



**KLASIFIKASI SIMILARITAS DOKUMEN MENGGUNAKAN
ALGORITMA WINNOWING DENGAN PENDEKATAN JACCARD
COEFFICIENT BERBASIS WEB**

LAPORAN SKRIPSI

Uzdha Zachrias
41520010243
UNIVERSITAS
MERCU BUANA

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA 2023**



**KLASIFIKASI SIMILARITAS DOKUMEN MENGGUNAKAN
ALGORITMA WINNOWER DENGAN PENDEKATAN JACCARD
COEFFICIENT BERBASIS WEB**

LAPORAN SKRIPSI

**Uzdha Zachrias
41520010243**

**UNIVERSITAS
MERCU BUANA**

Diajukan sebagai salah satu syarat untuk memperoleh gelar sarjana

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA 2023**

HALAMAN PENYATAAN KARYA SENDIRI

Saya yang bertanda tangan di bawah ini:

Nama : Uzdha Zachrias
NIM : 41520010243
Program Studi : Teknik Informatika

Judul Laporan Skripsi : KLASIFIKASI TINGKAT SIMILARITAS DOKUMEN
MENGUNAKAN ALGORITMA WINNOWER DENGAN
PENDEKATAN JACCARD COEFFICIENT BERBASIS
WEB

Menyatakan bahwa Laporan Skripsi ini adalah hasil karya saya sendiri dan bukan plagiat, serta semua sumber baik yang dikutip maupun dirujuk telah saya nyatakan dengan benar. Apabila ternyata ditemukan di dalam Laporan Skripsi saya terdapat unsur plagiat, maka saya siap mendapatkan sanksi akademis yang berlaku di Universitas Mercu Buana.

Jakarta, 28 Desember 2023

UNIVERSITAS
MERCU BUANA



Uzdha Zachrias.

HALAMAN PENGESAHAN

Laporan Skripsi ini diajukan oleh:

Nama Mahasiswa : Uzdha Zachrias
NIM : 41520010243
Program Studi : Teknik Informatika
Judul Laporan Skripsi : KLASIFIKASI SIMILARITAS DOKUMEN
MENGGUNAKAN ALGORITMA WINNOWER
DENGAN PENDEKATAN JACCARD COEFFICIENT
BERBASIS WEB

Telah berhasil dipertahankan pada sidang di hadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Strata 1 pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer Universitas Mercu Buana.

Disahkan oleh:

Pembimbing : Wawan Gunawan, S.Kom, M.T
NIDN : 0424108104
Ketua Penguji : Dr. Hadi Santoso, S.Kom., M.Kom
NIDN : 0225067701
Penguji 1 : Saruni Dwiasnati, ST, MM., M.Kom
NIDN : 0325128802
Penguji 2 : Lukman Hakim, ST., M.Kom
NIDN : 0327107701

(Wawan Gunawan)
(Hadi Santoso)
(Saruni Dwiasnati)
(Lukman Hakim)

Jakarta, 28 Desember 2023

Mengetahui,

Dekan



(Dr. Bambang Jokonowo, S.Si., M.T.I)

Ketua Program Studi



(Dr. Hadi Santoso, S.Kom., M.Kom)

KATA PENGANTAR

Puji syukur saya panjatkan kepada Tuhan Yang Maha Esa, karena atas berkat dan rahmat-Nya, saya dapat menyelesaikan Laporan Skripsi ini. Penulisan Laporan Skripsi ini dilakukan dalam rangka memenuhi salah satu syarat untuk mencapai gelar Sarjana Komputer pada Fakultas Ilmu Komputer Universitas Mercu Buana. Saya menyadari bahwa, tanpa bantuan dan bimbingan dari berbagai pihak, dari masa perkuliahan sampai pada penyusunan Skripsi ini, sangatlah sulit bagi saya untuk menyelesaikan Laporan Skripsi ini. Oleh karena itu, saya mengucapkan terima kasih kepada:

1. Prof. Dr. Ir. Andi Adriansyah, M.Eng selaku Rektor Universitas Mercu Buana.
2. Dr. Bambang Jokonowo, S.Si., M.T.I selaku Dekan Fakultas Ilmu Komputer.
3. Dr. Hadi Santoso, S.Kom., M.Kom selaku Ketua Program Studi Teknik Informatika.
4. Umniy Salamah, S.Kom., MMSI selaku Dosen Pembimbing Akademik yang telah membimbing saya selama perkuliahan;
5. Wawan Gunawan, S.Kom, M.T selaku Dosen Pembimbing yang telah menyediakan waktu, tenaga, dan pikiran untuk mengarahkan saya dalam penyusunan Laporan Skripsi ini;
6. Saruni Dwiasnati, ST, MM., M.Kom selaku Dosen Penguji 1 atas koreksi dan arahan serta masukannya.
7. Lukman Hakim, ST, M.Kom selaku Dosen Penguji 2 atas koreksi dan arahan serta masukannya.
8. Teman-teman yang memberikan bantuan dan dukungan selama penyusunan Laporan Skripsi ini.
9. Kepada Orang Tua dan keluarga saya yang telah mendoakan, mendukung, dan memberikan semangat disetiap langkah yang saya lalui.

Akhir kata, saya berharap Tuhan Yang Maha Esa berkenan membalas segala kebaikan semua pihak yang telah membantu. Semoga Laporan Skripsi ini membawa manfaat bagi pengembangan ilmu.

Jakarta, 28 Desember 2023



Uzdha Zachrias

HALAMAN PENYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH

Saya sebagai sivitas akademik Universitas Mercu Buana, saya yang bertanda tangan di bawah ini:

Nama : Uzdha Zachrias
NIM : 41520010243
Program Studi : Teknik Informatika

Judul Laporan Skripsi : KLASIFIKASI SIMILARITAS DOKUMEN
MENGUNAKAN ALGORITMA WINNOWER DENGAN
PENDEKATAN JACCARD COEFFICIENT BERBASIS
WEB

Demi pengembangan ilmu pengetahuan, dengan ini memberikan izin dan menyetujui untuk memberikan kepada Universitas Mercu Buana **Hak Bebas Royalti Non-Esklusif** (*Non-exclusive Royalty-Free Right*) atas karya ilmiah saya yang berjudul di atas beserta perangkat yang ada (jika diperlukan).

Dengan Hak Bebas Royalti Non-Esklusif ini Universitas Mercu Buana berhak menyimpan, mengalihmedia/format-kan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan Laporan Magang/Skripsi/Tesis/Disertasi saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

UNIVERSITAS
MERCU BUANA

Jakarta, 28 Desember 2023

Yang menyatakan,



Uzdha Zachrias.

ABSTRAK

Nama : Uzdha Zachrias
NIM : 41520010243
Program Studi : Teknik Informatika
Judul Laporan Skripsi : **KLASIFIKASI TINGKAT SIMILARITAS DOKUMEN
MENGUNAKAN ALGORITMA WINNOWER
DENGAN PENDEKATAN JACCARD COEFFICIENT
BERBASIS WEB**

Pembimbing : Wawan Gunawan, S.Kom, MT

Dalam era kemajuan teknologi informasi, akses mudah terhadap berbagai sumber informasi melalui internet telah mengubah cara mahasiswa melakukan penelitian. Meskipun memberikan manfaat signifikan, kemudahan ini juga membawa masalah plagiarisme, yang merupakan tindakan yang merugikan dalam dunia akademik. Plagiat adalah tindakan penyalinan atau pengambilan ide dari karya orang lain tanpa memberikan kredit yang pantas, yang bertentangan dengan pedoman akademik. Penelitian ini bertujuan mengembangkan sistem pendeteksi plagiarisme yang efektif dan sesuai dengan bahasa Indonesia. Sistem ini menggunakan algoritma Winnowing dengan pendekatan Jaccard Coefficient dan teknik eliminasi kata-kata tidak deskriptif (stopword) dalam bahasa Indonesia. Sampel dokumen berbahasa Indonesia diambil dari tugas akhir mahasiswa Universitas Mercu Buana. Data dikumpulkan dari repository universitas dan akan dianalisis untuk mengukur tingkat similaritas antar dokumen dan performa algoritma Winnowing dalam mendeteksi plagiarisme. Hasil penelitian ini menunjukkan bahwa pengembangan sistem deteksi plagiarisme dengan menggunakan algoritma Winnowing dan pendekatan Jaccard Coefficient berhasil mencapai nilai akurasi yang tinggi. Sistem pendeteksi similarity index mampu memberikan hasil yang akurat dan relevan pada dokumen berbahasa Indonesia.

Kata Kunci : Jaccard Coefficient, Plagiarisme, Stopword, Winnowing

ABSTRACT

Name : Uzdha Zachrias
NIM : 41520010243
Study Program : Informatics Engineering
Title Research Proposal : *CLASSIFICATION OF DOCUMENT SIMILARITY LEVELS USING WINNOWING ALGORITHM WITH JACCARD COEFFICIENT WEB-BASED*

In the era of information technology advancement, easy access to various sources of information through the internet has transformed the way students conduct research. While providing significant benefits, this convenience also brings the problem of plagiarism, which is a detrimental act in the academic world. Plagiarism involves copying or taking ideas from someone else's work without proper attribution, which contradicts academic guidelines. This research aims to develop an effective plagiarism detection system tailored to the Indonesian language. The system utilizes the Winnowing algorithm with a Jaccard Coefficient approach and the technique of eliminating non-descriptive words (stopwords) in the Indonesian language. Samples of Indonesian-language documents are taken from the final assignments of Mercu Buana University students. Data is collected from the university's repository and will be analyzed to measure the similarity level between documents and the performance of the Winnowing algorithm in detecting plagiarism. The results of this research indicate that the development of a plagiarism detection system using the Winnowing algorithm and the Jaccard Coefficient approach has achieved a high accuracy value. The similarity index detection system is capable of providing accurate and relevant results for documents in the Indonesian language.

Keywords: *Jaccard Coefficient, Plagiarism, Stopword, Winnowing*

UNIVERSITAS
MERCU BUANA

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PENYATAAN KARYA SENDIRI.....	ii
HALAMAN PENGESAHAN.....	iii
KATA PENGANTAR.....	iv
HALAMAN PENYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH.....	v
ABSTRAK	vi
ABSTRACT.....	vii
DAFTAR ISI.....	viii
DAFTAR TABEL	x
DAFTAR GAMBAR.....	xi
DAFTAR LAMPIRAN	xii
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Perumusan Masalah.....	4
1.3 Tujuan Penelitian.....	5
1.4 Manfaat Penelitian	5
1.5 Batasan Penelitian	6
1.6 Sistematika Penulisan.....	7
BAB 2 TINJAUAN PUSTAKA.....	8
2.1 Teori Utama.....	8
2.2 Teori Pendukung.....	11
2.3 Penelitian Terdahulu	19
2.4 Gap penelitian.....	46
BAB 3 METODE PENELITIAN.....	52
3.1 Pendekatan Penelitian.....	52
3.2 Desain Penelitian	52
3.3 Subjek Penelitian	52
3.4 Teknik Pengumpulan Data	52
3.5 Analisis Data	53
3.6 Prosedur Penelitian	54
3.7 Evaluasi Hasil Penelitian.....	55
3.8 Timeline Penelitian	57

BAB 4 HASIL DAN PEMBAHASAN	58
4.1 Use Case Diagram.....	58
4.2 Activity Diagram	58
4.3 Sequence Diagram	63
4.4 User Interface.....	68
4.5 Dataset	71
4.6 Pre-Processing.....	71
4.7 Pembuatan Model.....	73
4.8 Pengujian.....	82
4.9 Analisis Hasil.....	84
BAB 5 KESIMPULAN DAN SARAN.....	85
5.1 Kesimpulan	85
5.2 Saran.....	85
DAFTAR PUSTAKA	86
LAMPIRAN.....	88



UNIVERSITAS
MERCU BUANA

DAFTAR TABEL

Tabel 2.1 Penelitian Terdahulu ke-1	19
Tabel 2.2 Penelitian Terdahulu ke-2	20
Tabel 2.3 Penelitian Terdahulu ke-3	22
Tabel 2.4 Penelitian Terdahulu ke-4	24
Tabel 2.5 Penelitian Terdahulu ke-5	26
Tabel 2.6 Penelitian Terdahulu ke-6	28
Tabel 2.7 Penelitian Terdahulu ke-7	30
Tabel 2.8 Penelitian Terdahulu ke-8	31
Tabel 2.9 Penelitian Terdahulu ke-9	33
Tabel 2.10 Penelitian Terdahulu ke-10	35
Tabel 2.11 Penelitian Terdahulu ke-11	37
Tabel 2.12 Penelitian Terdahulu ke-12	39
Tabel 2.13 Penelitian Terdahulu ke-13	41
Tabel 2.14 Penelitian Terdahulu ke-14	42
Tabel 2.15 Penelitian Terdahulu ke-15	44
Tabel 2.16 Jumlah Jurnal Relevan	46
Tabel 5.1 Proses Case Folding.....	71
Tabel 5. 2 Proses Tokenizing	72
Tabel 5.3 Proses Stemming.....	72
Tabel 5. 4 Proses Stopword.....	73
Tabel 5.5 Pembentukan N-Gram.....	74
Tabel 5.6 Nilai ASCII	75
Tabel 5.7 Ilustrasi Klasifikasi Dengan Confusion Matrix	83

DAFTAR GAMBAR

Gambar 1.1 Indikasi Plagiarisme Tingkat Mahasiswa.....	2
Gambar 2.1 Proses Winowwing.....	9
Gambar 2.2 Proses Case Folding.....	13
Gambar 2.3 Proses Tokenizing.....	14
Gambar 2.4 Proses Stopword.....	15
Gambar 2.5 Proses Stemming.....	16
Gambar 2.6 Rolling Hash.....	17
Gambar 2.7 Ilustrasi Proses N-Gram.....	18
Gambar 2.8 Visualisasi Menggunakan VOSviewer.....	46
Gambar 3.1 Prosedur Penelitian.....	54
Gambar 3.2 Confusion Matrix.....	56
Gambar 3.3 Timeline Penelitian.....	57
Gambar 5.1 Use Case Diagram.....	58
Gambar 5.2 Activity Diagram Melihat Detail Dokumen.....	58
Gambar 5.3 Activity Diagram Mengganti Dokumen.....	59
Gambar 5.4 Activity Diagram Training Data.....	60
Gambar 5.5 Activity Diagram Membandingkan Dokumen.....	60
Gambar 5.6 Activity Diagram Menambahkan Stopword.....	61
Gambar 5.7 Activity Diagram Mengganti Stopword.....	62
Gambar 5.8 Activity Diagram Menghapus Stopword.....	62
Gambar 5.9 Sequence Diagram Melihat Detail Dokumen.....	63
Gambar 5.10 Sequence Diagram Mengganti Dokumen.....	64
Gambar 5.11 Sequence Diagram Training Data.....	65
Gambar 5.12 Sequence Diagram Membandingkan Dokumen.....	66
Gambar 5.13 Sequence Diagram Menambahkan Stopword.....	66
Gambar 5.14 Sequence Diagram Mengganti Stopword.....	67
Gambar 5.15 Sequence Diagram Menghapus Stopwordssss.....	68
Gambar 5.17 Halaman Training Data.....	68
Gambar 5.18 Halaman Edit Dokumen.....	69
Gambar 5.19 Halaman Detail Dokumen.....	69
Gambar 5.20 Halaman Check Similarity.....	70
Gambar 5.21 Halaman Stopwords.....	70
Gambar 5.22 Isi Dokumen 1 dan 2.....	80
Gambar 5.23 Tampilan Hasil Perbandingan.....	82
Gambar 5.24 Nilai Confussion Matrix.....	83

DAFTAR LAMPIRAN

Lampiran 1 Bimbingan	88
Lampiran 2 Bukti Submit.....	89
Lampiran 3 Naskah Artikel Jurnal	90
Lampiran 4 Curriculum Vitae	99
Lampiran 5 Surat Pernyataan HKI.....	100
Lampiran 6 Sertifikat BNSP	102

