

**PENDETEKSI KESAMAAN DOKUMEN PADA SISTEM INFORMASI  
PENDAFTARAN PROPOSAL SKRIPSI DENGAN PENDEKATAN  
ALGORITMA RABIN-KARP**

***DOCUMENT EQUALITY DETECTION IN INFORMATION SYSTEM OF THESIS  
PROPOSAL REGISTRATION WITH RABIN-CARP ALGORITHM APPROACH***

Pandu Pratama Putra<sup>1</sup>, Afriansyah<sup>2</sup>, Muhammad Syaifullah<sup>3</sup>  
STMIK AMIK Riau<sup>1,3</sup>, Universitas Lancang Kuning<sup>2</sup>  
[pandupratamaputra91@gmail.com](mailto:pandupratamaputra91@gmail.com)<sup>1</sup>

**ABSTRACT**

*Plagiarism is a significant problem in many areas, including universities. plagiarism is usually performed on digital content is to copy-paste of the original document. To anticipate, we need a way to analyze the technique of plagiarism. There are several approaches that can be taken, for example by using a search algorithm Rabin-Karp string because these algorithms can be used to detect plagiarism in a text document. In the testing phase, the test documents used were three documents with similarity level categories of low, medium and high. From some of the testing that has been done, this approach can find the longest quote the same between the two text documents and measure the similarity of text documents. With this system will help prevent acts of plagiarism in the thesis proposal registration so that the absence of the same thesis.*

**Keywords:** *Plagiarism, Rabin-Karp algorithm, Similarity, Document*

**ABSTRAK**

Plagiarisme adalah sebuah masalah yang cukup signifikan pada banyak bidang termasuk universitas. plagiarisme yang biasanya dilakukan terhadap konten digital adalah melakukan copy-paste terhadap dokumen asli. Untuk mengantisipasinya, dibutuhkan suatu cara yang dapat menganalisis teknik plagiarisme. Ada beberapa pendekatan yang bisa diambil, contohnya dengan mempergunakan algoritma pencarian string Rabin-Karp karena algoritma ini dapat digunakan untuk mendeteksi plagiarisme pada sebuah dokumen teks. Pada tahap pengujian, dokumen uji yang digunakan berjumlah tiga dokumen dengan kategori tingkat kemiripan rendah, sedang dan tinggi. Dari beberapa pengujian yang telah dilakukan, pendekatan ini dapat menemukan kutipan terpanjang yang sama di antara dua dokumen teks serta mengukur kemiripan dokumen teks. Dengan sistem ini akan membantu mencegah aksi plagiarisme pada pendaftaran proposal skripsi sehingga tidak adanya skripsi yang sama.

**Kata Kunci :** Plagiarisme, algoritma Rabin-Karp, Kesamaan, Dokumen

**PENDAHULUAN**

Plagiarisme menurut Kamus Besar Bahasa Indonesia (KBBI) berarti menggambar atau menulis garis-garis gambaran atau tulisan yang telah tersedia (dengan menempelkan kertas kosong pada gambar atau tulisan yang akan ditiru), mencontoh atau meniru tulisan atau pekerjaan orang lain, mencuri karangan orang lain dan mengakui sebagai karangan sendiri,

mengutip karangan orang lain tanpa seizin penulisnya (KBBI, 1997).

Kemampuan dosen di dalam melakukan kontrol dan penyeleksian masih terkendala dengan suatu database yang harus dicek dan diketahui dan dengan kemampuan daya ingat dari masing-masing dosen atau pembimbing, yang mungkin terbatas, sehingga terkadang satu dua judul terkadang lolos dan luput dari pantauan sehingga terjadi kemiripan atau bisa disebut plagiat

karena sebagian isinya hampir sama, terkadang hanya dibedakan sedikit pada judulnya saja. Pendaftaran proposal skripsi pada masing-masing institusi pendidikan harusnya dilakukan secara cermat dengan melakukan program aplikasi yang mudah untuk dioperasikan sehingga pada saat yang tepat dan cepat dapat diketahui bahwa judul tersebut sudah ada yang mirip sampai berapa persen dari masing-masing proposal skripsi yang dilakukan pengujian. Melalui pengolahan data dokumen proposal skripsi ini maka bagaimana suatu proses pendaftaran proposal skripsi yang akan diajukan ada di dalam proposal skripsi yang sudah terdokumentasi secara database dapat diketemukan, Pencarian dokumen proposal skripsi yang mirip dapat diketahui secara cepat.

Pembandingan kesamaan dokumen pada dokumen proposal skripsi ini dilakukan dengan membandingkan dokumen teks dengan seluruh dokumen teks yang tersimpan di basis data. Hasil pembandingan dengan aplikasi deteksi kesamaan dokumen ini tidak menunjukkan bahwa mutlak telah atau tidak terjadi plagiarisme antara dokumen teks yang dibandingkan. Hasil tersebut hanya memberikan informasi berupa persentase kesamaan antara dokumen teks, karena hingga saat ini belum ada definisi universal tentang persentase dalam pengambilan tulisan milik orang lain.

Aksi plagiat dalam karya tulis sangatlah mungkin terjadi, untuk itu perlu dilakukan upaya-upaya sebagai pencegahan maupun pendeteksian, sehingga dalam tulisan ini akan dibahas mengenai pendeteksian plagiarisme dari sebuah dokumen dengan membandingkannya dengan sebuah dokumen lainnya.

Permasalahan yang muncul yaitu :

1. Sering terjadinya kesamaan dokumen proposal skripsi.
2. Banyaknya judul yang berbeda namun isi dan maksud dari konsep pembahasannya sama dengan yang sudah ada sebelumnya.
3. Kemampuan dosen di dalam melakukan kontrol dan penyeleksian masih terkendala dengan kemampuan daya ingat dari masing-masing dosen atau pembimbing yang terbatas.

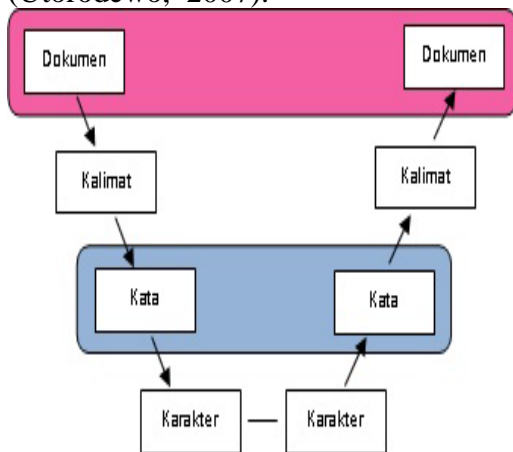
Batasan permasalahan penelitian ini adalah :

1. Aplikasi yang dibangun sebagai alat bantu untuk menemukan kesamaan didalam dokumen, sedangkan keputusan bahwasanya dokumen teks tersebut diduga sama atau tidak berada di tangan pihak dosen atau pembimbing.
2. Dokumen teks yang diidentifikasi yaitu pada latar belakang masalah, rumusan masalah, batasan masalah, usulan, langkah penyelesaian dan konsentrasi skripsi.
3. Data simulasi yang digunakan sebagai pembandingan hanya 3 dokumen proposal skripsi terdahulu yang sudah tersimpan di dalam database sistem.
4. Diterapkan pada sistem informasi pendaftaran proposal skripsi.
5. Format file yang dapat di input hanya file berekstensi doc.

### **Plagiarisme**

Plagiarisme adalah penjiplakan atau pengambilan karangan, pendapat, dan sebagainya dari orang lain dan menjadikannya seolah karangan dan pendapat sendiri (KBBI, 1997). Menurut peraturan Menteri Pendidikan Nasional RI No. 17 Tahun 2010 tentang pencegahan dan penanggulangan plagiat di perguruan tinggi. Bab 1 pasal 1 ayat 1, plagiat adalah 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 pihak lain yang diakui sebagai karya ilmiahnya, tanpa menyatakan sumber secara tepat dan memadai (PMPN RI, 2010). Yang digolongkan plagiarisme yaitu menggunakan tulisan orang lain secara mentah, tanpa memberikan tanda jelas (misalnya dengan menggunakan tanda kutip atau blok alinea yang berbeda) bahwa teks tersebut diambil persis dari tulisan lain mengambil gagasan orang lain tanpa memberikan anotasi yang cukup tentang sumbernya (Utorodewo, 2007).



**Gambar 1. Hirarki Pengukuran Kesamaan Dokumen**

#### 1. Karakter

Permasalahan yang terjadi pada pengukuran kemiripan karakter atau huruf adalah hanya dapat mengukur karakter atau huruf yang sama persis. Huruf a dan huruf A dianggap tidak sama. Pengukuran seperti ini banyak digunakan untuk mengukur keamanan password, memeriksa kata (*spell checker*) (Kurniawati et al. 2012).

#### 2. Kata

Permasalahan yang terjadi pada pengukuran kemiripan kata adalah hanya dapat mengukur kata yang sama persis. Fenomena yang sering kali terjadi adalah masyarakat mengambil tulisan orang lain dan kemudian mengganti kata dengan sinonim

katanya. Bentuk seperti ini tidak dapat diukur kemiripan dokumen apabila menggunakan pengukuran kemiripan kata karena tidak mempertimbangkan sinonim kata tersebut (Kurniawati et al. 2012).

#### 3. Kalimat

Permasalahan pengukuran kemiripan yang juga terjadi adalah mengambil penulisan orang lain dengan cara hanya mengganti subjek kalimat atau hanya mengganti predikat atau objek ataupun keterangan. Solusi yang harus dipertimbangkan untuk permasalahan di atas adalah menentukan struktur kalimat (Kurniawati et al. 2012).

#### 4. Dokumen

Permasalahan lain yang sering terjadi, selain mengganti kata dengan sinonim kata tersebut adalah posisi kalimat. Masyarakat yang mengambil penulisan orang lain, sering melakukan perubahan pada posisi kalimat. Contoh pada penulisan orang lain, kalimat yang diambil terletak pada kalimat pertama. Kemudian kalimat tersebut diletakkan pada kalimat kelima. Karena posisi berbeda maka tidak dapat diukur kemiripannya. Solusi yang harus dipertimbangkan adalah pengukuran semua kalimat pada dokumen satu terhadap semua kalimat pada dokumen kedua (Kurniawati et al. 2012).

#### **Pemrosesan Dokumen**

Dalam ilmu sistem temu kembali informasi khususnya pada algoritma pendeteksi plagiarisme dokumen teks ada beberapa istilah yang terdapat dalam hal pemrosesan dokumen, diantaranya (Rafles, 2013):

1. Preprocessing atau pembersihan teks merupakan tahapan yang dilakukan untuk mengubah data mentah menjadi data berkualitas yaitu data yang telah memenuhi persyaratan untuk dieksekusi pada sebuah algoritma. Bentuk pembersihan teks ini, dapat berupa menghilangkan

- spasi, tanda baca, simbol simbol, mengubah huruf kapital menjadi huruf kecil dan menghilangkan karakter-karakter yang tidak relevan lainnya.
2. Tokenizing merupakan tahap pemotongan kalimat menjadi kata pada sistem temu kembali informasi. Pemotongan kata ini dapat berbentuk satu kata (unigram atau uniword), dua kata (bigram atau biword), tiga kata (trigram atau triword), empat kata (quadgram atau quadword) dan seterusnya.
  3. Irisan (intersection) merupakan tahapan untuk menemukan kata bertalian yang sama di antara dua dokumen teks.

#### **Algoritma Rabin-Karp**

Algoritma Rabin Karp adalah algoritma pencarian kata yang mencari sebuah pola berupa substring dalam sebuah teks menggunakan hashing. Algoritma ini sangat efektif untuk pencocokan kata dengan pola banyak. Salah satu aplikasi praktis dari algoritma Rabin Karp adalah dalam pendeteksian plagiarisme. Dalam ilmu komputer, algoritma Rabin-Karp adalah algoritma pencarian string yang dibuat oleh Michael O. Rabin dan Richard M. Karp pada tahun 1987 yang menggunakan hashing untuk menemukan salah satu dari satu set string pola dalam teks (Salmuasih & Sunyoto, 2013).

Berikut ini adalah langkah-langkah yang perlu diperhatikan dalam pencocokan kata dengan algoritma Rabin-Karp :

1. Case folding  
Mengubah semua huruf dalam dokumen menjadi huruf kecil, hanya huruf "a" sampai dengan huruf "z" yang diterima.
2. Tokenizing  
Tahap pemotongan string input berdasarkan tiap kata yang

menyusunnya.

3. Menerapkan fungsi K-Gram  
Digunakan untuk mencari nilai similarity (kemiripan).
4. Menerapkan fungsi Hashing  
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.

#### **PHP**

PHP dikembangkan oleh Rasmus Lerdorf pada tahun 1994 yang pada awalnya mengembangkan sebuah perkakas yang digunakan sebagai engine parsing sebagai penterjemah beberapa macro. PHP (Hypertext Preprocessor) adalah bahasa pemrograman yang berjalan dalam sebuah wenserver dan berfungsi sebagai pengolah data pada sebuah server. Dengan menggunakan program PHP, sebuah website akan lebih interaktif dan dinamis (Andi, 2011).

#### **MySQL**

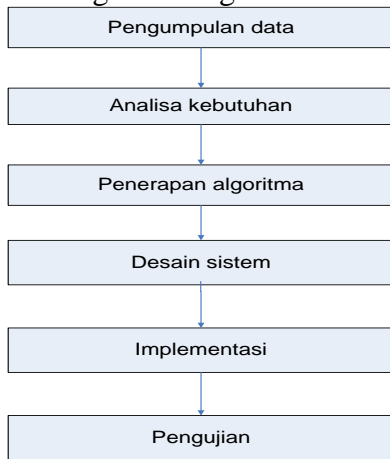
SQL (Structure Query Language) adalah suatu bahasa komputer yang mengikuti standar American National Standard Institute (ANSI), yaitu sebuah bahasa standar yang digunakan untuk mengakses dan melakukan manipulasi sistem database (Muslih & Purnama, 2013).

#### **Xampp**

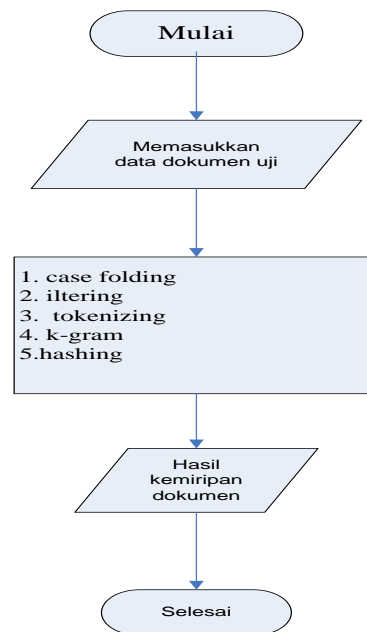
Xampp adalah perangkat lunak bebas, yang mendukung banyak sistem operasi, merupakan kompilasi beberapa program. Berfungsi sebagai server yang berdiri sendiri (localhost), yang terdiri atas program Apache, MySQL database, dan penterjemah bahasa yang ditulis dengan bahasa pemrograman PHP (Agustina, 2015)

## METODE PENELITIAN

Metode yang penulis gunakan adalah metode *waterfall*. Secara garis besar metode *waterfall* mempunyai langkah-langkah sebagai berikut :



Gambar 2. Waterfall Model



Gambar 3. Diagram Alir

## Perhitungan manual

Langkah – langkah penyelesaian :

### a. Case folding

Mengubah semua huruf dalam dokumen menjadi huruf kecil, hanya huruf “a” sampai dengan huruf “z” yang diterima kemudian menggabungkannya.

### b. Filtering

Membuang kata yang tidak penting.

**algoritma adalah kumpulan perintah.**

Contoh : kata **adalah** dibuang

Jadi: **algoritma kumpulan perintah**

### c. Tokenizing k gram dengan basis 5

Memecahkan kata menjadi potongan – potongan dimana setiap potongan mengandung karakter sebanyak k gram yang ditentukan

Contoh : **algor lgori gorit oritma ritma itma{} tma{}k ma{}ku a{}kum {}kump kumpu umpul mpula pulan ulan{} lan{}p an{}pe n{}per {}peri perin erint rinta intah**

### d. Nilai k gram kenilai hashing

Cara perhitungan hash menggunakan basis 9 :

Contoh :

Hashing=**algor**

## HASIL DAN PEMBAHASAN

### Analisa Kebutuhan

#### a. Hardware

Tabel 1. Spesifikasi Hardware

No	Hardware	Spesifikasi
1	Prosesor	Intel core i3
2	Memory	2 gb

#### b. Software

Tabel 2. Spesifikasi Software

No	Software	Spesifikasi
1	Sistem operasi	Windows 7.0
2	Database	MySQL
3	Web Server	Xampp
4	Aplikasi Editor	Sublime
5	Bahasa pemograman	PHP

### Penerapan Algoritma Rabin-Karp

Diagram alir penerapan algoritma rabin karp dapat dijelaskan seperti Gambar berikut:

$$\begin{aligned}
 &= \\
 &(a*9^4)+(1*9^3)+(g*9^2)+(o*9^1) \\
 &+(r*9^0) \\
 &=(97*9^4)+(108*9^3)+(103*9^2) \\
 &+(111*9^1)+(114*9^0) \\
 &=724605 \\
 &\text{Hashing}=\text{lgori} \\
 &= \\
 &(1*9^4)+(g*9^3)+(o*9^2)+(r*9^1) \\
 &+(i*9^0) \\
 &=(108*9^4)+(103*9^3)+(111*9^2) \\
 &+(114*9^1)+(105*9^0) \\
 &=793797
 \end{aligned}$$

Hashing=gorit

$$\begin{aligned}
 &= \\
 &(g*9^4)+(o*9^3)+(r*9^2)+(i*9^1)+(t*9^0) \\
 & \\
 &=(103*9^4)+(111*9^3)+(114*9^2)+(105*9^1) \\
 &+(116*9^0) \\
 &=766997
 \end{aligned}$$

Dst.....  
 e. Menghitung nilai kemiripan (similarity)

Menghitung kemiripan dengan menggunakan rumus Dice's similarity coefficient.

$$S = 2 * C / (A+B) * 100$$

Dimana :

S= nilai similarity

C= jumlah k gram dengan nilai yang sama dari hasil banding dokumen

A dan B=jumlah dari kumpulan k grams dalam dokumen input dan banding

Hashing yang sama antar dokumen akan kita beri nama fingerprinting, adapun nilai hashing yang sama sebanyak 8 hashing yaitu:

**724605 793797 766997 821035 834973 783203 799177 756349**

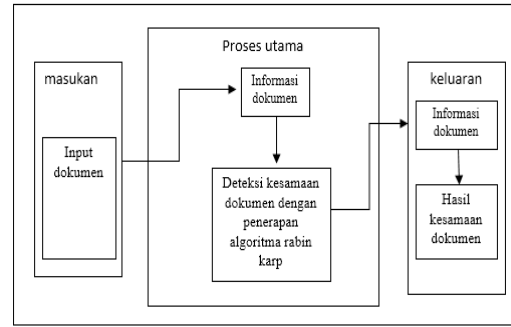
$$S = 2 * 8 / (22+19) * 100$$

$$= 16/41 * 100$$

$$= 39,02439 \% \text{ similarity}$$

**Desain Sistem**

**Gambaran umum sistem**

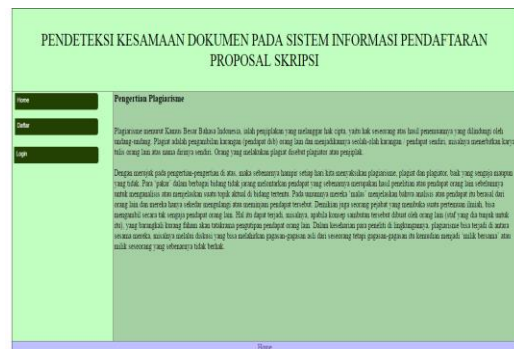


**Gambar 4. Gambaran Sistem**

**Implementasi sistem**

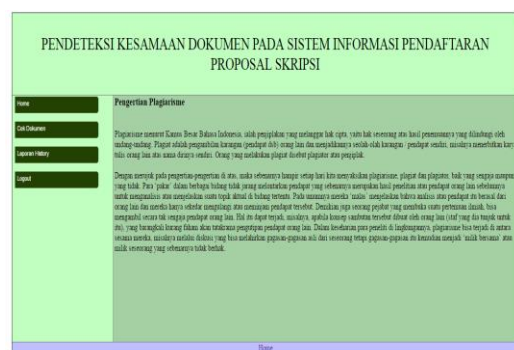
Dalam hal ini, implementasi dan pengujian sistem merupakan tahap dimana sistem siap dioperasikan pada keadaan yang sebenarnya, sehingga diketahui apakah sistem yang sudah diperbaharui dapat menghasilkan tujuan yang diinginkan. Berikut adalah desain dari penerapan pada tahap implementasi yang terdiri dari halaman – halaman aplikasi sistem sebagai berikut:

**Tampilan menu login**



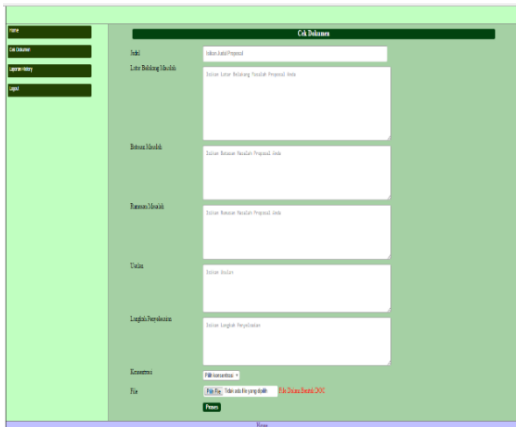
**Gambar 5. Menu Login**

**Tampilan menu beranda mahasiswa**



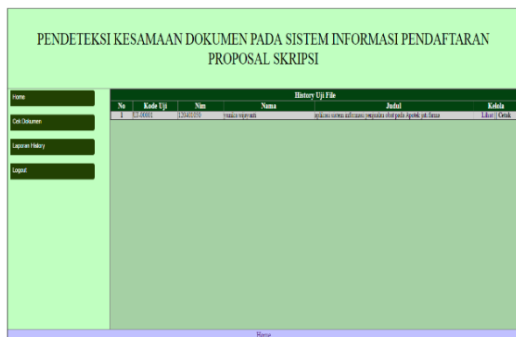
Gambar 6. Menu Beranda Mahasiswa

Tampilan Menu Cek dokumen Mahasiswa



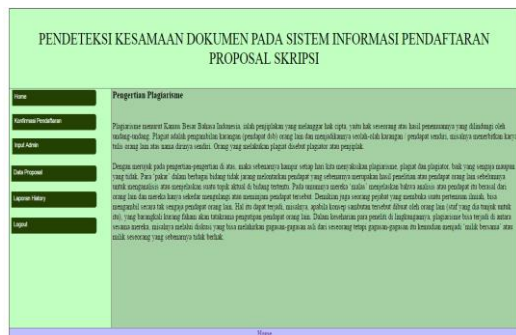
Gambar 7. Menu Cek Dokumen

Tampilan Menu Laporan History Mahasiswa



Gambar 8. Menu Laporan History Mahasiswa

Tampilan Menu Beranda Admin



Gambar 9. Menu Beranda Admin

**PENUTUP**  
**Simpulan**

1. Sistem pendeteksi kesamaan dokumen pendaftaran proposal skripsi ini dapat membantu mencegah terjadinya aksi plagiarisme pada bidang akademik. Karena mampu membandingkan satu dokumen dengan dokumen lain berdasarkan tingkat kesamaan isi dokumen yang telah ada.
2. Sistem pendeteksi kesamaan dokumen pendaftaran proposal skripsi ini mampu menghasilkan berapa persen kesamaan dari suatu dokumen proposal skripsi.
3. Mampu mendeteksi kesamaan dokumen dengan menerapkan algoritma Rabin-Karp dalam pembuatan sistem pendeteksi kesamaan dokumen ini.

**Saran**

Saran-saran yang diberikan penelitian ini adalah:

1. Untuk pengembangan selanjutnya sistem dapat di dikembangkan dengan menambahkan pendeteksian sinonim karena tindak plagiarisme sering kali dilakukan dengan menggunakan persamaan kata.
2. Mengembangkan penerapan algoritma Rabin karp dengan menambahkan tahap stemming .
3. Penelitian dilanjutkan dengan mencoba penggunaan metode pencarian yang berbeda untuk menemukan metode yang lebih baik untuk diterapkan.

**DAFTAR PUSTAKA**

Agustina, R., (2015). Journal Speed – Sentra Penelitian Engineering dan Edukasi – Volume 7 No 3 - 2015 - ijns.org. , 7(3): 7–12.  
Andi. (2011). Andi dan MADCOMS, 2011, aplikasi web database dengan dreamweaver dan

- PHPMYSQL*, Yogyakarta; Andi, Utorodewo, F. (2007). *Sebuah Pengantar Penulisan Ilmiah* Penerbit Fakultas Ekonomi Universitas Indonesia, ed.,
- KBBI, 1997. *Kamus Besar Bahasa Indonesia (KBBI)*,
- Kurniawati, A., Sekarwati, A.K. & Wicaksana, I.W.S. (2012). *Arsitektur Untuk Aplikasi Deteksi Kesamaan Dokumen. Konferensi Nasional Sistem Informasi, 297–302.*
- Muslih, M.T. & Purnama, B.E. (2013). *Pengembangan Aplikasi Sms Gateway Untuk Informasi Pendaftaran Peserta Didik Baru Di Sman 1 Jepara. Indonesian Journal on Networking and Security (IJNS), 50–55.*
- PMPN RI. (2010). *Peraturan Menteri Pendidikan Nasional RI No.17, Pencegahan dan Penanggulangan Plagiat di Perguruan Tinggi, Bab 1 pasal 1 ayat 1.*
- Rafles, A. (2013). **PLAGIARISME DOKUMEN DENGAN PENDEKATAN K-GRAM BERBASIS FRASA K-GRAM BERBASIS FRASA.**
- Salmuasih & Sunyoto, A. (2013). *Implementasi Algoritma Rabin Karp untuk Pendeteksian Plagiat Dokumen Teks Menggunakan Konsep Similarity. Seminar Nasional Aplikasi Teknologi Informatika, 23–28.*