

# IMPLEMENTASI ALGORITMA RABIN KARP DAN OPTIMASI DENGAN ALGORITMA STEMMER SASTRAWI DALAM DETEKSI PLAGIAT PADA JURNAL SKRIPSI MAHASISWA

**Herman**

Program Studi Teknik Informatika, STMIK TIME, Medan, Indonesia

Email: [hrman\\_ang@yahoo.com](mailto:hrman_ang@yahoo.com)

DOI: <https://doi.org/10.46880/methoda.Vol14No1.pp1-7>

## ABSTRACT

*The problem of plagiarism is a serious problem in tertiary institutions, given the high number of plagiarism acts that do not uphold the principle of honesty. However, previous studies that were conducted only focused on checking thesis similarities, but had not implemented the stemming algorithm in processing the data used in the research. Therefore, in this study a plagiarism detection system will be developed in thesis journals by implementing algorithm of Rabin Karp optimized with Literary Stemmer so that the results of the plagiarism detection rate are more accurate. Tests results in detecting plagiarism of student thesis journals, an accuracy rate of 80% was obtained in detecting similarities in student thesis journals in this study and the average speed of execution time was 25.6934 seconds.*

**Keyword:** *Plagiarism, Scientific Journal, Rabin Karp, Literary Stemmer.*

## ABSTRAK

*Salah satu masalah dalam perguruan tinggi adalah plagiarisme, mengingat banyaknya tindakan plagiarisme yang tidak menjunjung prinsip kejujuran. Permasalahan terkait plagiarisme di bidang akademis sudah banyak dibahas pada penelitian-penelitian sebelumnya. Namun penelitian-penelitian sebelumnya yang dilakukan hanya berfokus pada pengecekan kemiripan skripsi, namun belum menerapkan algoritma stemming dalam mengolah data-data yang digunakan pada penelitian. Perlu dikembangkan penelitian mengenai deteksi plagiat pada jurnal skripsi dengan mengimplementasikan algoritma Rabin Karp dan optimasi dengan algoritma Stemmer Sastrawi sehingga diharapkan dapat meningkatkan akurasi dari hasil pendeteksian plagiat. Berdasarkan hasil pengujian yang dilakukan melalui kombinasi algoritma Rabin Karp dan optimasi dengan Stemmer Sastrawi dalam mendeteksi plagiat jurnal skripsi mahasiswa mendapatkan tingkat keakuratan 80% dalam mendeteksi kemiripan jurnal skripsi mahasiswa pada penelitian ini serta rata-rata kecepatan waktu eksekusi yaitu 25,6934 detik.*

**Kata Kunci:** *Plagiarisme, Jurnal Ilmiah, Rabin Karp, Stemmer Sastrawi.*

## PENDAHULUAN

Plagiarisme merupakan suatu tindakan mengambil alih informasi, data ataupun ilmu pengetahuan yang telah ada serta kemudian mengakuinya sebagai karya pribadi, sehingga penemu yang asli tidak dikutip dalam informasi ataupun data yang dipakai. Salah satu kendala

dalam perguruan tinggi adalah tingginya tingkat plagiarisme yang tidak menjunjung prinsip kejujuran. Setiap tahunnya, dunia pendidikan dikejutkan oleh banyaknya pemberitaan mengenai kasus terkait plagiarisme terhadap sejumlah karya ilmiah. Tindakan plagiarisme yang dibiarkan terus menerus dapat membunuh

ide dan gagasan serta menurunkan tingkat kreativitas dari peneliti maupun akademisi .

Tindakan plagiat sering dilakukan di bidang akademisi adalah berkaitan dengan penulisan skripsi dan jurnal penelitian. Sebuah skripsi yang telah selesai akan dipublikasi dalam bentuk sebuah jurnal yang berkisar antara 1 sampai 20 halaman sebagai bukti bahwa hasil penelitian adalah milik hak cipta si peneliti. Plagiarisme telah menjadi salah satu permasalahan serius di era digital ini terutama di bidang akademik. Kemudahan dalam mengakses informasi melalui internet, diantaranya karya ilmiah, membuat aksi salin-tempel (copy-paste) sangat mudah dilakukan. Plagiarisme bukan saja merugikan penulis asli tetapi merugikan perkembangan ilmu pengetahuan juga. Merupakan hal yang penting untuk menciptakan lingkungan akademis yang dapat mendukung integritas intelektual dan menghargai karya asli. Tantangan plagiarisme di akademik telah banyak dilakukn di penelitian terdahulu, contohnya penelitian pada tahun 2021 membangun sebuah sistem deteksi plagiarisme skripsi menggunakan algoritma Winnowing. Hasil pengujian menunjukkan bahwa dalam mendeteksi plagiat secara baik pada algoritma Winnowing menggunakan K-Gram dan *window size* terbaik adalah 12 dan 4. Penelitian berikutnya yaitu pada tahun 2021 membahas mengenai pendeteksian plagiarisme pada bagian abstrak skripsi menggunakan Rabin Karp. Keakuratan dari hasil penelitian ini adalah nilai rata-rata kemiripan 86,84 % pada 2K-Gram; dan 69,56 % pada 3K-Gram; dan 56,06 % pada 4K-Gram; dan 48,71 % pada 5K-Gram; dan 44,30 % pada 6K-Gram.

Penelitian-penelitian sebelumnya yang dilakukan sudah cukup baik serta mampu mendeteksi plagiarisme pada skripsi dengan cukup baik. Namun penelitian-penelitian sebelumnya yang dilakukan hanya berfokus pada pengecekan kemiripan skripsi, namun belum menerapkan algoritma *stemming* dalam mengolah data-data yang digunakan pada penelitian (Hermawan & Ismiati, 2020). Berdasarkan penelitian yang dilakukan dengan penerapan teknik *stemming* dan *stopword*

*removal* pada tahap *preprocessing* tidak berpengaruh signifikan terhadap akurasi analisis sentimen dengan selisih akurasi tertinggi dan terburuk hanya sebesar 0,93% dan 0,8% baik pada *Confusion Matrix* maupun *K-Fold*. Meskipun peningkatan tidak signifikan, namun penerapan teknik *stemming* terbukti dapat meningkatkan akurasi pengecekan. Berdasarkan hal tersebut, akan dikembangkan aplikasi yang dapat mendeteksi plagiat pada jurnal skripsi dengan mengimplementasikan algoritma Rabin Karp dan optimasi dengan algoritma Stemmer Sastrawi sehingga diharapkan dapat meningkatkan akurasi dari hasil pendeteksian plagiat.

## TINJAUAN PUSTAKA

### Plagiarisme

Yang dimaksud dengan plagiarisme adalah suatu aksi dan kegiatan pengambilan karangan/pendapat orang lain dan menjadikannya seolah-olah karangan/pendapat sendiri, atau lebih dikenal dengan nama jiplakan (Bahri et al., 2018).

### Karya Ilmiah

Karya tulis yang berisi pembahasan secara ilmiah oleh penulis/peneliti disebut Karya Ilmiah. Tujuan karya ilmiah adalah memberikan jawaban atas sesuatu hal yang aktual dan belum pernah ditulis dan membuktikan kebenaran atas hal tersebut dalam tulisan, yang dijelaskan secara sistematis dan logis. Apabila penelitian atas tulisan tersebut telah pernah dilakukan, maka tujuan penelitian adalah untuk mengembangkan hasil penelitian sebelumnya

### Jurnal Skripsi

Salah satu syarat agar mahasiswa memperoleh gelar sarjana adalah mahasiswa membuat jurnal skripsi yaitu hasil yang berisi laporan penelitian atau tulisan ilmiah. Jurnal skripsi telah menjadi bentuk komunikasi ilmiah dan menjadi tolak ukur keberhasilan seorang mahasiswa dalam menyelesaikan tugas akademiknya. Jurnal skripsi biasanya hanya berupa 10-30 halaman dan merupakan ringkasan dari penelitian yang telah dilakukan untuk dipublikasikan.

membandingkan *token* dengan teks secara terus menerus dari awal sampai akhir teks

### Text Mining

*Text mining* adalah sebuah metode yang dipergunakan untuk mengambil suatu kesimpulan dari data text yang tidak terstruktur tetapi saling terkait. Proses yang dilakukan pada *text mining* bertujuan memproses data yang tadinya tidak terstruktur untuk diambil kesimpulan atas suatu pola yang lebih berarti sehingga dapat mengambil keputusan yang lebih tepat.

Tahap pertama adalah *pre-processing*, yaitu teknik membersihkan teks dari karakter yang tidak diinginkan, seperti tanda baca dan karakter khusus. Selanjutnya tahap tokenisasi berfungsi untuk memecah text/teks menjadi kata atau token. Tahap normalisasi berfungsi untuk mengubah kata yang mempunyai bentuk dan/atau dasar yang serupa menjadi kata dasar standard. Kemudian tahap *filtering* untuk membuang kata yang tidak mempunyai hubungan dari teks. Sedangkan analisis dilakukan untuk mengekstrak informasi dari teks yang telah diolah.

*Stemming* adalah proses untuk mencari dan menemukan kata dasar yang dapat dipergunakan dalam menentukan fitur di dalam teks. Sastrawi adalah modul *library* yang memungkinkan menghilangkan suffix dan prefix dalam proses *stemming*.

### Algoritma Rabin Karp

Algoritma *Rabin Karp* merupakan algoritma pencocokan atau pencarian *string*, diambil dari nama yang mengembangkan algoritma ini yaitu Michael O. Rabin dan Richard M. Karp. Algoritma ini menggunakan teknik *hash* untuk mencari sebuah string atau pola (m) diantara kumpulan substring pada teks yang lebih besar (n). Perbandingan dilakukan sekali lagi bila hash value sama, dan bila tidak sama substring akan bergerak ke arah kanan sebanyak (n-m) kali. Tahapan algoritma Rabin Karp terbagi menjadi:

#### 1. K-Gram

*K-Gram* merupakan perwakilan *token* dengan panjang k, kemudian membaca dan

#### 2. Hashing

Fungsi *hashing* adalah mengubah *string* menjadi *integer*. Salah satu proses *hashing* menggunakan fungsi *rolling hash*.

$$C_{(k-1)} * b^k + C_K) \text{mod } q \quad (1)$$

$$H(c_1 \dots c_k) = (c_1 * b^{(k-1)} + C_2 * b^{(k-2)} + \dots +$$

Keterangan:

H : substing

c : nilai ascii per-karakter

b : konstan bilangan prima

k : banyak karakter

q : modula bilangan prima

### Stemmer Sastrawi

Stemmer Sastrawi adalah library sederhana dari PHP yang digunakan untuk mengurangi kata yang terinfleksi ke bentuk dasarnya dengan mengikuti aturan Bahasa Indonesia.

## METODE PENELITIAN

### Metode Pengumpulan Data

Penelitian ini menggunakan metode:

#### a. Observasi

Observasi dilakukan secara langsung ke kampus STMIK TIME Medan untuk mengumpulkan data-data jurnal skripsi mahasiswa yang akan digunakan proses *training* dan *testing*.

#### b. Eksperimen

Melakukan eksperimen pada aplikasi deteksi plagiat untuk mengetahui kebutuhan-kebutuhan yang diperlukan pada aplikasi yang akan dibangun.

#### c. Studi Pustaka

Studi pustaka dilakukan dengan membaca referensi dari internet, buku dan jurnal.

### Analisis Sistem

Dibagi menjadi tiga tahapan yaitu:



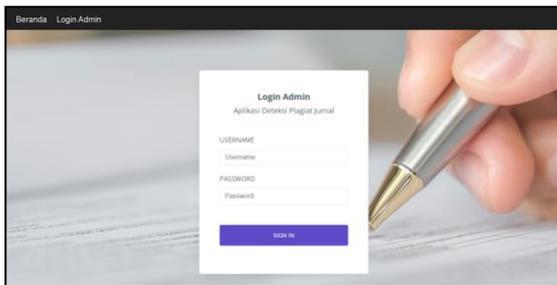


**Gambar 3.** Tampilan Detail Informasi Pengecekan Plagiarisme

Berikut ini hasil tampilan dari aplikasi pengecekan kemiripan teks yang digunakan oleh *user* admin antara lain.

### Tampilan Login Admin

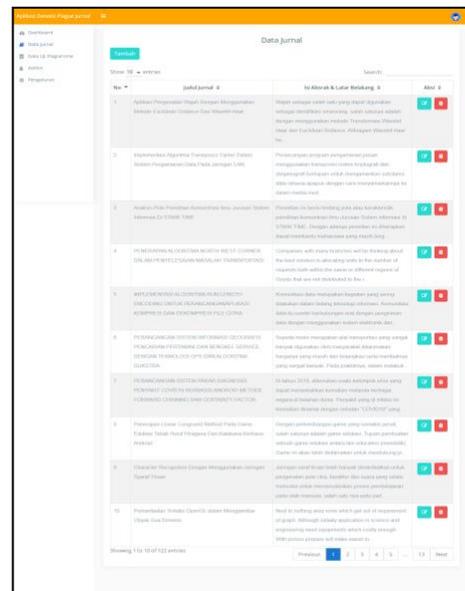
Tampilan pada Gambar 4 merupakan tampilan Login Admin yang berfungsi untuk autentikasi agar dapat masuk ke *dashboard* admin.



**Gambar 4.** Tampilan Login Admin

### Tampilan Kelola Data Jurnal Skripsi Mahasiswa

Tampilan kelola data jurnal skripsi mahasiswa yang ditunjukkan pada Gambar 5 merupakan tampilan yang berisikan *list* data-data jurnal skripsi mahasiswa yang telah ditambahkan sebelumnya. Pada tampilan ini juga tersedia adanya tombol tambah untuk menambah jurnal, tombol ikon pensil untuk mengubah jurnal, dan tombol ikon tong sampah untuk menghapus jurnal.



**Gambar 5.** Tampilan Kelola Data Jurnal Skripsi Mahasiswa

Selanjutnya, akan dilakukan pembahasan terkait hasil penelitian pada penelitian ini yaitu dengan melakukan pembahasan terkait implementasi dari algoritma Rabin Karp dan Stemmer Sastrawi dalam melakukan pendeteksian plagiat pada jurnal skripsi mahasiswa. Pada penelitian ini, dimasukkan data latih sebanyak 100 data jurnal skripsi dari *repository* STMIK TIME kemudian akan dilakukan pengujian sebanyak 20 kali pendeteksian untuk mengukur keakuratan dari kedua algoritma tersebut dalam berbagai perubahan yaitu tanpa perubahan, minor, sedang, dan mayor.

1. Tanpa perubahan sama sekali tidak melakukan perubahan artinya 100% kata mirip dari total kata yang ada.
2. Perubahan minor mengubah sebanyak 25% dari total kata yang ada.
3. Perubahan sedang mengubah sebanyak 50% dari total kata yang ada.
4. Perubahan mayor mengubah sebanyak 75% dari total kata yang ada.

Persentase yang diambil adalah persentase yang paling dekat dari skenario perubahan yang ditentukan. Ukuran nilai *K-Gram* yang digunakan adalah 5 dengan nilai basis *hash* yang digunakan adalah 7. Berikut ini Tabel 2 menunjukkan cara penalaran kesesuaian dari hasil pengujian.

**Tabel 1.** Penalaran Hasil Pengujian

Jenis Pengujian	Range Persen Plagiarisme Dianggap Sesuai (%)
Tanpa Perubahan	100
Minor	75-99,9
Sedang	25-74,9
Mayor	0-24,9

Selanjutnya, proses ini dilakukan sebanyak 20 kali dengan skenario dan perubahan yang berbeda. Hasil pengujian ditampilkan pada Tabel 2.

**Tabel 2.** Hasil Pengujian Keakuratan Deteksi Plagiarisme

Percobaan	Perubahan	Persen Plagiarisme	Waktu Eksekusi (Detik)	Keterangan
1	Tanpa Perubahan	100	24,0088	Sesuai
2	Tanpa Perubahan	100	26,1211	Sesuai
3	Tanpa Perubahan	100	25,0321	Sesuai
4	Tanpa Perubahan	100	23,2151	Sesuai
5	Tanpa Perubahan	100	25,6233	Sesuai
6	Minor	95,70	25,2207	Sesuai
7	Minor	88,43	21,2949	Sesuai
8	Minor	94,62	24,3211	Sesuai
9	Minor	79,63	23,4511	Sesuai
10	Minor	85,67	27,4232	Sesuai
11	Sedang	80,45	26,1113	Tidak Sesuai
12	Sedang	72,20	23,6821	Sesuai
13	Sedang	82,53	27,8932	Tidak Sesuai
14	Sedang	55,83	26,3851	Sesuai
15	Sedang	45,99	27,9999	Sesuai
16	Mayor	41,85	29,3214	Tidak Sesuai
17	Mayor	43,55	27,8621	Tidak Sesuai
18	Mayor	23,22	24,3211	Sesuai
19	Mayor	16,57	28,9004	Sesuai
20	Mayor	5,43	25,6792	Sesuai

$$\begin{aligned}
 \text{Akurasi Pendeteksian} &= (\text{Jumlah Percobaan Sesuai} / \text{Jumlah Percobaan}) * 100 \\
 &= (16/20) * 100 \\
 &= 0,8 * 100 \\
 &= 80\%
 \end{aligned}$$

$$\begin{aligned}
 \text{Rata-Rata Waktu Eksekusi} &= (\text{Jumlah Waktu Eksekusi} / \text{Jumlah Percobaan}) \\
 &= 513.8672/20 \\
 &= 25,6934
 \end{aligned}$$

Berdasarkan hasil pengujian dapat disimpulkan bahwa algoritma *Rabin Karp* yang dioptimasi dengan *Stemmer Sastrawi* memiliki tingkat keakuratan 80% dalam mendeteksi kemiripan jurnal skripsi mahasiswa pada

penelitian ini dengan rata-rata kecepatan waktu eksekusi yaitu 25,6934 detik.

## KESIMPULAN

Tahapan akhir dari sebuah penelitian adalah menguraikan kesimpulan yang didapatkan yaitu:

1. Aplikasi ini dapat mencegah tindakan plagiat yang dilakukan oleh para peneliti khususnya terkait jurnal skripsi melalui fitur pengecekan plagiarisme yang tersedia.
2. Berdasarkan hasil pengujian yang dilakukan melalui kombinasi algoritma Rabin Karp dan optimasi dengan Stemmer Sastrawi dalam mendeteksi plagiat jurnal skripsi mahasiswa mendapatkan tingkat keakuratan 80% dalam mendeteksi kemiripan jurnal skripsi mahasiswa pada penelitian ini serta rata-rata kecepatan waktu eksekusi yaitu 25,6934 detik.

## DAFTAR PUSTAKA

- Amr, Y., Asnita, D., Febriandi, Y., & Indrayani, D. (2021). *Buku Saku Panduan Anti Plagiarisme*, Langsa: Fakultas Syariah Institut Agama Islam Negeri.
- Bahri, S., Kana Trisnawati, I., & Bahri UIN Ar-Raniry Banda Aceh, S. (2018). Persepsi Mahasiswa Prodi Pendidikan Bahasa Inggris Uin Ar-Raniry Tentang Plagiarisme Tugas Kuliah. *Jurnal Ilmiah Didaktika*, 18(2), 205–224. <https://doi.org/10.22373/JID.V18I2.3243>
- Gunawan, I., Enri & Umaidah, Y. (2021). Deteksi Plagiarisme Abstrak Skripsi dengan Menggunakan Algoritma Rabin Karp (Studi Kasus: Fakultas Ilmu Komputer Universitas Singaperbangsa Karawang). *Jurnal Ilmu Komputer dan Teknologi Informasi*, 6(2), 75-81.
- Hermawan, L., & Ismiati, M. B. (2020). Aplikasi Pengecekan Dokumen Digital Tugas Mahasiswa Berbasis Website. *Jurnal Buana Informatika*, 11(2), 94–103.
- Kania, R., Solihati, T. I., & Arzaqi, F. N. (2022). Thesis Similarity Detection Application at Banten Jaya University. *Jurnal Sistem Informasi dan Informatika (Simika)*, 5(1), 78-89.
- Manalu, D. R., Rajagukguk, E., Siringoringo, R., Siahaan, D. K., & Sihombing, P.

- (2019, November). The Development of Document Similarity Detector by Jaccard Formulation. In *2019 International Conference of Computer Science and Information Technology (ICoSNIKOM)* (pp. 1-4). IEEE.
- Mawanta, I., Gunawan, T. S., & Wanayumini. (2021). Uji Kemiripan Kalimat Judul Tugas Akhir dengan Metode Cosine Similarity dan Pembobotan TF-IDF, *Jurnal Media Informatika Budidarma*, 5(2), 726-738.
- Mustikasari, D., Widaningrum, R., Arifin & Putri, W. H. E. (2020). Comparison of Effectiveness of Stemming Algorithms in Indonesian Documents, in *Proceedings of the 2nd Borobudur International Symposium on Science and Technology (BIS-STE 2020)*.
- Pahlevi, R. & Ziveria, M. (2022). Perancangan Deteksi Kemiripan pada Abstrak Artikel Ilmiah Informatika, *KALBISIANA : Jurnal Mahasiswa Institut Teknologi dan Bisnis Kalbis* 8(2), 1576-1586.
- Pradana, A. W. & Hayaty, M. (2019). The effect of stemming and removal of stopwords on the accuracy of sentiment analysis on indonesian-language texts," *Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control*, 4(4), 375-380.
- Pramudita, W., Tomasouw, B. P., Leleury, Z. A., & Rijoly, M. E. (2021) Perancangan Sistem Deteksi Plagiarisme Skripsi (Judul Dan Abstrak) Berbasis Matlab Menggunakan Algoritma Winnowing, *Pure and Applied Mathematics Journal (TENSOR)*, 2(2), 67-76.
- Prasidhatama, A., & Suryaningrum, K. (2018).Perbandingan Algoritma Nazief & Adriani Dengan Algoritma Idris Untuk Pencarian Kata Dasar, *Jurnal Teknologi & Manajemen Informatika*, 4(1), 192-195.
- Rahim, R. (2020). *Cara Praktis Penulisan Karya Ilmiah*, Yogyakarta: Zahir Publishing.
- Rosyida, I. A. & Alim, M. N. (2022). Publikasi Ilmiah dan Plagiarisme Dengan Locus of Control Sebagai Moderasi, *Journal of Management and Accounting (J-MACC)*, 5(1), 28-38.
- Satria, R. D., Suarjaya, I. M. A. D., & Pratama, I. P. A. E. (2022). Sentiment Analisa Antusias Masyarakat Terhadap Sampah Plastik Dengan Menggunakan Metode Support Vector Machine (SVM), *JITTER- Jurnal Ilmiah Teknologi dan Komputer*, 3(1).
- Siahaan, A. P. U., Aryza, E., Hariyanto, Rusiadi, Lubis, A. H., Ikhwan, A. & Kan, P. L. E. (2018). Combination of levenshtein distance and rabin-karp to improve the accuracy of document equivalence level, *International Journal of Engineering & Technology*, 7(2), 17-21.
- Siswanto, E. & Giap, Y. (2020) Implementasi Algoritma Rabin-Karp dan Cosine Similarity Untuk Pendeteksi Plagiarisme Pada Dokumen, *Jurnal ALGOR*, 1(2), 16-21.
- Temesvari, N. A. & Qomariana, W. Z. (2022). Pendampingan Menghindari Plagiarisme Dalam Penulisan Karya Tulis Bagi Mahasiswa, *Jurnal Abdimas*, 8(5), 281-286.
- Wali, M. & Safrizal. (2018). Similar text sebagai Pengkodean Aplikasi Plagiarisme, *Jurnal Teknologi Informasi dan Komunikasi (JTIK)*, 2(1), 11-19.
- Wibowo, R. K. & Hastuti, K. (2016) Penerapan Algoritma Winnowing Untuk Mendeteksi Kemiripan Teks Pada Tugas Akhir Mahasiswa, *Techno.COM*, 15(4), 303-311.
- Yulianto, M. A. & Nurhasanah. (2021) Pengaruh Stemming Nazief & Adriani terhadap Performa Algoritma RabinKarp dalam Mendeteksi Kemiripan Teks, *Jurnal Informatika Universitas Pamulang*, 6(4), 880-886.