

# SISTEM DETEKSI DINI PLAGIARISME TUGAS AKHIR MAHASISWA MENGGUNAKAN ALGORITMA N- GRAMS DAN WINNOWER

Rocky Yefrenes Dillak<sup>1</sup>, Folkes Lauma<sup>2</sup>, Ledy Jen Kadja<sup>3</sup>

## Abstrak :

The problem of plagiarism has been widely found in the various area of academic research. It became more complicated when the resources used were taken from a different language with a suspicious document. The aim of this paper is to develop a system that can be used as early warning system in plagiarism detection. The method work as follows: do some preprocessing steps, such as: removing punctuation, removing numbers, removing stopwords, removing repeated words, and doing a process called lemmatization of words. The next process is to extract its characteristic features using improved n-Gram method and the last step is to compare the content from translated document against downloaded documents. Based on experiments, the result is indicated that the method can be used as an early warning system for plagiarism detection.

**Kata Kunci :** Identifikasi iris mata, sistem biometrika, GLCM, jaringan syaraf tiruan, levenberg-marquardt algorithm.

## PENDAHULUAN

Tugas akhir (TA) merupakan tugas paling terakhir dalam aktivitas perkuliahan yang wajib dikerjakan oleh setiap mahasiswa Teknik Elektro Politeknik Negeri Kupang untuk menyelesaikan studi pada perguruan tinggi, tugas akhir yang dimaksud dalam pedoman ini skripsi rancangan yang dihasilkan pribadi individu.

Tahap-tahap yang dilakukan sebelum menyusun tugas akhir yaitu mahasiswa diharapkan untuk memasukkan judul-judul yang diinginkan sesuai dengan jurusan yang sedang ditekuni, judul-judul tersebut dimasukkan ke jurusan akademik elektro kemudian judul tersebut akan diserahkan kepada tim yang bertugas memeriksa kembali dan akan dibahas dalam rapat bersama untuk menentukan apakah judul yang dibuat layak untuk diterima, jika judul tersebut diterima maka akan segera dibagikan

kepada dosen pembimbing untuk mendampingi mahasiswa menyusun tugas akhir.

Tahap penyimpanan data-data tugas akhir pada jurusan teknik elektro masih manual sehingga seorang mahasiswa dengan bebas melakukan *copy-paste* proposal atau laporan tugas akhir dari awal sampai akhir tanpa diketahui oleh jurusan dan dosen. Dengan demikian Plagiarisme adalah tindakan mengambil ide orang lain, mengambil tulisan orang lain, dan mengambil teks secara keseluruhan dan mengakuinya sebagai miliknya sendiri.

Penelitian tentang deteksi kemiripan sebagai indikasi plagiat telah banyak dilakukan namun permasalahan yang sering muncul dengan berbagai algoritma yang telah diteliti adalah bagaimana meningkatkan keakurasian sistem yang dapat mempresentasikan kenyataan yang sebenarnya. Algoritma *N-gram* digunakan untuk mengambil potongan-potongan karakter huruf



sejumlah  $n$  dari sebuah kata yang secara kontinuitas dibaca dari teks sumber sampai akhir dari sebuah dokumen. Keunggulan Algoritma *N-gram* adalah tidak akan terlalu sensitif terhadap kesalahan penulisan yang terdapat pada suatu dokumen, sedangkan *Winnowing* membuang seluruh karakter-karakter yang tidak relevan misalnya tanda baca dan spasi.

Berdasarkan pada permasalahan diatas maka penulis mengusulkan sebuah sistem “Deteksi Dini Plagiarisme Pada Konten Teks Digital Tugas Akhir Mahasiswa Jurusan Teknik Elektro Menggunakan Algoritma *N-Gram* dan *Winnowing*”

## TINJAUAN PUSTAKA

Penelitian ini dilakukan tidak terlepas dari hasil penelitian-penelitian terdahulu yang pernah dilakukan sebagai bahan pertimbangan dan kajian.

Adapun hasil-hasil penelitian yaitu sebagai berikut :

Deteksi kemiripan teks yang dilakukan oleh Syahputra (2015) dengan menggunakan algoritma *winnowing* yaitu Aplikasi yang dirancang merupakan aplikasi yang dapat mendeteksi plagiat pada suatu teks. Plagiat atau penjiplakan merupakan tindakan mencontoh atau meniru tulisan atau pekerjaan tanpa izin dari pemiliknya dan mengakui sebagai hasil karya pribadi. Algoritma *Winnowing* digunakan dalam mendeteksi penjiplakan. Mendeteksi duplikasi menggunakan sinonim mendapatkan hasil persentase yang lebih tinggi dari pada tanpa menggunakan sinonim. Sistem yang dibangun telah dapat mendeteksi duplikasi dengan pendekatan sinonim dengan perbedaan  $\pm 0.82$  % lebih besar menggunakan synonym recognition dari pada tanpa synonym recognition.

Deteksi keberadaan kalimat sama sebagai indikasi penjiplakan yang dilakukan oleh Purwitasari (2013) dengan menggunakan algoritma Hashing berbasis *N-Gram*, Pendeteksian ini dirancang untuk deteksi keberadaan kalimat sama sebagai hasil *copy-paste* dan penjiplakan lebih lanjut seperti adanya penyebutan referensi yang baku. algoritma *Winnowing* sebagai

algoritma untuk deteksi kalimat sama sebagai indikasi terjadinya penjiplakan. Uji coba telah dilakukan untuk melihat kemampuan mendeteksi kalimat sama sebagai indikasi penjiplakan dengan perubahan nilai – nilai tertentu. Parameter – parameter yang telah diamati adalah nilai  $n$  dari *n-gram*, bilangan prima  $b$  yang menjadi basis dalam proses *hash*, ukuran *window*  $w$  dan nilai ambang batas penentuan penjiplakan. Parameter nilai  $n$  pada algoritma *Winnowing* digunakan saat mengubah teks sepanjang *n-gram* menjadi sekumpulan nilai-nilai *hash* dengan fungsi *hashing* yang membutuhkan bilangan prima  $b$  tertentu.

Perancangan sistem deteksi plagiat pada dokumen teks dengan konsep similarity yang dilakukan oleh Salmuasih (2013) dengan menggunakan algoritma Rabin Karp, pendeteksian ini dirancang untuk membandingkan *file* memberikan hasil berupa prosentase *similarity*. faktor yang mempengaruhi performa algoritma Rabin Karp karena Nilai *modulo* berpengaruh pada waktu proses, tetapi tidak pada nilai *similarity*, Penggunaan *stemming* berpengaruh pada keakuratan nilai *similarity* dan waktu proses. Dengan menggunakan *stemming* waktu proses cenderung lebih lama, tetapi tingkat akurasi *similarity*nya lebih tinggi. Semakin kecil *k-gram* menghasilkan akurasi nilai *similarity* yang lebih baik, dibandingkan *k-gram* yang lebih besar.

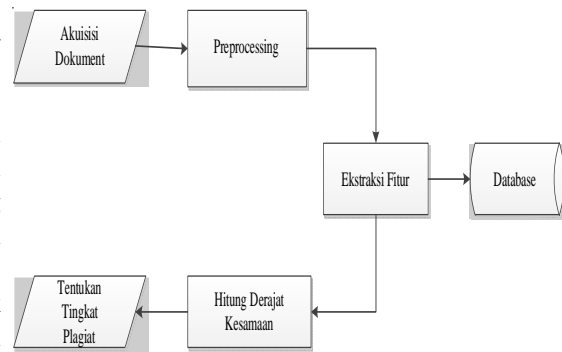
Sistem deteksi plagiarisme dokumen bahasa Indonesia yang dilakukan oleh Oktalina (2010) dengan menggunakan metode vector space model adapun dokumen yang diuji tingkat persentase kemiripannya yaitu dokumen jurnal program studi Teknik Informatika dan Sistem Informasi, dimana proses deteksi plagiarismenya melalui tahapan preprocessing, yaitu proses tokenisasi, penghapusan *stopwords*, dan *stemming*, tahapan selanjutnya perhitungan pembobotan dan *cosine similarity*. Dalam pengembangan sistem ini penulis menggunakan bahasa pemrograman Java. Metodologi pengembangan sistem menggunakan pendekatan model iteratif pengembangan inkremental. Tujuan utama sistem ini adalah untuk mengetahui

tingkat kemiripan atau plagiat suatu dokumen jurnal. Aplikasi ini diharapkan mampu mendeteksi dan memberikan persentase kemiripan dokumen dari proses tindakan plagiarisme mahasiswa dalam menyelesaikan tugas akhir.

Sistem deteksi plagiat pada dokumen bahasa Indonesia yang dilakukan oleh Suyanto (2012) menggunakan algoritma SCAM. Plagiat merupakan tindak kecurangan yang sering kali dilakukan. Baik terkadang tidak menyadari bahwa tindakan yang mereka lakukan tergolong plagiat ataukah dengan sengaja melakukan tindak plagiat tersebut. Dengan kemajuan teknologi, tindak plagiat saat ini lebih mudah untuk dilakukan karena banyaknya dokumen-dokumen yang diunggah di internet tanpa adanya suatu pengaman khusus sehingga dengan sangat mudah dapat di-copy oleh orang lain. Semakin meningkat kecanggihan manusia dalam melakukan plagiat, maka dibutuhkan juga sistem yang dapat membantu mendeteksi plagiat dalam sebuah dokumen. Oleh karena itu, penulis menggagas untuk dibangunnya sebuah sistem deteksi plagiat (Plagiarism detection system) yang menggunakan SCAM (Stanford Copy Analysis Mechanism). SCAM adalah sebuah mekanisme untuk menghitung tingkat kemiripan antara dua atau lebih dokumen. SCAM baik dalam mendeteksi dokumen yang overlap tanpa bergantung pada length dokumen tersebut. Dan juga SCAM mampu mengenali dengan baik untuk dokumen yang merupakan subset atau superset dari dokumen lainnya. SCAM baik dalam mendeteksi dokumen yang overlap tanpa bergantung pada length dokumen tersebut. Dan juga SCAM mampu mengenali dengan baik untuk dokumen yang merupakan subset atau superset dari dokumen lainnya. Dalam penelitian tugas akhir ini dilakukan pendeteksian terhadap 5 jenis plagiat, yaitu sinonim, pasif aktif, carbon copy, ubah susunan dan juga tambah kata. Dimana dokumen yang digunakan berupa abstrak Tugas Akhir mahasiswa IT Telkom berbahasa Indonesia.

## METODE PENELITIAN

Metode yang digunakan dalam penelitian ini adalah seperti pada gambar Arsitektur Sistem di bawah



Gambar 1. Arsitektur Sistem

Penjelasan arsitektur sistem yaitu user memasukkan file-file yang ingin di lakukan pendeteksian kemudian file-file yang sudah diinput akan diproses dengan cara analisis kebenaran arti dan kebenaran susunan terhadap teks. Tujuan dari pemrosesan ini adalah untuk mempersiapkan teks menjadi data yang akan mengalami pengolahan lebih lanjut, tahap lanjutannya diinput ke ekstraksi fitur di sinilah terdapat *N-Gram* untuk menghitung bobot setiap token yang ada dalam dokumen kemudian di simpan dalam dokumen kemudian barulah hitung derajat kesamaan dengan *cosine similarity* akan menghitung kemiripan vector dari artikel yang akan diuji.

### 1. Akuisisi Dokumen

Akuisisi dokumen merupakan langkah awal yang dilakukan yaitu mempersiapkan file-file yang ingin saya masukkan dalam akuisisi dokumen dan saya menyiapkan file sebanyak 20 untuk dibandingkan. Berikut ini adalah potongan kode program untuk akuisisi dokumen.

```

//Checking if the file is plain text or not
//echo "<center><span id=' Content'>
Contents of ".$FILES['file']['name']."
File</span></center>";
//Getting and storing the temporary file name
of the uploaded file
  
```



```

$fileName = $_FILES['filex']['tmp_name'];
$fileName1 = $_FILES['file']['tmp_name'];
//Throw an error message if the file could not
be open
$file = fopen($fileName,"r") or exit("Unable
to open file!");
$file1 = fopen($fileName1,"r") or exit("Unable
to open file!");

```

## 2. Preprocessing

Tahap ini dilakukan untuk mencari kebenaran arti dan kebenaran susunan terhadap suatu teks. Tujuan dari pemrosesan awal ini untuk mempersiapkan teks menjadi data yang akan mengalami pengolahan lebih lanjut. Dalam preprocessing ini terdapat 3 tahap yaitu *tokenizing*, *eliminasi stopwords*, dan *stemming*.

### a. Tokenizing

Tujuan dari *tokenizing* adalah mengubah semua huruf dalam dokumen menjadi huruf kecil ( a sampai z) selain huruf di hilangkan. Selain itu dalam *tokenizing*, seluruh kalimat dalam suatu dokumen akan dikonversi menjadi kata lalu disimpan dalam suatu array, potongan kode program untuk *tokenizing* seperti terlihat dibawah ini.

```

$x=explode(" ",$words);$kata="";
for($i=0;$i<count($x);$i++)
{
    $kata .= $x[$i];
}

```

### b. Eliminasi Stopwords

Tujuan dari *eliminasi stopwords* adalah membuang kata-kata yang tidak relevan seperti yang telah dijelaskan pada bab sebelumnya. Seluruh kata yang telah diidentifikasi sebagai *stopwords* disimpan dalam bentuk tabel pada data base. Berikut ini merupakan sintaks dari proses eliminasi stopwords.

```

function eliminasi_stopwords($input){
mysql_connect("localhost","root","");
mysql_select_db("word-stemming");
$query=mysql_query("select kata from
stopword");

```

```

$commonWords=array();
while($hasil=mysql_fetch_row($query))
{
    $commonWords[]=$hasil[0];
}
return preg_replace('^b(implode(' ',
$commonWords).')b/',,$input);
}

```

### c. Stemming

Setelah proses *eliminasi stopwords*, maka langkah selanjutnya adalah mencari akar kata atau bentuk kata dasar dari setiap kata (*stemming*). Tujuan dari tahap ini adalah tahap mencari root kata dari tiap kata *filtering*, tahap ini juga dilakukan proses pengembalian berbagai bentuk kedalam representasi yang sama. Berikut ini merupakan sintaks *stemming*.

```

function lakukan_stemming($data)
{
    $ss="";$z="";$x="";
    $datax=explode(" ", $data);
    for($k=0;$k<count($datax);$k++)
    {
        $x=(stemming(strtolower($datax[$k])););
        if(!empty($x))
        {
            $z=$x;
        }
        else
        {
            $z=strtolower($datax[$k]);
        }
        $ss .= $z. " ";
    }
    return $ss;
}

```

## 3. Ekstraksi Fitur

*Ekstraksi fitur* merupakan proses menghitung keunikan dari setiap dokumen yang dapat digunakan sebagai pembandingan terjadinya plagiarisme. Dalam ekstraksi fitur pada tugas akhir ini, digunakan dua algoritma yaitu algoritma *N-Gram* dan *Winnowing* untuk menghitung bobot setiap token yang ada dalam

dokumen (n=5) lalu disimpan dalam data base. Berikut ini adalah potongan kode program untuk ekstraksi fitur

```
function getNgrams($words, $n)
{
    $x=explode(" ", $words); $kata="";
    for($i=0; $i<count($x); $i++)
    {
        $kata .= $x[$i];
    }
    $word=trim($kata);
    $ngrams = array();
    $len = strlen($word);
    //$word=trim($word);
    for($i = 0; $i < $len; $i++) {
        if($i > ($n - 2)) {
            $ng = '';
            for($j = $n-1; $j >= 0; $j--) {
                $ng .= $word[$i-$j];
            }
            $ngrams[] = $ng;
        }
    }
    return $ngrams;
}
```

Setelah pencarian *N-Gram*, maka langkah selanjutnya adalah menghitung nilai hash dari setiap kata (*winnowing*) menggunakan rumus *rolling hash*. Potongan kode program seperti di bawah ini.

```
function hitung_hash($data, $k=5, $b=5)
{
    $data_temp="";
    $data_temp1=array();
    for($i=0; $i<count($data); $i++)
    {
        $h=array(); $h[$i]=0;
        $data_temp=$data[$i];
        $kk=$k;
        for($j=0; $j<strlen($data_temp); $j++)
        {
            $kk--; $h[$i] += ord($data_temp[$j]) *
            pow($b, $kk);
        }
        $data_temp1[$i]=$h[$i];
    }
    return $data_temp1;
}
```

#### 4. Menghitung Derajat Kesamaan

Proses selanjutnya setelah mencari nilai hash adalah memilih *fingerprint* berdasarkan nilai suatu *window w*, dimana nilai hash yang dipilih adalah nilai hash yang terkecil. Potongan program untuk menghitung nilai yang terkecil adalah sebagai berikut

```
function hitung_nilai_min($data)
{
    $temp_min=array();;
    for($i=0; $i<count($data); $i++)
    {
        $pecah=explode(" ", $data[$i]);
        //$sint=min($pecah); $temp_min[$i]=$sint;
        $temp_min[$i]=min ($pecah[0], $pecah[1],
        $pecah[2], $pecah[3]);
    }
    return $temp_min;
}
```

#### 5. Menghitung Derajat Kesamaan

Pada penelitian ini, perhitungan derajat kesamaan menggunakan *cosine similarity* dari teks yang telah diproses pada ekstraksi fitur. Tujuan dari proses ini adalah mengetahui tingkat kemiripan antara kedua dokumen dimana bila kemiripan kedua dokumen sangat tinggi maka kemungkinan plagiatnya akan semakin tinggi, demikian pun sebaliknya. Proses perhitungan derajat kesamaan pada dasarnya merupakan proses mencari kesamaan nilai hash dua buah dokumen. Potongan kode program untuk menghitung tingkat kemiripan dua buah dokumen seperti di bawah ini

```
$d=getNgrams($satu, 5);
$x=hitung_hash($d);
$y=winn($x);
$z=array_unique(hitung_nilai_min($y));
$d1=getNgrams($dua, 5);
$x1=hitung_hash($d1);
$y1=winn($x1);
$z1=array_unique(hitung_nilai_min($y1));
$hasil= array_diff($z, $z1);
$final=array_diff($z, $hasil);
$a=(count($final)/count($z))*100;
```



## HASIL DAN PEMBAHASAN

Pengujian sistem menggunakan Microsoft office word 2007 hasil upload yang terjadi keseluruhan file menjadi warning, simbol-simbol dan hasil akhir dalam bentuk digital



Gambar 2. Interface pengujian

Untuk menentukan jenis plagiarisme antara dokumen yang diuji ada 5 jenis penilaian persentase *similarity*:

- 0% : Hasil uji 0% berarti kedua dokumen tersebut benar-benar berbeda baik dari segi isi dan kalimat secara keseluruhan.
- <15% : Hasil uji 15% berarti kedua dokumen tersebut hanya mempunyai sedikit kesamaan.
- 15-50% : Hasil uji 15-50% berarti menandakan dokumen tersebut termasuk plagiat tingkat sedang.
- >50% : Hasil uji lebih dari 50% berarti dapat dikatakan bahwa dokumen tersebut mendekati plagiarisme.
- 100% : Hasil uji 100% menandakan bahwa dokumen tersebut adalah plagiat karena dari awal sampai akhir mempunyai isi yang sama.

Tabel 1. Pengujian Derajat Kesamaan

Plagiat	Tidak Plagiat
82.978723404255	6.2706270627063
100	14.569536423841
90.878378378378	27.40492170022
85.714285714286	32.642487046632
93.079584775087	36.312849162011

Sesuai dengan program yang telah dilakukan pendeteksian membuktikan bahwa pendeteksian yang telah dilakukan mampu membaca hingga tingkat plagiarism ( 100 %)

## KESIMPULAN

Berdasarkan percobaan-percobaan yang telah dilakukan, dapat diambil kesimpulan sebagai berikut :

- a. Sistem Plagiarisme dalam membandingkan file memberikan hasil berupa presentase similarity
- b. Berikut faktor yang mempengaruhi berjalannya algoritma N-Gram dan Winnowing :
  - i. Semakin banyak isi sebuah file yang ingin dideteksi, waktu prosesnya akan semakin lama (running time)
  - ii. Waktu mendeteksi file paling lambat 3.600 jika melewati batas maka akan terjadi eror
- c. Apabila yang dirancang merupakan aplikasi yang dapat mendeteksi plagiat pada suatu teks

Adapun saran yang diberikan oleh penulis untuk pengembangan aplikasi pendeteksian duplikasi ini, yaitu :

- a. Aplikasi ini tidak hanya dipakai untuk mendeteksi tugas akhir tapi juga dipakai untuk mendeteksi file lain
- b. Untuk kedepannya aplikasi ini dapat dikembangkan dengan menambah format dokumen dan lainnya yang dibandingkan dan tidak hanya berupa tulisan saja yang dibandingkan akan tetapi dapat membandingkan gambar, grafik, table dan lainnya.

## DAFTAR PUSTAKA:

- Aditya, Alan Nur, "Jago php & MySql", Penerbit dunia komputer: Bekasi.
- Anonim, Perancangan Sistem Deteksi Plagiat Pada Dokumen teks.
- Dengan Konsep Somilariti Menggunakan Algoritma Rabin Karp. <http://yudiagusta.wordpress.com/2008/05/13/similarity-measure/> diakses Kamis, 7 Maret 2013 jam 21.52 wib.
- Anonim., Deteksi Plagiarisme Dokumen Teks Menggunakan Algoritma Scam. <http://>

- elib.unikom.ac.id/files/disk1/596/jbptunik  
ompp-gdl-dennisahma-29785-9-unikom\_d-  
i.pdf, diakses 11 Januari 2012.
- Atmopawiro, Alasian.2006. Pengkajian dan  
Analisis Tiga Algoritma Efisien Rabin-Karp,  
Knuth-Morris-Pratt, dan Boyer-Moore  
dalam Pencarian Pola dalam Suatu Teks.  
Makalah tidak Terpublikasi. Program Studi  
Teknik Informatika Institut Teknologi  
Bandung.
- B.Klug, "The Plagiarism Checker", 2002, [http://  
www.dustball.com/cs/plagiarism.checker/](http://www.dustball.com/cs/plagiarism.checker/),  
diakses tanggal 10 Januari 2011.
- Even-Zohar, Y. 2002. Introduction to Text  
Mining.
- G.Greenspan, "Copyscape", 2006, <http://copyscape.com>,  
diakses tanggal 10 Januari 2011.
- Hearst, M. 2003. What is Text Mining?. [http : /  
/ www . sims . berkeley . edu / ~hearts t/ text-  
mining.html](http://www.sims.berkeley.edu/~hearts/text-mining.html).
- Information Retrieval Tutorial : Stemming Po-  
rter untuk Bahasa Indonesia.
- "[http://kabulkurniawan.com/2012/03/17/infor-  
mation-retrieval-tutorial-stemming-porter-  
untuk-bahasa-indonesia/](http://kabulkurniawan.com/2012/03/17/information-retrieval-tutorial-stemming-porter-untuk-bahasa-indonesia/)" diakses Selasa, 3  
juli 2012 jam 10.49 wib.
- Kadir, Abdul, Dasar Pemrograman Web ASP,  
cetakan pertama, PT Excel Media Kom-  
putation, Yakarta, 2005
- KBBI, 1997: 775.
- Kosinov, Serhiy. 2002. Evaluation of N-Grams  
Conflation Approach in Text-Based Infor-  
mation Retrieval. University of Al-  
berta.Canada.
- Kristanto, Andri, 2004, Rekayasa Perangkat  
Lunak : KOnsep Dasar, YOgyakarta : Gava  
Media.
- Kurniawati, Ana, Wicaksana, I Wayan Simri.  
Perbandingan Pendekatan Deteksi  
Plagiarism Dokumen Dalam Bahasa Inggris.  
Fakultas Ilmu Komputer dan Teknologi  
Informasi, Universitas Gunadarma, 2008.
- Madcoms, 2009, "Menguasai XHTML, CSS,  
PHP, & MySQL melalui dreamweaver",  
Penerbit ANDI:Yogyakarta.
- Mudafiq Riyan Pratama, Aplikasi Pendeteksi  
Duplikasi Dokumen Teks Bahasa Indonesia  
Menggunakan Algoritma WInnowing  
Dengan Metode K-Gram Dan Synonym  
Recognition, Jurusan Teknik Informatika  
Universitas Muhammadiyah Malang.
- Mutiara, Benny; Agustina, Sinta. 2008. Anti  
Plagiarsm Application with Algorithm Karp-  
Rabin at Thesis in Gunadarma.  
University. Gunadarma University. Depok,  
Indonesia.
- Oktalina,2010. Sistem Deteksi Plagiarisme  
Dokumen Bahasa Indonesia dengan Algo-  
ritma Metode Vector Space. Program Studi  
Teknik Informatika Stimik Gi Mdp.
- Purwitasari, 2013. Deteksi Keberadaan Kalimat  
Sama sebagai Indikasi Penjiplakan dengan  
Algoritma Hashing Berbasis N-Gram. Lab  
Semantik Web - Teknik Informatika, Institut  
Teknologi Sepuluh Nopember Surabaya.
- Rabin Karp Matching. [http://www. cedar-  
ville.edu/ personal/ personalpages/ shom-  
per/ cs3410\\_web/ resources/ rabin\\_karp\\_  
matching.ppt](http://www.cedarville.edu/personal/personalpages/shomper/cs3410_web/resources/rabin_karp_matching.ppt) diakses Senin, 5 November  
2012 jam 14.12 wib.
- Salmuasih, 2013. Perancangan Sistem Deteksi  
Plagiat Pada Dokumen Teks Dengan  
Konsep Similarity Menggunakan Algoritma  
Rabin Karp. Sekolah Tinggi Manajemen  
Informatika Dan Komputer Amokom  
Yogyakarta.
- Schelimer., Saul, et. Al, 2003 WInnowing: Local  
Algorithms for Document Fingerprinting,  
diakses 13 November 2012.
- Suyanto,2012. Sistem Deteksi Plagiat Pada  
Dokumen Bahasa Indonesia.  
Universitas Dian Nuswantoro Program Studi  
Teknik Informatika.
- Syahputra, Pelita Informatika Budi Darma,  
Volume : IX, Nomor: 1, Maret 2015. Imple-  
mentasi Algoritma WInnowing Untuk De-  
teksi Kemiripan teks. Program Studi Teknik  
Informatika STMIK Budidarma Medan.