

# PERBANDINGAN ALGORITMA WINNOWER DENGAN ALGORITMA RABIN KARP UNTUK MENDETEKSI PLAGIARISME PADA KEMIRIPAN TEKS JUDUL SKRIPSI

Fakultas Teknologi Informasi

Universitas Islam Kalimantan Muhammad Arsyad Al Banjari Banjarmasin

**Nur Alamsyah**

[uniskalam@gmail.com](mailto:uniskalam@gmail.com)

## ABSTRAK

Plagiarisme biasanya sering terjadi di dunia Pendidikan, baik itu dalam pengerjaan skripsi atau karya tulis ilmiah, dimana mahasiswa dituntut untuk menyelesaikan skripsinya sebagai syarat kelulusan dan mendapatkan gelar sarjana. Deadline untuk segera menyelesaikan karya tulis ilmiah inilah memicu mereka untuk mencari jalan pintas untuk melakukan copy-paste karya ilmiah atau skripsi orang lain sehingga terjadinya plagiarisme.

Untuk membantu mengurangi tingkat plagiarisme telah banyak dibuat sistem pendeteksi plagiarisme untuk memeriksa tingkat kemiripan teks dokumen. Pada umumnya, sistem pendeteksi plagiarisme dikembangkan untuk data teks seperti essay, artikel, jurnal, penelitian dan dokumen teks yang lebih terstruktur seperti *source code* bahasa pemrograman.

Untuk mendeteksi plagiarisme atau kemiripan teks dokumen dengan metode *fingerprinting* dapat dilakukan dengan perbandingan algoritma yang berkaitan pada bidang text mining misal perbandingan pendekatan Algoritma Rabin Karp, Algoritma Manber dan Algoritma Winnowing.

Pendekatan Algoritma Winnowing lebih baik daripada pendekatan algoritma Rabin Karp karena menghasilkan tingkat presentase yang lebih kecil dan waktu proses yang lebih cepat. Berdasarkan hasil Pengujian terhadap perbandingan pendekatan algoritma winnowing dengan algoritma rabin karp dapat dilihat kemungkinan kemiripan teks judul skripsi yang terkecil adalah dengan menggunakan pendekatan algoritma winnowing yaitu pada ujicoba ke 8 dengan nilai  $n\text{-gram} = 9$  dan  $\text{window} = 3$ , proses waktu 0.0257 dengan tingkat kemiripan terkecil yaitu 32.6 %.

**Kata Kunci :** *Algoritma Winnowing, Rabin Karp, fingerprint, n-gram, Hashing, plagiarisme*

## PENDAHULUAN

Plagiarisme tampaknya sudah membudaya, mudah terjadi terutama pada jenjang pendidikan S1. Kurangnya kontrol dari pembimbing yang relatif rendah, jumlah pembimbing yang tidak proporsional dengan jumlah mahasiswa dan sumber pustaka yang tidak mutakhir.

Menurut Abdi et al, 2015, plagiarisme adalah penggunaan ulang dari ide atau karya orang lain tanpa menyantumkan sumber aslinya.

Menurut Permendiknas, (Pencegahan dan Penanggulangan Plagiat di Perguruan Tinggi,

No 7, Pasal 1 ayat 1 2010) Plagiat merupakan perbuatan secara sengaja atau tidak sengaja dalam memperoleh atau mencoba memperoleh kredit atau nilai untuk suatu karya ilmiah, dengan mengutip sebagian atau seluruh karya dan/atau karya ilmiah orang lain, tanpa menyatakan sumber secara tepat dan memadai.

Untuk membantu mengurangi tingkat plagiarisme telah banyak dibuat sistem pendeteksi plagiarisme untuk memeriksa tingkat kemiripan teks dokumen. Pada umumnya, sistem pendeteksi plagiarisme dikembangkan untuk data teks seperti essay, artikel, jurnal, penelitian dan dokumen teks

yang lebih terstruktur seperti *source code* bahasa pemrograman.

Menurut Telepovska dan Gajdos (2010) aplikasi pendeteksian tersebut dapat menggunakan system local (terisolasi secara offline dan hanya menggunakan database individu) ataupun menggunakan system global (berjalan secara online dan menggunakan server internet sehingga dokumen yang dicari lebih beragam).

### Bentuk Plagiarisme

Menurut Clough (2003) bentuk plagiarism yang bisa terjadi yaitu:

1. Plagiarisme kata per kata, merupakan penyalinan kalimat secara langsung dari sebuah teks dokumen tanpa adanya pengutipan atau perizinan.
2. Plagiarisme paraphrase, merupakan penulisan ulang dengan mengubah kata atau sintaksis, tetapi teks aslinya masih dikenali.
3. Plagiarisme sumber sekunder, merupakan perbuatan mengutip kepada sumber asli yang didapat dari sumber sekunder dengan menghiraukan teks asli dari sumber sebenarnya.
4. Plagiarisme struktur sumber, merupakan penyalinan/ penjiplakan struktur suatu argument dari sebuah sumber.
5. Plagiarisme ide, merupakan penggunaan ulang suatu gagasan / pemikiran asli dari sebuah sumber teks tanpa bergantung bentuk teks sumber.
6. Plagiarisme authorship, merupakan pembubuhan nama sendiri secara langsung pada hasil karya orang lain.

Dari berbagai macam bentuk-bentuk plagiarisme diatas, dapat disimpulkan bahwa tindakan plagiarisme yang sering terjadi dalam dunia Pendidikan lebih cenderung kepada kegiatan menggunakan kembali suatu dokumen teks berupa kata/ kalimat dari suatu sumber yang tidak mengikuti kaidah tata aturan hak cipta, seperti aturan pengutipan

(citation) ataupun ketidakjelasan sumber/pengarang asli(bibliography).

### Kriteria Plagiarisme

Hal-hal yang dapat digolongkan sebagai plagiarisme adalah sebagai berikut:

1. Auto Plagiarisme, yakni seseorang yang memuat karyanya sendiri yang sama persis didua atau lebih media.
2. Mengakui karya tulis orang lain sebagai karya tulis sendiri.
3. Mengakui penemuan orang lain sebagai penemuan sendiri.
4. Mengakui gagasan/ide orang lain sebagai gagasan/ide sendiri
5. Meringkas atau mengutip karya orang lain tanpa menyebutkan sumbernya
6. Mengakui karya kelompok sebagai kepunyaan atau hasil sendiri, menyajikan tulisan yang sama dalam rangka yang berbeda tanpa menyebutkan darimana asalnya.

Sedangkan yang bukan termasuk tindak plagiarism adalah:

1. Mengutip karya orang lain (dengan merubah kalimat atau paraphrase) dengan menyebutkan sumbernya.
2. Menggunakan informasi yang merupakan fakta atau hal yang umum digunakan.
3. Menuliskan kembali gagaasan orang lain dengan menyebutkan sumbernya.

### Ambang Batas (*Threshold*) Plagiarisme

Menurut Sastroasmoro, threshold plagiarisme merupakan nilai ambang batas untuk menentukan apakah dokumen tersebut termasuk kedalam tindak plagiarisme atau tidak, terdapat 3 klasifikasi dalam menentukan ambang batas plagiarism, yaitu:

1. Plagiarisme ringan : < 30%
2. Plagiarisme sedang : < 30% - 70%
3. Plagiarism berat : > 70%

Namun angka-angka tersebut dapat disesuaikan dan ditentukan sendiri oleh lembaga atau universitas atau fakultas masing-masing.

Untuk mengatasi praktik plagiarisme, tidak cukup hanya mengingatkan kepada mahasiswa bahwa tindakan tersebut tidak boleh dilakukan, maka harus ada sistem atau metode untuk mendeteksi plagiarisme agar meminimalkan kecurangan dalam praktik plagiarisme.

Metode untuk mendeteksi plagiarisme dapat diklasifikasikan menjadi tiga metode yaitu metode perbandingan teks lengkap, metode dokumen fingerprinting dan metode kesamaan kata kunci.

### RUMUSAN MASALAH

Berdasarkan uraian dari latar belakang, maka dapat dirumuskan masalahnya adalah:

1. Terdapat banyak algoritma dalam mendeteksi plagiarisme, beberapa algoritma yang termasuk dalam metode fingerprinting yaitu algoritma Rabin Karp, algoritma winnowing dan algoritma Mannber untuk deteksi plagiarisme.
2. Membandingkan dua algoritma yaitu algoritma Rabin Karp dengan algoritma Winnowing dalam pendekatan deteksi plagiarisme.
3. Algoritma apa yang paling tepat dan efektif untuk mendeteksi plagiarisme pada teks dokumen khususnya judul skripsi.

### TUJUAN PENELITIAN

Tujuan penelitian ini adalah dengan melakukan perbandingan algoritma Rabin Karp dengan algoritma Winnowing agar dapat diketahui algoritma mana yang paling baik untuk mendeteksi plagiarisme pada dokumen teks, khususnya judul skripsi.

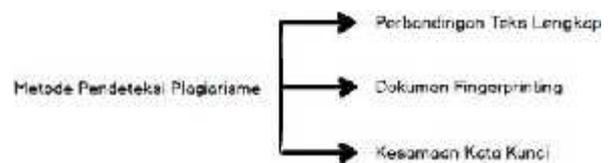
### TARGET LUARAN DAN MANFAAT PENELITIAN

Target Luaran pada penelitian ini adalah setelah nantinya didapatkan mana algoritma terbaik untuk mendeteksi plagiarisme maka dapat diterapkan oleh Ketua Program Studi dalam menentukan tingkat plagiarisme pada judul-judul skripsi yang telah diajukan mahasiswa agar terhindar dari plagiarisme.

Sedangkan manfaat dari penelitian ini adalah untuk memudahkan mendeteksi tingkat plagiarisme pada judul-judul skripsi yang ada difakultas dengan judul skripsi yang diajukan dengan menggunakan algoritma winnowing.

### METODE PENELITIAN

Metode Pendeteksi Plagiarisme di bagi menjadi tiga bagian yaitu metode perbandingan teks lengkap, metode dokumen fingerprinting, dan metode kesamaan kata kunci. Klasifikasi metode pendeteksi plagiarisme dapat digambarkan sebagai berikut:



Gambar Klasifikasi Metode Pendeteksi Plagiarisme

Ketiga metode diatas dapat diuraikan sebagai berikut (Stein, 2006) :

#### 1. Perbandingan Teks Lengkap

Metode ini di terapkan dengan membandingkan semua isi dokumen. Dapat diterapkan untuk dokumen yang besar. Pendekatan ini membutuhkan waktu yang lama tetapi cukup efektif, karena kumpulan dokumen yang diperbandingkan adalah dokumen yang di simpan pada penyimpanan lokal. Metode perbandingan teks lengkap tidak dapat diterapkan untuk kumpulan dokumen yang tidak terdapat pada dokumen lokal. Algoritma yang digunakan pada metode ini adalah algoritma brute force, algoritma edit distance, algoritma boyer moore dan algoritma lavenshtein distance

#### 2. Dokumen Fingerprinting

Dokumen fingerprinting merupakan metode yang digunakan untuk mendeteksi keakuratan salinan antar dokumen, baik semua teks yang terdapat di dalam dokumen atau hanya sebagian teks saja. Prinsip kerja dari metode dokumen

fingerprinting ini adalah dengan menggunakan teknik hashing. Teknik hashing adalah sebuah fungsi yang mengkonversi setiap string menjadi bilangan.

### 3. Kesamaan Kata Kunci.

Prinsip dari metode ini adalah mengekstrak kata kunci dari dokumen dan kemudian di bandingkan dengan kata kunci pada dokumen yang lain. Pendekatan yang digunakan pada metode ini adalah teknik dot.

### Metode Dokumen Fingerprinting

Metode fingerprinting adalah metode yang menelusuri karakter satu persatu pada deret karakter. Prinsip kerja dari metode dokumen fingerprinting ini adalah dengan menggunakan teknik hashing. Teknik hashing adalah sebuah fungsi yang mengkonversi setiap string menjadi bilangan. Bilangan-bilangan tersebut menghasilkan nilai-nilai fingerprints sebagai acuan dalam perhitungan kesamaan dokumen. Kelebihan dari metode fingerprinting adalah waktu proses lebih cepat dibandingkan dari metode perbandingan teks lengkap dan metode kesamaan kata kunci. Pengujian similaritas dokumen fingerprinting memiliki tahapan pencarian nilai hashing dari setiap kata, pengambilan nilai fingerprints dan pembobotan persentasi kemiripan.

Beberapa algoritma yang termasuk dalam metode fingerprinting yaitu Algoritma Rabin Karp, Algoritma Winnowing dan Algoritma Rabin Karp . Fokus penelitian ini menggunakan dua algoritma yaitu Algoritma Rabin Karp dan Algoritma Winnowing. Analisis dilakukan dengan tujuan mendapatkan informasi persentase plagiarisme dari kedua algoritma tersebut dalam penerapannya pada dokumen teks khususnya judul-judul skripsi.

### Algoritma Winnowing

*Winnowing* adalah algoritma yang digunakan untuk melakukan proses document fingerprinting (Schleimer, Wilkerson, & Aiken, 2003: 4). Document fingerprinting

merupakan metode yang digunakan untuk mendeteksi keakuratan salinan antar dokumen atau hanya sebagian teks saja. Prinsip kerja dari metode *document fingerprinting* ini adalah dengan menggunakan teknik *hashing*. Teknik *hashing* adalah sebuah fungsi yang mengkonversi setiap string menjadi bilangan.

Algoritma *winnowing* melakukan penghitungan nilai-nilai *hash* dari setiap *k-gram*, untuk mencari nilai *hash* selanjutnya digunakan fungsi *rolling hash*. Kemudian dibentuk *window* dari nilai-nilai *hash* tersebut. Dalam setiap *window* dipilih nilai *hash* minimum. Jika ada lebih dari satu *hash* dengan nilai minimum, dipilih nilai *hash* yang paling kanan. Kemudian semua nilai *hash* terpilih disimpan untuk dijadikan *fingerprint* dari suatu dokumen.

Input dari proses document fingerprinting adalah file teks. Kemudian outputnya akan berupa sekumpulan nilai *hash* yang disebut *fingerprint*. *Fingerprint* ini yang akan dijadikan dasar pembandingan kesamaan antara teks yang telah dimasukkan.

Syarat dari algoritma deteksi plagiarisme (Schleimer, Wilkerson, & Aiken, 2003: 2) yaitu;

- a. ***whitespace insensitivity***,  
yaitu pencocokan teks *file* seharusnya tidak terpengaruh oleh spasi, jenis huruf kapital, tanda baca dan sebagainya.
- b. ***noise surpression***,  
yaitu menghindari pencocokan teks *file* dengan panjang kata yang terlalu kecil atau kurang relevan dan bukan merupakan kata yang umum digunakan.
- c. ***position independence***,  
yaitu pencocokan teks *file* seharusnya tidak bergantung pada posisi kata-kata sehingga kata dengan urutan posisi berbeda masih dapat dikenali jika terjadi kesamaan.

*Winnowing* telah memenuhi syarat-syarat tersebut dengan cara membuang seluruh karakter-karakter yang tidak relevan misal: tanda baca, spasi dan juga karakter lain,

sehingga nantinya hanya karakter-karakter yang berupa huruf atau angka yang akan diproses lebih lanjut.

Langkah-langkah dalam penerapan Algoritma Winnowing adalah sebagai berikut (Sariyanti Astutik, Andharini Dwi Cahyani, Mochammad Kautsar Sophan, 2014):

**1. Langkah pertama**

**Pembuangan Karakter yang Tidak Relevan.**

Yaitu penghapusan tanda baca, spasi dan symbol-simbol seperti =, #, %, &, (, ), -, \_, \$, @, !, /, ", seperti contoh dibawah ini:

*Aplikasi Deteksi Source Code C++*

Akan dirubah menjadi

*aplikasideteksisourcecodec*

**2. Langkah kedua**

**Pembentukan Rangkaian n-gram.** Yaitu dengan cara membentuk rangkaian karakter sepanjang *n* dari hasil pembuangan karakter yang tidak relevan, dari teks diatas yang telah dibersihkan dengan ukuran *k*, ukuran *k* = 7 (gram terbaik pada penelitian sebelumnya)

aplikas plikasi likasid ikaside kasidet asidete sidetek ideteks deteksi eteksis teksiso eksisou ksisour sisourc isource sourcec ourceco urcecod rcecode cecodec

**3. Langkah ketiga**

**Perhitungan Fungsi Hash untuk tiap n-gram.**

Yaitu melakukan perhitungan-perhitungan nilai-nilai *hash* dari setiap *gram*, fungsi yang digunakan untuk menghasilkan *nilai hash* dari rangkaian *gram* dalam algoritma Winnowing adalah *rolling hash*.

*Rolling Hash* adalah suatu cara untuk mentransformasi sebuah *string* menjadi suatu nilai yang unik dengan panjang tertentu (*fixed-length*) yang berfungsi sebagai penanda *string* tersebut. Fungsi

untuk menghasilkan nilai ini disebut *fungsi hash*, sedangkan nilai yang dihasilkan disebut *nilai hash*.

*Fungsi hash*  $H(c_{1..k})$  didefinisikan sebagai berikut:

$$H_{(ck)} = c_1 * b^{(k-1)} + c_2 * b^{(k-2)} + ..... + c_k * b^{(k-k)}$$

Keterangan :

*c* = nilai ascii karakter

*b* = basis (bilangan prima)

*k* = banyak karakter

hasil *rolling hash* dari kalimat diatas adalah

26194 27766 27060 26674 26700  
 26210 27802 26394 26118 26866  
 28098 26734 27840 28486 27734  
 28786 28354 28492 27234 25482

**4. Langkah keempat**

**Pembentukan Window dari Nilai Hash.**

Pembentukan *nilai hash* dari window dengan ukuran *w* = 9 yaitu :

W-1 : {26194 27766 27060 26674 26700 26210 27802 26394 26118} W-2 : {27766 27060 26674 26700 26210 27802 26394 26118 26866} W-3 : {27060 26674 26700 26210 27802 26394 26118 26866 28098} W-4 : {26674 26700 26210 27802 26394 26118 26866 28098 26734} W-5 : {26700 26210 27802 26394 26118 26866 28098 26734 27840} W-6 : {26210 27802 26394 26118 26866 28098 26734 27840 28486} W-7 : {27802 26394 26118 26866 28098 26734 27840 28486 27734} W-8 : {26394 26118 26866 28098 26734 27840 28486 27734 28786} W-9 : {26118 26866 28098 26734 27840 28486 27734 28786 28354} W-10 : {26866 28098 26734 27840 28486 27734 28786 28354 28492} W-11 : {28098 26734 27840 28486 27734 28786 28354 28492 27234} W-12 : {26734 27840 28486 27734 28786 28354 28492 27234 25482}

**5. Langkah kelima  
Pemilihan Fingerprint dari Setiap Window**

Langkah terakhir yaitu memilih nilai terkecil dari setiap window untuk dijadikan *fingerprint*, hasil dari *nilai fingerprint*nya sebagai berikut:

26118 26118 26118 26118 26118  
26118 26118 26118 26118 26734  
26734 25482

**6. Persamaan Jaccard Coefficient.**

*Nilai fingerprint* yang dibentuk dari algoritma winnowing digunakan untuk mengukur prosentase kemiripan teks pada persamaan *Jaccard Coefficient*.

Persamaan *Jaccard Coefficient* digunakan untuk menghitung kemiripan (*similarity*) dari kumpulan kata-kata yang telah dihitung nilai *hash* nya. Berikut ini rumus persamaan *Jaccard Coefficient*

$$\text{Similarity} = \frac{\text{Jumlah\_fingerprint\_sama}}{\text{Total seluruh fingerprint}} \times 100$$

**Algoritma Rabin Karp**

Algoritma Rabin-Karp diciptakan oleh Michael O. Rabin dan Richard. Karp pada tahun 1987 yang menggunakan fungsi hashing untuk menemukan pattern di dalam string teks.

Karakteristik Algoritma Rabin-Karp :

1. Menggunakan sebuah fungsi hashing
2. Fase preprocessing menggunakan kompleksitas waktu  $O(m)$
3. Untuk fase pencarian kompleksitasnya :  $O(mn)$
4. Waktu yang diperlukan  $O(n+m)$

Algoritma Rabin-Karp adalah algoritma pencocokan string yang menggunakan fungsi hash sebagai pembanding antara string yang dicari ( $m$ ) dengan substring pada teks ( $n$ ). Apabila hash value keduanya sama maka akan dilakukan perbandingan sekali lagi terhadap karakter-karakternya. Apabila hasil keduanya

tidak sama, maka substring akan bergeser ke kanan. Pergeseran dilakukan sebanyak ( $n-m$ ) kali. Perhitungan nilai hash yang efisien pada saat pergeseran akan mempengaruhi performa dari algoritma ini

Langkah-langkah dalam algoritma Rabin Karp:

1. Menghilangkan tanda baca dan mengubah ke teks sumber dan kata yang ingin dicari menjadi kata-kata tanpa huruf kapital.

Contoh:

Aplikasi Deteksi Source Code C++

Akan dirubah menjadi

aplikasi deteksi source code c

2. Membagi teks ke dalam gram-gram yang ditentukan nilai k-gram nya. Misalkan ukuran  $k = 7$  maka hasilnya seperti dibawah ini:

{aplikas} {plikasi} {likasi} {ikasi d}  
{kasi de} {asi det} {si dete} {i detek} {  
deteks} {deteksi} {eteksi} {teksis} {eksi  
so} {ksi sou} {si sour} {i sourc} { source}  
{source} {ource c} {urce co} {rce cod}  
{ce code} {e code} { code c}

3. Mencari nilai hash dengan fungsi rolling hash dari tiap gram yang terbentuk.
4. Mencari nilai hash yang sama antara 2 teks.

191614241	217501169	209949739
204832761	207007517	191955506
221255080	192781322	74441702
195269355	199245837	223500051
197988836	209673042	221276279
193014503	77006679	223484029
217299753	227221413	219436636
192265603	185691736	74404924

- Menentukan persamaan 2 buah teks dengan persamaan Dice's Similarity Coefficient.

Untuk menghitung kemiripan dapat digunakan *Dice Similarity Coefficients*. dengan cara menghitung jumlah K-Gram yang digunakan pada kedua teks nanti yang diuji. Nilai kemiripan tersebut dapat dihitung dengan rumus :

$$S = \frac{2C}{A + B}$$

Dimana:

**S** = nilai kemiripan  
**C** = jumlah K-Gram yang sama  
**A** serta **B** = jumlah K-Gram dari masing-masing string

## ANALISIS HASIL DAN PEMBAHASAN

### Analisis Algoritma Wnnowing dengan Algoritma Rabin Karp

Dalam penelitian ini akan dilakukan analisis performansi dari dua buah algoritma yang termasuk dalam metode fingerprinting yaitu Algoritma Wnnowing dan Algoritma Rabin Karp yang dapat diterapkan untuk mendeteksi similaritas pada dokumen teks yakni pada judul-judul skripsi.

Analisis performansi dilakukan dengan tujuan mendapatkan informasi dari kedua algoritma yaitu berupa informasi kecepatan, ketepatan, dan jumlah langkah dari Algoritma Wnnowing dan Algoritma Rabin Karp dalam penerapannya pada sebuah sistem.

### Analisis Input dan Output

Analisis Input dan Output sistem menjelaskan tentang bagaimana sistem menerima inputan dan menghasilkan output dari proses yang telah dilakukan. Berikut ini merupakan analisis input dan output dari algoritma yang akan dibuat.

### Analisis Input dan Output Algoritma Wnnowing

Algoritma Wnnowing memiliki input dan output data sebagai berikut.

### Analisis Input

Pada penerapan Algoritma Wnnowing dalam sebuah sistem membutuhkan beberapa inputan meliputi:

- Dokumen Teks berupa judul-judul skripsi yang berektensi .txt
- Jumlah Karakter Minimal 100 Huruf
- Jumlah Gram akan dimulai dengan gram terkecil
- Jumlah Window akan dimulai dengan window terkecil

### Analisis Output

Output yang dihasilkan dari sistem berupa:

- Persentasi Kemiripan dari judul teks yang dibandingkan
- Informasi Waktu Proses

### Analisis Input dan Output Algoritma Rabin Karp

Algoritma Rabin Karp memiliki input dan output data sebagai berikut:

### Analisis Input

Pada penerapan Algoritma Rabin Karp dalam sebuah sistem membutuhkan beberapa inputan meliputi:

- Dokumen Teks berektensi .txt
- Jumlah Karakter Minimal 100 Huruf
- Jumlah Gram yang dimulai dengan nilai terkecil

### Analisis Output

Output yang dihasilkan dari sistem berupa:

- Persentasi Kemiripan dari dokumen yang dibandingkan
- Informasi Waktu Proses

### Implementasi Perbandingan Algoritma

Untuk dapat mengetahui algoritma mana yang terbaik dalam mendeteksi kemiripan teks dokumen khususnya kemiripan teks judul-judul skripsi yang diajukan oleh mahasiswa dengan menerapkan proses algoritma terbaik untuk menghasilkan presentasi kemiripan dan waktu proses. Tujuan utama dari deteksi kemiripan judul adalah untuk menentukan

presentase kemiripan judul satu dengan judul lainya dengan waktu proses yang lebih cepat.



Skema deteksi Kemiripan judul skripsi

Langkah-langkah deteksi kemiripan judul skripsi dapat digambarkan sebagai berikut:

1. Masukan judul skripsi pada masing-masing algoritma yaitu algoritma winnowing dan algoritma Rabin Karp nantinya akan dilihat tingkat presentase dengan judul skripsi yang sudah ada sebelumnya.
2. Memasukan nilai  $n$ -gram, untuk memebentuk rangkaian gram pada judul yang dimasukan dan judul yang dibandingkan.
3. Masukan nilai Window hanya pada algoritma Winnowing, untuk menentukan pembentukan window dari nilai Hash.
4. Deteksi kemiripan judul skripsi akan diproses dari masing-masing algoritma, dari proses deteksi kemiripan judul tersebut akan menampilkan tingkat presentase dengan judul-judul yang sudah ada sebelumnya.

### Implementasi Perbandingan Algoritma

Tahapan-tahapan untuk mengimplementasikan perbandingan algoritma winnowing dengan algoritma rabin karp untuk mendeteksi presentase plagiarisme pada judul skripsi adalah sebagai berikut;

1. **Ambil 2 buah teks judul skripsi yang akan dibandingkan.**

Teks Judul 1:

“SISTEM PENDUKUNG KEPUTUSAN PENENTUAN KARYAWAN BENGKEL BERPRESTASI MENGGUNAKAN METODE SAW”

Teks Judul 2:

“SISTEM PENDUKUNG KEPUTUSAN BEASISWA PADA MAHASISWA BERPRESTASI MENGGUNAKAN METODE AHP”

### 2. Pembuangan Karakter yang Tidak Relevan

Hapus semua huruf yang bukan A-Z, a-z, 0-9. Kemudian ubah menjadi huruf kecil semua.

Teks Judul 1 akan terbentuk menjadi seperti berikut:

Sistem pendukung keputusan penentuan karyawan bengkel berprestasi menggunakan metodesaw

dan Teks Judul 2 akan terbentuk menjadi seperti berikut:

sistem pendukung keputusan beasiswa pada mahasiswa berprestasi menggunakan me tode ahp

### 3. Pembentukan Rangkaian $n$ -gram

pembentukan  $n$ -gram untuk masing-masing teks judu dan jumlah data pengelompokan  $n$ -gram ini dimulai dari 2, 3, 4, 5, 6, 7, 8, 9, 10 pada masing-masing algoritma.

### 4. Pengujian algoritma Winnowing dan algoritma Rabin Karp

Selanjutnya tahap ujicoba dengan membandingkan kedua algoritma sesuai dengan langkah-langkah yang dimiliki oleh algoritma winnowing dan algoritma Rabin Karp dalam menentukan hasil presentase kemiripan teks dokumen khususnya pada judul-judul skripsi.

**Pendekatan Algoritma winnowing**

berikut tabel serangkaian hasil ujicoba yang telah dilakukan berdasarkan masing-masing algoritma.

**ALGORITMA WINNOWING**

uji	n-gram (n)	window (w)	waktu proses (second)	kemiripan %
1	2	3	0.0247	73.3
	2	5	0.0249	71.4
	<b>2</b>	<b>7</b>	<b>0.0250</b>	<b>66.6</b>
	2	9	0.0253	66.6
2	3	3	0.0281	57.5
	3	5	0.0303	55.1
	<b>3</b>	<b>7</b>	<b>0.0260</b>	<b>50</b>
3	4	3	0.0253	48.8
	<b>4</b>	<b>5</b>	<b>0.0257</b>	<b>50</b>
	4	7	0.0270	45.5
4	4	9	0.0261	56.2
	5	3	0.0255	44.6
	5	5	0.0282	45.1
5	<b>5</b>	<b>7</b>	<b>0.0311</b>	<b>42.8</b>
	5	9	0.0275	43.7
	<b>6</b>	<b>3</b>	<b>0.0268</b>	<b>40.4</b>
6	6	5	0.0263	43.3
	6	7	0.0270	42.8
	6	9	0.0263	43.7
7	<b>7</b>	<b>3</b>	<b>0.0252</b>	<b>39.5</b>
	7	5	0.0294	44.8
	7	7	0.0260	40
8	7	9	0.0305	43.7
	<b>8</b>	<b>3</b>	<b>0.0265</b>	<b>36.7</b>
	8	5	0.0325	40
9	8	7	0.0289	40
	8	9	0.0258	43.7
	<b>9</b>	<b>3</b>	<b>0.0257</b>	<b>32.6</b>
10	9	5	0.0263	33.3
	9	7	0.0255	33.3
	9	9	0.0263	35.2
11	<b>10</b>	<b>3</b>	<b>0.0263</b>	<b>32.6</b>
	10	5	0.0263	33.3
	10	7	0.0274	33.3
12	10	9	0.0308	35.2

Berdasarkan dari tabel diatas dapat dilihat bahwa algoritma Winnowing melakukan ujicoba sebanyak 9 kali dengan n-gram dimulai dengan n=2 sampai batas akhir n-gram= 10 kemudian dapat dilihat waktu proses yang berbeda-beda tiap n-gram dan untuk window mulai dari 3, 5, 7, dan 9 window. Dari hasil ujicoba maka dapat ditentukan tiap-tiap n-gram dan window terbaiknya.

Tabel Hasil window dan kemiripan terbaik

**ALGORITMA WINNOWING**

Jumlah Ujicoba	n-gram (n)	window (w)	waktu proses (second)	kemiripan %
1	2	7	0.0250	66.6
2	3	7	0.0260	50
3	4	5	0.0257	50
4	5	7	0.0311	42.8
5	6	3	0.0268	40.4
6	7	3	0.0252	39.5
7	8	3	0.0265	36.7
8	9	3	0.0257	32.6
9	10	3	0.0263	32.6

**Pendekatan Algoritma Rabin Karp**

Setelah melakukan ujicoba terhadap nilai n-gram mulai dari 2 sampai dengan 10 maka dapat dilihat hasil pendekatan algoritma Rabin Karp dalam menentukan jumlah n-gram terbaiknya.

**ALGORITMA RABIN KARP**

Jumlah Ujicoba	n-gram (n)	waktu proses (second)	kemiripan %
1	2	0.0256	84.03
2	3	0.0264	71.52
3	4	0.0245	67.90
4	5	0.0254	64.20
5	6	0.0262	60.87
6	7	0.0253	58.75
7	8	0.0260	56.60
8	9	0.0255	54.78
9	10	0.0280	52.90

Berdasarkan dari ujicoba yang telah dilakukan terhadap masing-masing algoritma maka dapat dilihat hasil pengujian seperti tabel dibawah ini:

NO	ALGORITMA WINNOWER				ALGORITMA RABIN KARP			
	n-gram (n)	window (w)	waktu proses (second)	kemiripan %	n-gram (n)	waktu proses (second)	kemiripan %	
1	2	7	0.0250	66.6	2	0.0256	84.03	
2	3	7	0.0260	50	3	0.0264	71.52	
3	4	5	0.0257	50	4	0.0245	67.90	
4	5	7	0.0311	42.8	5	0.0254	64.20	
5	6	3	0.0268	40.4	6	0.0262	60.87	
6	7	3	0.0252	39.5	7	0.0253	58.75	
7	8	3	0.0265	36.7	8	0.0260	56.60	
8	9	3	0.0257	32.6	9	0.0255	54.78	
9	10	3	0.0263	32.6	10	0.0280	52.90	

Dari hasil perbandingan pendekatan algoritma winnowing dengan algoritma rabin karp dapat dilihat kemungkinan kemiripan teks judul skripsi yang terkecil adalah dengan menggunakan pendekatan algoritma winnowing yaitu pada ujicoba ke 8 dengan nilai n-gram = 9 dan window = 3 , proses waktu 0.0257 dengan tingkat kemiripan terkecil yaitu 32.6 %

**PENUTUP**

**Kesimpulan**

Berdasarkan hasil pengujian yang telah dilakukan pada masing-masing algoritma yaitu algoritma winnowing dengan algoritma rabin karp terhadap teks judul skripsi, maka dapat ditarik kesimpulan sebagai berikut;

1. Pendekatan Algoritma Winnowing lebih baik daripada pendekatan algoritma Rabin Karp karena menghasilkan tingkat presentase yang lebih kecil dan waktu proses yang lebih cepat.
2. Untuk algoritma winnowing memiliki inputan yaitu dokumen teks berekstensi .txt, jumlah karakter, jumlah Gram dan jumlah Window, sedangkan output yang akan dihasilkan berupa Persentase kemiripan dari teks dokumen yang dibandingkan.
3. Untuk algoritma rabin karp memiliki inputan yaitu dokumen teks berekstensi .txt, jumlah karakter dan jumlah Gram, sedangkan output yang akan dihasilkan berupa Persentase kemiripan dari teks dokumen yang dibandingkan.
4. Terdapat perbedaan mendasar dari kedua algoritma yaitu pada algoritma winnowing memiliki window sedangkan algoritma rabin karp hanya memiliki k-gram

5. Berdasarkan hasil Pengujian terhadap perbandingan pendekatan algoritma winnowing dengan algoritma rabin karp dapat dilihat kemungkinan kemiripan teks judul skripsi yang terkecil adalah dengan menggunakan pendekatan algoritma winnowing yaitu pada ujicoba ke 8 dengan nilai n-gram = 9 dan window = 3 , proses waktu 0.0257 dengan tingkat kemiripan terkecil yaitu 32.6 %

**Saran**

Berdasarkan dari hasil perbandingan dengan pendekatan algoritma winnowing dengan pendekatan algoritma rabin karp pada deteksi plagiarisme kemiripan teks Judul skripsi masih memiliki kekurangan dan kelemahan.

Oleh Karena itu perlu perbaikan dan perkembangan untuk penelitian selanjutnya dengan saran sebagai beriku:

1. Untuk mendeteksi plagiarisme atau kemiripan teks dokumen dengan metode *fingerprinting* dapat dilakukan dengan menambah perbandingan algoritma lainnya yang berkaitan pada bidang text mining misal perbandingan Algoritma Rabin Karp, Algoritma Manber dan Algoritma Winnowing.
2. Dapat dikembangkan dengan mendeteksi dokumen teks lengkap jadi tidak hanya teks judul skripsi saja.
3. Dapat dibuatkan aplikasi deteksi plagiarisme secara online yang sudah memiliki database judul-judul skripsi mahasiswa yang telah lulus, kemudian mahasiswa yang ingin mengambil judul bisa langsung cek di aplikasi tersebut sebelum mengusulkan judulnya.

**DAFTAR PUSTAKA**

Analisis String Matching Pada Judul Skripsi dengan Algoritma Knuth-Morris Pratt(KMP)Agustus 2017*ILKOM Jurnal Ilmiah Volume 9 Nomor 2 ISSN*

*Cetak 2087-1716 ISSN Online 2548-7779*168 - 172

*(InfoTekJar) Vol 2, No 1 e- ISSN : 2540-7600p-ISSN : 2540-7597*

Aplikasi Pendeteksi Duplikasi Dokumen Teks Bahasa Indonesia Menggunakan Algoritma Winnowing dengan metode K-gram dan Synonym Recognition. 2011 *Jurnal Tugas Akhir Jurusan Teknik Informatika Universitas Muhammadiyah Malang*

Pencocokan Pola Majemuk Dengan Algoritma Rabin Karp 2015 *Makalah IF2211 Startegi Algoritma- Sem.II* Halaman 1 - 6

Christian 2013 Perancangan Sistem Deteksi Plagiarisme Dokumen Teks Menggunakan Algoritma Damerau Levenshtein Distance *Jurnal Mahasiswa PTIIK UB Vol 1*

Penerapan algoritma Boyes Moore Pada Aplikasi Pengajuan Judul Skripsi Berbasis Web 2014 *Seminar Informasi dan Teknologi Ilmiah (INTI)*

*Plagiarisme , Pelanggaran Hak cipta dan Etika* 2011 Jakarta Kanisius

Implementasi Algoritma Winnowing Untuk Deteksi Kemiripan Judul Skripsi Studi Kasus STIMIK Budidarma Januari 2017 *Majalah Ilmiah INTI Volume: XII, Nomor : 1 ISSN : 2339-210X*

2013 *Rancang Bangun Aplikasi Pendeteksi Penjiplakan Dokumen Menggunakan Algoritma Biword Winnowing Pekanbaru Riau UNIVERSITAS ISLAM NEGERI SULTSN SYARIF KASIM*

Implementasi Algoritma Winnowing Untuk Deteksi Kemiripan Text Maret 2015 *Pelita Informatika Budi Darma, Volume ; IX, Nomor; 1 ISSN 2301-9425*134-138

Sistem Penilaian Esai Otomatis pada E-Learning dengan Algoritma Winnowing November 2014 *Jurnal Informatika, Vol 12, No.2 ISSN 1411-0105*47-52

Implementasi Metode Rabin Karp Untuk Mendeteksi Tingkat Kesamaan Dua Dokumen 2013 *Tugas Akhir Fakultas Sains dan Teknologi Universitas Islam negeri Sultan Syarif Kasim Riau*

Imuasih 2013 Implementasi Algoritma Rabin Karp Untuk Pendeteksian Plagiat Dokumen Teks Menggunakan Konsep Similarity *Seminar Nasional Aplikasi Teknologi Informasi (SNATI) ISSN : 1907 - 5022*F-23 - F28

Nurdin September 2017 Sistem Pendeteksian Kemiripan Judul Skripsi Menggunakan Algoritma Winnowing *Jurnal Nasional Informatika dan Teknologi Jaringan*