token Kode Sumber B.....	21
Tabel 3.4 Nilai Similaritas VSM	23
Tabel 3.5 Nilai Similaritas RKRGSST <i>average penalty</i>	23
Tabel 3.6 Nilai Similaritas <i>Local Alignment</i>	23
Tabel 3.7 Nilai Similaritas VSM- <i>string matching</i> RKRGSST <i>average penalty</i>	23
Tabel 3.8 Nilai Similaritas VSM- <i>string matching Local Alignment</i>	23
Tabel 3.9 <i>Use Case Scenario</i> Aplikasi	28
Tabel 5.1 Nilai <i>Pearson Correlation Dataset</i> Terkontrol.....	49
Tabel 5.2 Nilai Korelasi <i>Dataset</i> Empirik	51
Tabel 5.3 Pengujian Aspek Plagiarisme	60
Tabel 5.4 Waktu Proses Algoritma <i>Introductory Programming Plagiarism Dataset Based on Project RKRGSST</i>	61
Tabel 5.5 Waktu Perbandingan Algoritma (ms) <i>Dataset Real Plagiarism Based on Project Local Alignment</i>	62
Tabel 5.6 Perbandingan Waktu Proses Algoritma <i>Introductory Programming Plagiarism Dataset Based on Project RKRGSST dan Local Alignment</i>	63
Tabel 5.7 Jumlah <i>Token Course Dataset</i>	64
Tabel 5.8 Waktu Proses Algoritma (ms) <i>Course Dataset RKRGSST</i>	65
Tabel 5.9 Waktu Proses Algoritma (ms) <i>Course Dataset LA</i>	65
Tabel 5.10 Waktu Proses Algoritma (ms) <i>Course Dataset RKRGSST dan LA</i>	66
Tabel 5.11 Pengujian <i>Black Box</i> Aplikasi <i>Prototype</i>	67
Tabel 5.12 Nilai Rata-rata dan Standar Deviasi Pernyataan.....	73

DAFTAR KODE SUMBER

Kode Sumber 2.1 Target Plagiarisme	6
Kode Sumber 2.2 Hasil Plagiarisme	7
Kode Sumber 2.3 <i>Top-Level Algorithm</i>	10
Kode Sumber 2.4 <i>Scan Pattern Algorithm</i>	11
Kode Sumber 2.5 <i>Markstring Algorithm</i>	11
Kode Sumber 3.1 Target Plagiarisme A	18
Kode Sumber 3.2 Kode Sumber B Hasil Plagiarisme Kode Sumber A	19
Kode Sumber 3.3 Kode Sumber C Hasil Plagiarisme Kode Sumber A	21
Kode Sumber 4.1 VSM <i>Generate Attribute Index</i>	39
Kode Sumber 4.2 VSM <i>Cosine Similarity</i>	40
Kode Sumber 4.3 RKRGS <i>Top-Level Algorithm</i>	41
Kode Sumber 4.4 <i>Scanpattern</i> Bagian 1.....	42
Kode Sumber 4.5 <i>Scanpattern</i> Bagian 2.....	43
Kode Sumber 4.6 <i>Markstring</i>	44
Kode Sumber 4.7 RKRGS <i>Calculation Average Similarity Standard</i>	44
Kode Sumber 4.8 <i>Local Alignment Calculation Similarity Matrix Score</i>	45
Kode Sumber 4.9 <i>Local Alignment Traceback</i>	46
Kode Sumber 5.1 Kode Sumber A Hasil Plagiarisme Kasus 1	53
Kode Sumber 5.2 Kode Sumber B Hasil Plagiarisme Kasus 1.....	54
Kode Sumber 5.3 Kode Sumber Original Kasus 1	54
Kode Sumber 5.4 Kode Sumber A Hasil Plagiarisme Kasus 2	55
Kode Sumber 5.5 Kode Sumber B Hasil Plagiarisme Kasus 2.....	56
Kode Sumber 5.6 Kode Sumber Original Kasus 2	57
Kode Sumber 5.7 Kode Sumber A Hasil Plagiarisme Kasus 3	58
Kode Sumber 5.8 Kode Sumber B Hasil Plagiarisme Kasus 3.....	59
Kode Sumber 5.9 Kode Sumber Original Kasus 3	59