

**IMPLEMENTASI ALGORITMA WINNOWER UNTUK
MENDETEKSI KEMIRIPAN PADA DOKUMEN TEKS**

Skripsi



Oleh

OBED KHARISMAN

22084401

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS KRISTEN DUTA WACANA

2012

**IMPLEMENTASI ALGORITMA WINNOWER UNTUK
MENDETEKSI KEMIRIPAN PADA DOKUMEN TEKS**

Skripsi



©
Diajukan kepada Program Studi Teknik Informatika Fakultas Teknologi Informasi

Universitas Kristen Duta Wacana

Sebagai Salah Satu Syarat dalam Memperoleh Gelar

Sarjana Komputer

Disusun Oleh

OBED KHARISMAN

22084401

PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS TEKNOLOGI INFORMASI

UNIVERSITAS KRISTEN DUTA WACANA

2012

PERNYATAAN KEASLIAN SKRIPSI

Saya menyatakan dengan sesungguhnya bahwa skripsi dengan judul :

IMPLEMENTASI ALGORITMA WINNOWER UNTUK MENDETEKSI KEMIRIPAN PADA DOKUMEN TEKS

yang saya kerjakan untuk melengkapi sebagian persyaratan menjadi Sarjana Komputer pada pendidikan Sarjana Program Studi Teknik Informatika Fakultas Teknologi Informasi Universitas Kristen Duta Wacana, bukan merupakan tiruan atau duplikasi dari skripsi keserjanaan di lingkungan Universitas Kristen Duta Wacana maupun di Perguruan Tinggi atau instansi manapun, kecuali bagian yang sumber informasinya dicantumkan sebagaimana mestinya.

Jika dikemudian hari didapati bahwa hasil skripsi ini adalah hasil plagiasi atau tiruan dari skripsi lain, saya bersedia dikenai sanksi yakni pencabutan gelar keserjanaan saya.

Yogyakarta, 7 Desember 2012



OBED KHARISMAN

22084401

HALAMAN PERSETUJUAN

Judul Skripsi : IMPLEMENTASI ALGORITMA WINNOWING
UNTUK MENDETEKSI KEMIRIPAN PADA
DOKUMEN TEKS

Nama : OBED KHARISMAN

NIM : 22084401

Matakuliah : Skripsi (Tugas Akhir)

Kode : TIW276

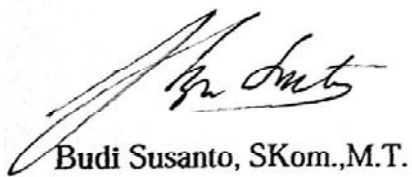
Semester : Gasal

Tahun Akademik : 2012/2013

Telah diperiksa dan disetujui di
Yogyakarta,
Pada Tanggal, 10 Desember 2012




Dosen Pembimbing I



Budi Susanto, SKom.,M.T.

Dosen Pembimbing II



Ir. Sri Suwarno, M. Eng

HALAMAN PENGESAHAN

IMPLEMENTASI ALGORITMA WINNOWER UNTUK MENDETEKSI KEMIRIPAN PADA DOKUMEN TEKS

Oleh: OBED KHARISMAN / 22084401

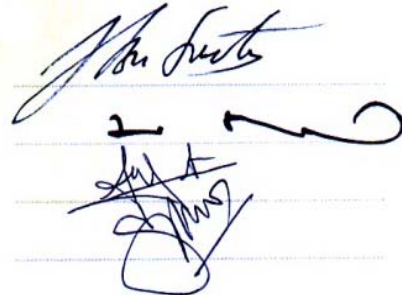
Dipertahankan di depan Dewan Penguji Skripsi
Program Studi Teknik Informatika Fakultas Teknologi Informasi
Universitas Kristen Duta Wacana - Yogyakarta
Dan dinyatakan diterima untuk memenuhi salah satu syarat memperoleh gelar
Sarjana Komputer
pada tanggal 26 November 2012

Yogyakarta, 10 Desember 2012

Mengesahkan,

Dewan Penguji:

1. Budi Susanto, SKom.,M.T.
2. Ir. Sri Suwarno, M.Eng.
3. Aditya Wikan Mahastama, S.Kom
4. Hendro Setiadi, M.Eng



Dekan

(Drs. Wimmie Handiwidjojo, MIT.)

Ketua Program Studi

(Nugroho Agus Haryono, M.Si)

UCAPAN TERIMAKASIH

Dengan penuh kerendahan hati dan rasa yang tulus, saya mengucapkanterimakasih berlimpah kepada:

1. *Special thanks to my God Almighty Jesus Christ* yang senantiasa memberikan kekuatan, berkat yang melimpah selama penulis mejalani masa perkuliahan hingga pada akhirnya. Terima kasih buat suka dan duka yang Tuhan ijinkan dalam kehidupan penulis.
2. Bapak Budi Susanto S.Kom., M.T dan bapak Ir. Sri Suwarno, M. Eng yang dengan kesabaran dan kasihnya mau membimbing penulis untuk menyelesaikan setiap lembaran yang ada dalam skripsi ini.
3. Untuk seluruh dosen Fakultas Teknologi Informasi Program Studi Teknik Informatika beserta seluruh staff administratif .
4. Terima kasih yang sedalam-dalamnya buat setiap tetesan keringat dari Papa (Alm) dan Mama. Ada begitu banyak pergumulan yang dihadapi oleh penulis, akan tetapi sangat banyak juga untaian doa dari Papa dan Mama demi keberhasilan anak-anak. “makasi ee.. mama, te bosan-bosan trima kita pe telpon2 yang bikin saki telinga apalagi kalo soal ‘*Kombahina Bonggoli ku da a’.. makasi papa untuk semua yang sudah papa berikan..selalu sa ingat papa punya sindiran terakhir -mopigi sama papa baru di kupang..? haha*”☺
5. Terima kasih buat Om Asi, Me’Lin yang selalu senantiasa memberikan dukungan dan perhatian kepada penulis selama masa perkuliahan hingga pada akhirnya.
6. Untuk Kakak : terima kasih k’Ity (Sally Christine Cholly Puahadi S.Si., Apt – *permintaan khusus!*) untuk segala dukungannya yang selama ini sudah diberikan. terima kasih buat K’Icha atas segala kebaikannya selama ini. “kalo te ada kamu dua, hmm.. beraaaat! Te ada yg mo di gate-gate kalo so akhir bulan.. hehehe”. Tak lupa terima kasih buat K’donny dan K’dewa untuk semua dukungannya selama ini.

7. Terima kasih buat teman-teman saudara “Green” Ezra Immanuel (AP), Horas, Ady, Lindan (adi saputra) untuk semua perhatiannya selama empat tahun bersama-sama, ada banyak suka dan duka yang terjadi yang membuat kita semua bisa saling memahami menjadi keluarga.
8. Terima kasih untuk para Sahabat “Heinekers” ada Tonce, terima kasih untuk tidak pernah bosan mendengar setiap cerita, terima kasih buat nasehat dan masukannya. “Adit Kasim, makasi ee kawand.. dari kecil torang so sama-sama sampe skarang kuliah torang juga sama-sama.. tapi kau lulus duluan lea..”. Terima kasih juga untuk Arfie, Beby, Fera, Tiwi, Venty, Dian, Felix, Eldon, Horis, k’pey, David, Chandra, Fely. Terima kasih untuk kebersamaan selama empat tahun ini, terima kasih buat cerita kita yang luar biasa. Terima kasih buat Rizal, Sury, Echi, Dito untuk kebersamaannya selama perkuliahan dan selalu memberikan semangat apalagi saat-saat saya lagi jenuh. Kalian sahabat-sahabat yang sangat luar biasa. “Minta maaf eee.. kalo selama ini kita ada kase saki hati ngoni-ngoni samua... ☺”.
9. Terima kasih buat teman-teman “Multimedia GKI Gejayan”. terima kasih untuk semua supportnya & doanya selama ini yang selalu mendukung dan memberikan masukan dan pelajaran yang luar biasa. Terima kasih juga buat teman-teman di GKI Gejayan & TIM Doa GKI Gejayan yang selalu mendukung dalam doa.
10. Terima kasih untuk k’Yelly Nabuasa untuk bantuannya dalam penyusunan skripsi ini sehingga penulis bisa menyelesaikannya dengan baik. “makasi ee kk sudah suru mahasiswanya kumpul tugas di b pu email..”
11. Terima kasih buat k’Marshall Daryance Faah. S.Si yang banyak memberi nasehat-nasehat dengan cara yang sedikit “makan hati” tetapi membuat penulis bisa jadi seperti sekarang ini. “makasiii oo kk nyonk”. Terima kasih buat Onal & JuNiaRson.. “makasi ee kawand...!”
12. Terima kasih untuk teman-teman & saudara-saudara yang tidak dapat saya sebutkan satu persatu dalam ucapan terima kasih ini. Terima kasih untuk semua dukungan doanya. Tuhan Yesus Memberkati.

KATA PENGANTAR

Puji dan syukur ke hadirat Tuhan Yang Maha Esa atas berkat, rahmat dan himat sehingga penulis dapat menyelesaikan Tugas Akhir dengan judul Implementasi Algoritma *Winnowing* Untuk Mendeteksi Kemiripan Pada Dokumen Teks dengan baik.

Penulisan laporan ini merupakan kelengkapan dan pemenuhan dari salah satu syarat dalam memperoleh gelar Sarjana Komputer. Selain itu bertujuan melatih mahasiswa untuk dapat menghasilkan suatu karya yang dapat dipertanggungjawabkan secara ilmiah, sehingga dapat bermanfaat bagi penggunaannya.

Dalam menyelesaikan pembuatan Tugas Akhir ini baik dalam pembuatan program dan penulisan laporan, penulis telah banyak mendapat banyak bimbingan, masukan dan saran dari berbagai pihak, baik secara langsung maupun tidak langsung. Untuk itu dengan segenap kerendahan hati, pada kesempatan ini penulis menyampaikan ucapan terima kasih kepada :

1. Bapak/Ibu Dosen Pembimbing I yang telah memberikan bimbingannya dengan sabar dan baik kepada penulis.
2. Bapak/Ibu Dosen pembimbing II atas bimbingan, petunjuk dan masukan yang diberikan selama pengerjaan tugas ini sejak awal hingga akhir.
3. Semua Dosen Fakultas Teknologi Informasi Program Studi Teknik Informatika Universitas Kristen Duta Wacana.
4. Staf administrasi Fakultas Teknologi Informasi Program Studi Teknik Informatika Universitas Kristen Duta Wacana.
5. Pihak-pihak lain yang tidak dapat penulis sebutkan satu per satu, sehingga Tugas Akhir ini dapat terselesaikan dengan baik.

Penulis menyadari program dan laporan Tugas Akhir ini masih jauh dari sempurna. Oleh karena itu, penulis sangat mengharapkan kritik dan saran yang

membangun dari pembaca sekalian. Sehingga suatu saat penulis dapat memberikan karya yang lebih baik lagi.

Akhir kata penulis ingin meminta maaf bila ada kesalahan baik dalam penyusunan laporan maupun yang pernah penulis lakukan sewaktu penyusunan Tugas Akhir ini. Sekali lagi penulis mohon maaf yang sebesar-besarnya. Dan semoga ini dapat berguna bagi kita semua.

Yogyakarta, November 2012

Penulis



© UKDWN

INTISARI

IMPLEMENTASI ALGORITMA WINNOWER UNTUK MENDETEKSI KEMIRIPAN PADA DOKUMEN TEKS

Semakin mudahnya untuk saling bertukar dan mendapatkan informasi tidak hanya memberikan dampak positif bagi kemajuan teknologi, tetapi juga membawa dampak negatif yang hampir tidak dapat dihindari seperti tindakan *copy paste* yang dapat berujung pada plagiarisme yang sengaja maupun tidak sengaja dilakukan. Tindakan tersebut dapat mematikan kreatifitas seseorang karena sudah terbiasa mengambil sesuatu yang bukan hasil karyanya dengan mudah dan merupakan pelanggaran hak cipta.

Beberapa metode yang dapat digunakan untuk mendeteksi atau menganalisa plagiarisme seperti pencocokan *substring*, kesamaan kata kunci dan dokumen *fingerprintings*. Algoritma *Winnower* merupakan algoritma dokumen *fingerprint* yang dapat digunakan untuk mendeteksi plagiarisme. *Input* dari algoritma *winnower* adalah *string* dari sebuah dokumen dan *output* dari algoritma tersebut adalah nilai-nilai *hash* yang disebut sebagai *fingerprints* dari dokumen tersebut.

Hasil pada penelitian ini berupa sebuah sistem untuk mendeteksi kemiripan pada dokumen teks dan melihat pengaruh dari nilai-nilai parameter yang digunakan terhadap prosentase hasil perhitungan kemiripan dokumen uji.

Kata Kunci : [Algoritma *Winnower*, *Plagiarisme*, *Kemiripan*, *Dokumen Fingerprinting*, *Stopword*]

DAFTAR ISI

SAMPUL	i
HALAMAN JUDUL	ii
PERNYATAAN KEASLIAN TUGAS AKHIR	iii
HALAMAN PERSETUJUAN	iv
HALAMAN PENGESAHAN	v
UCAPAN TERIMA KASIH	vi
KATA PENGANTAR	viii
INTISARI	x
DAFTAR ISI	xi
DAFTAR TABEL	xiv
DAFTAR GAMBAR	xv
DAFTAR LAMPIRAN	xviii
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Perumusan Masalah	2
1.3 Batasan Masalah	2
1.4 Tujuan Penelitian	2
1.5 Metode Penelitian	3
1.6 Sistematika Penulisan	3
BAB 2 TINJAUAN PUSTAKA	5
2.1 Tinjauan Pustaka	5
2.2 Landasan Teori	6
2.2.1 Plagiarisme	5
2.2.1.1 Metode Pendeteksi Plagiarisme	8
2.2.2 Text <i>Preprocessing</i>	9

2.2.3 Hashing	10
2.2.4 K-Grams	11
2.2.5 Rolling Hash	12
2.2.6 Algoritma <i>Winnowing</i>	14
2.2.7 Pengukuran Nilai <i>Similarity</i>	16
BAB 3 ANALISIS DAN PERANCANGAN SISTEM	17
3.1 Spesifikasi Sistem	17
3.2 Gambaran Kerja Sistem	17
3.2.1 Proses <i>Preprocessing</i> Dokumen	18
3.2.2 Proses deteksi kemiripan dokumen menggunakan Algoritma <i>Winnowing</i>	19
3.3 Perancangan Antar Muka Sistem	24
3.3.1 <i>Input</i> Sistem	24
3.3.2 <i>Output</i> Sistem	25
3.3.3 <i>Output</i> Tambahan	26
3.3.3.1 <i>Output</i> Detail Dokumen	27
3.3.3.2 <i>Output</i> Detail Perbandingan Antar Dokumen	28
3.4 Perancangan Database	29
3.4.1 Tabel <i>Stopword</i>	29
3.4.2 Tabel Dokumen Uji	29
3.4.3 Tabel Dokumen <i>Compare</i>	30
3.4.2 Tabel Uji <i>Session</i>	30
3.5 Perancangan Pengujian Deteksi Kemiripan Pada Dokumen teks	31
3.5.1 Tujuan Pengujian	31
3.5.2 Bahan dan Alat Pengujian	31
3.5.2.1 Perangkat Keras	32
3.5.2.2 Perangkat Lunak	32
3.5.3 Mekanisme Pengujian	32
BAB 4 IMPLEMENTASI DAN ANALISIS SISTEM	33
4.1 Implementasi Sistem	33
4.1.1 Implementasi Rancangan Tampilan	33

4.1.1.1 Halaman Beranda	33
4.1.1.2 Halaman <i>Stopword</i>	35
4.1.1.3 Halaman Riwayat Pengujian	35
4.1.1.4 Halaman Hasil Pengujian	36
4.1.1.5 Halaman Detail Dokumen Uji	38
4.1.1.6 Halaman Detail Pengujian Dokumen	39
4.1.2 Implementasi Rancangan Proses	28
4.2 Analisis Sistem	40
4.2.1 Implementasi Algoritma <i>Winnowing</i> Untuk Mendeteksi Kemiripan Pada Dokumen Teks	40
4.3 Evaluasi Aplikasi Sistem Pendeteksi Kemiripan Pada Dokumen Teks	55
BAB 5 KESIMPULAN DAN SARAN	57
5.1 Kesimpulan	57
5.2 Saran	57
DAFTAR PUSTAKA	58



DAFTAR TABEL

TABEL	KETERANGAN	HALAMAN
Tabel 3.1	Tabel <i>Stopword</i>	29
Tabel 3.2	Tabel Dokumen_uji	29
Tabel 3.3	Tabel Dokumen_compare	30
Tabel 3.4	Tabel Uji_session	30
Tabel 4.1	Pengolahan Dokumen Uji	54

© UKDW

DAFTAR GAMBAR

GAMBAR	KETERANGAN	HALAMAN
Gambar 2.1	Pembersihan karakter-karakter yang tidak relevan	15
Gambar 2.2	Pembentukan rangkaian gram	15
Gambar 3.1	<i>Flowchart</i> program secara umum	17
Gambar 3.2	<i>Flowchart</i> proses <i>preprocessing</i>	18
Gambar 3.3	<i>Flowchart</i> Proses kerja pendeteksian kemiripan dokumen	20
Gambar 3.4	<i>Flowchart</i> proses pembentukan <i>k-grams</i>	21
Gambar 3.5	<i>Flowchart Hashing K-grams</i>	22
Gambar 3.6	<i>Flowchart</i> proses pembuatan <i>window</i> dan pemilihan <i>fingerprints</i>	23
Gambar 3.7	<i>Flowchart</i> perhitungan kemiripan antar dokumen	24
Gambar 3.8	Antarmuka <i>input</i> sistem	25
Gambar 3.9	Antarmuka <i>output</i> sistem	26
Gambar 3.10	Antarmuka <i>Output</i> Detail Dokumen	27
Gambar 3.11	Antarmuka Detai Perbandingan Dokumen.	28
Gambar 4.1	Halaman Beranda	34
Gambar 4.2	Halaman Stopword	35
Gambar 4.3	Halaman Riwayat Pengujian Sistem	36
Gambar 4.4	Halaman Hasil Pengujian	37
Gambar 4.5	Halaman Detai Dokumen Uji	38
Gambar 4.6	Halaman Detail Pengujian Dokumen	39
Gambar 4.7	Grafik perbandingan persentase kemiripan dengan pengujian pada nilai <i>k-gram</i> tanpa <i>filtering stopwords</i>	45
Gambar 4.8	Grafik perbandingan waktu proses dengan pengujian pada nilai <i>k-gram</i> tanpa <i>filtering stopwords</i>	45

DAFTAR GAMBAR

GAMBAR	KETERANGAN	HALAMAN
Gambar 4.9	Grafik perbandingan persentase kemiripan dengan pengujian pada nilai basis tanpa <i>filtering stopword</i>	46
Gambar 4.10	Grafik perbandingan waktu proses dengan pengujian pada nilai basis tanpa <i>filtering stopword</i>	46
Gambar 4.11	Grafik perbandingan persentase kemiripan dengan pengujian pada nilai <i>window</i> tanpa <i>filtering stopword</i>	47
Gambar 4.12	Grafik perbandingan waktu proses dengan pengujian pada nilai <i>window</i> tanpa <i>filtering stopword</i>	47
Gambar 4.13	Grafik perbandingan persentase kemiripan dengan pengujian pada nilai <i>k-gram</i> dengan <i>filtering</i>	48
Gambar 4.14	Grafik perbandingan waktu proses dengan pengujian pada nilai <i>k-gram</i> dengan <i>filtering stopword</i>	48
Gambar 4.15	Grafik perbandingan persentase kemiripan dengan pengujian pada nilai basis dengan <i>filtering stopword</i>	49
Gambar 4.16	Grafik perbandingan waktu proses dengan pengujian pada nilai basis dengan <i>filtering stopword</i>	49
Gambar 4.17	Grafik perbandingan persentase kemiripan dengan pengujian pada nilai <i>window</i> dengan <i>filtering</i>	50
Gambar 4.18	Grafik perbandingan waktu proses dengan pengujian pada nilai <i>window</i> dengan <i>filtering stopword</i>	50
Gambar 4.19	Gambar <i>visual</i> kemiripan pada dokumen sebesar 12%.	51
Gambar 4.20	Grafik Perbandingan Persentase Kemiripan dengan K-Gram = 8, Basis = 7 dan Window 7	52
Gambar 4.21	Grafik Perbandingan Waktu Proses dengan K-Gram = 8, Basis = 7 dan Window 7	53
Gambar 4.22	Grafik Perbandingan Persentase Kemiripan Dokumen Plagiarisme	53
Gambar 4.23	Grafik Perbandingan Waktu Proses Pengujian Dokumen Plagiarisme	54

DAFTAR GAMBAR

GAMBAR	KETERANGAN	HALAMAN
Gambar 4.24	visualisasi kemiripan kedua tesis dengan parameter k-gram=9, basis=19, window=9 dan menggunakan <i>filtering stopwords</i> .	55
Gambar 4.25	Hasil <i>Rolling Hash</i> nilai INF	56

© UKDW

DAFTAR LAMPIRAN

KETERANGAN	HALAMAN
Daftar Stopword	A
Kode Program	B

© UKDW

INTISARI

IMPLEMENTASI ALGORITMA WINNOWING UNTUK MENDETEKSI KEMIRIPAN PADA DOKUMEN TEKS

Semakin mudahnya untuk saling bertukar dan mendapatkan informasi tidak hanya memberikan dampak positif bagi kemajuan teknologi, tetapi juga membawa dampak negatif yang hampir tidak dapat dihindari seperti tindakan *copy paste* yang dapat berujung pada plagiarisme yang sengaja maupun tidak sengaja dilakukan. Tindakan tersebut dapat mematikan kreatifitas seseorang karena sudah terbiasa mengambil sesuatu yang bukan hasil karyanya dengan mudah dan merupakan pelanggaran hak cipta.

Beberapa metode yang dapat digunakan untuk mendeteksi atau menganalisa plagiarisme seperti pencocokan *substring*, kesamaan kata kunci dan dokumen *fingerprintings*. Algoritma *Winnowing* merupakan algoritma dokumen *fingerprint* yang dapat digunakan untuk mendeteksi plagiarisme. *Input* dari algoritma *winnowing* adalah *string* dari sebuah dokumen dan *output* dari dari algoritma tersebut adalah nilai-nilai *hash* yang disebut sebagai *fingerprints* dari dokumen tersebut.

Hasil pada penelitian ini berupa sebuah sistem untuk mendeteksi kemiripan pada dokumen teks dan melihat pengaruh dari nilai-nilai parameter yang digunakan terhadap prosentase hasil perhitungan kemiripan dokumen uji.

Kata Kunci : [Algoritma *Winnowing*, *Plagiarisme*, *Kemiripan*, *Dokumen Fingerprinting*, *Stopword*]

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Semakin mudahnya pertukaran informasi dewasa ini tidak hanya membawa dampak positif bagi kemajuan teknologi khususnya dalam hal penyebaran informasi dan akses terhadap informasi-informasi tersebut, tetapi juga membawa dampak negatif yang hampir tidak dapat dihindari seperti tindakan plagiarisme yang sengaja maupun tidak sengaja dilakukan dimana dapat mematikan kreatifitas seseorang karena sudah terbiasa mengambil sesuatu yang bukan miliknya dengan mudah.

Saat ini terdapat beberapa algoritma *dokumen fingerprinting* yang digunakan dan dikembangkan menjadi sebuah aplikasi untuk mendeteksi plagiarisme. Algoritma *dokumen fingerprinting* ada bermacam-macam, antara lain *Running Karp Robin Matching and Greddy String Tiling*, Pendekatan *Manber*, Pendekatan *Heintze* dan Algoritma *Winnowing* (Stein, 2006).

Berdasarkan masalah diatas, pada tugas akhir ini akan melakukan perancangan aplikasi dengan membandingkan kemiripan dokumen asli dengan dokumen yang ingin diuji. Dengan mengetahui persentase kemiripan kedua dokumen tersebut dapat dijadikan bahan pertimbangan apakah dokumen yang diuji tersebut merupakan hasil menjiplak karya seseorang atau tidak.

1.2. Perumusan Masalah

Bagaimana membuat sebuah aplikasi berbasis web yang dapat mengukur tingkat kemiripan dokumen teks dan bagaimana pengaruh nilai *gram*, basis dan *window* terhadap persentase kemiripan dokumen teks berdasarkan Algoritma *Winnowing*.

1.3. Batasan Masalah

1. Algoritma yang digunakan untuk pengujian dokumen teks menggunakan Algoritma *Winnowing*.
2. Data yang digunakan untuk diuji hanya berupa teks.
3. Data yang diuji bertipe Teks dengan ekstensi *.txt*.
4. Nilai *k-gram* dan *window* yang digunakan dibatasi antara 2 sampai 10.
5. Nilai basis *hash* yang digunakan dibatasi antara bilangan prima 2 sampai 30
6. Nilai *k-gram*, Basis *hash* dan *window* saat pengujian ditentukan oleh user.
7. Pada proses *Rolling Hash*, menggunakan nilai ASCII dari karakter.
8. Algoritma *Winnowing* tidak dapat mengenali makna dari urutan kata atau kalimat dalam dokumen uji.
9. Maksimal karakter dalam dokumen uji sebanyak 10.000 karakter.

1.4. Tujuan Penelitian

Tujuan dari penelitian ini untuk mengimplementasikan algoritma *winnowing* dalam mendeteksi kemiripan pada dokumen teks dan mendapatkan nilai parameter yang baik untuk digunakan pada algoritma tersebut.

1.5. Metode / Pendekatan

1. Studi pustaka/literatur mengenai *dokumen fingerprinting* dengan algoritma *winning* serta literatur lain yang berhubungan dengan Tugas Akhir ini.
2. Melakukan analisis teori algoritma *winning* dan bagaimana menerapkannya pada sistem yang akan dibuat.
3. Mendesain antarmuka sistem dan menguji sistem untuk memastikan bahwa hasil implementasi berjalan sesuai dengan algoritma.
4. Menganalisis hasil dan menarik kesimpulan dari Tugas Akhir yang sudah dibuat.

1.6. Sistematika Penulisan

Laporan tugas akhir ini disusun menjadi 5 bab, yaitu : Pendahuluan, Landasan Teori, Analisa dan Perancangan Sistem, Implementasi dan Pengujian, dan Kesimpulan.

Bab 1 adalah Pendahuluan, yang berisi latar belakang masalah dilakukannya penelitian, perumusan masalah yang mencakup implementasi metode yang digunakan, batasan masalah, tujuan penelitian yang dilakukan, metode yang digunakan dalam penelitian yaitu suatu kumpulan prinsip atau teknik yang digunakan untuk memecahkan suatu masalah secara ilmiah, serta sistematika penulisan.

Bab 2 berisi tinjauan pustaka dan landasan teori. Tinjauan pustaka menguraikan berbagai teori yang didapatkan dari berbagai sumber pustaka. Landasan teori memuat penjelasan tentang konsep dan prinsip utama yang diperlukan untuk memecahkan masalah.

Bab 3 merupakan perancangan sistem yaitu analisis teori-teori yang digunakan, yaitu perancangan sistem secara keseluruhan termasuk input, output dan algoritma.

Bab 4 menjelaskan implementasi dan analisis sistem. Bab ini berisi pembahasan implementasi dan analisis dari sistem yang dibangun.

Bab 5 memuat kesimpulan dan saran. Bab ini berisi kesimpulan dari hasil analisis kegiatan implementasi yang dilakukan dan saran-saran ke depan untuk kegiatan penelitian serupa agar kinerja sistem lebih baik dari sistem yang dibangun.

© UKDW

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Algoritma *Winnowing* dapat mengukur kemiripan antar dokumen. Berdasarkan penelitian yang dilakukan pada Bab sebelumnya, prosentase kemiripan dokumen lebih dipengaruhi oleh besaran nilai *k-gram* dan nilai basis yang digunakan. Penggunaan nilai *k-gram* lebih dari 2 akan memberikan prosentase kemiripan yang sesuai jika dilihat secara *visual* dan penggunaan nilai basis yang lebih besar akan memberikan pengaruh terhadap nilai hashing yang dihasilkan sehingga akan berdampak pada nilai kemiripan yang yang didapatkan oleh sistem.

Perubahan nilai *window* dan penggunaan *filteringstopword* cenderung hanya mempengaruhi waktu proses yang digunakan dalam pengujian dokumen. Penggunaan *filteringstopword* juga berfungsi untuk mengurangi waktu proses yang lama.

5.2 Saran

Saran pada pengembangan aplikasi yang selanjutnya agar aplikasi dapat digunakan untuk membaca dokumen-dokumen teks tidak terbatas hanya pada dokumen berekstensi *.txt* tetapi dapat membaca format dokumen teks yang pada umumnya digunakan seperti *pdf* dan *doc* atau *docx*. Pengembangan aplikasi juga diharapkan dapat mendeteksi kemiripan pada file *code* program, sehingga nantinya aplikasi ini tidak hanya dapat mendeteksi plagiarisme pada dokumen teks tetapi juga plagiarisme pada file-file *code* program.

DAFTAR PUSTAKA

- Ceska, Zdenek, & Fox, Chris (2011). The Influence of Text Pre-processing on Plagiarism Detection. Dalam Angelova, Galia and Bontcheva, Kalina and Mitkov, Ruslan and Nicolov, Nicolas and Nikolov, Nikolai, (Eds.) *International Conference on Recent Advances in Natural Language Processing 2009*. Association for Computational Linguistics, 55-59.
- Cornic, Piere. (2008). *Software Plagiarism Detection Using Model-Driven Software Development in Eclipse Platform*. Computer Science, Faculty Engineering and Physical Sciences, University of Manchester.
- Elbegbayan, Norzima. 2005. *Winnowing, a Document Fingerprinting Algorithm*. Linkoping University.
- Kamus Besar Bahasa Indonesia . 2005. Edisi ketiga. Departemen Pendidikan Nasional. Balai Pustaka.
- Nugroho, Eko. 2011. *Perancangan Sistem Deteksi Plagiarisme Dokumen Teks Dengan Menggunakan Algoritma Rabin-Karp*. Program Studi Ilmu Komputer, Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Brawijaya.
- Peraturan Menteri Pendidikan Nasional Republik Indonesia Nomor 17 Tahun 2010. *Pencegahan dan Penanggulangan Plagiat di Perguruan Tinggi*.
- Salim, Peter, Yenny Salim. 1991. *Kamus Bahasa Indonesia Kontemporer*. Modern English Press.
- Sastroasmoro, Sudigdo. (2007). *Beberapa Catatan Tentang Plagiarisme*. Majalah Kedokteran Indonesia, Vol 57 (8).

Scheleimer, Saul, & Daniel S. Wilkerson, & Alex Aiken. *Winnowing : Local Algorithms for Document Fingerprinting*. Chicago.

Soelistyo, Budi, Henry (2011). *Plagiarisme : Pelanggaran Hak Cipta dan Etika*. Penerbit Kanisius.

Stein, Benno, & Eissen, Sven M.Z. (2006). Near Similarity Search and Plagiarism Analysis. Dalam Spiliopoulou et al (Eds). *Data and Information Analysis to Knowledge Engineering Selected Papers from the 295th Annual Conference of German Classification Society (GFKI)*. Magdeburg : Springer

© UKDW