PROGRAM KREATIVITAS MAHASISWA

MODEL MATRIKS DARI DOKUMEN SERTA SUDUT ANTARA DUA SUBRUANG UNTUK MENDUGA DINI PLAGIARISME DOKUMEN

BIDANG KEGIATAN:

PKM PENULISAN ILMIAH

Diusulkan oleh :

MUHAMMAD ILYAS	G54051625/2005
VERAWATI	G54052035/2005
AHDIANI FEBRIYANTI	G54104020/2004

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

INSTITUT PERTANIAN BOGOR

BOGOR

2008

HALAMAN PENGESAHAN PROGRAM KREATIVITAS MAHASISWA

1. Judul Kegiatan : Model matriks dari dokumen serta sudut antara dua subruang untuk menduga dini plagiarisme dokumen
2. Bidang Ilmu (Pilih salah satu) :
   ( ) Kesehatan ( ) Pertanian
   (X) MIPA ( ) Teknologi dan Rekayasa
   ( ) Sosial Ekonomi ( ) Humaniora
   ( ) Pendidikan
3. Ketua Pelaksana Kegiatan/Penulis Utama
   a. Nama Lengkap : Muhammad Ilyas
   b. NIM : G54051625
   c. Departemen : Matematika
   d. Universitas : Institut Pertanian Bogor
   e. Alamat Rumah dan No Tel./HP : Villa Kebun Raya Blok S/23 Ciomas, Bogor. (0251) 631845 / 085691449698
4. Anggota Pelaksana Kegiatan : 2 orang
5. Dosen Pendamping
   a. Nama Lengkap dan Gelar : Dra. Farida Hanum, M.Si.
   b. NIP : 131 956 709
   c. Alamat Rumah dan No Tel./HP : KPP IPB Baranangsiang IV Blok D-33 (0251) 385135

Bogor, 06 Maret 2008

Menyetujui
Sekretaris Departemen/Program Studi

Dr. Ir. I Wayan Mangku, M.Sc.
NIP. 131 663 020

Ketua Pelaksana Kegiatan

Muhammad Ilyas
NIM. G54051625

Wakil Rektor Bidang
Akademik dan Kemahasiswaan,

Prof. Dr. Ir. Yonny Koesmaryono, MS
NIP. 131 473 999

Dosen Pendamping

Dra. Farida Hanum, M.Si.
NIP. 131 956 709

LEMBAR PENGESAHAN SUMBER PENULISAN ILMIAH PKMI

1. Judul Tulisan yang Diajukan : Model matriks dari dokumen serta sudut antara dua subruang untuk menduga dini plagiarisme dokumen.
2. Sumber Penulisan (beri tanda X yang dipilih)

( ) Kegiatan Praktek Lapang / Kerja dan sejenisnya, KKN, Magang, Kegiatan Kewirausahaan (pilih salah satu), dengan keterangan lengkap :

____

(X) Kegiatan Ilmiah lainnya (sebutkan) dengan keterangan lengkap :

1. P.P.Rosliyanti, M.Alban, dan R.Agustian. 2005. Kumpulan Tugas Komputasional Kapita Selekta Matematika Komputasi. Bogor.
2. Hasdiani, R.Agustian. 2007. Kumpulan Tugas Kerja Mandiri Terpantau. Bogor.

Keterangan ini kami buat dengan sebenarnya.

Bogor, 06 Maret 2008

Mengetahui
Sekretaris Departemen/Program Studi

Dr. Ir. I Wayan Mangku, M.Sc.
NIP. 131 663 020

Ketua Pelaksana Kegiatan

Muhammad Ilyas
NIM. G54051625

MODEL VEKTOR DAN MATRIKS DARI DOKUMEN SERTA SUDUT ANTARA DUA VEKTOR DAN DUA SUBRUANG UNTUK MENDUGA DINI PLAGIARISME DOKUMEN

Muhammad Ilyas, Verawati, Ahdiani Febriyanti

Departemen Matematika FMIPA Institut Pertanian Bogor

ABSTRAK

Prinsip kerja mesin pencari (search engine) memanfaatkan konsep dasar aljabar linear yaitu ruang vektor dan dekomposisi matriks. Berdasarkan konsep dasar aljabar linear yang digunakan tersebut, dikembangkan pemodelan dokumen untuk mendeteksi dini terjadinya tindak plagiarisme. Representasi vektor kolom dimanfaatkan dalam pengonversian dokumen input. Konsep lain yang digunakan ialah besar sudut antara dua vektor. Dengan cara menghitung besarnya sudut yang dibentuk oleh dua vektor menjadi pertimbangan untuk menentukan indikasi tindak plagiarisme.

Dalam karya tulis ini diajukan sebuah model alternatif dimana dokumen direpresentasikan sebagai matriks, dengan baris merepresentasikan banyaknya kata berbeda dan kolom merepresentasikan banyaknya paragraf dalam dokumen. Plagiarisme dokumen dapat dideteksi dengan menentukan sudut antara dua subruang yang dibangkitkan oleh vektor kolom masing-masing dokumen yang akan dibandingkan. Diajukan algoritme penghitungan sudut antar subruang yang ekivalen menurut definisi sudut antara dua subruang.

Kata kunci : vektor, matriks, sudut, subruang, aplikasi aljabar linear, text mining, plagiarisme.

PENDAHULUAN

Penerapan mesin pencari (*search engine*) sering digunakan untuk mencari (*surfing*) situs internet tertentu, men*down-load* data maupun gambar. Pengguna (*user*) cukup memasukkan kata kunci (*keyword*) yang diinginkan kemudian situs penyedia *search engine* akan bekerja dan kemudian menampilkan hasil (*result*) yang diinginkan. Dengan adanya mesin pencari seperti ini memudahkan orang untuk memperoleh informasi berupa data maupun dokumen dari internet. Perlu diketahui bahwa prinsip kerja mesin pencari (*search engine*) ini menggunakan konsep dasar aljabar linear, yaitu ruang vektor dan dekomposisi matriks ortogonal (Berry & Browne 1992).

Prinsip kerja mesin pencari (*search engine*) tersebut ternyata dapat pula dimanfaatkan sebagai penduga terjadinya tindak plagiarisme (Rosliyanti et al. 2006). Fakta bahwa banyaknya pengolah kata elektronik (*word processor*) semakin melengkapi fenomena tindak plagiarisme. Orang-orang dapat dengan mudah melakukan *cut-paste* pada suatu dokumen tertentu kemudian membuat dokumen baru dengan hanya melakukan sedikit revisi atau bahkan tidak sama sekali. Fakta inilah yang mendasari dikembangkannya suatu cara yang dapat menduga terjadinya tindak plagiarisme.

Tindakan plagiarisme merupakan tindakan merugikan baik di bidang akademik maupun bidang umum dan terapan lainnya. Jika dikaji lebih lanjut, plagiarisme adalah tindak kriminal meniru hak cipta orang lain. Hal ini tentu saja menimbulkan sejumlah kekhawatiran. Untuk membantu dan memudahkan pekerjaan tim pemeriksa untuk menduga terjadinya tindak plagiarisme maka karya tulis ini mencoba menggali lebih lanjut pemanfaatan konsep dasar aljabar linear yang serupa dengan prinsip kerja mesin pencari (*search engine*) yaitu ruang vektor dan dekomposisi matriks ortogonal (Berry & Browne 1992). Karya tulis ini juga mengembangkan hasil karya ilmiah sebelumnya yang juga mengkaji mengenai cara menduga terjadinya tindak plagiarisme dengan konsep dasar aljabar linear (Rosliyanti et al. 2006). Pada karya ilmiah tersebut dokumen dimodelkan secara matematis sebagai vektor, pada karya tulis ini dokumen dimodelkan dalam bentuk matriks. Dengan kasus yang sama, karya tulis ini membatasi objek amatan pada dokumen berupa teks.

TELAAH PUSTAKA

Penerapan konsep dasar aljabar linear untuk menduga terjadinya plagiarisme bukanlah hal yang baru karena karya tulis ini mengacu pada penulisan ilmiah sebelumnya. Penulisan ilmiah sebelumnya mengacu pada pendekatan teoritik yang mengadopsi konsep representasi vektor, dalam hal ini dokumen yang akan diperiksa, dikonversi terlebih dahulu oleh suatu program menjadi vektor berukuran k x 1, dimana k adalah banyaknya kata yang berbeda pada seluruh dokumen yang akan diperiksa tersebut. Berikut adalah contoh representasi dokumen yang berisi kalimat judul sebagai vektor.

MODEL VEKTOR DAN MATRIKS DARI DOKUMEN SERTA SUDUT ANTARA DUA VEKTOR DAN DUA SUBRUANG UNTUK MENDUGA DINI PLAGIARISME DOKUMEN

Kata	Frekuensi
Model	1
Vektor	2
Dan	2
Matriks	1
Dari	1
Dokumen	2
Serta	1
Sudut	1
Antara	1
Dua	2
Subruang	1
Untuk	1
Menduga	1
Dini	1
Plagiarisme	1

Secara matematis :

$$j = \begin{pmatrix} 1 \\ 2 \\ 2 \\ 1 \\ 1 \\ 2 \\ 1 \\ 1 \\ 1 \\ 2 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

Banyaknya kata berbeda, yaitu 15, artinya representasi kalimat judul sebagai vektor misal j dengan ukuran *(15* x *1)* merupakan subruang di R^{15}.

Untuk menduga bahwa sebuah dokumen telah meniru dokumen yang lain yaitu dengan mengukur sudut antara dua dokumen tersebut (Rosliyanti et al. 2006). Dalam hal ini, ukuran kedua vektor haruslah sama. Pengukuran sudut antara dua buah dokumen diidentikkan dengan pengukuran sudut antara dua buah vektor yaitu dengan rumus sebagai berikut :

$$\cos\theta = \frac{<x, y>}{\|x\|\|y\|}$$

dimana :

- $<x, y> = x^T y$ ialah hasil perkalian dalam vektor **x** dan vektor **y**,

  $x^T y = x_1 y_1 + x_2 y_2 + \ldots + x_n y_n$ (Leon 1998)

- **$\|x\|$ dan $\|y\|$ masing-masing adalah panjang *Euclidean* (*norm*) dari vektor** **x** dan vektor **y**, panjang *Euclidean* (*norm*) suatu vektor diperoleh dari akar penjumlahan kuadrat elemen vektor tersebut (notasi : $\|x\| = \sqrt{a^2 + b^2}$; $x = (\, a\ b\,)^T$) (Leon 1998).
- θ ialah besar sudut yang diapit oleh kedua vektor x dan y.

Dugaan menyontek atau plagiarisme dapat diamati dari besar sudut yang dibentuk oleh kedua vektor tersebut. Jika sudut yang dibentuk kecil maka diduga telah terjadi plagiarisme (tetap dibutuhkan pemeriksaan secara manual setelah pendugaan ini). Tetapi kekurangan pendekatan teoritik semacam ini yaitu terletak pada dokumen input yang dipandang sebagai satu vektor secara keseluruhan. Paragraf dalam dokumen tidak terwakili oleh representasi vektor, sehingga dua paragraf identik dari dua buah dokumen tidak dapat terdeteksi.

Karya tulis ini mencoba memberikan ide baru bahwa dokumen bisa dipandang menurut kumpulan paragraf yang terdapat di dalamnya sehingga representasi vektor yang semula digunakan pada penulisan ilmiah terdahulu (Rosliyanti et al. 2006), dapat ditingkatkan menjadi representasi matriks. Karya tulis ini mengeksplorasi lebih jauh mengenai model ruang matriks yang merupakan perluasan bagi model ruang vektor, seperti dalam *information retrieval model* (Antonellis et al. 2006). Karya tulis ini mencoba membuat representasi dokumen berdasarkan ekstraksi paragraf dalam dokumen. Pada dasarnya suatu dokumen teks terdiri dari satu atau beberapa paragraf. Dalam *Information Retrieval Model* (IRM), suatu dokumen teks dapat "diterjemahkan" ke dalam bentuk suatu vektor. Dokumen yang memuat paragraf tersebut akan dapat dinyatakan sebagai kumpulan vektor yang berupa ruang vektor, atau matriks jika representasi paragraf tersebut berupa vektor kolom. Data atau dokumen terkait akan diidentifikasi melalui operasi vektor sederhana (*simple vector operations*). Prosedur dalam karya tulis ini meliputi pemeriksaan dokumen, dan konversi menjadi vektor berukuran k x d_n, dimana k adalah banyaknya kata yang berbeda pada seluruh dokumen yang akan diperiksa tersebut dan d_n adalah banyaknya paragraf pada dokumen ke-n, n=1,2. Secara matematis dapat ditulis :

$$D1 = \{\, \underline{v}_1^{(1)}, \underline{v}_2^{(1)}, \ldots, \underline{v}_{d1}^{(1)} \,\} := [\, \underline{v}_1^{(1)}\ \underline{v}_2^{(1)} \ldots \underline{v}_{d1}^{(1)} \,]$$

$$D2 = \{\, \underline{v}_1^{(2)}, \underline{v}_2^{(2)}, \ldots, \underline{v}_{d2}^{(2)} \,\} := [\, \underline{v}_1^{(2)}\ \underline{v}_2^{(2)} \ldots \underline{v}_{d2}^{(2)} \,]$$

Untuk menduga bahwa pada dua buah dokumen terdapat plagiarisme maka dapat diukur sudut antara dua dokumen tersebut (Rosliyanti et al. 2006). Berikut diberikan definisi sudut antara dua buah subruang di ruang vektor.

Definisi : (Sudut antara dua Subruang, Gunawan dan Neswan 2005, Gunawan, Neswan dan Setya-Budhi 2005)

$(X, <\,,\,>)$ adalah ruang hasil kali dalam berdimensi 2 atau lebih, $U = \mathrm{span}\{u_1, u_2, \ldots, u_p\}$ dan $V = \{v_1, v_2, \ldots, v_q\}$ adalah 2 subruang di X dimana $1 \le p \le q < \infty$. Asumsikan bahwa $\{u_1, u_2, \ldots, u_p\}$ dan $\{v_1, v_2, \ldots, v_q\}$ adalah ortonormal, maka θ adalah sudut antara 2 subruang U dan V, yang diberikan sebagai berikut.

$$\cos^2 \theta := \det(M^T M)$$

dengan $M := [<u_i, v_k>]^T$ adalah matriks berukuran q x p.

PENDEKATAN TEORITIK

Dalam pembuatan penduga terjadinya plagiarisme dengan menggunakan konsep dasar aljabar linear, ada beberapa pendekatan teoritik yang digunakan. Berikut adalah pendekatan teoritik yang diajukan.

1. Dokumen input dinyatakan sebagai suatu matriks.

   Dasar prinsip kerja suatu mesin pencari (*search engine*) yaitu sebuah vektor mendefinisikan sebuah dokumen untuk setiap termin/kata pada dokumen tersebut (Berry & Browne 1992). Pada karya tulis ini, input berupa dokumen teks akan dikonversi menjadi suatu matriks dan bukan lagi berupa sebuah vektor k x 1 (Rosliyanti et al. 2006), melainkan sebagai Subruang di R^k. Dokumen yang digunakan sebagai input dipandang sebagai sebuah matriks. Misalkan dua buah dokumen direpresentasikan sebagai matriks D1 dan D2, maka :

   $D1 = \{ \underline{v}_1^{(1)}, \underline{v}_2^{(1)}, \ldots, \underline{v}_{d1}^{(1)} \} := [\, \underline{v}_1^{(1)}\ \underline{v}_2^{(1)} \ldots \underline{v}_{d1}^{(1)} \,]$

   $D2 = \{ \underline{v}_1^{(2)}, \underline{v}_2^{(2)}, \ldots, \underline{v}_{d2}^{(2)} \} := [\, \underline{v}_1^{(2)}\ \underline{v}_2^{(2)} \ldots \underline{v}_{d2}^{(2)} \,]$

   dimana:

   $\underline{v}_i^{(n)}$ = representasi paragraf ke-i pada dokumen ke-n, $i = 1,2,3,\ldots,d_n$, $n = 1,2$.

   $\underline{v}_i^{(n)}$ = vektor berukuran k x 1 dimana k merupakan banyaknya kata yang berbeda pada keseluruhan dokumen n yang dijadikan input.

Dalam hal ini nilai k pada dua dokumen input harus sama. Elemen dalam vektor $\underline{v}_i^{(n)}$ berupa frekuensi dari setiap kata k yang berbeda yang terdapat pada tiap paragraf ke-i, i = 1,2,3,..., d_n pada dokumen ke-n, n = 1,2.

2. " Kemiripan" dua buah dokumen dapat diukur dengan "sudut".

Untuk mendapatkan himpunan ortonormal maka dilakukan faktorisasi QR. Teorema yang menguatkan yaitu sebagai berikut.

Teorema : (Faktorisasi QR, Leon 1998)

Jika A adalah sebuah matriks m x n dengan rank n, maka A dapat difaktorkan ke dalam sebuah hasil kali QR, dimana Q adalah sebuah matriks m x n dengan kolom-kolom ortonormal dan R adalah sebuah matriks m x n yang merupakan matriks segitiga atas dan invertible.

Untuk itu, sebelum menentukan besar sudut antara 2 dokumen, dilakukan faktorisasi QR terhadap matriks D1 dan D2 untuk memperoleh himpunan ortonormal. Melalui faktorisasi QR dengan serangkaian transformasi *Householder* akan diperoleh suatu matriks ortogonal yang akan digunakan untuk menentukan besar sudut antara 2 subruang. Langkah-langkahnya sebagai berikut.

➢ Dilakukan faktorisasi QR terhadap matriks D1 untuk memperoleh himpunan ortonormal.

  - Menurut teorema diatas maka D1 dapat dinyatakan sebagai :
    D1 = Qa Ra , dimana $Qa^T = H_{d1}H_{d1-1}\ldots H_1$
  - Dapat diketahui rank penuh dari Ra, yaitu dengan melihat dari banyaknya elemen diagonal utama yang tidak nol. Misal banyaknya elemen diagonal utama yang tidak nol = s1, $1 \le s1 \le d1$.
  - Maka didefinisikan matriks ortogonal Qa1 :
    Qa1 adalah matriks berukuran k x s1 yang diambil dari matriks Qa dengan k baris dan s1 kolom pertama. Vektor-vektor kolom dari Qa1 merupakan basis ortonormal bagi ruang vektor dokumen D1.

➢ Dilakukan faktorisasi QR terhadap matriks D2 untuk memperoleh himpunan ortonormal.

  - Serupa dengan matriks D1 maka D2 dapat dinyatakan sebagai :
    D2 = Qb Rb , dimana $Qb^T = H_{d2}H_{d2-1}\ldots H_1$

- Dapat diketahui rank penuh dari Rb, yaitu dengan melihat dari banyaknya elemen diagonal utama yang tidak nol. Misal banyaknya elemen diagonal utama yang tidak nol = s2, $1 \leq s2 \leq d2$.
- Maka didefinisikan matriks ortogonal Qb2 :

  Qb2 adalah matriks berukuran k x s2 yang diambil dari matriks Qb dengan k baris dan s2 kolom pertama. Vektor-vektor kolom dari Qb2 merupakan basis ortonormal bagi ruang vektor dokumen D2.

➢ Secara matematis :

$Qa1 = \{ \underline{p}_1^{(1)}, \underline{p}_2^{(1)}, \ldots, \underline{p}_{s1}^{(1)} \} := [\, \underline{p}_1^{(1)}\, \underline{p}_2^{(1)} \ldots \underline{p}_{s1}^{(1)} \,]$

$Qb2 = \{ \underline{p}_1^{(2)}, \underline{p}_2^{(2)}, \ldots, \underline{p}_{s2}^{(2)} \} := [\, \underline{p}_1^{(2)}\, \underline{p}_2^{(2)} \ldots \underline{p}_{s2}^{(2)} \,]$

Maka Qa1 adalah himpunan s_1 vektor dan Qb2 adalah himpunan s_2 vektor.

Pandang : $U = \text{Span}\ \{ p_1^{(1)}, p_2^{(1)}, p_3^{(1)}, \ldots, p_{s1}^{(1)} \} \subset R^k$ berdimensi s_1

$V = \text{Span}\ \{ p_1^{(2)}, p_2^{(2)}, p_3^{(2)}, \ldots, p_{s2}^{(2)} \} \subset R^k$ berdimensi s_2

($\subset$ = Subruang)

Karena Qa1 dan Qb2 adalah matriks ortogonal yang diperoleh dari faktorisasi QR dengan serangkaian transformasi Householder, maka himpunan $U = \{ p_1^{(1)}, p_2^{(1)}, p_3^{(1)}, \ldots, p_{s1}^{(1)} \}$ dan $V = \{ p_1^{(2)}, p_2^{(2)}, p_3^{(2)}, \ldots, p_{s2}^{(2)} \}$ merupakan basis ortonormal, maka dapat dihitung besar sudut θ antara dua subruang U dan V dengan menggunakan rumus berikut (Gunawan & Neswan 2005) :

$$\cos^2 \theta := \det\left(M^T M\right) \qquad (1)$$

dimana $M := \left[< p_i^{(1)}, p_j^{(2)} >\right]^T$ adalah matriks berukuran s_2 x s_1, M^T adalah transpose dari matriks M dan < , > melambangkan perkalian dalam (*inner product*) di R^k yaitu $< p_i^{(1)}, p_j^{(2)} >= \left(p_i^{(1)}\right)^T p_j^{(2)}$. (Leon 1998)

Jadi dua dokumen direpresentasikan menjadi dua buah matriks yang berbeda, masing-masing berukuran k x d_1 dan k x d_2, dimana d_1 menyatakan banyaknya paragraf pada dokumen 1 dan d_2 menyatakan banyaknya paragraf pada dokumen 2. Kemudian dilakukan faktorisasi QR terhadap D1 dan D2 untuk memperoleh himpunan ortonormal dan selanjutnya perhitungan dengan formula (1) untuk memperoleh sudut θ yang dibentuk oleh kedua dokumen, dalam hal ini $0 \leq \theta \leq \pi/2$ (Gunawan & Neswan 2005).

3. Dugaan terjadinya plagiarisme diamati dari besar sudut yang dibentuk oleh dua dokumen yang dijadikan input.
   Jika sudut yang dibentuk antara subruang U dan V memiliki nilai yang kecil, maka dapat diduga telah terjadi plagiarisme. Jika $\theta = 0$, maka dua dokumen dikatakan serupa/identik, jika $\theta = \pi/2$, maka dua dokumen dikatakan sama sekali berbeda, dan jika $0 < \theta < \pi/2$, maka diduga bahwa terjadi plagiarisme antara dua dokumen tersebut. Studi komputasional untuk model vektor kriteria plagiarisme terjadi jika $\theta \leq \pi/6$. Hasil berupa besar sudut yang diperoleh masih berstatus dugaan, untuk memastikan telah terjadi atau tidaknya plagiarisme perlu dilakukan lagi uji empiris berupa pemeriksaan manual. (Rosliyanti et al. 2006).

HASIL

Karya tulis ini menggali lebih lanjut penerapan konsep dasar aljabar linear untuk menduga terjadinya plagiarisme. Jika pada awalnya digunakan representasi vektor sebagai acuan (Rosliyanti et al. 2006) maka kali ini representasi matriks dan perhitungan sudut antara subruang di R^n digunakan sebagai ide yang dipandang lebih baik. Pada karya tulis ini diajukan sebuah model yang lebih baik tanpa membuang konsep yang digunakan pada model representasi vektor, atau dengan kata lain model terdahulu masih tetap digunakan.

PEMBAHASAN

Konsep dasar aljabar linear berupa ruang vektor, matriks, dan sudut merupakan alat yang cukup efektif digunakan sebagai model untuk penduga awal terjadinya plagiarisme. Konsep awal yang menggunakan representasi vektor sebagai alat utama penduga diganti dengan representasi matriks, hal ini ditinjau dari hasil dengan penggunaan representasi vektor mengakibatkan tidak terwakilinya paragraf yang terdapat pada sebuah dokumen, kekurangan konsep awal semacam ini yaitu terletak pada dokumen input yang dipandang sebagai satu vektor secara keseluruhan, oleh karena itu diperlukan konsep baru sehingga suatu dokumen dipandang sebagai kumpulan dari beberapa paragraf. Kemiripan dua buah dokumen dapat dilihat dari sudut yang dibentuk antara dua buah Subruang di

R^k, pasangan dokumen yang sudutnya kurang dari sudut batas yang ditentukan diduga memiliki kecenderungan terjadi plagiarisme. Jika $\theta = 0$, maka dua dokumen dikatakan serupa/identik, jika $\theta = \pi/2$, maka dua dokumen dikatakan sama sekali berbeda, dan jika $0 < \theta < \pi/2$, maka diduga bahwa terjadi plagiarisme antara dua dokumen tersebut. Hasil berupa besar sudut yang diperoleh masih berstatus dugaan, untuk memastikan telah terjadi atau tidaknya plagiarisme perlu dilakukan lagi uji empiris berupa pemeriksaan manual.

KESIMPULAN

Penerapan konsep dasar aljabar linear yang digunakan pada prinsip kerja mesin pencari (*search engine*) dapat juga digunakan sebagai penduga terjadinya plagiarisme. Model yang diajukan dimana dokumen direpresentasikan sebagai matriks dengan kolomnya mewakili paragraf pada dokumen tersebut lebih mendekati keadaan sebenarnya jika dibandingkan dengan model dimana dokumen direpresentasikan sebagai vektor. Pada model terakhir ini sudah terbukti secara empiris bahwa penyaringan (*screening*) awal untuk pendugaan terjadinya tindak plagiarisme cukup efektif. Diharapkan bahwa model yang telah diajukan ini dapat menghasilkan kinerja lebih baik menilik model ini jauh lebih detail dari model sebelumnya.

DAFTAR PUSTAKA

Antonellis, I., Gallopoulos, E. (2006). *Exploring Term-Document Matrices from Matrix Models in Text Mining.* Greece: University of Patras.

Berry, M.W., Browne, M. (1992). *Understanding Search Engines: Mathematical Modelling and Text Retrieval.* Philadelphia: Society for Industrial and Applied Mathematics.

Gunawan, H., Neswan, O. (2005). *On Angles Between Subspaces Of Inner Product Spaces. Vol. 11, No. 2.* Indonesia: Math. Society (MIHMI).

Gunawan, H., Neswan, O., Setya-Budhi W. (2005). *"A Formula for Angles Between Two Subspaces of Inner Product Spaces", Beiträge Algebra Geom. Vol. 46(2), 311-320.*

Leon, S.J. (1998). *Aljabar Linear dan Aplikasinya.* Edisi ke-5. A. Bondan, Penerjemah; Erlangga. Terjemahan dari *Linear Algebra with Applications.*

Rosliyanti, P.P., Alban, M., Agustian, R. (2006). *Penerapan Konsep Dasar Aljabar Linear Untuk Menduga Terjadinya Plagiarisme. Makalah PKMI.* Dir.Penelitian dan Pengabdian kepada Masyarakat, Depdiknas. Bogor: IPB.