

# DETEKSI PLAGIARISME KARYA ILMIAH DENGAN PEMANFAATAN DAFTAR PUSTAKA DALAM PENCARIAN KEMIRIPAN TEMA MENGGUNAKAN METODE COSINE SIMILARITY (Studi Kasus: Di Universitas Muhammadiyah Magelang)

Fauzi Bayu Sejati, Purwono Hendradi<sup>2</sup>, Bambang Pujiarto<sup>3</sup>

Prodi Teknik Informatika Universitas Muhammadiyah Magelang

## ABSTRAK

*Plagiarisme merupakan salah satu masalah yang harus ditanggulangi dalam dunia akademik karena dapat mencemarkan integritas akademik. UM Magelang merupakan perguruan tinggi yang mempertimbangkan plagiarisme sebagai masalah penting yang harus dicermati. Namun, pada saat ini di UM Magelang belum ada sistem pendeteksian tindakan plagiarisme yang terkomputerisasi. Sistem deteksi plagiat yang ada pada saat ini menggunakan perbandingan teks antar dokumen secara keseluruhan. Penelitian ini bertujuan untuk merancang dan membangun aplikasi deteksi plagiarisme dengan pemanfaatan daftar pustaka dengan menggunakan metode cosine similarity yang merupakan ide atau gagasan baru dalam melakukan pendeteksian tindakan plagiarisme. Deteksi dilakukan dengan mengecek kesamaan daftar pustaka terlebih dulu sebelum mengecek keseluruhan isi dokumen. Hasil dari penerapan metode cosine similarity dengan memanfaatkan daftar pustaka dalam mencari kemiripan tema mampu mencari kemiripan antar dokumen dengan tepat dan cepat. Threshold 0,6 digunakan sebagai batas parameter tingkat plagiarisme yang patut dicurigai*

**Kata Kunci:** *plagiarisme, aplikasi deteksi plagiarisme dengan pemanfaatan daftar pustaka, cosine similarity.*

## A. PENDAHULUAN

Perkembangan sistem teknologi informasi pada era saat ini berkembang dengan pesat. Penggunaan sistem teknologi informasi untuk memenuhi pekerjaan manusia tentu juga semakin berkembang. Salah satunya *information retrieval* atau temu kembali informasi yang mempunyai pengertian proses menemukan kembali dokumen-dokumen relevan untuk memenuhi kebutuhan informasi para pengguna. Proses tersebut bertujuan untuk menemukan kembali dokumen-

dokumen yang relevan dengan membandingkan kata-kata yang ada pada dokumen yang dicari untuk diketahui informasi yang relevan dari hasil penelusuran.

Berdasarkan dari manfaat yang terkandung di bidang *information retrieval* maka, hal tersebut dapat digunakan dalam kasus untuk melakukan pendeteksian unsur plagiarisme pada sebuah karya ilmiah. Karya ilmiah merupakan hasil dari kegiatan intelektual yang melalui proses penelitian, peninjauan dan pemikiran yang mendalam terhadap suatu topik atau pokok

suatu permasalahan tertentu. Di kalangan akademis, baik dosen maupun mahasiswa karya ilmiah dapat dijadikan sebagai sarana untuk menyampaikan gagasannya kepada masyarakat mengenai suatu topik yang ditelitinya secara mendalam. Karya ilmiah yang dipublikasikan melalui media teknologi informasi berdampak positif dalam penyediaan referensi untuk penelitian-penelitian selanjutnya. Tetapi salah satu dampak negatifnya adalah memberikan kemungkinan adanya tindakan copy paste yang berujung pada plagiarisme.

Selain itu permasalahan yang muncul adalah bagaimana agar dapat mengidentifikasi tindakan plagiat pada karya ilmiah dengan cepat dan mudah. Sebagai contoh untuk kasus dalam penyusunan laporan ini, penulis melakukan study kasus pendeteksian plagiarisme di Universitas Muhammadiyah Magelang pada skripsi mahasiswa. Alasan penulis melaksanakan study kasus di Universitas Muhammadiyah Magelang, karena belum adanya sistem yang terkomputerisasi dalam pengecekan tentang unsur plagiarisme pada karya ilmiah. Sistem yang saat ini berjalan dalam pengecekan plagiarisme adalah dengan membuka dokumen secara langsung jika didapati sebuah dokumen karya ilmiah terlihat memiliki kecurigaan unsur plagiarisme dari judul dan isi laporan yang tidak sengaja ditemukan oleh akademis seperti dosen atau karyawan perpustakaan. Oleh karena itu perlunya sistem yang terkomputerisasi untuk membantu pengecekan unsure plagiarisme pada karya ilmiah. Hal tersebut diharapkan dapat lebih bermutu dan bebas dari unsur

plagiat untuk di-upload atau tidaknya ke dalam situs portal pendidikan nasional yang tentunya bertujuan untuk dapat menambah rating Universitas Muhammadiyah Magelang di dunia maya.

Salah satu solusi yang dapat dilakukan dalam pendeteksian unsure plagiat pada dokumen, penulis mengusulkan solusi dengan menggunakan sistem temu kembali informasi yang memanfaatkan daftar pustaka dalam mencari kecocokan tema terlebih dahulu dalam mendeteksi sekian banyak skripsi dari berbagai jurusan yang ada pada Universitas Muhammadiyah Magelang. Hal ini dilakukan karena daftar pustaka memiliki pengaruh dan hubungan tentang isi atau topic yang dibahas dalam sebuah laporan. Oleh karena itu pemanfaatan daftar pustaka digunakan agar proses pendeteksian antara dokumen uji dengan dokumen pembanding cocok sesuai kesamaan tema, karena sangat jauh kemungkinan seorang mahasiswa di bidang Teknik Informatika melakukan adanya tindakan plagiat dalam dokumen karya ilmiah di bidang Hukum.

Dalam melakukan filter dengan memanfaatkan daftar pustaka sebagai indikator dari kesamaan tema terlebih dahulu sebelum dilakukan pendeteksian isi dalam dokumen, digunakan metode *cosine similarity*. Metode *cosine similarity* adalah metode yang banyak digunakan untuk menghitung dua buah dokumen. Metode cosine similarity juga merupakan salah satu ukuran untuk mendeteksi kemiripan teks atau dokumen yang populer. Hal ini dikarenakan metode cosine similarity adalah alat ukur untuk menghitung nilai cosines sudut antara dua vector. Sehingga

nantinya akan diketahui berapa banyak jumlah kemiripan teks yang sama antara dokumen yang dibandingkan.

Berdasarkan penjelasan tersebut, pemanfaatan daftar pustaka dengan metode cosine similarity dinilai cocok untuk digunakan dalam mendeteksi kemiripan teks dan mengidentifikasi adanya unsure plagiarisme antar dokumen pada karya ilmiah di Universitas Muhammadiyah Magelang.

## B. TINJAUAN PUSTAKA

### 1. PLAGIARISME

Menurut Adimihardja (2005), plagiarisme adalah pencurian dan penggunaan gagasan atau tulisan orang lain (tanpa cara-cara yang sah) dan diakui sebagai miliknya sendiri. Plagiarisme juga didefinisikan sebagai kegiatan dengan sengaja menyalin pemikiran atau kerja orang lain tanpa cara-cara yang sah (Adimihardja, 2002). Pelaku plagiarisme dikenal juga dengan sebutan plagiat (Rosyidi, 2007). Secara garis besar, tindakan yang termasuk plagiarisme akademis antara lain (Rosyidi, 2007):

- a. Menyalin tulisan orang lain mentah-mentah, tanpa memberikan penjelasan bahwa tulisan tersebut diambil dari tulisan lain dan atau tanpa menyebutkan sumbernya.
- b. Mengambil gagasan orang lain tanpa membrikan keterangan yang cukup tentang sumber gagasan tersebut.

Menurut Adimihardja (2005) terdapat banyak sekali jenis plagiarisme, antara lain sebagai berikut:

- a. Mengutip atau mengulang gagasan orang lain dalam suatu percakapan tanpa merujuk kepada yang mempunyai gagasan, tanpa memberi penghargaan atau ucapan terima kasih kepada yang mempunyai gagasan tersebut. Mencuri gagasan orang lain dalam suatu percakapan kemudian menuliskannya tanpa izin sah dari yang mempunyai gagasan tersebut termasuk plagiarisme dan ini merupakan kesalahan ilmiah.
- b. Semua pendapat atau pernyataan orang lain secara tertulis yang dikutip tanpa memberi penghargaan kepada yang punya pendapat melalui catatan kaki orang atau daftar pustaka.
- c. Melakukan kutipan tak langsung dari pendapat atau pernyataan orang lain secara tertulis tanpa melakukan rephrase (parafrase).
- d. Mengutip tabel dan gambar tanpa menyebutkan sumbernya.
- e. Dua tulisan berjudul dan berisi sama, maka yang keluar belakangan merupakan hasil plagiat.
- f. Menyalin seluruh hasil karya orang lain dan salinan itu diakui sebagai tulisan sendiri walaupun pemilik karya tulis mengizinkan secara tulus. Hasil karya yang dimaksudkan meliputi yang dipublikasi (buku, artikeel dalam jurnal/prosiding/majalah) dan yang tidak dipublikasi (makalah untuk seminar, laporan penelitian, skripsi,

- laporan penelitian, tesis, disertasi, diklat, buku ajar)
- g. Tulisan mahasiswa yang dipublikasi tanpa menuliskan nama mahasiswa sebagai penulis pertama.
  - h. Penulis yang dengan sengaja mengirimkan tulisan berjudul sama pada dua jurnal atau lebih.
  - i. Menerjemahkan suatu tulisan orang lain dan menulis dirinya sebagai penulis.
  - j. Tulisan orang lain dimodifikasi baik organisasi maupun frasenya tanpa mencantumkan nama penulis aslinya.

## 2. DETEKSI PLAGIARISME

Ada 3 macam teknik yang dibangun untuk menentukan nilai kemiripan dokumen (Zaka, 2009) antara lain:

1. Distance-based similarity measure, yaitu mengukur tingkat kesamaan dua buah objek dari segi jarak geometris dari variabel-variabel yang tercakup di dalam kedua objek tersebut. Metode Distance-based similarity measure ini meliputi Minkowski Distance, Manhattan/CityBlock Distance, Euclidean Distance, Jaccard Distance, Dice's Coefficient, Cosine Similarity dan Sounddex Distance.
2. Feature-based similarity measure, yaitu melakukan perhitungan tingkat kemiripan dengan mempresentasikan objek ke dalam bentuk feature-feature yang ingin diperbandingkan. Featured-based similarity banyak digunakan dalam

melakukan pengklasifikasian atau pattern matching untuk gambar dan teks.

3. Probabilistic-based similarity measure, yaitu menghitung tingkat kemiripan dua objek dengan merepresentasikan dua set objek yang dibandingkan dalam bentuk probability. Meliputi Kullback Leibler Distance dan Posterior Probability.

## 3. STEMMING

Stemming merupakan proses dasar dalam pengolahan teks yang sering digunakan untuk meningkatkan efisiensi dan efektivitas text retrieval, machine translation, peringkasan dokumen serta klasifikasi teks. Stemmer menghilangkan affiks dari sebuah kata untuk mengkategorikannya ke dalam bentuk kata dasarnya. Sebagai contoh, kata "bekerja", "pekerjaan" dan "mengerjakan" akan dikategorikan ke dalam kata dasar "kerja".

Mengenali kata berdasarkan kata dasarnya akan meningkatkan kemampuan pencarian dokumen relevan dalam bentuk document retrieval, tetapi juga dapat menurunkan selektivitas karena makna asli dari kata dapat hilang. Sebagai contoh, mengubah kata "makanan" menjadi "makan" akan mengubah makna asli dari kata benda (nomina) menjadi kerja (verba). Stemming bermaksud untuk meningkatkan pengenalan, tetapi bisa menjadi mengurangi presisi.

## 4. COSINE SIMILARITY

Cosine Similarity digunakan untuk mengukur kedekatan antara dua vektor. Cosine Similarity merupakan hasil dot produk kedua vektor tersebut yang dinormalisasikan dengan dibagi dengan Euclidean Distance antara kedua vektor tersebut. Cosine Similarity dapat diimplementasikan untuk menghitung nilai kemiripan antar kalimat dan menjadi salah satu teknik untuk mengukur kemiripan teks yang populer. Contoh penggunaan Cosine Similarity dalam menguji kemiripan dua buah kalimat adalah sebagai berikut:

A : Lucie loves me more than  
Linda loves me

B : Junie likes me more than  
Lucie loves me

Uji kemiripan teks dilakukan dengan rumus:

$$\text{Kemiripan} = \frac{\sum_{n=1}^j (nA \times nB)}{\sqrt{\sum_{n=1}^j (nA)^2} \times \sqrt{\sum_{n=1}^j (nB)^2}}$$

Dengan:  $j = |A \cup B|$

$nA$  = jumlah kemunculan kata indeks ke-n dari daftar kata pada kalimat A.

$nB$  = jumlah kemunculan kata indeks ke-n dari daftar kata pada kalimat B.

Tabel 2.1 Contoh Uji kemiripan Teks

Indeks	Daftar Kata	Jumlah Kemunculan Kata	
		A	B
1	Lucie	1	1
2	Loves	2	1
3	Me	2	2
4	More	1	1

5	Than	1	1
6	Linda	1	0
7	Junie	0	1
8	Likes	0	1

Berdasarkan rumus tersebut diatas dilakukan perhitungan seperti berikut:

Tingkat uji kemiripan teks =

$$\begin{aligned} & \frac{\sum_{n=1}^j (nA \times nB)}{\sqrt{\sum_{n=1}^j (nA)^2} \times \sqrt{\sum_{n=1}^j (nB)^2}} \\ &= \frac{(1 \times 1) + (2 \times 1) + (2 \times 2) + (1 \times 1) + (1 \times 1) + (1 \times 0) + (0 \times 1) + (0 \times 1)}{\sqrt{1^2 + 2^2 + 2^2 + 1^2 + 1^2 + 1^2 + 0^2 + 0^2} \times \sqrt{1^2 + 1^2 + 1^2 + 2^2 + 1^2 + 1^2 + 0^2 + 1^2}} \\ &= \frac{(1) + (2) + (4) + (1) + (1) + (0) + (0) + (0)}{\sqrt{1+4+4+1+1+1+0+0} \times \sqrt{1+1+4+1+1+0+1+1}} \\ &= \frac{9}{\sqrt{12} \times \sqrt{10}} = 0.821584 \end{aligned}$$

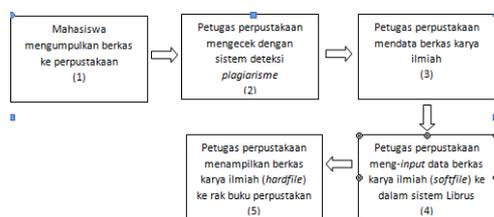
### C. METODOLOGI

Aplikasi deteksi plagiarisme dengan pemanfaatan daftar pustaka dalam pencarian kemiripan tema menggunakan metode cosine similarity ini menggunakan metode pemrograman berorientasi objek.

### D. ANALISIS DAN PERANCANGAN

Hasil perhitungan data jalur algoritma Universitas Muhammadiyah Magelang adalah salah satu perguruan tinggi yang ada di Kota Magelang tepatnya di daerah Mertoyudan Kabupaten Magelang. Selain proses belajar mengajar, para dosen dan mahasiswanya masing-masing juga melakukan penelitian dan membuat suatu karya ilmiah. Hasil dari karya ilmiah tersebut kemudian dikumpulkan ke perpustakaan dan tentunya dimasukkan ke dalam database perpustakaan dan ada juga yang sampai ke DIKTI. Pada saat ini belum ada sistem untuk mendeteksi keaslian dari karya ilmiah tersebut untuk mencegah dan mengetahui adanya tindakan plagiarisme pada

karya ilmiah. Alur sistem yang sedang berjalan pada pengumpulan karya ilmiah (skripsi) mahasiswa di Universitas Muhammadiyah Magelang sebagai berikut pada gambar 3.1



Gambar 3.1 Usulan Rancangan Alur Sistem

Pada gambar 3.2 diatas diusulkan alur rancangan sistem pengumpulan karya ilmiah (skripsi) dengan penambahan proses pengecekan dengan sistem deteksi plagiarisme pada langkah nomor 2 sebelum berkas dilakukan proses lebih lanjut.

**1. Rancangan Sistem**

Sistem deteksi plagiarisme yang akan dibuat menggunakan metode cosine similarity untuk mendeteksi kemiripan antar dokumen. Untuk meningkatkan keakuratan, dilakukan penyaringan dengan tidak memproses daftar lampiran dan daftar pustaka. Hal ini dilakukan untuk mempercepat proses dalam pendeteksian kemiripan antar dokumen. Alur kerja sistem secara umum dapat dilihat pada Gambar 3.3.



Gambar 3.3 Rencana Alur Kerja Sistem

Dari Gambar 3.3 merupakan alur berjalannya sistem aplikasi deteksi plagiarisme pada karya ilmiah di Universitas Muhammadiyah Magelang sebagai berikut :

1. Proses pertama dimulai dari memasukkan data dokumen. Terdapat dokumen uji dan dokumen pembanding yang akan dideteksi kemiripannya.
2. Proses kedua isi teks secara keseluruhan diubah menjadi baris teks yang kemudian disimpan.
3. Proses ketiga yaitu penentuan batas wilayah dokumen yang akan dideteksi kemiripannya, dalam hal ini daftar pustaka akan dibatasi untuk dijadikan filter yang akan dideteksi terlebih dulu saat proses pendeteksian nantinya
4. Proses keempat melakukan pemecahan baris teks menjadi per kalimat menggunakan API yang sudah disediakan JAVA yaitu OPENNLP.
5. Proses kelima melakukan text preprocessing yaitu menyeleksi dan membuang kata-kata yang tidak memiliki arti dan stemming untuk menghilangkan imbuhan pada kalimat sehingga menjadi kata dasar.
6. Langkah keenam melakukan pengkodean baris teks daftar pustaka.
7. Proses ketujuh melakukan proses filter daftar pustaka dalam proses pencarian kecocokan berdasarkan kode yang

- sudah dilakukan sebelumnya.
8. Proses kedelapan melakukan pengolahan kemiripan pada content dokumen menggunakan cosine similarity.
  9. Proses kesembilan adalah laporan hasil dari pengujian atau pendeteksian kemiripan dari dokumen-dokumen yang sudah diproses.

Alur proses global sistem deteksi plagiarisme dengan Cosine similarity murni terdiri dari 4 proses utama yaitu Proses Input Data Dokumen Teks Karya Ilmiah (yang terdiri dari proses nomor 1 s.d 6 gambar di atas), Proses Filter Dokumen (proses nomor 7), Perbandingan Kalimat content dokumen (proses nomor 8), dan Laporan Hasil Proses (nomor 9).

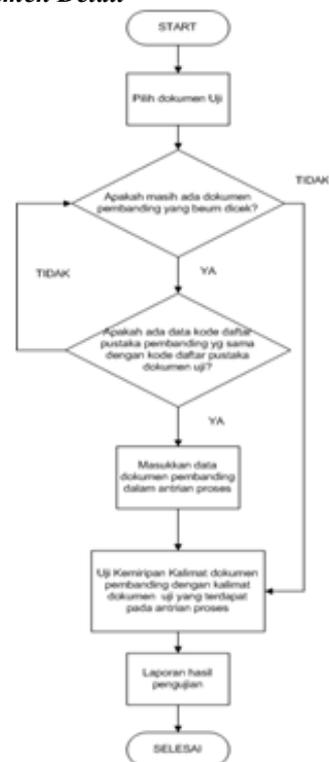
Sebelum dokumen karya ilmiah diuji tingkat kemiripannya dengan dokumen karya ilmiah lain, terlebih dahulu dilakukan mekanisme Input Data Dokumen Teks Karya Ilmiah. Selanjutnya dilakukan filter dokumen untuk menentukan prioritas dokumen yang akan dibandingkan terlebih dahulu berdasarkan kemiripan data-data baris teks daftar pustaka yang dimiliki oleh dokumen. Dokumen perbandingan yang memiliki kemiripan tertinggi pada data kode teks daftar pustaka dokumen uji akan diuji lebih dulu.

**1. Flowchart Teknis Perbandingan Dokumen**



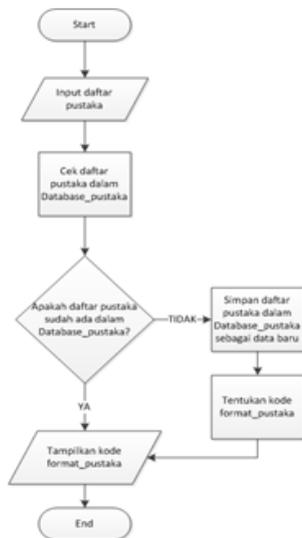
Gambar 3.4 flowchart Teknis Perbandingan Dokumen

**2. Flowchart Teknis Perbandingan Dokumen Detail**



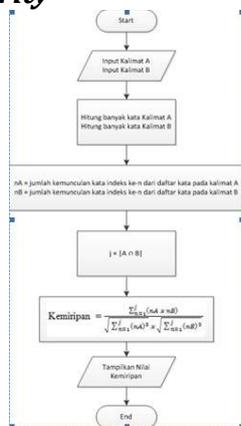
Gambar 3.5 flowchart Teknis Perbandingan Dokumen Detai

### 3. Flowchart Teknis Pengolahan Daftar Pustaka



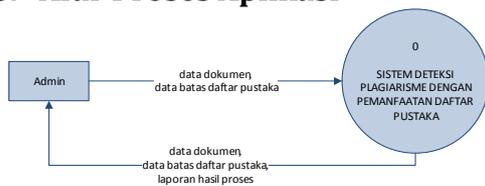
Gambar 3.6 Flowchart Teknis Pengolahan Daftar Pustaka.

### 4. Flowchart Dokumen dengan Cosine Similarity



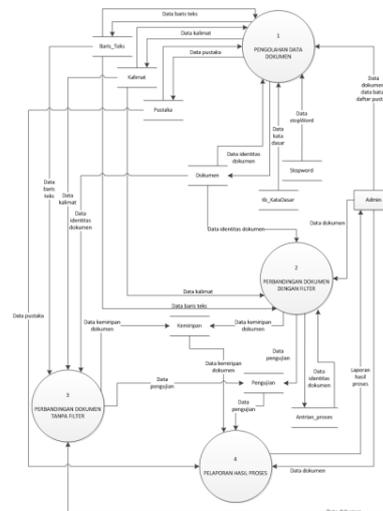
Gambar 3.7 Flowchart pengolahan dokumen dengan cosine similarity

### 5. Alur Proses Aplikasi

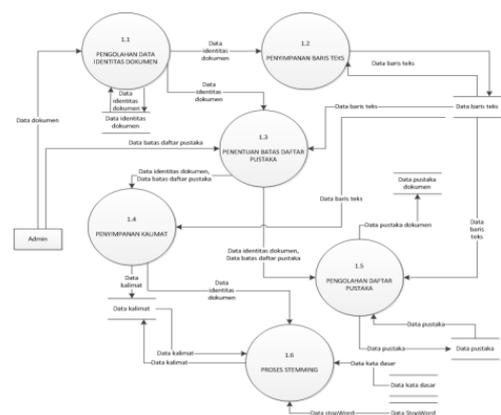


Gambar 3.8 Diagram Konteks

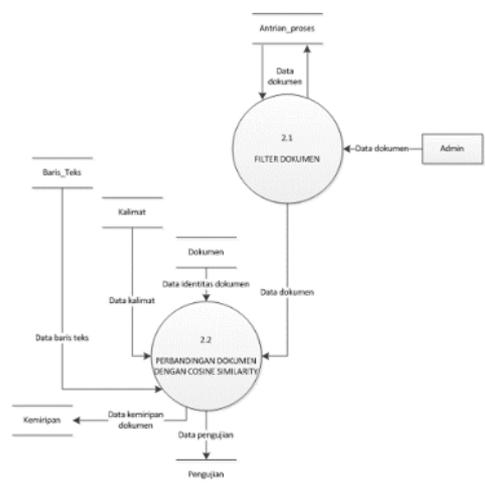
### Aplikasi Deteksi plagiarisme dengan cosine similarity.



Gambar 3.9 DFD level 0 deteksi plagiarisme dengan cosine similarity.



Gambar 3.10 DFD level 1 pengolahan deteksi aplikasi plagiarisme





dari *user* untuk mengambil tindakan selanjutnya dari hasil dokumen yang telah diketahui mempunyai kemiripan yang tinggi atau mencurigakan.

## 2. Saran

Berdasarkan hasil penelitian ada beberapa saran yang dapat dilakukan untuk pengembangan sistem menjadi lebih baik sebagai berikut :

- a. Untuk penelitian ke depan diharapkan metode *cosine similarity* dapat

dipergunakan dalam sebuah sistem lebih lanjut untuk kepentingan akademik yang bermanfaat.

- b. Perlu adanya pendeteksian sinonim atau persamaan kata dalam pengecekan kata yang berbeda namun memiliki arti yang sama.

Perlu adanya penelitian lebih lanjut dengan metode lain agar diketahui kelebihan dan kelemahan masing-masing metode yang digunakan

## G. DAFTAR PUSTAKA

- [1] Adriani, M., J.Asian, B. Nazief, S.M.M. Tahaghogi, and H.E Williams. 2007. *Stemming Indonesian. AcmTransactions On Asian Language Information Processing*, 6(4), p.1-33.  
<http://portal.acm.org/citation.cfm?doid=1316457.1316459>, diakses tanggal 10 September 2016.
- [2] Hasan, A., S Dardjowidjojo, H.Lapoliwa, A.M.Moeliono.2003. *Tata Bahasa Baku Bahasa Indonesia (edisi ketiga)*. Jakarta: Balai Pustaka.
- [3] Oktavian, R. 2011. Perancangan Sistem Deteksi [Plagiarisme](#) Dokumen Teks Dengan Menggunakan Multithreading. *Skripsi*. FMIPA Universitas Gadjah Mada. Yogyakarta.
- [4] Taufik, U. 2015. Pendekatan Deteksi Plagiarisme Berbasis Kutipan Dan Algoritme Kang untuk Teks Berbahasa Indonesia. *Thesis*. FMIPA Universitas Gadjah Mada.Yogyakarta.
- [5] Susanti, M. 2009. Rancang Bangun Aplikasi Berbasis Web untuk Deteksi Plagiarisme. *Skripsi*. FMIPA Universitas Gadjah Mada, Yogyakarta.
- [6] Nugroho, E. 2001. Perancangan Sistem Deteksi Plagiarisme Dokumen Teks Dengan Menggunakan Algoritma Rabin-Karp. *Skripsi*. Program Studi Ilmu Komputer. Jurusan Matematika. Fakultas Matematika dan ilmu Pengetahuan Alam. Universitas Brawijaya.
- [7] Rikky Wisnu Nugraha dan Astri Budiarti. 2013. Perangkat Lunak Pendeteksi Kemiripan File dengan Metode *Text Mining* Berbasis Web. *Jurnal*. Konsentrasi Teknik Informatika.Program Studi Teknik Informatika. STMIKLPKIA.
- [8] Zaka, Bilal. 2009.*Theory and Application of Similarity Detection Technique. Dissertation*. Institute for [Information](#) Systems and Computer Media (IISCM). Graz University of Technology Austria.
- [9] <https://sites.google.com/site/nirajatweb/home>, diakses tanggal 10 Juni 2016