

APLIKASI ALJABAR LINEAR UNTUK MENDETEKSI ADANYA PLAGIARISME SERTA DALAM METODOLOGI *SEARCH ENGINE*

Sahat Pandapotan Nainggolan^{1,*}, Mukhammad Solikhin², Andi Muhammad Anwar³, Nanda Arista Rizki⁴

¹ Faculty of Vocation, Del Institute of Technology

² Universitas Negeri Malang

³ Universitas Hasanuddin

⁴ Universitas Mulawarman

Email : sahat.nainggolan@del.ac.id (S. P. Nainggolan), mukhammad.solikhin.fmipa@um.ac.id (M. Solikhin), andimuhammadanwar@unhas.ac.id (A. M. Anwar), nanda.arista@fkip.unmul.ac.id (N. A. Rizki)

*Corresponding Author

Abstract

The theory of algebra is highly beneficial and used in daily life. Identifying plagiarism is one of many activities that can be solved with the concept of linear algebra. Even today, plagiarism is still a crucial issue to discuss, preferably in academic scope, both in assignment papers and academic theses. Identifying plagiarism aims to provide a document as a vector. Even the basic concept of linear algebra can be applied in plagiarism detection applications. This paper proposed a modeling example of plagiarism checking on a document with a matrix representation and the calculation of angles among subspaces of each compared document. Finally, the results can be used as one of the considerations to determine the similarity index.

Keywords: plagiarism, linear algebra, vector, matrix representation

Submitted: 10 October 2022; Revised: 17 December 2022; Accepted Publication: 5 January 2023;

Published Online: January 2023

DOI: 10.17977/um055v4i1p1-5

PENDAHULUAN

Teori aljabar pertama kali dikemukakan oleh Abu Abdullah Muhammad Ibn Musa al-Khwarizmi. Aljabar sendiri merupakan salah satu cabang matematika yang merupakan wujud dari dunia formal aksiomatik berdasarkan pembagian dunia dalam matematika menurut Tall, sehingga memerlukan cara berpikir matematik tingkat tinggi (*advanced mathematical thinking*) (Sugiman, 2011: 166). Teori aljabar ini terus berkembang sejalan dengan perkembangan ilmu pengetahuan dan teknologi. Salah satu penggunaannya dalam bidang teknologi yaitu mendeteksi adanya plagiarisme. Salah satu pendeteksiannya dengan menggunakan Algoritma Rabin Karp, Algoritma ini digunakan karena sangat efektif untuk pencarian lebih dari satu kata (multi pattern) [1].

Plagiarisme masih merupakan isu yang sering dibahas, apalagi dalam ruang lingkup akademis, baik dalam pembuatan makalah tugas mahasiswa, maupun skripsi mahasiswa. Dengan semakin berkembangnya teknologi yang semakin memudahkan kita untuk saling bertukar informasi, hal ini menyebabkan mudahnya kita mendapatkan informasi yang kita inginkan melalui mesin pencari. Melalui kemudahan inilah, timbul tindakan copy dan paste dari berbagai informasi yang didapat, dan diakui sebagai karya sendiri. Plagiarisme dapat mematikan kreatifitas seseorang karena sudah terbiasa mengambil sesuatu yang bukan miliknya. Hal ini akan menimbulkan kecenderungan sikap malas dan tidak mau berfikir. Praktik menyalin beberapa bagian atau keseluruhan tulisan tanpa mencantumkan sumber menjadi hal biasa yang sering ditemukan pada pengumpulan makalah ataupun skripsi. Pencegahan dan pendeteksiannya merupakan cara yang dapat dilakukan untuk mengurangi plagiarisme.

Pencegahan berarti menghalangi munculnya plagiarisme yang lebih ditekankan kepada moral masyarakat dan sistem pendidikan. Cara ini akan memberikan efek jangka panjang yang luar biasa. Pendeteksian berarti cara untuk mengungkapkan plagiarisme. Hal ini lah yang menjadi ide untuk mencegah makalah makalah atau skripsi mahasiswa dari praktik plagiarisme.

Berangkat dari hal itu, salah satu upaya untuk mendeteksi secara dini tindak plagiarisme adalah dengan menggunakan konsep aljabar linier khususnya materi *Vector Space Modelling* (Model Ruang Vektor). Di sisi lain, dalam proses pencarian menggunakan search engine kita juga bisa memodelkan sistem search engine tersebut menggunakan konsep yang serupa. Sehingga menggunakan *Vector Space Modelling* kita tidak hanya bisa mengetahui secara dini terkait dengan tindak plagiarisme namun juga bisa mengetahui bagaimana suatu sistem *search engine* bekerja.

METODE

Secara singkat, cara kerja untuk mendeteksi adanya plagiarisme dari suatu dokumen dibandingkan dengan dokumen lain adalah dengan mengimplementasikan dokumen sebagai vektor. Pertama, mencari frekuensi kemunculan suatu kata pada dokumen, kemudian melakukan hal yang sama dengan dokumen lain yang akan diuji plagiarismenya. Setelah itu, sajikan hasil kemunculan kata tadi dalam sebuah vektor berukuran sama. Lalu kita cari *cosine similarity*-nya. Lalu, kesamaan dua vektor tersebut dapat dilihat dari cosines sudut atau sudut yang dibentuk dari dua vektor tersebut. Jika sudut yang dibentuk adalah 90° maka dua vektor tersebut tegak lurus, atau artinya dua dokumen tersebut berbeda. Namun jika sudut yang dibentuk adalah 0° , maka dua vektor tersebut berhimpit artinya dokumen tersebut sama persis atau plagiat total.

HASIL DAN PEMBAHASAN

Salah satu penerapan konsep aljabar linear adalah model ruang vektor. Model ruang vektor ini awalnya diaplikasikan pada sistem temu-balik informasi (*Information Retrieval System*).

Dalam hal ini, dokumen yang akan diperiksa dikonversi dulu menjadi vektor berukuran $n \times 1$, dengan n adalah banyaknya kata berbeda yang ada dalam dokumen sebagai kamus kata (*vocabulary*) atau indeks kata (*term index*). Kata-kata tersebut akan membentuk suatu vektor berdimensi n dalam ruang vektor R^n . Setiap dokumen atau *query* dinyatakan sebagai vektor $w = (w_1, w_2, w_3, \dots, w_n)$ dalam R^n .

Berikut adalah contoh representasi dokumen sebagai berikut:

SAYA CINTA SUMATERA DAN SAYA ADALAH MAHASISWA INSTITUT TEKNOLOGI DEL YANG CINTA TEKNOLOGI

Kata	Frekuensi
Saya	2
Cinta	2
Sumatera	1
Dan	1
Adalah	1
Mahasiswa	1
Institut	1
Teknologi	2
Del	1
Yang	1

Secara matematis, Q dapat ditulis dengan:

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

Banyaknya kata berbeda ada 10 , artinya representasi kalimat judul sebagai vektor misal Q dengan ukuran 10×1 . Vektor Q merupakan elemen di R^{10} . Selanjutnya diperlukan dokumen uji sebagai patokan tingkat plagiasinya, misalkan D dengan ukuran yang sama yaitu 10×1 . Setelah itu, dilakukan perhitungan terhadap sudut yang terbentuk dari kedua vektor tersebut menggunakan *cosine similarity*.

Misalkan ada 2 dokumen uji baru, sebut saja D_1 dan D_2 dengan detail sebagai berikut:

$$D_1 = \begin{pmatrix} 1 \\ 2 \\ 5 \\ 3 \\ 1 \\ 2 \\ 3 \\ 2 \\ 1 \\ 1 \end{pmatrix}, D_2 = \begin{pmatrix} 0 \\ 5 \\ 3 \\ 2 \\ 0 \\ 0 \\ 1 \\ 1 \\ 5 \\ 2 \end{pmatrix}$$

Seperti yang telah dipaparkan sebelumnya, jika $D_1 = (1, 2, 5, 3, 1, 2, 3, 2, 1, 1)$ artinya dalam dokumen tersebut terdapat kata “saya” sebanyak 1 buah, kata “cinta” sebanyak 2 buah, dan seterusnya.

Hasil perhitungan didapatkan sebagai berikut:

$$Q \cdot D_1 = (2)(1) + (2)(2) + (1)(5) + \dots + (1)(1) + (1)(1) = 26$$

$$Q \cdot D_2 = (2)(0) + (2)(5) + (1)(3) + \dots + (1)(5) + (1)(2) = 25$$

$$\|Q\| = \sqrt{2^2 + 2^2 + 1^2 + \dots + 1^2 + 1^2} = \sqrt{19}$$

$$\|D_1\| = \sqrt{1^2 + 2^2 + 5^2 + \dots + 1^2 + 1^2} = \sqrt{58}$$

$$\|D_2\| = \sqrt{0^2 + 5^2 + 3^2 + \dots + 5^2 + 2^2} = \sqrt{69}$$

$$sim(Q, D_1) = \cos \theta_1 = \frac{Q \cdot D_1}{\|Q\| \cdot \|D_1\|} = \frac{26}{\sqrt{19}\sqrt{58}} = \frac{26}{\sqrt{1102}}$$

$$\theta_1 = \cos^{-1}\left(\frac{26}{\sqrt{1102}}\right) = 38.44^\circ$$

$$sim(Q, D_2) = \cos \theta_2 = \frac{Q \cdot D_2}{\|Q\| \cdot \|D_2\|} = \frac{25}{\sqrt{19}\sqrt{69}} = \frac{25}{\sqrt{1311}}$$

$$\theta_2 = \cos^{-1}\left(\frac{25}{\sqrt{1311}}\right) = 46.33^\circ$$

Berdasarkan hasil perhitungan diperoleh bahwa dokumen Q memiliki beda sudut yang dibentuk dengan dokumen 1 sebesar 38.44° , sedangkan dengan dokumen 2 memiliki beda sudut sebesar 46.33° . Ini artinya dokumen Q lebih mirip dengan dokumen D_1 dibandingkan dengan dokumen D_2 .

Vector Space Model pada *Search Engine*

Model ini menerjemahkan *articles* dan *query* sebagai vektor dengan menggunakan keywords. Kemudian dianalisis hubungan antara vektor *articles* dan vektor *query* untuk memunculkan hasil yang diminta *user*. Secara konsep menggunakan hal yang sama pada *cosine similarity* pada aljabar linier. Namun pada *Vector Space Model* menggunakan sebuah vektor berukuran $m \times n$ untuk mengkategorisasikan dokumen, dengan m adalah banyak *keywords* yang ada dalam koleksi dan n adalah banyak dokumen yang ada dalam koleksi.

Contoh:

Matriks berikut disusun berdasarkan artikel dalam majalah “*Biking Monthly*”

$$\begin{matrix} & d_1 & d_2 & d_3 & d_4 & d_5 & d_6 & d_7 & d_8 \\ \text{Road} & (1 & 0 & .5 & .25 & 0 & .33 & 0 & .7) \\ \text{Mountain} & (0 & .33 & 0 & .5 & 1 & .33 & 0 & .2) \\ \text{Touring} & (0 & .66 & .5 & .25 & 0 & .33 & 1 & .1) \end{matrix}$$

Dokumen 1 (d_1) artinya hanya tentang “*road bikes*”, sedangkan dokumen 6 (d_6) berisi tentang semua jenis sepeda. *Vector Space Model* menggunakan nilai di setiap jenis dokumen di atas untuk membuat vektor pada ruang 3 dimensi. Misalkan dengan menuliskan *keywords* “*road biking*” maka akan diterjemahkan dalam suatu vektor, misalkan $\mathbf{b} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$.

Vektor dalam ruang yang paling dekat dengan vektor \mathbf{b} tersebut adalah vektor yang paling berhubungan (dokumen yang paling relevan). *Vector Space Model* dibuat dengan menghitung sudut antara masing-masing vektor dokumen dengan vektor *query*. Perhitungan sudut tersebut menggunakan *cosine similarity* seperti yang telah dibahas di atas. Ketika sudut yang dihasilkan adalah 0° maka dokumen dan *query* benar-benar sangat relevan. Begitupun sebaliknya, jika sudut yang didapat adalah 90° maka kedua dokumen tersebut sangat tidak relevan atau berbeda.

PENUTUP

Teori aljabar linear dapat diaplikasikan sebagai dasar untuk melihat kecocokan antara dua buah dokumen terkait dengan plagiarisme. Selain itu juga, dapat diaplikasikan untuk metodologi *seacrh engine* dalam menemukan dokumen yang relevan. Dalam materi plagiarism, jika sudut yang dibentuk adalah 0^0 maka dua dokumen tersebut identik, sebaliknya jika sudut yang dibentuk adalah 90^0 maka kedua dokumen tersebut sangat berbeda. Jika sudut θ berada pada $0^0 < \theta < 90^0$ maka dikategorikan terdapat plagiarisme. Dengan catatan semakin kecil sudutnya maka tingkat plagiarismenya semakin besar. Pada *search engine*, jika sudut yang dibentuk adalah 0^0 maka antara *query* dan dokumen yang dicari adalah relevan. Sebaliknya jika sudut yang dibentuk adalah 90^0 maka antara *query* dan dokumen yang dicari sangat berbeda atau tidak relevan.

DAFTAR RUJUKAN

Karp, Richard M., Michael O. Rabin. 1987. Efficient Randomized Pattern-Matching Algorithms. *IBM Journal of Research and Development*, 31(2), 249-260.

Ridhatillah, Ardini . 2003. Dealing with Plagiarism in the Information System Research Community: A Look at Factors that Drive Plagiarism and Ways to Address Them, *MIS Quarterly*; 27(4), 511- 532.

KBBI, 1997: 775.

- Marsigit, Erliani, E., Dhoruri, A. & Sugiman. (2011). *Matematika 2*. Jakarta: PT. Quadra Inti Solusi.
- Sastroasmoro, Sudigdo (2014). *Dasar-Dasar Metodologi Penelitian Klinis*. Jakarta: Sagung Seto.
- Sentosa, Johan. *Aplikasi Model Ruang Vektor dan Matriks untuk Mendeteksi Adanya Plagiarisme*. 2015. STEI-ITB.
- Solikhin, M., Nainggolan, S.P., & Fitriyaningsih, I.(2022). Aplikasi Invers Matriks Diperluas (Pseudoinvers) Pada Kriptografi Cipher Hill Atas Lapangan \mathbb{Z}_97 . *JKMA: Jurnal Kajian Matematika dan Aplikasinya*, 3(2), 26-32.
- Anton, Howard & Rorres, Chris. *Elementary Linier Algebra 11th-edition*. 2013. ISBN: 978-1-119-43441-3. Library of Congress Cataloging-in-Publication data: USA.
- Munir, Rinaldi. 2015. *Aplikasi Aljabar Vektor pada Sistem Temu- balik Informasi*. Bahan Kuliah IF2123 Aljabar Geometri. Bandung: Program Studi Teknik Informatika Sekolah Teknik Elektro dan Informatika Institut Teknologi Bandung.
- Sudigdo, Sastroasmoro. 2007. *Beberapa Catatan Tentang Plagiarisme*, Majalah Kedokteran Indonesia.