PERANCANGAN SISTEM DETEKSI PLAGIARISME KARYA ILMIAH BERBAHASA INDONESIA MENGGUNAKAN ALGORITMA *WINNOWING FINGERPRINT*

Feri Okto Sinaga
(05.04.1.1.1.00652)

Pembimbing I : Hermawan, S.T.,M.Kom
Pembimbing II : Fika Hastarita R, S.T, M.Eng

ABSTRAK

Pada era teknologi informasi ini banyak hal yang seharusnya baik mejadi disalah gunakan, salah satunya adalah plagiarisme. Plagiarisme merupakan masalah penting, terutama dalam lingkungan akademisi, dimana terjadi penjiplakan karya tulis ilmiah yang makin marak sekarang ini.

Salah satu cara untuk mengurangi plagiarisme yaitu dengn membangun sebuah sistem deteksi plagiarisme. Adapun algoritma yang digunanak dalam proses deteksi plagiarisme yaitu dengan menggunakan algoritma winnowing. Dokumen yang akan dibandingkan diproses dan mendpatkan *fingerprint* dari dokumen tadi, dimana *fingerprint* tersebut berupa ekumpulan nilai hash. *Fingerprint* yang sama antara dokumen yang dibandingkan kemiripannya merupakan kunci dari penghitungan nilai *similarity*-nya.

Hasil penghitungan kemiripan antar 2 dokumen dengan 2 *gram* menghasilkan *similarity* mencapai 75,56%, sedangkan dengan percobaan dengan 10 gram, beberapa dokumen tidak mendapatkan kemiripan, yaitu dengan nilai 0%. Dari hasil uji coba sistem deteksi plagiarisme menggunakan algoritma *winnowing fingerprint*, nilai *similarity*-nya bergantung pada besarnya nilai *gram*, basis dan *window* dari sistem tersebut. Semakin kecil nilai *gram*, maka hasil dari *similarity* semakin besar.

***Kata Kunci** : similarity, plagiarisme, winnowing, document fingerprint*

DAFTAR ISI

Hal

HALAMAN JUDUL
LEMBAR PENGESAHAN
LEMBAR PERSEMBAHAN
KATA PENGANTAR iii
ABSTRAK v
DAFTAR ISI vi
Daftar Gambar viii
Daftar Tabel ix
Daftar Lampiran x

BAB I PENDAHULUAN 1
1.1 Latar Belakang 1
1.2 Perumusan Masalah 1
1.3 Tujuan penelitian 2
1.4 Manfaat penelitian 2
1.5 Batasan Masalah 2
1.6 Metode Penelitian tugas akhir 2
1.7 Sistematika Penulisan 3

BAB II STUDI PUSTAKA 5
2.1 Peneletiian Sebelumya 5
2.2 Plagiarisme 7
2.3 *Text Mining* 8
2.3.1 Ruang Lingkup *Text Mining* 10
2.3.2 Ekstraksi dokumen 12
2.4 Algoritma Winnowing 15
2.4.1 *Whitespace Insensitivity* 16
2.4.2 Pembentukan rangkaian *gram* 17
2.4.3 *Hashing* 18
2.4.4 pembentukan window dari nilai *hash* 22
2.4.5 *string matching* 26
2.5 ASCII 27

BAB III ANALISA DAN RANCANGAN SISTEM..... 29
3. 1 Analisa sistem……..……. 29
3.1.1 Kebutuhan Perangkat Keras .....……... 29
3.1.2 Kebutuhan Perangkat Lunak …..... 30
3. 2 Analisa masalah..... 30
3. 3 Analisa data 30
3. 4 Analisa input 31
3. 5 Analisa output …… 31
3. 6 Perancangan Sistem ……...……. 31
3.6.1 Gambaran Umum ……..…..…… 31
3.6.2 *Flowchart Sistem* 32
3.6.3 *Usecase Diagram*..... 32
3.6.4 *Activity Diagram* 33

BAB IV IMPLEMENTASI……..... 37
4.1 Lingkungan implementasi……..…….. 37
4.2 Skenario pengujian fungsional…….. 37
4.2.1 main page ……..…… 37
4.2.2 browse data/ dokumen……..... 38
4.2.3 input nilai *gram*..... 38
4.2.4 penghitungan *similarity* 2 dokumen 39
4.3 Skenario uji coba aplikasi……..... 39
4.3.1 uji coba 10 dokumen dengan gram yang sama.. 39
4.3.2 uji coba 2 file dengan gram yang berbeda…… 40
4.4 Evaluasi 45
4.5.1 Evaluasi uji coba 45
BAB V PENUTUP 47
5.1 Kesimpulan …….. 47
5.2 Saran……. 47
DAFTAR PUSTAKA ……. 48

DAFTAR GAMBAR

Gambar 2.1 tahap *preprocessing*..... 13
Gambar 2.2 *tokenizing* .....……. 14
Gambar 2.3 *filtering* …….....……..… 14
Gambar 2.4 *stemming*…….. 15
Gambar 3.1 desain rancangan sistem .....… 31
Gambar 3.2 *flowchart sistem deteksi plagiarisme dengan metode winnowing*.....…. 32
Gambar 3.3 *use case diagram* .....…… 33
Gambar 3.4 *Activity diagram pengujian similarity*.....…. 34
Gambar 4.1 tampilan utama dari program.....…… 37
Gambar 4.2 browse data.....…… 38
Gambar 4.3 input nilai *gram*..... 38
Gambar 4.4 hasil uji 2 dokumen .....……. 39

DAFTAR TABEL

Tabel	2.1 Tabel pembuangan karakter yang tidak relevan ……	16
Tabel	2.2 Tabel pembentukan rangkaian 5-gram……...	18
Tabel	2.3 Tabel pembentukan nilai hash.....…….	20
Tabel	2.4 Tabel pembentukan nilai window dari nilai hash ……..	22
Tabel	2.5 Tabel fingerprint dari susunan window …….....……	25
Tabel	4.1 Tabel uji coba file dengan 5-gram……..	39
Tabel	4.2 Tabel uji coba dengan 2-gram ……	40
Tabel	4.3 Tabel uji coba dengan 3-gram ……	41
Tabel	4.4 Tabel uji coba dengan 4-gram ……	41
Tabel	4.5 Tabel uji coba dengan 5-gram ……	42
Tabel	4.6 Tabel uji coba dengan 6-gram ……	42
Tabel	4.7 Tabel uji coba dengan 7-gram ……	43
Tabel	4.8 Tabel uji coba dengan 8-gram ……	43
Tabel	4.9 Tabel uji coba dengan 9-gram ……	44
Tabel	4.10 Tabel uji coba dengan 10-gram ……	44

Universitas Trunojoyo

DAFTAR LAMPIRAN

Lampiran 1 Data dokumen .....…… 51
Lampiran 2 hasil proses *similarity* …….. 53