

# PENDETEKSI TINGKAT PLAGIASI TEKS BERSEGMENTASI PENDEK

Tugas Akhir



Oleh:

Yanuar Septian

22 08 4484



Program Studi Teknik Informatika Fakultas Teknologi Informasi

Universitas Kristen Duta Wacana

Tahun 2012

# PENDETEKSI TINGKAT PLAGIASI TEKS BERSEGMENTASI PENDEK

Tugas Akhir



Diajukan kepada Fakultas Teknologi Informasi  
Universitas Kristen Duta Wacana  
Sebagai salah satu syarat dalam memperoleh gelar  
Sarjana Komputer



Disusun oleh:

Yanuar Septian

22 08 4484

Program Studi Teknik Informatika Fakultas Teknologi Informasi

Universitas Kristen Duta Wacana

Tahun 2012

## PERNYATAAN KEASLIAN TUGAS AKHIR

Saya mengatakan dengan sesungguhnya bahwa tugas akhir dengan judul:

### **PENDETEKSI TINGKAT PLAGIASI TEKS BERSEGMENTASI PENDEK**

Yang saya kerjakan untuk melengkapi sebagian persyaratan menjadi Sarjana Komputer pada pendidikan sarjana Program Studi Teknik Informatika, Fakultas Teknologi informasi Universitas Kristen Duta Wacana, bukan merupakan tiruan atau duplikasi dari skripsi kesarjanaaan di lingkungan Universitas Kristen Duta Wacana maupun di Perguruan Tinggi atau instansi manapun, kecuali bagian yang sumber informasinya dicantumkan sebagaimana mestinya.

Jika dikemudian hari didapati bahwa hasil skripsi ini adalah hasil plagiasi atau tiruan dari skripsi lain, saya bersedia dikenai sanksi yakni pencabutan gelar kesarjanaaan saya.

Yogyakarta, 30 April 2012



Yanuar Septian

22 08 4484



**HALAMAN PERSETUJUAN**

Judul : **PENDETEKSI TINGKAT PLAGIASI TEKS  
BERSEGMENTASI PENDEK**

Nama : Yanuar Septian

NIM : 22 08 4484

Mata Kuliah : Tugas Akhir

Kode : TIW276

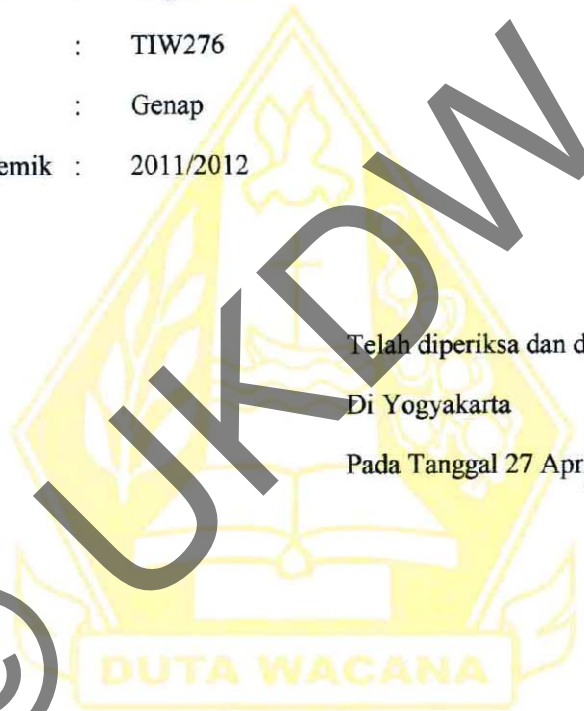
Semester : Genap

Tahun akademik : 2011/2012

Telah diperiksa dan disetujui

Di Yogyakarta

Pada Tanggal 27 April 2012



Dosen Pembimbing I

A handwritten signature in black ink, appearing to read "Lucia", is written over the name of the first supervisor.

Lucia Dwi Krisnawati, S.S., M.A.

Dosen Pembimbing II

A handwritten signature in black ink, appearing to read "R. Gunawan Santosa", is written over the name of the second supervisor.

Drs. R. Gunawan Santosa, M.Si

**HALAMAN PENGESAHAN**

**SKRIPSI**

**PENDETEKSI TINGKAT PLAGIASI TEKS BERSEGMENTASI PENDEK**

Oleh : Yanuar Septian/22 08 4484

Dipertahankan di depan dewan penguji Tugas Akhir/Skripsi  
Program Studi Teknik Informatika Fakultas Teknologi Informasi

Universitas Kristen Duta Wacana – Yogyakarta

Dan dinyatakan diterima untuk memenuhi salah satu

Syarat memperoleh gelar

Sarjana Komputer

Pada tanggal

21 Mei 2012

Yogyakarta, 24 Mei 2012

Mengesahkan,

Dewan penguji:

1. Drs. R. Gunawan Santosa, M.Si.
2. Lucia Dwi Krisnawati, S.S., M.A.
3. Junius Karel Tampubolon, S.Si., M.T.
4. Theresia Herlina Rochadiani, S.Kom., M.T.

  
.....  
  
.....  
.....  
  
.....

Dekan



Drs. Wimmie Handiwidjojo, MIT.

Ketua Program Studi



Nugroho Agus Haryono, S.Si, M.Si.

## UCAPAN TERIMA KASIH

Puji syukur kehadiran Tuhan Yang Maha Esa karena atas rahmat-Nya Tugas Akhir berjudul Pendeteksi Tingkat Plagiasi Teks Bersegmentasi Pendek ini dapat diselesaikan dengan baik. Penulisan laporan dan pembuatan tugas akhir ini diharapkan dapat berguna dan menjadi suatu karya ilmiah yang nantinya terus dikembangkan menjadi lebih baik lagi demi kemajuan teknologi informasi.

Pembuatan tugas akhir ini tidak lepas dari dukungan dan bantuan banyak pihak. Oleh karena itu dengan kerendahan hati, penulis menyampaikan ucapan terima kasih kepada:

1. **Ibu Lidwina** di surga yang telah menghantarku dari sejak aku dalam kandungan hingga masuk gerbang perguruan tinggi. Semangat dan cintanya yang berkobar selalu kurasakan hingga detik ini. Beliauah wanita terhebat bagiku dan malaikatku.
2. **Ayah Harijanto Sugondo** sumber inspirasiku dan **Paskalia** adik tercinta yang selalu menularkan keceriaan.
3. **Ibu Lucia Dwi Krisnawati, S.S., M.A.** selaku dosen pembimbing I yang telah membimbing dengan sabar dan penuh semangat
4. **Drs. R. Gunawan Santosa, M.Si.** selaku dosen pembimbing II atas petunjuk, saran, dan masukan yang membangun
5. Keluarga besar dan kerabat yang menjadi kekuatan
6. Teman-teman terkasih yang tak henti-hentinya mengobarkan semangat dan keluarga seataiku di Yogya: Adrian, Rudi, Popo, Wawan, Dani, Alex, Tegar, Fengky, Robin, Anto, Hengky, Hendra, dan Hugo.
7. Semua pihak yang telah mendukung baik pembuatan tugas akhir ini secara langsung ataupun tidak langsung

Saran dan kritik yang membangun sangat penulis harapkan demi pengembangan dan perbaikan karya ilmiah ini. Jika terdapat kesalahan dalam pembuatan ataupun penulisan laporan tugas akhir ini, penulis memohon maaf. Semoga tugas akhir ini berguna dan mendatangkan manfaat.

Yogyakarta, 30 April 2012

Penulis

© UKDW

## INTISARI

Plagiasi menjadi isu yang memanas seiring dengan kemajuan teknologi informasi dimana pengaksesan karya ilmiah dan intelektual terutama dalam bentuk dokumen teks menjadi semakin mudah. Maurer et al. (2006) mengemukakan bahwa salah satu solusi memecahkan plagiasi yang umum digunakan adalah perbandingan dokumen, yakni melakukan pengecekan kata demi kata antara dokumen yang dicurigai dengan dokumen-dokumen yang dimungkinkan menjadi sumber plagiasi.

Sistem Pendeteksi Tingkat Plagiasi Teks Bersegmentasi Pendek akan menyeleksi maksimal sepuluh dokumen dalam korpus yang diduga menjadi dokumen sumber plagiasi melalui pendekatan *vector space model* dan perhitungan *cosine similarity* guna menghitung kedekatan masukan paragraf dengan dokumen-dokumen dalam korpus. Korpus terdiri dari abstraksi-abstraksi mahasiswa Teknik Informatika Universitas Kristen Duta Wacana. Setelah itu, proses *exhaustive search* dilakukan dengan membandingkan tiap *gram* di masing-masing kalimat masukan dengan tiap *gram* dalam dokumen-dokumen terseleksi guna memperoleh prosentase plagiasi masukan terhadap masing-masing dokumen terseleksi. Nilai  $n$  dalam *gram* yang digunakan adalah lima.

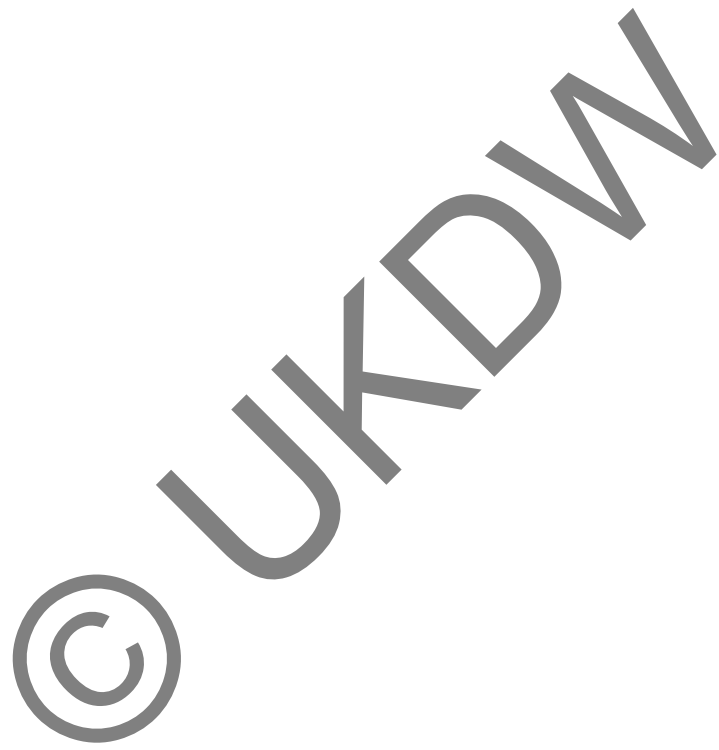
Pendeteksi Plagiasi Teks Bersegmentasi Pendek mampu mendeteksi plagiasi yang tersebar ataupun terpusat dimana fragmen plagiasi terdiri dari lima kata atau lebih secara efektif ketika jumlah dokumen sumber kurang dari sama dengan lima. Proses seleksi awal kurang efektif dalam menunjang keakuratan pendeteksian dan dipengaruhi oleh beberapa parameter, antara lain: proporsi fragmen plagiasi dalam masukan, ada tidaknya kata atau istilah yang khusus yang merujuk ke dokumen sumber tertentu. Pengkajian proses seleksi awal secara lebih lanjut perlu dilakukan.



## DAFTAR ISI

HALAMAN JUDUL.....	
PERNYATAAN KEASLIAN SKRIPSI.....	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN.....	iii
UCAPAN TERIMA KASIH.....	iv
INTISARI.....	vi
DAFTAR ISI.....	vii
DAFTAR TABEL.....	ix
DAFTAR GAMBAR.....	x
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang.....	1
1.2. Perumusan Masalah.....	2
1.3. Batasan Masalah.....	2
1.4. Hipotesis.....	3
1.5. Tujuan Penelitian.....	3
1.6. Metode/Pendekatan.....	3
1.7. Sistematika Penulisan.....	4
BAB II TINJAUAN PUSTAKA.....	6
2.1. Tinjauan Pustaka.....	6
2.2. Landasan Teori.....	8
BAB III ANALISIS DAN PERANCANGAN SISTEM.....	15
3.1. Alat dan Bahan.....	15
3.2. Rancangan Database.....	15
3.3. Pembuatan Korpus.....	16
3.4. Preproses.....	16
3.5. Proses.....	17
3.6. Evaluasi.....	21
3.7. Arsitektur Sistem.....	21

BAB IV IMPLEMENTASI DAN ANALISIS SISTEM..... 23  
4.1. Implementasi..... 23  
4.2. Eksperimen dan Analisis..... 30  
4.3. Kekurangan dan Kelebihan Sistem..... 42  
BAB V KESIMPULAN DAN SARAN..... 43  
7.1.Kesimpulan..... 43  
7.2.Saran..... 43  
DAFTAR PUSTAKA..... 45



## DAFTAR TABEL

Tabel 3.1. Kamus Data Tabel_vsm	15
Tabel 4.1. Total waktu pemrosesan dan panjang masing-masing query	31
Tabel 4.2. Hasil percobaan pendeteksian tingkat plagiasi	33
Tabel 4.3. Rata-rata presisi dan recall untuk masing-masing jumlah dokumen sumber	41
Tabel 4.4. Rata-rata prosentase jumlah dokumen sumber yang terambil pada proses seleksi awal untuk masing-masing jumlah dokumen sumber	41



UKDWM

## DAFTAR GAMBAR

Gambar 2.1. Ilustrasi Dokumen Pelabelan Plagiasi	12
Gambar 2.2. Contoh Query dengan pelabelan	13
Gambar 2.3. Contoh Query dengan hasil pendeteksian sistem	13
Gambar 3.1. Flowchart Preproses Pendeteksi Tingkat Plagiasi Teks Bersegmentasi Pendek	17
Gambar 3.2. Flowchart Perencanaan Proses Pendeteksi Tingkat Plagiasi Teks Bersegmentasi Pendek	20
Gambar 3.3. Arsitektur Sistem Pendeteksi Tingkat Plagiasi Teks Bersegmentasi Pendek	22
Gambar 4.1. Implementasi Pembuatan Korpus Sistem Pendeteksi Tingkat Plagiasi Teks Bersegmentasi Pendek	23
Gambar 4.2. Pseudocode implementasi filterasi Stopword	24
Gambar 4.3. Pseudocode Pengukuran Cosine Similarity sebuah query terhadap dokumen-dokumen sumber	25
Gambar 4.4. Pseudocode Pengukuran Exhaustive Search sebuah query terhadap dokumen-dokumen sumber	26
Gambar 4.5. Pseudocode Evaluasi	28
Gambar 4.6. Website Pendeteksi Tingkat Plagiasi Teks Bersegmentasi Pendek	30

# BAB I

## PENDAHULUAN

### 1.1. Latar Belakang Masalah

Kemajuan teknologi informasi memberikan kemudahan akses atas berbagai informasi termasuk hasil karya intelektual dalam bentuk dokumen, seperti karya penulisan ilmiah. Plagiasi menjadi suatu isu yang muncul sebagai dampak kemudahan pengaksesan karya ilmiah, terutama pada kalangan akademisi dan pelajar, meskipun tidak hanya terbatas pada lingkup tersebut.

Menurut *IEEE PSPB Operations Manual (from section 8.2.1.B.7)* , plagiasi adalah “ ... menggunakan ide , proses, hasil, atau kata-kata orang lain yang telah ada lebih dulu tanpa menyatakan penulis dan sumber aslinya.” (IEEE,2008). Salah satu bentuk plagiasi paling sederhana adalah *verbatim plagiarism* yakni mengambil suatu kalimat atau bagian kalimat secara apa adanya atau identik tanpa mencantumkan penulis dan sumber.

Maurer et al. (2006) mengemukakan bahwa salah satu solusi plagiasi yang umum digunakan adalah perbandingan dokumen, yakni melakukan pengecekan kata demi kata antara dokumen yang dicurigai dengan dokumen-dokumen yang dimungkinkan menjadi sumber plagiasi. Deteksi plagiasi merupakan salah satu upaya untuk mengurangi tingkat plagiasi.

Pendeteksian tingkat plagiasi bagian teks atau teks bersegmentasi pendek menjadi langkah awal yang akan diteliti dalam waktu satu semester dan nantinya diharapkan dapat dikembangkan menjadi pendeteksian tingkat plagiasi lingkup dokumen yang lebih panjang dan luas. Kata Plagiasi pada penelitian ini merujuk pada deret kumpulan kata berurutan yang sama persis dengan deret kumpulan kata berurutan serupa pada sumber tertentu. Pendeteksi tingkat plagiasi teks bersegmentasi pendek akan menghitung tingkat kesamaan *verbatim plagiarism* dalam suatu dokumen berupa paragraf atau bagian teks dari karya ilmiah mahasiswa yang dibuat.

Penelitian ini menggabungkan model pencarian *Vector Space Model* untuk memperoleh pembobotan dan perhitungan *cosine similarity* guna memperoleh peringkat dokumen yang memiliki kesamaan dengan *query* masukan. Selanjutnya, proses *exhaustive search* melalui penggunaan *n-gram* digunakan sebagai sarana untuk menghitung tingkat prosentase kesamaan masukan dengan dokumen-dokumen terseleksi. Penggunaan *n-gram* mampu menangani fragmen teks yang dicampur atau dimodifikasi.

## 1.2. Perumusan Masalah

Penelitian ini berfokus pada seberapa akuratkah *exhaustive search* dengan penggunaan *n-gram* mampu mendeteksi tingkat plagiasi masukan bagian teks atau paragraf dengan dokumen yang ada dalam korpus. Hal ini didukung oleh proses seleksi awal dengan bantuan model pencarian *Vector Space Model* dan pemberian peringkat melalui pengukuran *cosine similarity* terhadap dokumen-dokumen dalam korpus

## 1.3. Batasan Masalah

Pendeteksi tingkat plagiasi bagian teks merupakan pendeteksi *verbatim plagiarism* eksternal atau kesamaan identik dengan masukan berupa satu paragraf pendek. Kata Plagiasi pada penelitian ini merujuk pada deret kumpulan kata berurutan yang sama persis dengan deret kumpulan kata berurutan serupa pada sumber tertentu. Sistem ini akan mengambil sepuluh dokumen dalam korpus sebagai sumber rujukan yang dinilai memiliki kedekatan tertinggi dengan masukan dan menghitung besar prosentase plagiasi masukan terhadap masing-masing dokumen terseleksi tersebut. Korpus merupakan kumpulan dokumen abstraksi tugas akhir mahasiswa Teknik Informatika Universitas Kristen Duta wacana. Sistem diimplementasikan dalam bentuk aplikasi berbasis web dengan masukan dan korpus berupa dokumen teks yang hanya berekstensi . txt.

Pengolahan sistem terbatas pada teks murni, tidak termasuk gambar, tabel, ataupun segala jenis format teks.

#### **1.4. Hipotesis**

Hipotesis Penelitian ( $h_0$ )

Pendeteksi tingkat plagiasi bagian teks mampu mendeteksi tingkat plagiasi dengan akurasi tinggi yakni menghasilkan nilai presisi dan recall lebih dari sama dengan 70% dimana panjang masukan paragraph akan berbanding lurus dengan waktu pemrosesan.

#### **1.5. Tujuan Penelitian**

Penelitian ini bertujuan untuk mendeteksi tingkat plagiasi bagian teks dengan dokumen dalam korpus yang ada secara akurat dengan terlebih dahulu menyeleksi dokumen dalam korpus melalui perankingan. Selain itu, keefektifan metode pendeteksian yang digunakan dalam mendeteksi tingkat plagiasi menjadi fokus penelitian ini. Jika  $m$  adalah masukan paragraf atau bagian teks dan  $K$  adalah korpus berisi kumpulan dokumen abstraksi tugas akhir mahasiswa Teknik Informatika Universitas Kristen Duta Wacana yang pernah dibuat, maka tujuan penelitian ini adalah menjawab perumusan masalah : “Berapa besar prosentase plagiasi  $m$  terhadap  $k \in K$  ?”

#### **1.6. Metode/ Pendekatan**

Berikut ini adalah beberapa metode dan pendekatan yang digunakan dalam Pendeteksi tingkat plagiasi teks bersegmentasi pendek

1.6.1. Penyeleksian dokumen yang memiliki nilai kedekatan tertinggi dengan masukan.

Model pencarian *Vector Space Model* digunakan untuk mencari sepuluh dokumen dalam korpus yang memiliki kedekatan tertinggi dengan *query* masukan paragraf/ bagian teks melalui perhitungan *cosine similarity* sebagai pengukur jarak *query* dan dokumen. Penyeleksian sepuluh dokumen nantinya diharapkan mampu mengurangi waktu komputasi pada proses *exhaustive search*. Selain itu, sepuluh dokumen merupakan jumlah referensi yang cukup banyak untuk masukan sebuah paragraf.

#### 1.6.2. *Exhaustive Search*

Pengukuran tingkat plagiasi masukan paragraf/ bagian teks terhadap sepuluh dokumen terseleksi yang memiliki nilai kedekatan tertinggi dengan *query* masukan dilakukan melalui proses *exhaustive search* berupa perhitungan *n-gram* dengan nilai  $n = 5$ . Prosentase kesamaan diukur dengan rumus [2.5] yang mengadaptasi rumus perhitungan Cedeño, A.B. dan Rosso, P. (2009) ( rumus [2.4]).

#### 1.6.3. Metode evaluasi presisi dan recall

Sejumlah *query* masukan dengan variasi tingkat kemiripan terhadap dokumen-dokumen korpus akan dijadikan sampel uji evaluasi sistem pendeteksi plagiasi teks bersegmentasi pendek. Perhitungan evaluasi yang digunakan adalah *macro averaged / user oriented* (Potthast et.al,2009) melalui perhitungan presisi (rumus [2.6]) dan recall (rumus [2.7]).

### 1.7. **Sistematika penulisan**

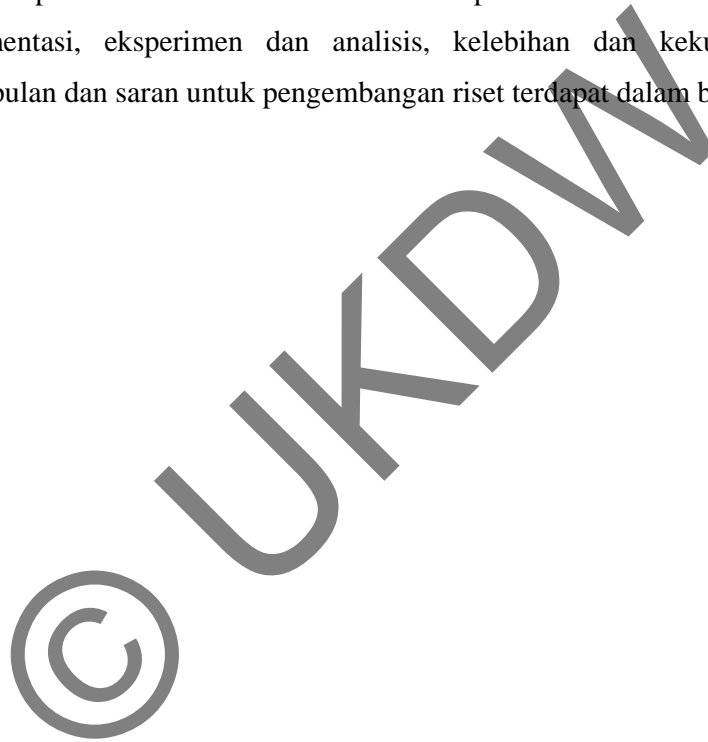
Pendahuluan pada Bab I berisi tentang latar belakang masalah, perumusan masalah, batasan masalah, hipotesis, tujuan penelitian, metode/pendekatan, dan sistematika penulisan.



Tinjauan Pustaka pada Bab II memuat mengenai pembahasan penelitian terdahulu dan sumber pustaka yang menjadi referensi. Landasan teori yang terdapat dalam bab tersebut menjelaskan berbagai teori pendukung, konsep, dan prinsip utama dalam penelitian ini.

Analisis dan Perancangan Sistem pada Bab III berisi perencanaan serta cara perancangan sistem yang meliputi alat dan bahan, rancangan database, pembuatan corpus, preproses, proses, evaluasi, dan arsitektur sistem. Gambaran kerja, cara perancangan dan pembuatan sistem juga dijelaskan dalam bab ini.

Implementasi dan Analisis Sistem pada Bab IV memaparkan hasil implementasi, eksperimen dan analisis, kelebihan dan kekurangan sistem. Kesimpulan dan saran untuk pengembangan riset terdapat dalam bab akhir.



## BAB V

### KESIMPULAN DAN SARAN

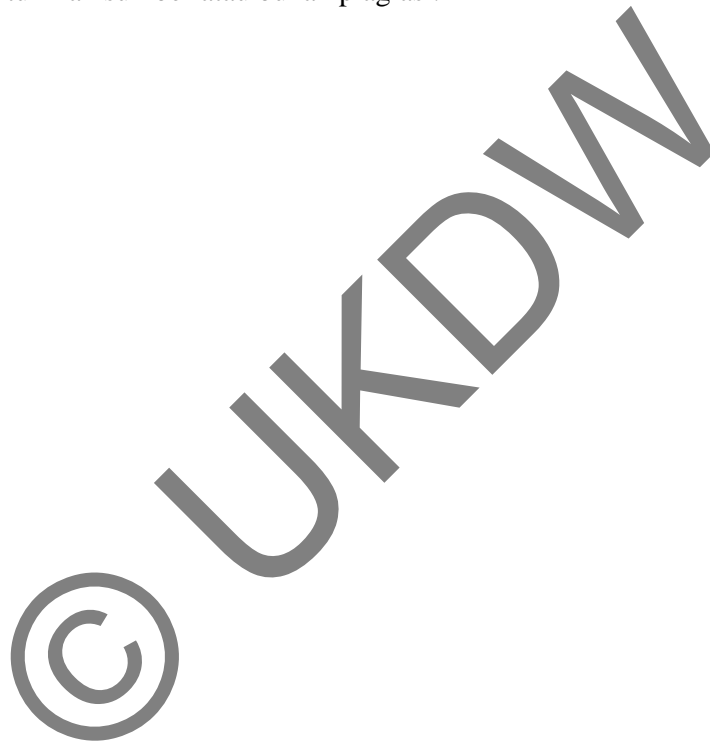
#### 5.1. Kesimpulan

Berdasarkan penelitian dan eksperimen yang dilakukan terhadap sistem pendeteksi tingkat plagiasi teks bersegmentasi pendek, dapat disimpulkan bahwa: (1). Dengan *exhaustive search* menggunakan *5-gram* tanpa mengindahkan pencantuman sumber, Pendeteksi Tingkat Plagiasi Teks Bersegmentasi Pendek mampu mendeteksi tingkat plagiasi secara efektif dengan evaluasi rata-rata nilai presisi sebesar 1 dan rata-rata nilai recall sebesar 0.89 dari tiga puluh empat *query* uji dimana tiap kalimat plagiasi terdiri dari minimal lima kata dan panjang *query* berbanding lurus dengan waktu pemrosesan; (2). Proses seleksi awal dan pembatasan maksimal sepuluh dokumen dengan pendekatan *vector space model* melalui pengukuran *cosine similarity* ternyata kurang efektif terutama saat dokumen sumber berjumlah banyak (lebih dari lima dokumen sumber) dan berpengaruh pada nilai recall yang cenderung menurun; (3). Ketidakefektifan proses seleksi awal dipengaruhi oleh beberapa parameter, yakni: banyaknya *stopword* dalam fragmen plagiasi, kehadiran istilah atau kata khusus yang merujuk pada dokumen sumber tertentu, dan proporsi suatu fragmen plagiasi dalam masukan;

#### 5.2. Saran

Berdasarkan hasil penelitian, proses seleksi awal dengan pembobotan/pendekatan *vector space model* dan penggunaan *cosine similarity* sebagai pengukur jarak kedekatan *query* dan dokumen sumber masih kurang efektif dalam menyeleksi dokumen sumber. Oleh karena itu, pengkajian pada proses seleksi awal untuk meningkatkan nilai recall dapat menjadi upaya yang dilakukan pada penelitian selanjutnya dengan diikuti optimalisasi keseluruhan

algoritma guna mempersingkat waktu komputasi dan pemrosesan. Saat waktu komputasi dapat diminimalisasi, maka proses seleksi awal dapat digunakan untuk mengambil dokumen secara fleksibel sesuai dengan jumlah dokumen sumber dari fragmen-fragmen plagiasi *query*. Pendeteksi Tingkat Plagiasi teks bersegmentasi pendek dapat dikembangkan menjadi Pendeteksi Plagiasi dalam lingkup yang lebih luas, seperti Pendeteksi Plagiasi dokumen dengan lebih efektif. Nilai presisi =1 masih labil dan perlu adanya *post-processing* yang menyaring fragmen-fragmen plagiasi sebenarnya yang tidak mencantumkan sumber dan yang mencantumkan sumber atau bukan plagiasi.



## Daftar Pustaka

- Basile,C., Benedetto, D., Caglioti, E., Cristadoro,G.,& Degli Esposti, M. (2009). A Plagiarism detection procedure in three steps: selection, matches and “squares”. Dalam Stein,B., Rosso, P., Stamatou,E., Koppel, M., Agirre,E. (eds.) *SEPLN 2009 Workshop on Uncovering Plagiarism, Authorship, and Social Software Misuse (PAN 2009)*.CEUR-WS.org. 1-9.
- Cedeño, Alberto B., & Rosso,P.(2009). On Automatic Plagiarism Based on n-Grams Comparison. Dalam *ECIR '09 Proceedings of the 31th European Conference on IR Research on Advances in Information Retrieval*. Heidelberg: Springer-Verlag Berlin, 696-700.
- Dehak, N., Dehak, R., Glass, J., Reynolds, D. & Kenny, P.(2010). Cosine Similarity Scoring without Score Normalization Techniques. Dalam *Proceedings of Odyssey 2010 - The Speaker and Language Recognition Workshop (Odyssey 2010)*. Czech Republic: Bmo, 71-75.
- Dreher,H. (2007). Automatic Conceptual Analysis for Plagiarism Detection. Dalam Eli Cohen (Ed.), *Journal of Issues in Informing Science and Information Technology*. Santa Rosa: The Informing Science Institute, Vol 4, 601-614
- IEEE Publications Services and Product Board Operations Manual.(2010, 12 Februari). IEEE Publications. hlm 58.
- Knight, A., Almeroth, K. & Bimber, B. (2004). An Automated System for Plagiarism Detection Using the Internet. Dalam L. Cantoni & C. McLoughlin (Eds.), *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 2004*. Chesapeake, VA: AACE, 3619-3625.
- Lee, Dik L., Chuang, H., & Seamons, K. (1997). Document Ranking and the Vector-Space Model. Dalam Spark Jones (Ed.), *Journal IEEE Software* . Los Amigos: IEEE Computer Society Press, Vol 14 (2), 39-46.

- Manning, Christopher D., Raghavan,P., & Schütze, H. (2008). *Introduction to Information Retrieval* .New York: Cambridge University Press.
- Maurer, H., Kappe, F. & Zaka, B. (2006). Plagiarism – A survey. *Journal of Universal Computer Science*, Vol 12(8), 1050-1084.
- Pothast, M., Stein, B., Eiselt, A., Barrón-Cedeño, A., Rosso, P. (2009). Overview of the 1st International Competition on Plagiarism Detection. Dalam: Stein B., Rosso P., Stamatatos E., Koppel M., and Agirre E. (Eds.), *SEPLN Workshop on Uncovering Plagiarism, Authorship, and Social Software Misuse*, PAN-2009, 1-9.

