

Rancang Bangun Sistem Deteksi Plagiarisme Dokumen Teks Dengan Algoritma Winnowing Dan Rabin-Karp

Yusup Nurhidayat*¹, Ir. Sofa Sofiana, S.Kom, M.Kom²

^{1,2}Universitas Pamulang Jl. Raya Puspatek No. 46, Serpong, Kota Tangerang Selatan, Banten 15316, Indonesia, 021-74709855 ext.1042

^{1,2}Jurusan Teknik Informatika, Fakultas Teknik, Universitas Pamulang

e-mail: *¹yusup.nhidayat@gmail.com, ²dosen00407@unpam.ac.id

Abstrak

Plagiarisme merupakan tindakan penjiplakan yang semakin marak terjadi, terutama dalam lingkungan akademik. Untuk mengatasi hal tersebut, diperlukan sistem yang mampu mendeteksi kesamaan isi antar dokumen teks secara otomatis dan akurat. Penelitian ini bertujuan untuk merancang dan membangun sistem pendeteksi plagiarisme berbasis web dengan memanfaatkan kombinasi algoritma Winnowing dan Rabin-Karp. Algoritma Rabin-Karp digunakan pada tahap awal untuk melakukan pencocokan substring berbasis nilai hash, sedangkan algoritma Winnowing diterapkan untuk menentukan fingerprint teks yang representatif guna meningkatkan efisiensi dan akurasi proses deteksi.

Sistem dikembangkan menggunakan bahasa pemrograman Python serta dilengkapi dengan antarmuka pengguna yang sederhana dan responsif. Hasil pengujian menunjukkan bahwa sistem mampu mendeteksi tingkat kemiripan antar dokumen dengan akurasi yang cukup tinggi, serta dapat membedakan antara kemiripan substansial dan kebetulan. Dengan demikian, sistem ini dapat menjadi alat bantu yang efektif dalam mencegah dan mengidentifikasi praktik plagiarisme pada dokumen teks. Sistem ini menunjukkan akurasi deteksi 100% pada dokumen identik dan efisiensi waktu di bawah 3 detik per dokumen.

Manfaat praktik dari penelitian ini bagi institusi pendidikan adalah membantu menjaga integritas akademik dengan menyediakan sistem yang dapat mendeteksi plagiarisme secara otomatis, sehingga mempermudah dosen dalam memeriksa keaslian karya ilmiah mahasiswa secara efisien, objektif dan mendukung penerapan etika penulisan ilmiah.

Kata Kunci: *Plagiarisme, Deteksi Teks, Winnowing, Rabin-Karp, Fingerprinting, Python,*

Abstract

Plagiarism is an act of copying that has become increasingly prevalent, especially in academic environments. To address this issue, a system capable of automatically and accurately detecting text similarity between documents is required. This study aims to design and develop a web-based plagiarism detection system that utilizes a combination of the Winnowing and Rabin-Karp algorithms. The Rabin-Karp algorithm is used in the initial stage to perform substring matching based on hash values, while the Winnowing algorithm is applied to determine representative text fingerprints to improve detection efficiency and accuracy.

The system was developed using the Python programming language and equipped with a simple and responsive user interface. The test results show that the system can detect the similarity level between documents with high accuracy and distinguish between substantial and coincidental similarities. Therefore, this system can serve as an effective tool to prevent and identify plagiarism practices in text documents. This system demonstrated 100% detection accuracy on identical documents and time efficiency of under 3 second per document.

The practical benefit of this research for educational institutions is to help maintain academic integrity by providing a system capable of automatically detecting plagiarism, thereby facilitating lecturers in efficiently and objectively verifying the authenticity of students' academic works while supporting the implementation of ethical scientific writing practices

Keywords: *Plagiarism, Text Detection, Winnowing, Rabin-Karp, Fingerprinting, Python,*

I. PENDAHULUAN

Plagiarisme merupakan tantangan serius dalam lingkungan akademik karena dapat mengurangi orisinalitas karya ilmiah dan menurunkan mutu penelitian. Di Indonesia, peningkatan jumlah publikasi dan tugas akhir mahasiswa membuat kebutuhan terhadap mekanisme deteksi plagiarisme menjadi semakin penting. Pemeriksaan manual terhadap kesamaan teks seringkali memakan waktu dan rentan terhadap subjektivitas, sehingga diperlukan sistem otomatis yang mampu mendeteksi kemiripan teks secara cepat dan andal. Penelitian ini bertujuan merancang dan mengimplementasikan sistem pendeteksi plagiarisme berbasis algoritma *fingerpringting* (Winnowing) dan *hashing*/pencocokan pola (Rabin-Karp) untuk meningkatkan ketepatan dan efisiensi pemeriksaan dokumen teks berbahasa Indonesia.

Berbagai penelitian sebelumnya telah mengusulkan metode pendeteksian plagiarisme, antara lain teknik berbasis *n-gram*, *document fingerpringting*, dan algoritma pencocokan string berbasis *hashing*. Metode Winnowing dikenal efektif untuk menghasilkan *fingerprint* yang tahan terhadap perubahan minor, sedangkan Rabin-Karp menawarkan kecepatan dalam

pencocokan pola melalui *hashing*. sebagian besar penelitian sebelumnya masih berfokus pada korpus berbahasa Inggris dan berskala kecil, sehingga penerapan dan evaluasi algoritma Winnowing serta Rabin-Karp pada teks berbahasa Indonesia belum banyak dilakukan. Penelitian ini berupaya mengisi celah tersebut dengan merancang sistem deteksi plagiarisme yang menggabungkan dan menguji kedua algoritma pada dataset teks Bahasa Indonesia yang representatif.

Penelitian ini diharapkan memberikan manfaat teoritis maupun praktis. Secara teoritis, hasil penelitian menambah wawasan mengenai penerapan dan kinerja algoritma Winnowing dan Rabin-Karp pada korpus Bahasa Indonesia, serta menyediakan perbandingan empiris antara kedua pendekatan. Secara praktis, sistem yang dirancang dapat menjadi alat bantu bagi dosen, reviewer, dan institusi pendidikan untuk mendeteksi kesamaan dokumen secara lebih cepat dan konsisten, sehingga mendukung upaya peningkatan integritas akademik. Selain itu, dataset dan hasil eksperimen yang dihasilkan dapat menjadi basis bagi penelitian selanjutnya dalam pengembangan metode deteksi plagiarisme yang lebih tahan terhadap parafrase dan manipulasi teks.

II. METODE PELAKSANAAN

Penelitian ini menggunakan metode *research and development (R&D)* dengan pendekatan eksperimental karena bertujuan menghasilkan produk berupa sistem perangkat lunak yang dapat digunakan untuk mendeteksi plagiarisme, sekaligus menguji efektivitas algoritma yang diimplementasikan. Metode R&D dipilih karena sesuai untuk proses perancangan, pengembangan, dan penyempurnaan sistem, sedangkan pendekatan eksperimental digunakan untuk menguji kinerja sistem dalam berbagai kondisi melalui pengukuran tingkat akurasi dan efisiensi. Sistem yang dibangun melalui beberapa tahapan meliputi pengumpulan dokumen teks sebagai input, tahap *preprocessing* yang mencakup normalisasi, penghapusan tanda baca, dan tokenisasi, dilanjutkan dengan ekstraksi

fingerprint menggunakan algoritma Winnowing, pencocokan pola dengan algoritma Rabin-Karp, serta perhitungan persentase kesamaan dokumen yang hasilnya disajikan kepada pengguna. Rancangan ini disusun untuk mendukung proses deteksi kesamaan teks secara cepat dan akurat pada dokumen teks. Adapun teknik yang digunakan adalah sebagai berikut:

a. Algoritma Winnowing

Algoritma Winnowing digunakan untuk menghasilkan *fingerprint* dokumen dengan cara membagi teks menjadi potongan *k-gram*, menghitung nilai *hash*, dan memilih nilai minimum dalam jendela tertentu. Hasil *fingerprint* tersebut merepresentasikan ciri khas dokumen yang digunakan untuk deteksi kesamaan.

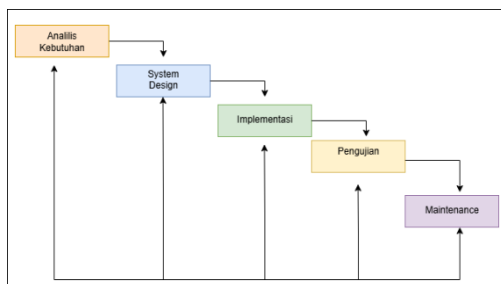
b. Algoritma Rabin-Karp

Algoritma Rabin-Karp diterapkan untuk mempercepat proses pencocokan pola dengan memanfaatkan fungsi hash rolling, sehingga dapat mengidentifikasi bagian teks yang memiliki kesamaan antar dokumen. Kombinasi

kedua algoritma ini dipilih karena Winnowing efektif dalam mempertahankan ciri teks meskipun terdapat sedikit perubahan, sementara Rabin-Karp efisien dalam pencarian pola dalam teks berukuran besar.

III. HASIL DAN PEMBAHASAN

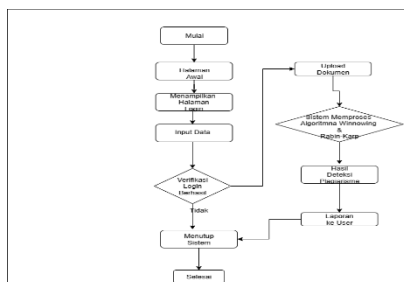
Metode pengembangan sistem perangkat lunak pada penelitian ini menggunakan model *Waterfall*, karena sesuai untuk penelitian akademik yang membutuhkan tahapan sistematis, terstruktur, dan terdokumentasi. Model ini dipilih untuk memastikan setiap fase pengembangan perangkat lunak berjalan secara berurutan dan dapat dievaluasi sebelum melanjutkan ke tahap berikutnya. Adapun tahapan pengembangan sistem meliputi:



Gambar 1. Pengembangan Sistem Perangkat Lunak Waterfall

a. Flowchart Sistem

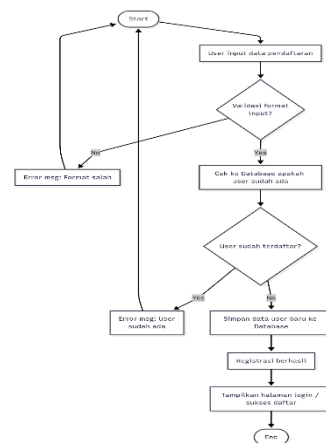
Flowchart ini menunjukkan alur kerja sistem pendeteksi plagiarisme dengan satu aktor, yaitu **User**. Proses dimulai dari login dan unggah dokumen, kemudian sistem melakukan preprocessing teks, ekstraksi fingerprint dengan algoritma Winnowing, serta pencocokan pola menggunakan algoritma Rabin-Karp. Hasil akhir berupa persentase kesamaan ditampilkan dalam laporan yang dapat digunakan user untuk menilai indikasi plagiarisme.



Gambar 2. Flowchart sistem

b. Flowchart User Daftar

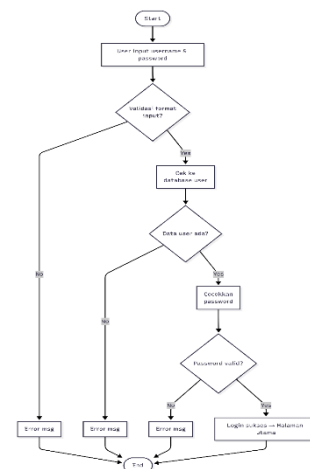
Proses pembuatan akun baru pada sistem, dimana diagram ini menekankan pada urutan atau langkah-langkah yang dilakukan user dan sistem dari awal hingga akhir.



Gambar 3. Flowchart User Daftar Akun

c. Flowchart User Login

Saat pengguna (user) melakukan proses login ke dalam sebuah sistem. Diagram ini menjelaskan langkah-langkah mulai dari user memasukkan username dan password, sistem ini melakukan validasi hingga muncul hasil berupa akses berhasil atau pesan error jika gagal.



Gambar 4. Flowchart User Login

Sistem pendeteksi plagiarisme yang dikembangkan berhasil diimplementasikan dengan antarmuka yang sederhana dan mudah digunakan. Pengguna dapat melakukan login, mengunggah dokumen dalam format teks, serta memperoleh hasil deteksi berupa persentase kesamaan antar dokumen. Proses deteksi berjalan melalui tahapan preprocessing, ekstraksi fingerprint menggunakan algoritma Winnowing, serta pencocokan pola dengan algoritma Rabin-Karp.

Pengujian sistem dilakukan pada kumpulan dokumen berbahasa Indonesia dengan berbagai tingkat kesamaan, mulai dari dokumen identik, dokumen dengan modifikasi kecil, hingga dokumen berbeda. Hasil pengujian menunjukkan bahwa sistem mampu mendeteksi tingkat kesamaan secara akurat. Pada dokumen identik, sistem menghasilkan nilai kesamaan mendekati 100%, sementara pada dokumen dengan sedikit modifikasi, persentase kesamaan berkisar antara 70%–85%. Adapun pada dokumen yang berbeda, nilai kesamaan turun signifikan hingga di bawah 20%. Hal ini menunjukkan bahwa kombinasi algoritma Winnowing dan Rabin-Karp efektif dalam mengidentifikasi kemiripan teks.

Selain akurasi, aspek efisiensi juga diuji dengan mengukur waktu pemrosesan. Hasilnya, proses deteksi pada dokumen dengan panjang rata-rata 5.000 kata dapat diselesaikan dalam waktu kurang dari 3 detik, sehingga sistem dinilai cukup efisien untuk penggunaan praktis. Pembahasan ini menegaskan bahwa penggunaan Winnowing sebagai metode pembentukan fingerprint mampu mengurangi kompleksitas data yang diproses, sedangkan Rabin-Karp mempercepat proses pencocokan pola melalui teknik rolling hash.

Secara keseluruhan, hasil penelitian ini menunjukkan bahwa sistem yang dibangun dapat digunakan sebagai alat bantu untuk mendeteksi plagiarisme pada dokumen teks berbahasa Indonesia. Meskipun demikian, terdapat beberapa keterbatasan, antara lain sensitivitas sistem terhadap sinonim dan parafrasa, sehingga penelitian selanjutnya dapat menggabungkan pendekatan semantik atau *natural language processing* (NLP) untuk meningkatkan kualitas deteksi.

a. Halaman Antar Muka Pengguna

Halaman website utama terdiri dari menu Login/Registrasi untuk autentikasi pengguna,

serta halaman Riwayat Terbaru yang menampilkan persentase kesamaan antar dokumen.

Selamat Datang di Aplikasi Deteksi Plagiarisme

Silakan login atau daftar untuk menggunakan fitur cek plagiarisme.

10 Riwayat Terbaru

File 1	File 2	Waktu	Kemiripan	Hasil
SKRPSI BAB V YUSUP NURHIDAYAT.pdf	contoh jurnal rabin karp.pdf	2025-08-30 09:57:22	29.02%	Unduh
Copy of SKRPSI BAB 1-5 YUSUP NURHIDAYAT.docx	contoh jurnal rabin karp.pdf	2025-08-30 08:41:03	34.8%	Unduh
SKRPSI asal mullah akbar RHOI4404012024 file setelah revisi.pdf	SKRPSI BAB 4 YUSUP NURHIDAYAT.pdf	2025-08-17 16:34:02	19.17%	Unduh
File 1.docx	File 2.docx	2025-06-18 08:42:17	34.78%	Unduh
File 1.docx	File 2.docx	2025-06-18 08:01:44	34.78%	Unduh

Gambar 5. Halaman Utama Website

b. Halaman Daftar Akun

Halaman ini menampilkan sebuah form ditengah layar dengan judul dibagian atas seperti “Daftar Akun”. Di bawah judul terdapat beberapa field input yang disusun secara vertikal, antara lain input untuk nama lengkap, email, password, untuk memudahkan pengguna memahami informasi yang harus diisi.

Daftar

Username:

Email:

Password:

Daftar Akun

Sudah punya akun? [Silakan login.](#)

Gambar 6. Halaman Daftar Akun

c. Halaman Login

Halaman login untuk memastikan akses yang mudah bagi pengguna. validasi input dilakukan secara otomatis untuk memastikan data yang dimasukan sesuai standar yang dibuat, jika terjadi kesalahan dalam input maka akan muncul “login gagal” untuk memberitahukan kepada pengguna karena salah dalam input.

Gambar 7. Halaman Login

d. Halaman Profil Pengguna

Halaman ini untuk mengidentifikasi tingkat kesamaan isi antara satu dokumen dengan dokumen lainnya guna memastikan keaslian karya tulis. Pada sistem ini, deteksi dilakukan melalui beberapa tahap, yaitu preprocessing teks, pembentukan *fingerprint* menggunakan algoritma Winnowing, serta pencocokan pola dengan algoritma Rabin-Karp. Hasil akhir berupa persentase kesamaan yang menunjukkan apakah suatu dokumen berpotensi mengandung plagiarisme atau tidak.

Gambar 8. Halaman Profil Pengguna

e. Halaman Hasil Perbandingan Deteksi Plagiarisme

Pada halaman ini pengguna dapat melihat hasil perbandingan dua dokumen yang sudah berhasil di analisa oleh sistem agar memudahkan pengguna. Kedua dokumen ditampilkan dalam

dua kolom dengan bagian-bagian teks yang memiliki kemiripan diberi warna yang sama sehingga pengguna dapat langsung melihat kesamaan antar paragraf atau kalimat.

Gambar 9. Halaman Hasil Perbandingan Deteksi Plagiarisme

f. Evaluasi Sistem

Evaluasi sistem dilakukan untuk mengukur kinerja dan efektivitas sistem pendeteksi plagiarisme yang dibangun. Pengujian dilakukan menggunakan metode *black box testing* untuk memastikan seluruh fungsi berjalan sesuai dengan kebutuhan, meliputi login, unggah dokumen, proses deteksi, dan penyajian hasil. Hasil pengujian menunjukkan bahwa sistem berjalan dengan baik dan sesuai yang diharapkan.

No	Proses Pengelolaan	Langkah	Hasil Yang diinginkan	Hasil Yang didapat	Keterangan
1	Daftar Akun	Daftar dan input data user	Berhasil daftar dan input data user	Berhasil daftar, data disimpan di serta tampil diwebsite halaman daftar	Berhasil
2	Login	Login dan Input data user	Berhasil Login data user	Berhasil login, data disimpan di halaman login	Berhasil
	Upload file dokumen	User mengUpl oad file dokumen	Berhasil Upload file 1 dan file 2 dokumen dan disimpan	Berhasil Upload dokumen dan disimpan didatabase untuk memproses deteksi plagiarisme	Berhasil
4	Daftar Riwayat deteksi Plagiarisme	Login sebagai user	Menampilka n daftar riwayat di website	Berhasil menampilkan daftar riwayat terbaru	Berhasil
5	Hasil	Perbandi n dua dokumen berbeda	Menampilka n dokumen 1 dan 2 tidak mirip	Berhasil menampilkan kemiripan sangat rendah	Berhasil

Tabel 1. Evaluasi sistem menggunakan *BlackBox*

IV. SIMPULAN

Penelitian ini berhasil merancang dan membangun sistem pendeteksi plagiarisme pada dokumen teks berbahasa Indonesia dengan memanfaatkan algoritma Winnowing dan Rabin-Karp. Hasil pengujian menunjukkan bahwa sistem mampu mendeteksi tingkat kesamaan dokumen secara akurat, dengan nilai kesamaan mendekati 100% pada dokumen identik, 70–85% pada dokumen yang dimodifikasi sebagian, serta di bawah 20% pada dokumen yang berbeda topik. Dari sisi efisiensi, sistem mampu memproses

dokumen berukuran ± 5.000 kata dalam waktu kurang dari 3 detik, sehingga dinilai cukup cepat untuk penggunaan praktis. Dengan demikian, sistem yang dikembangkan dapat menjadi alat bantu efektif dalam mendeteksi indikasi plagiarisme pada karya tulis. Namun, sistem masih memiliki keterbatasan dalam mendeteksi plagiarisme berbasis parafrasa dan sinonim, sehingga pada penelitian selanjutnya dapat dikembangkan dengan pendekatan semantik atau teknik *natural language processing* (NLP) untuk meningkatkan kualitas deteksi.

DAFTAR PUSTAKA

- Abdelhamid, M., Azouaou, F., & Batata, S. (2022). A survey of plagiarism detection systems: Case of use with English, French and Arabic languages. arXiv.
- Ardhiansyah, M. (n.d.). deteksi plagiarisme pada file dokumen berdasarkan tingkat kesamaan dengan menggunakan metode algoritma rabin-karp berbasis web. *Jurnal Ilmukomputer*, 2, 760–766.
- Dimas Dwi Ichtiarto. (2024). Penerapan Algoritma Rabin-Karp Pada Sistem
- Jupron, & Sutrisno. (2022). Penerapan Algoritma Winnowing untuk Mendeteksi Kemiripan Dua Teks yang Berbeda. In *Scientia Sacra: Jurnal Sains* (Vol. 2, Issue 3).
- Kamat, O., Ghosh, T., Kalaivani, J., Angayarkanni, V., & Rama, P. (2024). Plagiarism detection using machine learning. arXiv.
- Nasywa, A. (2024). Implementasi Kebijakan Anti Plagiarisme dan Etika Penulisan Ilmiah di Lingkungan Akademik Indonesia. *Jurnal Penelitian Ilmu Sosial*, 1, 231–234.
- Nova, M., & Utami, W. H. U. (t.t.). EFL students' perception of Turnitin for detecting plagiarism on academic writing. *International Journal of Education* (Nova).
- Ramadhani, S. (2015). Sistem Pencegahan Plagiarisme Tugas Akhir Menggunakan Algoritma Rabin-Karp (Studi Kasus: Sekolah Tinggi Teknik Payakumbuh). *Jurnal Teknologi Informasi & Komunikasi Digital Zone*, 6, 1, 44–52.
- Rian Amadeo Christofel Palandeng, D. O. S. R. M. (2023). aspek hukum plagiarisme sebagai pelanggaran integritas akademik di perguruan tinggi. *Jurnal Hukum*, 12.
- Saputro, H., & Budiyo, U. (2024). model penilaian esai otomatis menggunakan algoritma rabin-karp, dice coefficient similarity dan synonym recognition studi kasus pada universitas budi luhur. *SKANIKA: Sistem Komputer Dan Teknik Informatika*, 7(1), 1–11.
- Sunardi, A. Yudhana, Dan I. A. Mukaromah, 2018, Implementasi Deteksi plagiarisme Menggunakan Metode N-Gram Dan Jaccard Similarity Terhadap Algoritma Winnowing, *Transmisi*, 20, 3, 105–110 Tracer Study Fakultas Teknologi Informasi UNISBA Blitar Berbasis Web. *JOURNAL ZETROEM*, 6(1), 58–65
- Suryati, T., Y. Wibisono, Dan Y. Wihardi, 2018, Aplikasi Deteksi Plagiarisme Dokumen Skripsi Dengan Algoritma Rabin-Karp, *Jurnal Teori Dan Aplikasi Ilmu Komputer*, 1, 2, 91–95.
- Silalahi, E., Plagiarisme Sebagai Peningkatan, D., Irani Tarigan, M., & Veronica Sinaga, R. (2024a). deteksi plagiarisme sebagai peningkatan integritas akademik. 3. *Jurnal Pengabdian Pada Masyarakat*,
- Tantoni, A. Dan M. T. A. Zaen, 2018, Implementasi Double Caesar Cipher Menggunakan ASCII, *Jurnal Informatika & Rekayasa Elektronika (Jire)*, 1, 2, 24–32.
- Purba, A. H. Dan Z. Situmorang, 2017, Analisis Perbandingan Algoritma Rabin-Karp Dan Levenshtein Distance Dalam Menghitung Kemiripan Teks *Jurnal Teknik Informatika Unika St. Thomas (Jtiust)*, 2, 2, 24–32.

Teguh Prasetyo Utomo, S. (2024). Aplikasi Turnitin sebagai Sarana Cek Plagiarisme Tugas Akhir Mahasiswa Fakultas Hukum Universitas Islam Indonesia. 65–81.

Zulfikar, A. F, 2017, Pengembangan Algoritma Stemming Bahasa Indonesia Dengan Pendekatan Dictionary Base Stemming Untuk Menentukan Kata Dasar Dari Kata Yang Berimbuhan, Jurnal Informatika Universitas Pamulang, 2, 3, 143–146.