

PERBANDINGAN ALGORITMA *WINNOWER* DAN *BIWORD WINNOWER* UNTUK PENDETEKSAN PLAGIASI DOKUMEN

Idrus Tamam
(14.04.1.1.2.00149)

Pembimbing 1 : Firdaus Solihin, S.Kom., M.Kom
Pembimbing 2 : Yoga Dwitya Pramudita, S.Kom., M.Cs.

ABSTRAK

Plagiasi merupakan pengambilan ide atau karangan orang lain dan menjadikannya seolah-olah ide atau karangannya sendiri. Plagiasi merupakan salah satu pelanggaran hak cipta. Untuk mengatasi plagiasi yang sudah banyak terjadi bahkan dalam kalangan akademik maka dibuatlah aplikasi untuk mendeteksi plagiasi. Teknik untuk mendeteksi plagiasi sudah cukup banyak salah satunya adalah *winnower*. *Winnower* merupakan algoritma plagiasi yang berbasis karakter. Algoritma ini merupakan salah satu algoritma terbaik untuk mendapatkan nilai similarity baik dalam segi akurasi ataupun performansi. Kemudian muncul juga algoritma modifikasi atau pengembangan dari *winnower* yaitu *biword winnower*. *Biword winnower* merupakan algoritma yang berbasis frasa atau kata. Penelitian ini bermaksud untuk membandingkan algoritma *winnower* dan *biword winnower* untuk mencari tahu algoritma mana yang lebih efektif dan lebih efisien dalam mendeteksi plagiasi pada dokumen. Hasil dari perbandingan tersebut adalah algoritma *winnower* lebih efisien karena unggul dalam waktu komputasi sedangkan *biword winnower* lebih efektif karena unggul dalam hal keakuratan kemiripan dalam mendeteksi plagiasi dokumen.

Kata Kunci : deteksi plagiasi, *winnower*, *biword winnower*.



ABSTRACT

Plagiarism is taking other people's ideas or essay and making it as the idea or composed his own. Plagiarism is a violation of copyright. To resolve the plagiarism that has been going on even in academic circles then made an application to detect plagiarism. Techniques for detecting plagiarism is already pretty much one of them is Winnowing. Winnowing is an algorithm-based plagiarism character. This algorithm is one of the best algorithm to get a similarity value both in terms of accuracy or performance. Then came also the modification or development of winnowing algorithms that is biword Winnowing algorithms. Biword Winnowing is an algorithm based on words or phrases. This study intends to compare winnowing algorithms and biword Winnowing algorithms to find out which algorithm is more effective and more efficient in detecting plagiarism in documents. The results of these comparisons is winnowing algorithms is more efficient because of its superior in computing time and biword Winnowing is more effective because of its superior in terms of accuracy in detecting plagiarism document similarity.

Keywords : *plagiarism detection, winnowing, biword winnowing*



KATA PENGANTAR

Alhamdulillah puji syukur senantiasa penulis ucapkan atas nikmat, karunia, rahmat, serta hidayah Allah SWT, sehingga penulis dapat menyelesaikan Tugas Akhir yang berjudul :

“PERBANDINGAN ALGORITMA *WINNOWER* DAN *BIWORD* *WINNOWER* UNTUK PENDETEKSIAN PLAGIASI DOKUMEN”

Dalam penyusunan Tugas Akhir ini, penulis menyadari bahwa semua ini tidak lepas dari peran dan dukungan dari berbagai pihak yang telah dengan tulus senantiasa memberikan dukungan, bimbingan, dan motivasi, hingga akhirnya Tugas Akhir ini dapat terselesaikan. Oleh sebab itu penulis tak lupa menyampaikan terima kasih yang sebesar-besarnya kepada :

1. Allah SWT yang telah memberikan petunjuk, rahmat, serta hidayah-Nya dalam menyelesaikan seluruh proses dalam Tugas Akhir ini. Shalawat serta salam semoga selalu tercurahkan kepada junjungan Nabi Besar Muhammad SAW yang menjadi teladan dan telah menyampaikan petunjuk-Nya.
2. Kedua orang tuaku, serta seluruh keluarga besar yang selalu memberikan dukungan dan semangat, baik dalam bentuk materi, moril, maupun doa-doa yang selalu dipanjatkan kepada Allah SWT. Semoga semuanya senantiasa sehat dan selalu dalam lindungan Allah SWT.
3. Bapak Firdaus Solihin, S.Kom., M.Kom. selaku dosen pembimbing I dan Yoga Dwitya Pramudita, S.Kom., M.Cs. selaku dosen pembimbing II. Terimakasih karena bapak telah banyak meluangkan waktu untuk membimbing dan memberikan arahan kepada penulis, sehingga mampu terselesaikannya Tugas Akhir ini.



4. Bapak dan Ibu Dosen Fakultas Teknik, khususnya Prodi Teknik Informatika Universitas Trunojoyo Madura.
5. Teman-teman Teknik Informatika UTM dan Asisten Praktikum Lab Teknik Informatika atas bimbingan dan dukungannya selama ini.
6. Teman-teman yang senasib dan seperjuangan karena sama-sama mahasiswa alih jenjang Reyza, Bagus, Mukhlis, dll
7. Dan untuk semua pihak yang tidak dapat disebutkan satu persatu yang turut berperan dalam Tugas Akhir ini.

Akhir kata, penulis mengucapkan terima kasih dan mohon maaf apabila terdapat kesalahan dalam pengetikan, penulisan, dan penyampaian dalam Tugas Akhir ini. Semoga Tugas Akhir ini dapat bermanfaat dan dapat dikembangkan dengan lebih baik nantinya.

Bangkalan, 16 Februari 2017

Penulis

DAFTAR ISI

	Hal.
COVER	i
HALAMAN JUDUL	iii
HALAMAN PENGESAHAN	v
HALAMAN PERNYATAAN ORISINALITAS	vii
HALAMAN MOTTO	ix
HALAMAN PERSEMBAHAN	xi
ABSTRAK	xiii
KATA PENGANTAR	xv
DAFTAR ISI	xvii
DAFTAR GAMBAR	xix
DAFTAR TABEL	xxiii
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Perumusan Masalah	2
1.3. Tujuan dan Manfaat	2
1.3.1. Tujuan	2
1.3.2. Manfaat	3
1.4. Batasan Masalah	3
1.5. Metodologi Penelitian	3
1.6. Sistematika Penulisan	4
BAB II STUDI PUSTAKA	5
2.1. Penelitian Sebelumnya	5
2.2. Landasan Teori	7
2.2.1. <i>Plagiarisme</i>	7
2.2.2. Pendekatan Pengukuran Kemiripan	7
2.2.3. K-gram	8
2.2.4. Hashing	8
2.2.5. Window	9
2.2.6. Winnowing	9
2.2.7. Biword Winnowing	9
2.2.8. ASCII	9
2.2.9. MD5	10
2.2.10. Jaccard Coefficient	10
2.2.11. Contoh Penerapan Algoritma Winnowing	10
2.2.12. Contoh Penerapan Algoritma Biword Winnowing	13
BAB III PERANCANGAN SISTEM	17



3.1. Alur Program	17
3.1.1. Usecase Diagram Aplikasi	17
3.1.2. Flowchart Cek Plagiasi	18
3.2. Arsitektur Sistem	19
3.3. Kerangka Desain Interface Sistem.....	20
3.2.1. Interface Aplikasi	20
3.2.2. Interface Menu Home	21
3.2.3. Interface Menu Cek Plagiasi	21
3.2.4. Interface Help.....	22
3.4. Metode	23
3.3.1. Flowchart Winnowing.....	23
3.3.2. Flowchart Biword Winnowing.....	24
BAB IV IMPLEMENTASI SISTEM.....	27
4.1. Spesifikasi Perangkat.....	27
4.2. Implementasi <i>User Interface</i>	27
4.2.1. Halaman Home	27
4.2.2. Halaman Cek Plagiasi	28
4.2.3. Halaman Help	31
4.3. Kode Program.....	31
4.3.1. Kode Program Winnowing	32
4.3.2. Kode Program Biword Winnowing.....	35
4.4. Pengujian	38
4.4.1. Uji Coba Kemiripan Data 100%	40
4.4.2. Uji Coba Kemiripan Data 50%	50
4.4.3. Uji Coba Kemiripan Data 25%	71
BAB V PENUTUP	93
5.1. Kesimpulan	93
5.2. Saran.....	93
DAFTAR PUSTAKA.....	95
LAMPIRAN	97
Lampiran A :	97
TENTANG PENULIS	99



DAFTAR GAMBAR

	Hal.
Gambar 3. 1 usecase aplikasi	17
Gambar 3. 2 flowchart cek plagiasi	18
Gambar 3. 3 Arsitektur sistem	19
Gambar 3. 4 Interface Aplikasi	20
Gambar 3. 5 Interface Menu Home	21
Gambar 3. 6 Interface Menu Cek Plagiasi	21
Gambar 3. 7 Interface Help	22
Gambar 3. 8 Flowchart Winnowing	23
Gambar 3. 9 Flowchart Biword Winnowing	24
 Gambar 4. 1 Tampilan Menu Home	 28
Gambar 4. 2 Tampilan Menu Cek Plagiasi	29
Gambar 4. 3 Tampilan Tabel History	30
Gambar 4. 4 Tampilan Menu Help	31
Gambar 4. 5 Preprocessing	32
Gambar 4. 6 Tokenisasi	32
Gambar 4. 7 Hashing	33
Gambar 4. 8 Windowing	33
Gambar 4. 9 Pemilihan Fingerprint	34
Gambar 4. 10 Irisan	34
Gambar 4. 11 Gabungan	35
Gambar 4. 12 Preprocessing	35
Gambar 4. 13 Lowercase	35
Gambar 4. 14 Set Bigram	36
Gambar 4. 15 Hashing	37
Gambar 4. 16 Windowing	37
Gambar 4. 17 Pemilihan fingerprint	37
Gambar 4. 18 Alur data pengujian	38
Gambar 4. 19 Pengaruh k-gram terhadap waktu pada jurnal dengan kemiripan data 100% menggunakan algoritma winnowing	41
Gambar 4. 20 Pengaruh k-gram terhadap waktu pada abstrak dengan kemiripan 100% menggunakan algoritma biword winnowing	41
Gambar 4. 21 Pengaruh bilangan prima terhadap waktu pada jurnal dengan kemiripan 100% menggunakan algoritma winnowing	43



Gambar 4. 22 Pengaruh bilangan prima terhadap waktu komputasi pada jurnal dengan kemiripan 100% dengan algoritma biword winnowing ..	43
Gambar 4. 23 Pengaruh bilangan prima terhadap waktu komputasi pada abstrak dengan kemiripan 100% dengan algoritma winnowing	44
Gambar 4. 24 Pengaruh bilangan prima terhadap waktu komputasi pada abstrak dengan kemiripan 100% menggunakan algoritma biword winnowing	44
Gambar 4. 25 Pengaruh window terhadap waktu pada jurnal dengan kemiripan 100% menggunakan algoritma winnowing	46
Gambar 4. 26 Pengaruh window terhadap waktu pada jurnal dengan kemiripan 100% menggunakan algoritma biword winnowing	46
Gambar 4. 27 Pengaruh window terhadap waktu pada abstrak dengan kemiripan 100% menggunakan algoritma winnowing	47
Gambar 4. 28 Pengaruh window terhadap waktu komputasi pada abstrak dengan kemiripan 100% menggunakan algoritma biword winnowing..	47
Gambar 4. 29 Pengaruh k-gram terhadap similarity pada jurnal dengan kemiripan data 50%	51
Gambar 4. 30 Pengaruh k-gram terhadap similarity pada abstrak dengan kemiripan data 50%	52
Gambar 4. 31 Pengaruh k-gram terhadap waktu pada jurnal dengan kemiripan data 50%	53
Gambar 4. 32 Pengaruh k-gram terhadap waktu pada abstrak dengan kemiripan data 50%	54
Gambar 4. 33 Pengaruh bilangan prima terhadap similarity pada jurnal dengan kemiripan 50% menggunakan algoritma winnowing.....	55
Gambar 4. 34 Pengaruh bilangan prima terhadap similarity pada jurnal dengan kemiripan 50% menggunakan algoritma biword winnowing....	56
Gambar 4. 35 Pengaruh bilangan prima terhadap similarity pada abstrak dengan kemiripan 50% menggunakan algoritma winnowing.....	56
Gambar 4. 36 Pengaruh bilangan prima terhadap similarity pada abstrak dengan kemiripan 50% menggunakan algoritma biword winnowing....	57
Gambar 4. 37 Pengaruh bilangan prima terhadap waktu pada jurnal dengan kemiripan 50% menggunakan algoritma winnowing.	58
Gambar 4. 38 Pengaruh bilangan prima terhadap waktu pada jurnal dengan kemiripan 50% menggunakan algoritma biword winnowing....	59
Gambar 4. 39 Pengaruh bilangan prima terhadap waktu pada abstrak dengan kemiripan 50% menggunakan algoritma winnowing.....	59
Gambar 4. 40 Pengaruh bilangan prima terhadap waktu pada abstrak dengan kemiripan 50% menggunakan algoritma biword winnowing....	60



Gambar 4. 41 Pengaruh window terhadap similarity pada jurnal dengan kemiripan 50% menggunakan algoritma winnowing.....	61
Gambar 4. 42 Pengaruh window terhadap similarity pada jurnal dengan kemiripan 50% menggunakan algoritma biword winnowing	61
Gambar 4. 43 Pengaruh window terhadap similarity pada abstrak dengan kemiripan 50% menggunakan algoritma winnowing.....	62
Gambar 4. 44 Pengaruh window terhadap similarity pada abstrak dengan kemiripan 50% menggunakan algoritma biword winnowing	62
Gambar 4. 45 Pengaruh window terhadap waktu pada jurnal dengan kemiripan 50% menggunakan algoritma winnowing.....	64
Gambar 4. 46 Pengaruh window terhadap waktu pada jurnal dengan kemiripan 50% menggunakan algoritma biword winnowing	64
Gambar 4. 47 Pengaruh window terhadap waktu pada abstrak dengan kemiripan 50% menggunakan algoritma winnowing.....	65
Gambar 4. 48 Pengaruh window terhadap waktu pada abstrak dengan kemiripan 50% menggunakan algoritma biword winnowing	66
Gambar 4. 49 Pengaruh k-gram terhadap similarity pada jurnal dengan kemiripan 25%	72
Gambar 4. 50 Pengaruh k-gram terhadap similarity pada abstrak dengan kemiripan 25%	73
Gambar 4. 51 Pengaruh k-gram terhadap waktu pada jurnal dengan kemiripan 25%	74
Gambar 4. 52 Pengaruh k-gram terhadap waktu pada abstrak dengan kemiripan 25%	74
Gambar 4. 53 Pengaruh bilangan prima terhadap similarity pada jurnal dengan kemiripan 25% menggunakan algoritma winnowing	76
Gambar 4. 54 Pengaruh bilangan prima terhadap similarity pada jurnal dengan kemiripan 25% menggunakan algoritma biword winnowing ...	76
Gambar 4. 55 Pengaruh bilangan prima terhadap similarity pada abstrak dengan kemiripan 25% menggunakan algoritma winnowing	77
Gambar 4. 56 Pengaruh bilangan prima terhadap similarity pada abstrak dengan kemiripan 25% menggunakan algoritma biword winnowing ...	77
Gambar 4. 57 Pengaruh bilangan prima terhadap waktu pada jurnal dengan kemiripan 25% menggunakan algoritma winnowing	79
Gambar 4. 58 Pengaruh bilangan prima terhadap waktu pada jurnal dengan kemiripan 25% menggunakan algoritma biword winnowing ...	79
Gambar 4. 59 Pengaruh bilangan prima terhadap waktu pada abstrak dengan kemiripan 25% menggunakan algoritma winnowing	80



Gambar 4. 60 Pengaruh bilangan prima terhadap waktu pada abstrak dengan kemiripan 25% menggunakan algoritma biword winnowing....	80
Gambar 4. 61 Pengaruh window terhadap similarity pada jurnal dengan kemiripan 25% menggunakan algoritma winnowing	82
Gambar 4. 62 Pengaruh window terhadap similarity pada jurnal dengan kemiripan 25% menggunakan algoritma biword winnowing.	82
Gambar 4. 63 Pengaruh window terhadap similarity pada abstrak dengan kemiripan 25% menggunakan algoritma winnowing	83
Gambar 4. 64 Pengaruh window terhadap similarity pada abstrak dengan kemiripan 25% menggunakan algoritma biword winnowing	83
Gambar 4. 65 Pengaruh window terhadap waktu pada jurnal dengan kemiripan 25% menggunakan algoritma winnowing	85
Gambar 4. 66 Pengaruh window terhadap waktu pada jurnal dengan kemiripan 25% menggunakan algoritma biword winnowing	85
Gambar 4. 67 Pengaruh window terhadap waktu pada abstrak dengan kemiripan 25% menggunakan algoritma winnowing	86
Gambar 4. 68 Pengaruh window terhadap waktu pada abstrak dengan kemiripan 25% menggunakan algoritma biword winnowing	86



DAFTAR TABEL

	Hal.
Tabel 2. 1 Penelitian Sebelumnya	5
Tabel 4. 1 Spesifikasi Perangkat pada laptop.....	27
Tabel 4. 2 Tabel parameter.....	39
Tabel 4. 3 Pengaruh k-gram terhadap waktu komputasi pada data dengan kemiripan 100% dengan algoritma winnowing.....	40
Tabel 4. 4 Pengaruh bilangan prima terhadap waktu komputasi.....	42
Tabel 4. 5 Pengaruh window terhadap waktu komputasi.....	45
Tabel 4. 6 Nilai parameter dengan waktu terkecil pada jurnal menggunakan algoritma winnowing	48
Tabel 4. 7 Nilai parameter dengan waktu terkecil pada jurnal menggunakan algoritma biword winnowing.....	49
Tabel 4. 8 Nilai parameter dengan waktu terkecil pada abstrak menggunakan algoritma winnowing	49
Tabel 4. 9 Nilai parameter dengan waktu terkecil pada abstrak menggunakan algoritma biword winnowing.....	50
Tabel 4. 10 Pengaruh k-gram terhadap similarity pada kemiripan data 50%	51
Tabel 4. 11 Pengaruh k-gram terhadap waktu pada kemiripan data 50%	52
Tabel 4. 12 Pengaruh bilangan prima terhadap similarity pada kemiripan data 50%.....	54
Tabel 4. 13 Pengaruh bilangan prima terhadap waktu pada kemiripan data 50%	57
Tabel 4. 14 Pengaruh nilai window terhadap similarity pada kemiripan data 50%.....	60
Tabel 4. 15 Pengaruh nilai window terhadap waktu pada kemiripan data 50%	63
Tabel 4. 16 K-gram terbanyak dengan similarity tertinggi pada kemiripan 50%	66
Tabel 4. 17 K-gram terbanyak dengan waktu komputasi terkecil pada kemiripan 50%.....	67
Tabel 4. 18 Bilangan prima terbanyak dengan similarity tertinggi pada kemiripan 50%.....	68
Tabel 4. 19 Bilangan prima terbanyak dengan waktu komputasi terkecil pada kemiripan 50%.....	69



Tabel 4. 20 Nilai window terbanyak dengan similarity tertinggi pada kemiripan 50%.....	69
Tabel 4. 21 Nilai window terbanyak dengan waktu komputasi terkecil pada kemiripan 50%	70
Tabel 4. 22 Pengaruh k-gram terhadap similarity pada kemiripan data 25%.....	71
Tabel 4. 23 Pengaruh k-gram terhadap waktu pada kemiripan data 25%	73
Tabel 4. 24 Pengaruh bilangan prima terhadap similarity pada kemiripan data 25%	75
Tabel 4. 25 Pengaruh bilangan prima terhadap waktu pada kemiripan data 25%.....	78
Tabel 4. 26 Pengaruh window terhadap similarity pada kemiripan data 25%.....	81
Tabel 4. 27 Pengaruh window terhadap waktu pada kemiripan data 25%	84
Tabel 4. 28 k-gram terbanyak dengan similarity tertinggi pada kemiripan 25%.....	87
Tabel 4. 29 k-gram terbanyak dengan waktu komputasi terkecil pada kemiripan 25%.....	88
Tabel 4. 30 Bilangan prima terbanyak dengan similarity tertinggi pada kemiripan 25%.....	88
Tabel 4. 31 Bilangan prima terbanyak dengan waktu komputasi terkecil pada kemiripan 25%	89
Tabel 4. 32 Window terbanyak dengan similarity tertinggi pada kemiripan 25%.....	90
Tabel 4. 33 Window terbanyak dengan waktu komputasi terkecil pada kemiripan 25%.....	90
Tabel 4. 34 Perbandingan waktu komputasi	91
Tabel 4. 35 Perbandingan similarity	92

