

IMPLEMENTASI ALGORITMA WINNOWER UNTUK MENDETEKSI KEMIRIPAN TEKS PADA ARTIKEL

IMPLEMENTATION OF WINNOWER ALGORITHM TO DETECT TEXT SIMILARITY ON ARTICLE

Rizky Maulana Fauzi¹, Julian Chandra Wibawa²

^{1,2} Program Studi Sistem Informasi, Fakultas Teknik dan Ilmu Komputer, Universitas Komputer Indonesia
Email : rizky_mfz@emil.unikom.ac.id, maeztro_87@yahoo.co.id²

Abstrak - Plagiarisme (penjiplakan) sangat marak terjadi di era modern saat ini, dimana plagiarisme mudah dilakukan karena semakin majunya teknologi digital seperti pembuatan tugas akhir ilmiah di bidang akademik baik tingkat sekolah maupun perguruan tinggi. Algoritma winnower adalah salah satu algoritma yang dapat mendeteksi kemiripan teks dengan menggunakan metode *fingerprint*. Tujuan dari penelitian yang dilakukan ini adalah mengimplementasikan algoritma winnower untuk mendeteksi kemiripan teks pada artikel tugas akhir mahasiswa pada program studi manajemen informatika Universitas Komputer Indonesia, dan penelitian ini menggunakan metode deskriptif. Lalu hasil dari penelitian ini akan berupa suatu aplikasi yang dibangun dengan bahasa pemrograman PHP dan Mysql, dan keluaran dari aplikasi ini yaitu berupa persentase perbandingan artikel yang dibandingkan dengan artikel lain yang disimpan dalam sebuah basis data.

Kata kunci : plagiarisme, algoritma winnower, artikel.

Abstract - Plagiarism (plagiarism) is often happening in modern era today, where plagiarism easy to do because of progressively digital technology such as making scientific final project in the field of academics both school and college level. Winnower algorithm is one algorithm that can detect text similarity by using fingerprint method. Purpose of this research applies winnower algorithm to detect text similarity in articles of student's final assignment at Informatics management program of Indonesia Computer University. The results of this research is an application designed using PHP and Mysql programming languages, and the output of this application is a percentage of abstract comparison which compared with other abstracts stored in a database.

Keyword : plagiarism, algorithm winnower, articles.

I. PENDAHULUAN

Program Studi Manajemen Informatika di Universitas Komputer Indonesia Bandung merupakan salah satu program studi di dalam lingkungan Fakultas Teknik dan Ilmu Komputer. Prodi manajemen informatika telah memiliki berbagai fasilitas online untuk mahasiswa salah satunya yaitu perwalian online, nilai online, pengajuan proposal skripsi atau tugas akhir secara online (APTAMI), dan berbagai fasilitas online lainnya.

Permasalahan terjadinya plagiarisme pada bidang akademik khususnya perguruan tinggi disebabkan oleh kebiasaan malas dan tidak kreatif yang ingin cepat dalam menyelesaikan tugas-tugasnya dan masalah ini pun kerap terjadi di prodi manajemen informatika UNIKOM. Hal semacam itu akan dapat menimbulkan masalah terhadap hasil belajar mahasiswa. Saat ini telah tersedia beberapa *software* untuk deteksi plagiarisme seperti *WCOPYFIND*, *Turnitin* dan lain sebagainya, salah satu *software* yang digunakan Prodi manajemen informatika UNIKOM yaitu *turnitin*, akan tetapi untuk penggunaan *software* tersebut memakan biaya yang sangat mahal dan tidak terbuka atau bebas digunakan untuk mahasiswa. Masalah lainnya yaitu pada saat mahasiswa hendak mengajukan proposal tugas akhirnya, tidak sedikit mahasiswa yang bingung dalam pengajuan tersebut dikarenakan mahasiswa tersebut tidak mengetahui bahwa proposal yang akan diajukan itu telah ada atau tidak sebelumnya, judul yang sama tidak menutup kemungkinan isi dari judul tersebut juga sama, akan tetapi isi juga dapat berbeda walaupun dengan judul yang sama. Dan jika isi itu sama, mahasiswa akan diduga melakukan plagiat dan akan memungkinkan bahwa tugas akhir yang diajukan tersebut harus diubah dikemudian hari.

Untuk mengatasi masalah tersebut, diperlukan suatu fitur yang dapat menentukan kemiripan teks, fitur tersebut akan menentukan kemiripan proposal tugas akhir berdasarkan abstrak, karena abstrak adalah ringkasan dari seluruh isi tugas akhir, dan fitur yang akan dibuat akan menghasilkan sebuah persentase perbandingan kemiripan abstrak dengan abstrak yang telah ada sebelumnya. Deskripsi singkat diatas menjadi dasar penelitian bagi penulis untuk membuat suatu aplikasi pendeteksian plagiarisme dengan penerapan algoritma untuk menentukan kemiripan teks, dan salah satu algoritma untuk menentukan kemiripan teks yaitu algoritma *Winnower*.

Pada penelitian terdahulu yang dilakukan oleh Bambang Imam Hermawan dengan judul “ **Analisis Performansi Algoritma Winnower dan Algoritma Manber Untuk Mendeteksi Kesamaan Dokumen Teks Bahasa Indonesia**” bertujuan untuk mendapatkan informasi dari Algoritma Winnower dan Algoritma Manber berupa informasi

perbandingan kecepatan, ketepatan dan pengaruh jumlah parameter input dalam proses mendeteksi kesamaan dokumen. Terdapat kesimpulan dari penelitian ini yaitu bahwa “dalam hal kecepatan Manber lebih baik dari Winnowing karena membutuhkan proses yang lebih sedikit, akan tetapi dalam hal ketepatan Winnowing lebih baik dari Manber karena Winnowing memilih nilai *fingerprints* dari nilai terkecil setiap *window* yang dibuat, sehingga pemilihan *fingerprints* dokumen diperoleh dari banyak perbandingan dari nilai *window-window* yang beragam”[1]. Penelitian tersebut dijadikan alasan atau acuan oleh penulis dalam melakukan penelitian yang akan dilakukan.

perbedaannya adalah penelitian yang dilakukan oleh Bambang Imam Hermawan yaitu menganalisis performa dari algoritma winnowing dan algoritma manber dalam mendeteksi kemiripan dokumen teks, sedangkan penelitian yang dilakukan penulis yaitu merupakan penerapan algoritma winnowing untuk mengukur kemiripan dokumen teks pada abstrak.

Sedangkan Pada penelitian terdahulu yang dilakukan oleh Jarwati dan yang lainnya dengan judul “ **Penerapan Algoritma Winnowing Pada Sistem Rekomendasi untuk menentukan Dosen Pembimbing Skripsi (Studi Kasus Program Studi Sistem Informasi)** ” yang bertujuan untuk memilih dosen pembimbing skripsi yang sesuai dengan kompetensi, beban kerja dan topik yang dikerjakan oleh mahasiswa. terdapat pula kesimpulan pada penelitian ini yaitu “ implementasi algoritma winnowing digunakan untuk merekomendasi penentuan dosen pembimbing untuk memberikan perhitungan kesamaan antara topik penelitian mahasiswa dengan penelitian yang pernah dilakukan dosen dengan membandingkan kata kunci pada setiap penelitian dan topik “[2]

perbedaannya adalah penelitian yang dilakukan penulis yaitu merupakan penerapan algoritma *winnowing* untuk mengukur kemiripan dokumen teks pada abstrak, Dan menggunakan metode pengembangan *prototype*. Sedangkan penelitian yang dilakukan oleh Jarwati merupakan penerapan algoritma *winnowing* pada sistem rekomendasi untuk menentukan pemilihan dosen pembimbing skripsi, dan metode pengembangannya yaitu metode pengembangan *waterfall*

Maksud dari penelitian ini yaitu membuat sebuah sistem untuk mengukur tingkat kemiripan dokumen teks dengan menerapkan algoritma winnowing. Tujuan yang akan dicapai didalam penelitian ini adalah mengetahui bagaimana sistem untuk mengukur tingkat plagiarisme yang berjalan di prodi manajemen Informatika UNIKOM, menghasilkan perancangan sistem untuk mengukur tingkat plagiarisme dengan menerapkan algoritma winnowing, mengetahui bagaimana pengujian dari sistem untuk mengukur tingkat plagiarisme dengan menerapkan algoritma winnowing yang nantinya akan diterapkan di prodi manajemen informatika UNIKOM, mengetahui bagaimana implementasi sistem untuk mengukur tingkat plagiarisme dengan menerapkan algoritma winnowing.

II. KAJIAN PUSTAKA

2.1 Dokumen *fingerprint*

adalah metode yang dilakukan untuk mendeteksi kemiripan salinan antar dokumen, baik seluruh teks dari dokumen maupun sebagian teks. Metode *fingerprinting* ini menggunakan teknik *hash*, *hashing* merupakan fungsi yang mengkonversi tiap string menjadi bilangan. Algoritma yang menggunakan metode ini diantaranya *Rabin-karp*, *Winnowing*, dan *Manber*[1].

2.2 Pengertian Algoritma

Definisi algoritma itu sendiri adalah logika, metode dan tahap (urutan) atau langkah yang digunakan untuk memecahkan suatu masalah[3]

2.3 Algoritma *Winnowing*

merupakan algoritma dokumen *fingerprint* untuk melakukan proses mengecek kesamaan kata yang menghasilkan output berupa kumpulan nilai *hash* yang terbentuk dari perhitungan ASCII setiap karakter. Kumpulan *hash* itulah yang disebut *fingerprint* yang digunakan dalam mengecek penjiplakan[4]. Adapun langkah – langkah dari algoritma winnowing yaitu : membuang karakter yang tidak relevan, membentuk rangkaian *n-gram*, menghitung *hash* dari tiap gram (*rolling hash*), pembentukan *window*, pemilihan *fingerprint hash* terkecil dari tiap window, hitung persamaan dengan *Jaccard Coeficient*.

2.3.1 Membuang karakter yang tidak relevan

Pada tahap ini proses yang dilakukan yaitu menghapus tanda baca, simbol – simbol, spasi dan *case-sensitive*. Dengan contoh teks 1 = @”Abdul Malik”!! sehingga teks 1 = abdulmalik

2.3.2 N-Gram

Rangkaian n-gram pada winnowing dilakukan dengan merangkai panjang karakter = n. Rangkaian pertama dimulai dari karakter ke-1 sampai ke-n lalu rangkaian kedua dimulai dari karakter ke-2 sampai ke-n+1 dan seterusnya[4]. Setelah rangkaian terbentuk, rangkaian tersebut akan di proses menjadi rangkaian hash dengan proses *rolling hash*

2.3.3 *Rolling hash*

Menghitung nilai hash dengan ASCII tiap karakter berdasarkan rangkaian gram, adapun rumus dari *rolling hash* adalah sebagai berikut :

$$c_1 * b^{k-1} + c_2 * b^{k-2} * \dots + c_{k-1} * b + c_k$$

Keterangan : C = nilai ASCII karakter, b = bilangan prima, dan k = nilai gram

2.3.4 *Jaccard Coeficient Similarity*

mengukur kemiripan, ketidakmiripan, dan jarak dari dataset. Mengukur antara dua dataset adalah hasil dari pembagian antara jumlah data yang sama dari kedua dataset dibagi dengan jumlah semua data pada dataset[4]. Algoritma winnowing menggunakan persamaan *jaccard* untuk menghitung persamaan dari *fingerpint* yang dibentuk berdasarkan window. Berikut adalah rumus dari *jaccard* coefficient :

$$\text{Similarity}(d_1, d_2) = \frac{\sum H(d_1 \cap d_2)}{\sum H(d_1 \cup d_2)} \times 100\%$$

Keterangan : d1 = teks 1 d2 = teks 2 dan H = nilai hash

2.4 Pengertian plagiarisme

Plagiarisme adalah tindakan mengambil ide, menjiplak karya atau tulisan orang lain untuk dijadikan karya tulisan sendiri tanpa mencantumkan sumber asli ataupun nama penulisnya [5]

III. METODE PENELITIAN

Penelitian yang dilakukan oleh penulis menggunakan metode deskriptif, yang mana penelitian ini yaitu bertujuan untuk mendeskripsikan atau menggambarkan suatu keadaan dari objek yang sedang diteliti. Metode pengumpulan data didapatkan dari hasil wawancara, observasi dan dokumentasi yang mana sumber data primer didapatkan melalui (observasi dan wawancara) dan sumber data sekunder (dokumentasi). Metode pendekatan dan pengembangan yang digunakan yaitu pendekatan berorientasi objek dan pengembangan *prototype*. Dan alat bantu (*tools*) yang digunakan untuk pemodelan dalam analisis dan perancangan sistem adalah *UML (Unified Modeling language)*.

IV. HASIL DAN PEMBAHASAN

Dari hasil penelitian, penulis akan menggambarkan sebuah contoh penerapan algoritma winnowing dalam memproses sebuah teks seperti dibawah ini :

Dokumen 1

Saya Mahasiswa
UNIKOM

Dokumen 2

Saya Mahasiswa
UNPAD

Kedua teks tersebut akan diproses untuk membuang karakter yang tidak relevan menjadi :

sayamahasiswaunikom

sayamahasiswaunpad

Kemudian melakukan pembentukan rangkaian *n-gram*, dimana pada contoh ini nilai *n-gram* yang diberikan yaitu n = 3. Maka teks akan berubah menjadi rangkaian seperti berikut :

say aya yam ama mah aha has asi sis isw swa
wau aun uni nik iko kom

say aya yam ama mah aha has asi sis isw swa
wau aun unpad

Selanjutnya melakukan perhitungan hash dari masing – masing karakter berdasarkan rangkaian *n-gram (rolling hash)*

$$c_1 * b^{k-1} + c_2 * b^{k-2} * \dots + c_{k-1} * b + c_k$$

c = nilai ASCII karakter, b = bilangan prima dan k = *n-gram*

$$\begin{aligned} &= s * 11^2 + a * 11^1 + y * 11^0 \\ &= 115 * 121 + 97 * 11 + 121 * 1 \\ &= 15103 \text{ (dst).} \end{aligned}$$

Maka akan didapatkan nilai hash sebagai berikut :

15103 13165 15817 13033 14360 12978 13766 13107 15185 14089 15321
15583 13134 15472 14572 13993 14277

15103 13165 15817 13033 14360 12978 13766 13107 15185 14089 15321
15583 13134 15479 14639 14719

Selanjutnya yaitu pembentukan *window* dari rangkaian *hash (winnowing)* seperti pada tabel berikut :

Tabel 1 : rangkaian window

No	Dokumen 1	Dokumen 2
1	{ 15103 13165 15817 }	{ 15103 13165 15817 }
2	{ 13165 15817 13033 }	{ 13165 15817 13033 }
3	{ 15817 13033 14360 }	{ 15817 13033 14360 }
4	{ 13033 14360 12978 }	{ 13033 14360 12978 }
5	{ 14360 12978 13766 }	{ 14360 12978 13766 }
6	{ 12978 13766 13107 }	{ 12978 13766 13107 }
7	{ 13766 13107 15185 }	{ 13766 13107 15185 }
8	{ 13107 15185 14089 }	{ 13107 15185 14089 }
9	{ 15185 14089 15321 }	{ 15185 14089 15321 }
10	{ 14089 15321 15583 }	{ 14089 15321 15583 }
11	{ 15321 15583 13134 }	{ 15321 15583 13134 }
12	{ 15583 13134 15472 }	{ 15583 13134 15472 }
13	{ 13134 15472 14572 }	{ 13134 15472 14639 }
14	{ 15472 14572 13993 }	{ 15472 14639 14719 }
15	{ 14572 13993 14277 }	

Kemudian didapatkan nilai *fingerpint* sebagai berikut, yang didapatkan dari nilai *hash* terkecil dari rangkaian *window*:

Fingerprint dok 1				Fingerprint dok 2			
[13165,1]	[13033,3]	[12978,5]	[13107,7]	[13165,1]	[13033,3]	[12978,5]	[13107,7]
[14089,9]	[13134,12]	[13993,15]		[14089,9]	[13134,12]	[14639,14]	

Langkah terakhir untuk menghitung persentasi kemiripan (*similarity*) dari sebuah dokumen dilakukan menggunakan Jaccard's Similarity Coefficient dengan rumus:

$$D_{A,B} = \frac{|A \cap B|}{|A \cup B|} \times 100$$

Keterangan: A : nilai fingerprints dokumen 1
B : nilai fingerprints dokumen 2

Dimana $D(A,B)$ adalah nilai similaritas, $|A \cap B|$ adalah jumlah fingerprint yang sama dari dokumen 1 dan 2, $|A \cup B|$ adalah jumlah seluruh fingerprint dari dokumen 1 dan 2.

dimana $A = \{13165, 13033, 12978, 13107, 14089, 13134, 13993\}$

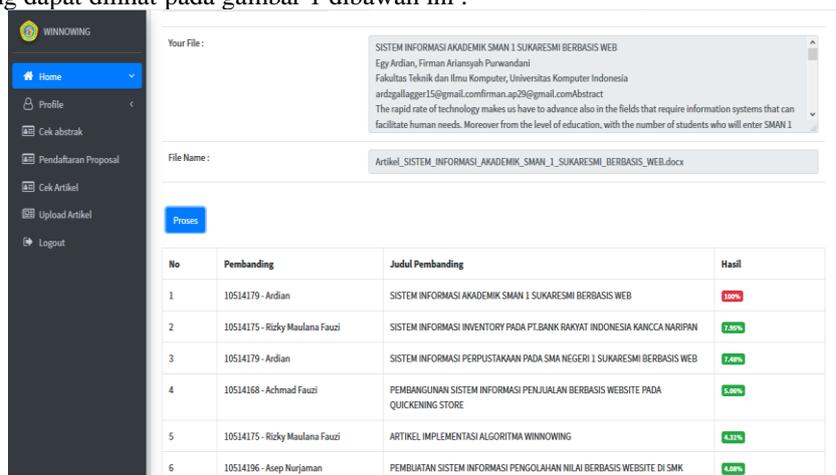
dan $B = \{13165, 13033, 12978, 13107, 14089, 13134, 14639\}$

maka $|A \cap B| = \{13165, 13033, 12978, 13107, 14089, 13134\}$

dan $|A \cup B| = \{13165, 13033, 12978, 13107, 14089, 13134, 13993, 14639\}$

sehingga, $D(a,b) = 6/8 * 100 = 75\%$

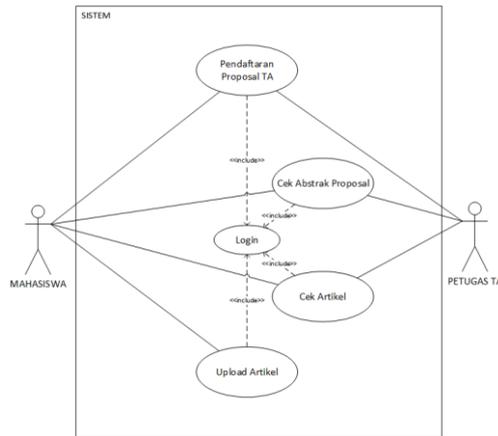
dari proses tersebut dapat dilihat bahwa teks 1 dan teks 2 memiliki persamaan sebesar 75%. Berdasarkan hasil tersebut penulis menerapkannya dalam sebuah aplikasi untuk membandingkan file artikel dengan artikel lain yang disimpan didalam basis data yang dapat dilihat pada gambar 1 dibawah ini :



Gambar 1 hasil deteksi perbandingan file artikel

A. Pendekatan Sistem yang digunakan

Pendekatan sistem yang digunakan dalam penelitian ini pendekatan berorientasi objek dengan menggunakan alat bantu *use case diagram* untuk menggambarkan sebuah prosedur sebuah sistem serta proses yang terjadi. Berikut *use case diagram* yang dapat dilihat pada gambar 2.



Gambar 2 Use case diagram

B. Implementasi perangkat lunak

Implementasi perangkat lunak yang penulis gunakan dalam pembuatan aplikasi pengecekan file abstrak pada Program studi manajemen informatika Universitas Komputer Indonesia adalah sebagai berikut :

1. *Operating System*, sistem operasi yang digunakan untuk membangun aplikasi ini adalah *Windows 10*.
2. *Sublime Text*, *software* yang penulis gunakan dalam membangun aplikasi dengan menggunakan bahasa pemrograman php.
3. *Browser*, *browser* yang digunakan penulis adalah *google chrome*, digunakan untuk menjalankan aplikasi.
4. *Database*, *database* dan *web server* yang digunakan penulis untuk membangun aplikasi ini adalah *database server XAMPP v3.2.1*

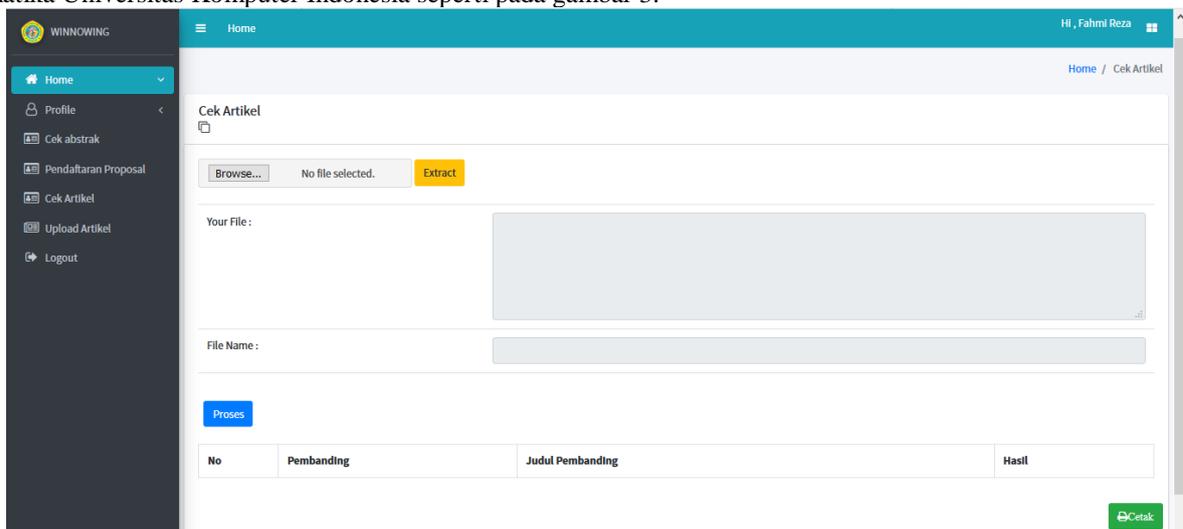
C. Implementasi perangkat keras

Implementasi perangkat keras berdasarkan (*minimum requirement*) kebutuhan minimal sistem yang dibutuhkan yaitu :

1. *Processor* Intel / AMD dan sekelasnya.
2. *Memory* 2GB atau lebih.
3. *Hardisk* 500GB atau lebih.
4. *Mouse, keyboard*
5. *Internet*

D. Implementasi antar muka

Berikut adalah implementasi antar muka dari aplikasi pengecekan file artikel pada Program studi manajemen informatika Universitas Komputer Indonesia seperti pada gambar 3.



Gambar 3 Implementasi antar muka

V.KESIMPULAN DAN SARAN

A.Kesimpulan

Berdasarkan dari hasil kegiatan penelitian yang sudah dilakukan oleh penulis, didapatkan berupa kesimpulan dan saran guna mendapatkan perbaikan dan masukan – masukan bagi penulis maupun bagi peneliti yang akan datang, dan kesimpulan yang didapatkan adalah sebagai berikut :

1. Aplikasi pendeteksian plagiat atau kemiripan teks yang telah dirancang ini dapat digunakan tanpa harus menggunakan biaya.
2. Aplikasi pendeteksian plagiat atau kemiripan teks yang telah dirancang ini dapat digunakan atau terbuka bagi mahasiswa.
3. Mahasiswa dapat menggunakan dan melihat hasil pengecekan secara langsung.
4. Mahasiswa dapat menggunakan aplikasi ini untuk mengecek artikelnya secara langsung sehingga dapat mengetahui bahwa artikel tersebut telah sesuai atau tidak dengan batas tingkat kemiripan yang telah ditentukan.

B.Saran

Sistem yang telah dibuat ini masih dapat berkembang dan masih memiliki banyak kekurangan, untuk itu penulis memiliki beberapa saran bagi pembaca maupun peneliti yang akan datang :

1. Mengembangkan aplikasi pengecekan file abstrak ini untuk dapat mendeteksi file berskala besar seperti tugas akhir dan lain sebagainya.
2. Algoritma winnowing pada saat ini belum dapat mendeteksi makna (sinonim) dari sebuah kata. Karena hal itu penulis mengharapkan penelitian yang akan datang dapat mengatasi kekurangan dari algoritma winnowing tersebut.
3. Aplikasi yang telah dibuat ini masih tidak dapat menampilkan kalimat atau kata mana yang mengandung unsur kemiripan atau plagiat. Maka dari itu penulis mengharapkan penelitian yang akan datang dapat mengatasi kekurangan dari aplikasi yang telah dibuat ini.

DAFTAR PUSTAKA

- [1] Bambang Imam Hermawan. 2015 “ *Analisis Performansi Algoritma Winnowing Dan Algoritma Manber Untuk Deteksi Kesamaan Dokumen Teks Berbahasa Indonesia* “. Teknik Informatika. Fakultas Teknik dan Ilmu Komputer. UNIKOM Bandung.
- [2] Jarwati et al., “ *Penerapan Algoritma Winnowing Pada Sistem Rekomendasi Penentuan Dosen Pembimbing Skripsi (Studi Kasus Program Studi Sistem Informasi)* “ Vol. 1, 2017.
- [3] Agus Setiawan, “ *Implementasi Algoritma Winnowing Untuk Deteksi Kemiripan Judul Skripsi Studi Kasus Stmik Budidarma* “ Vol. 12, No. 1, 2017.
- [4] Reynaldi Karisma, Khafiih Hastuti. “ *Penerapan Algoritma Winnowing Untuk Mendeteksi Kemiripan Teks Pada Tugas Akhir Mahasiswa* ” Vol. 15, No. 4, 2016.
- [5] Lako, Andreas, "Plagiarisme Akademik1," Jawa Pos Radar Semarang, Juni 2012.