

ABSTRAK

Teknologi mesin pencari yang semakin berkembang sangat membantu pengguna internet untuk mendapatkan informasi secara *online*. Dengan menggunakan situs mesin pencarian seperti Google, Bing, dan yahoo maka informasi yang dibutuhkan hampir semuanya tersedia, mulai dari berita, ilmu pengetahuan, atau bahkan *paper* dari penelitian orang lain. Komputer dan internet merupakan penyebab utama meningkatnya aksi plagiarisme. Plagiarisme adalah tindakan yang disengaja atau tidak disengaja untuk mendapatkan nilai untuk studi ilmiah. Mendeteksi plagiarisme sangatlah penting dalam dunia pendidikan karena kredibilitas lembaga ini sebagian besar didasarkan pada orisinalitas. Metode untuk mendeteksi plagiarisme dapat diklasifikasikan menjadi tiga metode yaitu metode perbandingan teks lengkap, metode dokumen *fingerprinting*, dan metode kesamaan kata kunci. Salah satu metode *fingerprint* yaitu *Winnowing* sedangkan untuk metode kesamaan kata kunci yaitu *Cosine similarity*. Oleh karena itu muncul pertanyaan metode manakah yang lebih baik dari kedua pendekatan tersebut jika menggunakan *stemming Enhanced Confix Stripping* untuk mendeteksi kemiripan dokumen. Berdasarkan pengujian yang dilakukan mengenai perbandingan algoritma *Winnowing* dan *Cosine Similarity* yang sama-sama dikombinasikan dengan *Enhanced Confix Stripping* maka didapatkan nilai rata-rata hasil untuk sistem yang dibuat dengan algoritma *Winnowing* sebesar **13.13%** dan *Plagiarism Checker X* sebesar **10.59%**. Dengan demikian hasil yang didapatkan dari sistem yang dibuat dan *Plagiarism Checker X* tidaklah jauh berbeda, yaitu dengan selisih sebesar **2.54%** dan waktu yang didapat **1.27** detik dalam skala 100%.

Kata Kunci : Plagiasi, Algoritma *Winnowing*, *Stemming*, *Enhanced Confix Stripping*, Algoritma *Cosine Similarity*.

DAFTAR ISI

HALAMAN JUDUL	ii
HALAMAN PENGESAHAN.....	iii
HALAMAN PERNYATAAN ORISINALITAS.....	iv
ABSTRAK.....	v
KATA PENGANTAR	vi
DAFTAR ISI.....	viii
DAFTAR GAMBAR.....	xi
DAFTAR TABEL.....	xii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Perumusan Masalah	3
1.2.1 Permasalahan.....	3
1.2.2 Metode Usulan	4
1.2.3 Pertanyaan Penelitian	4
1.3 Tujuan dan Manfaat	4
1.3.1 Tujuan Penelitian.....	4
1.3.2 Manfaat.....	4
1.4 Batasan Masalah	4
1.5 Sistematika Penulisan	5
BAB II KAJIAN PUSTAKA.....	6
2.1 Plagiasi.....	6
2.2 Algoritma <i>Winnowing</i>	8
2.3 Pembobotan TF-IDF (<i>Term Frequency Inverse Document Frequency</i>).....	11
2.4 <i>Cosine Similarity</i>	18
2.5 <i>Enhanced Confix Stripping</i>	19
2.6 Pengujian.....	21
2.7 Penelitian Terkait.....	22
BAB III METODE PENELITIAN	25
3.1 Dataset.....	25
3.2 Arsitektur Sistem	26

3.3	Algoritma <i>Enhanced Confix Stripping Stemmer</i>	28
3.4	Algoritma <i>Winnowing</i>	30
3.4.1	<i>K-Gram</i>	31
3.4.2	<i>Rolling Hash</i>	32
3.4.3	<i>Window</i>	33
3.4.4	<i>Jaccard Coefficient</i>	35
3.5	Pembobotan TF-IDF	36
3.5.1	Algoritma <i>Cosine Similarity</i>	36
3.6	Skenario Pengujian	38
BAB IV HASIL DAN PEMBAHASAN		39
4.1	Tahap Pembangunan Sistem	39
4.2	Menyiapkan <i>Tools</i> yang digunakan	39
4.3	Skenario Pengujian	39
4.4	Menyiapkan Dataset	41
4.5	Pembuatan Sistem	41
4.6	Pengujian dan Evaluasi	41
4.6.1	Pengujian 1	41
4.6.1.1	Hasil pengujian 1	41
4.6.2	Pengujian 2	42
4.6.2.1	Hasil pengujian 2	43
4.6.3	Pengujian 3	44
4.6.3.1	Hasil pengujian 3	45
4.6.4	Pengujian 4	46
4.6.4.1	Hasil pengujian 4	47
4.6.5	Evaluasi	48
BAB V PENUTUP		49
5.1	Kesimpulan	49
5.2	Saran	49
REFERENSI		50
LAMPIRAN		52
1.	Judul Dataset	52
2.	Lampiran hasil pengujian 1	54
3.	Lampiran hasil pengujian 2	61
4.	Lampiran hasil pengujian 3	68

5. Lampiran hasil pengujian 4.....	74
BIOGRAFI PENULIS	81

DAFTAR GAMBAR

Gambar 2. 1 Klasifikasi Metode Pendeteksi Plagiarisme [5]	7
Gambar 2. 2 Algoritma Winnowing dan Rabin-Karp [3]	8
Gambar 2. 3 Pre-processing	9
Gambar 2. 4 Pembentukan Rangkaian Gram	10
Gambar 3. 1 Arsitektur Sistem	26
Gambar 3. 2 Flowchart Algoritma Enhanced Confix Stripping Stemmer	28
Gambar 3. 3 Flowchart Algoritma Winnowing	30
Gambar 3. 4 Alur Proses dari Pembentukan K-Gram	31
Gambar 3. 5 Flowchart Rolling Hash	32
Gambar 3. 6 Flowchart Window	33
Gambar 3. 7 Flowchart Jaccard Coefficient	35
Gambar 3. 8 Pembobotan TF-IDF	36
Gambar 3. 9 Algoritma Cosine Similarity	37

DAFTAR TABEL

Tabel 2. 1 Hasil <i>Pre-Processing</i>	12
Tabel 2. 2 Perhitungan <i>TF dan DF</i>	13
Tabel 2. 3 Perhitungan <i>IDF</i>	14
Tabel 2. 4 Perhitungan <i>TF-IDF</i>	15
Tabel 2. 5 Perhitungan Panjang Dokumen.....	16
Tabel 2. 6 Hasil Perhitungan <i>Cosine Similarity</i>	19
Tabel 2. 7 Aturan Pemenggalan <i>Enhanced Confix Stripping Stemmer</i> [20].....	20
Tabel 2. 8 Penelitian Terkait Mengenai Deteksi Plagiasi	23
Tabel 4. 1 Skenario Pengujian	40
Tabel 4. 2 Contoh hasil pertama perbandingan Sistem dengan <i>Plagiarism Checker X</i>	41
Tabel 4. 3 Contoh hasil kedua perbandingan Sistem dengan <i>Plagiarism Checker X</i>	43
Tabel 4. 4 Contoh hasil ketiga perbandingan Sistem dengan <i>Plagiarism Checker X</i>	45
Tabel 4. 5 Contoh hasil keempat perbandingan Sistem dengan <i>Plagiarism Checker X</i> ...	47