

**SISTEM PENDETEKSI PLAGIARISME DOKUMEN TEKS
LINTAS BAHASA MENGGUNAKAN METODE WINNOWER**

SKRIPSI

Digunakan Sebagai Syarat Maju Ujian Diploma IV
Politeknik Negeri Malang

Oleh:

MARIA PUJI RAHAYU NIM. 1741720041



**PROGRAM STUDI TEKNIK INFORMATIKA
JURUSAN TEKNOLOGI INFORMASI
POLITEKNIK NEGERI MALANG
JULI 2021**

**SISTEM PENDETEKSI PLAGIARISME DOKUMEN TEKS
LINTAS BAHASA MENGGUNAKAN METODE WINNOWER**

SKRIPSI

Digunakan Sebagai Syarat Maju Ujian Diploma IV
Politeknik Negeri Malang

Oleh:

MARIA PUJI RAHAYU NIM. 1741720041



**PROGRAM STUDI TEKNIK INFORMATIKA
JURUSAN TEKNOLOGI INFORMASI
POLITEKNIK NEGERI MALANG
JULI 2021**



HALAMAN PENGESAHAN


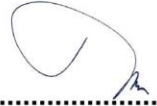


SISTEM PENDETEKSI PLAGIARISME DOKUMEN TEKS LINTAS BAHASA MENGGUNAKAN METODE WINNOWER

Disusun oleh:

MARIA PUJI RAHAYU **NIM. 1741720041**

Skripsi ini telah diuji pada tanggal 10 Juni 2021

Disetujui oleh:

- | | | | |
|--------------------------|---|--|---|
| 1. Pembimbing Utama | : | <u>Mustika Mentari, S.Kom., M.Kom</u>
NIP. 198806072019032016 |  |
| 2. Pembimbing Pendamping | : | <u>Imam Fahrur Rozi, ST., MT.</u>
NIP. 198406102008121004 |  |
| 3. Penguji Utama | : | <u>Ekojono, ST., M.Kom.</u>
NIP. 195912081985031004 |  |
| 4. Penguji Pendamping | : | <u>Annisa Puspa Kirana, S.Kom., M.Kom</u>
NIP. 198901232019032016 |  |

Mengetahui,


Ketua Jurusan
Teknologi Informasi
Rudy Ardyanto, S.T., M.CS.
NIP. 197111101999031002

Ketua Program Studi
Teknik Informatika



Imam Fahrur Rozi, ST., MT.
NIP. 198406102008121004

PERNYATAAN

Dengan ini saya menyatakan bahwa pada Skripsi ini tidak terdapat karya, baik seluruh maupun sebagian, yang sudah pernah diajukan untuk memperoleh gelar akademik di Perguruan Tinggi manapun, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis disitasi dalam naskah ini serta disebutkan dalam daftar sitasi/pustaka.

Malang, 26 April 2021



MP
Maria Puji Rahayu

ABSTRAK

Puji Rahayu, Maria. “Sistem Pendeteksi Plagiarisme Dokumen Teks Lintas Bahasa Menggunakan Metode *Winnowing*”. **Pembimbing: (1) Mustika Mentari, S.Kom., M.Kom, (2) Imam Fahrur Rozi, S.T., M.T.**

Skripsi, Program Studi Teknik Informatika, Jurusan Teknologi Informasi, Politeknik Negeri Malang, 2021.

Saat ini banyak sekali dokumen teks seperti jurnal yang tersebar di internet baik jurnal berbahasa Indonesia maupun berbahasa Inggris. Dengan hal tersebut memungkinkan tindakan plagiarisme dengan menyalin dari jurnal asing yang diterjemahkan ke dalam bahasa lain maupun menyalin langsung tanpa diubah dari bahasa aslinya. Salah satu cara yang dapat menekan tindakan tersebut adalah dengan membangun sistem pendeteksi plagiarisme dokumen teks lintas bahasa.

Metode yang dapat digunakan untuk mendeteksi plagiarisme dokumen adalah metode *Winnowing*. Metode *Winnowing* merupakan metode di mana inputan teks akan diproses sehingga menghasilkan nilai *hash* yang disebut dengan *fingerprint*. Penelitian ini bertujuan untuk membangun sistem yang dapat mendeteksi plagiarisme dokumen teks yang berbeda bahasa menggunakan metode *Winnowing*. Dokumen teks yang dapat diuji berupa inputan teks dan *file* PDF. Dokumen yang digunakan dalam pengujian sistem berupa jurnal yang memiliki topik yang sama. Hasil tingkat akurasi tertinggi yang dihasilkan antara perhitungan *Jaccard Coefficient* dengan aplikasi Plagiarism Checker X yaitu pada skenario keempat dengan nilai persentase 84,7%.

Kata Kunci : Plagiarisme, Metode *Winnowing*, *Jaccard Coefficient*

ABSTRACT

Puji Rahayu, Maria. "Cross Language Text Document Plagiarism Detection System Using Winnowing Method". Supervisors (1) Mustika Mentari, S.Kom., M.Kom, (2) Imam Fahrur Rozi, S.T., M.T.

Thesis, Informatics Engineering Study Program, Information Technology Department, State Polytechnic of Malang, 2021.

Currently, there are a lot of text documents such as journals available on the internet, both in Indonesian and English. This allows plagiarism by copying from foreign journals which are translated into other languages or direct copying without being changed from the original language. One way that can suppress this action is by building a plagiarism detection system for cross-language text documents.

The method that can be used to detect document plagiarism is the Winnowing method. The Winnowing method is a method where text input will be processed to produce a hash value which is called a fingerprint. This research aimed to build a system that could detect plagiarism of text documents of different languages using the Winnowing method. Text documents that could be tested from text input and PDF files. Documents used in system testing were journals having the same topic. The results of the highest accuracy rate between the calculation of the Jaccard Coefficient and Plagiarism Checker X application were in the fourth scenario with a percentage value of 84,7%.

Keywords: Plagiarism, Winnowing Method, Jaccard Coefficient

KATA PENGANTAR

Puji Syukur kami panjatkan kehadiran Tuhan YME atas segala rahmat dan hidayah-Nya penulis dapat menyelesaikan skripsi dengan judul “SISTEM PENDETEKSI PLAGIARISME DOKUMEN TEKS LINTAS BAHASA MENGGUNAKAN METODE WINNOWER”. Skripsi ini penulis susun sebagai persyaratan untuk menyelesaikan studi program Diploma IV Program Studi Teknik Informatika, Jurusan Teknologi Informasi, Politeknik Negeri Malang.

Kami menyadari bahwasannya dengan tanpa adanya dukungan dan kerja sama dari berbagai pihak, kegiatan laporan akhir ini tidak akan dapat berjalan baik. Untuk itu, kami ingin menyampaikan rasa terima kasih kepada:

1. Bapak Rudy Ariyanto, ST., M.Cs., selaku Ketua Jurusan Teknologi Informasi
2. Bapak Imam Fahrur Rozi, ST., MT., selaku Ketua Program Studi DIV Teknik Informatika sekaligus Pembimbing Pendamping
3. Ibu Mustika Mentari, S.Kom., M.Kom, selaku Pembimbing Utama
4. Kedua orang tua dan saudara yang selalu memberikan semangat
5. Puput, ireng, abang, *day6* dan *stray kids* yang selalu membantu menemani mengerjakan skripsi ini
6. Dan seluruh pihak yang telah membantu dan mendukung lancarnya pembuatan Laporan Akhir dari awal hingga akhir yang tidak dapat kami sebutkan satu persatu.

Penulis menyadari bahwa dalam penyusunan laporan akhir ini, masih banyak terdapat kekurangan dan kelemahan yang dimiliki penulis baik itu sistematika penulisan maupun penggunaan bahasa. Untuk itu penulis mengharapkan saran dan kritik dari berbagai pihak yang bersifat membangun demi penyempurnaan laporan ini. Semoga laporan ini berguna bagi pembaca secara umum dan penulis secara khusus. Akhir kata, penulis ucapkan banyak terima kasih.

Malang, 26 April 2021

Penulis

DAFTAR ISI

	Halaman
SAMPUL DEPAN	i
HALAMAN JUDUL.....	ii
HALAMAN PENGESAHAN.....	iii
PERNYATAAN.....	iv
ABSTRAK	v
<i>ABSTRACT</i>	vi
KATA PENGANTAR	vi
DAFTAR ISI.....	vii
DAFTAR GAMBAR	ix
DAFTAR TABEL.....	xi
DAFTAR LAMPIRAN.....	xii
BAB I. PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	2
1.3 Tujuan.....	3
1.4 Batasan Masalah	3
1.5 Sistematika Penulisan	3
BAB II. LANDASAN TEORI	5
2.1 Tinjauan Pustaka.....	5
2.2 Dasar Teori	7
2.2.1 Plagiarisme	7
2.2.2 <i>Text Mining</i> dan <i>Text Preprocessing</i>	7
2.2.3 Metode Winnowing.....	9
2.2.4 <i>Jaccard Coefficient</i>	10
2.2.5 Dasar Hukum Plagiarisme.....	11
BAB III. METODOLOGI PENELITIAN	13
3.1 Waktu dan Tempat Penelitian.....	13
3.2 Metode Pengumpulan Data	13
3.3 Metode Pengolahan Data.....	14
3.3.1 Input Dokumen	14
3.3.2 Menerjemahkan Dokumen Teks	14
3.3.3 Pembatasan Uji Dokumen.....	14
3.3.4 <i>Preprocessing</i>	14
3.3.5 Membentuk Nilai <i>N-Gram</i>	15
3.3.6 Membentuk Nilai <i>Hash/Rolling Hash</i>	16
3.3.7 Membentuk <i>Window</i>	16

3.3.8 Membentuk <i>Fingerprint</i>	16
3.3.9 Menghitung Kemiripan Menggunakan <i>Jaccard Coefficient</i>	17
3.4 Desain Sistem	17
3.4.1 Rancangan <i>Flowchart</i> Sistem	17
3.4.2 Analisis Kebutuhan Sistem dan Perangkat	18
3.5 Uji Coba Sistem.....	19
BAB IV. ANALISIS DAN PERANCANGAN SISTEM.....	20
4.1 Deskripsi Sistem.....	20
4.2 Analisis Pengguna	33
4.3 Analisis Kebutuhan Fungsional.....	33
4.4 Perancangan Detail Sistem	34
4.4.1 Perancangan Proses.....	34
4.4.2 Perancangan <i>Use Case Diagram</i>	34
4.4.3 Perancangan <i>Activity Diagram</i>	35
4.4.4 Perancangan <i>Entity Relationship Diagram (ERD)</i>	36
4.4.5 Perancangan Desain <i>Database</i>	37
4.4.6 Perancangan <i>Database</i>	38
4.4.7 Perancangan Antarmuka Pengguna.....	40
BAB V. IMPLEMENTASI DAN PENGUJIAN.....	48
5.1 Implementasi	48
5.1.1 Halaman <i>Home</i>	48
5.1.2 Halaman Daftar Dokumen	50
5.1.3 Halaman Dokumen Referensi	51
5.1.4 Halaman Dokumen Pribadi	52
5.1.5 Halaman Pilih Dokumen Perbandingan.....	53
5.1.6 Halaman Input Dokumen Berupa Teks.....	55
5.1.7 Halaman Input Dokumen Berupa PDF	56
5.1.8 Halaman Riwayat Perhitungan.....	57
5.1.9 Halaman Detail Hasil Perhitungan.....	58
5.2 Pengujian	60
5.2.1 Lingkungan Pengujian	61
5.2.2 Rencana Pengujian	61
5.2.3 Analisis Rencana Pengujian.....	72
BAB VI. HASIL DAN PEMBAHASAN	83
6.1 Hasil Penelitian.....	83
6.2 Pembahasan Hasil.....	96
BAB VII. KESIMPULAN DAN SARAN	103
7.1 Kesimpulan.....	103
7.2 Saran	103
DAFTAR PUSTAKA	104
LAMPIRAN.....	108

DAFTAR GAMBAR

	Halaman
Gambar 2. 1 Tahap Preprocessing	8
Gambar 2. 2 Proses Case Folding	8
Gambar 2. 3 Proses Tokenizing	8
Gambar 2. 4 Proses Filtering.....	9
Gambar 4. 1 Use Case Diagram	35
Gambar 4. 2 Activity Diagram.....	36
Gambar 4. 3 Entity Relationship Diagram (ERD) Sistem	37
Gambar 4. 4 Perancangan Desain Database.....	37
Gambar 4. 5 Antarmuka Halaman Beranda Utama User.....	41
Gambar 4. 6 Antarmuka Halaman Beranda Utama Admin	41
Gambar 4. 7 Antarmuka Halaman Daftar Dokumen User.....	42
Gambar 4. 8 Antarmuka Halaman Daftar Dokumen Admin	42
Gambar 4. 9 Antarmuka Halaman Pilih Dokumen Pembanding User.....	43
Gambar 4. 10 Antarmuka Halaman Pilih Dokumen Pembanding Admin	43
Gambar 4. 11 Antarmuka Halaman Input Dokumen Teks User	44
Gambar 4. 12 Antarmuka Halaman Input Dokumen PDF User	44
Gambar 4. 13 Antarmuka Halaman Input Dokumen Teks Admin	45
Gambar 4. 14 Antarmuka Halaman Input Dokumen Teks PDF Admin	45
Gambar 4. 15 Antarmuka Halaman Riwayat User.....	46
Gambar 4. 16 Antarmuka Halaman Riwayat Admin	46
Gambar 4. 17 Antarmuka Halaman Detail Hasil Perhitungan User	47
Gambar 4. 18 Antarmuka Halaman Detail Hasil Perhitungan Admin.....	47
Gambar 5. 1 Halaman Home User	48
Gambar 5. 2 Halaman Login User	49
Gambar 5. 3 Halaman Home User	49
Gambar 5. 4 Halaman Home Admin.....	50
Gambar 5. 5 Halaman Daftar Dokumen User	50
Gambar 5. 6 Halaman Daftar Dokumen Admin	51
Gambar 5. 7 Halaman Dokumen Referensi User.....	51

Gambar 5. 8 Halaman Daftar Referensi Admin.....	52
Gambar 5. 9 Halaman Dokumen Pribadi	52
Gambar 5. 10 Halaman Pilih Dokumen Pembanding Pribadi User	53
Gambar 5. 11 Halaman Pilih Dokumen Pembanding Referensi User	54
Gambar 5. 12 Halaman Pilih Dokumen Pembanding Admin	54
Gambar 5. 13 Halaman Pilih Dokumen Pembanding Admin	55
Gambar 5. 14 Halaman Input Dokumen Berupa Teks User	55
Gambar 5. 15 Halaman Input Dokumen Berupa Teks Admin.....	56
Gambar 5. 16 Halaman Input Dokumen Berupa PDF User.....	56
Gambar 5. 17 Halaman Input Dokumen Berupa PDF Admin	57
Gambar 5. 18 Halaman Riwayat Perhitungan User	57
Gambar 5. 19 Halaman Riwayat Perhitungan Admin.....	58
Gambar 5. 20 Halaman Detail Hasil Perhitungan User	58
Gambar 5. 21 Halaman Detail Hasil Perhitungan Admin.....	59
Gambar 5. 22 Potongan Kode Program Proses Winnowing	60
Gambar 6. 1 Grafik Persentase Kemiripan Dokumen pada Skenario 1	97
Gambar 6. 2 Grafik Persentase Kemiripan Dokumen pada Skenario 2.....	98
Gambar 6. 3 Grafik Persentase Kemiripan Dokumen Skenario 3	99
Gambar 6. 4 Grafik Persentase Kemiripan Dokumen Skenario 4	100
Gambar 6. 5 Grafik Persentase Kemiripan Dokumen Skenario 5	101
Gambar 6. 6 Grafik Hasil Kuesioner.....	102

DAFTAR TABEL

	Halaman
Tabel 3. 1 Tabel ASCII	15
Tabel 3. 2 Spesifikasi Perangkat Lunak.....	18
Tabel 3. 3 Spesifikasi Perangkat Keras.....	19
Tabel 4. 1 Isi Dokumen Uji dan Dokumen Pembanding	20
Tabel 4. 2 Proses Terjemahan Dokumen	21
Tabel 4. 3 Tahap Preprocessing	21
Tabel 4. 4 Contoh Hasil N-Gram dengan $n = 9$	21
Tabel 4. 5 Hasil Rolling Hash	23
Tabel 4. 6 Hasil Pembentukan Window.....	24
Tabel 4. 7 Proses Fingerprint	32
Tabel 4. 8 Analisis Pengguna.....	33
Tabel 4. 9 Tabel Admin	38
Tabel 4. 10 Tabel User	38
Tabel 4. 11 Tabel tb_dokumen.....	39
Tabel 4. 12 Tabel tb_hitung	40
Tabel 5. 1 Daftar Dokumen Uji Coba Sistem	61
Tabel 5. 2 Rencana Pengujian Fungsional	62
Tabel 5. 3 Data Pengujian	64
Tabel 5. 4 Pertanyaan Kuesioner	71
Tabel 5. 5 Perbandingan Hasil Perhitungan.....	72
Tabel 5. 6 Prosedur Pengujian Fungsional.....	75
Tabel 5. 7 Prosedur Pengujian Validitas Sistem	80

DAFTAR LAMPIRAN

Lampiran 1 Tabel Daftar Stopword

Lampiran 2 Format Cover CD

Lampiran 3 Format Punggung Halaman Sampul

Lampiran 4 Daftar Log Bimbingan

Lampiran 5 Form Verifikasi Abstrak dan Tata Tulis Buku Laporan Skripsi