



IMPLEMENTASI *LATENT SEMANTIC ANALYSIS METHOD* UNTUK MENDETEKSI KEMIRIPAN KALIMAT ESAI PADA UJIAN ONLINE TEKNIK INFORMATIKA UNIVERSITAS HALU OLEO

Muh Budi Dharmawan P.^{*1}, Bambang Pramono², Statiswaty³, Muhammad Ihsan Sarita⁴

^{1,2,3,4}Jurusan Teknik Informatika, Fakultas Teknik Universitas Halu Oleo, Kendari

e-mail: ^{*1}budidharmawanp@gmail.com, ²bambangpramono09@gmail.com, ³statiswaty@uho.ac.id,

⁴ihsansarita@yahoo.co.id

Abstrak

Dunia pendidikan sangat memperhatikan pembentukan dan penanaman karakter peserta didik, sebab hal ini merupakan tolak ukur keberhasilan sebuah metode pendidikan. Untuk mengetahui kecenderungan plagiarisme saat ujian, maka perlu penanaman sistem yang dapat mendeteksi kemiripan jawaban peserta didik saat melaksanakan ujian *online*.

Latent Semantic Analysis Method merupakan metode yang dapat memecahkan masalah plagiarisme dengan cara menghitung bobot jawaban peserta ujian dan pengukuran similaritas atau kesamaan jawaban peserta ujian terhadap peserta ujian lainnya. Penelitian ini menunjukkan lama kerja sistem yang telah diuji dengan parameter jumlah peserta ujian dan jumlah kata yang berbeda. Penelitian ini menunjukkan bahwa setiap penambahan 3 peserta dan 3 kata yang berbeda pada masing masing jawaban akan menghasilkan lama kerja sistem yang cenderung meningkat dengan rata-rata peningkatan berkisar antara 0.118 - 0.197 *second*.

Kata kunci; *E-Learning, Esai Test, Plagiarisme, Deteksi Kemiripan, Latent Semantic Analysis*

Abstract

The world of education is able to pay attention to the formation and planting of student characters, because this is a measure of the success of the educational method. To find out the plagiarism tendencies during the exam, it is necessary to plant a system that can detect similarities in student answers when trying online exams.

The Latent Semantic Analysis method is a method that can solve the problem of plagiarism by calculating the weight of the answers of the examinees and measuring the similarity or similarity of the answers of the examinees to other examinees. This study shows the duration of the system that has been tested with the parameters of the number of examinees and the number of different words. This study shows that each addition of 3 participants and 3 different words in each answer will result in a long working system that tends to increase with an average increase in the range of 0.118 - 0.197 seconds.

Keywords; *E-Learning, Essay Test, Plagiarism, Detection of Similarities, Latent Semantic Analysis*

1. PENDAHULUAN

Dalam perkembangannya, internet telah banyak membawa dampak perubahan sebagian besar aspek kehidupan, hal ini dibuktikan dengan tingginya angka penggunaan

internet di Indonesia. Pengguna internet di Indonesia tumbuh 8,9 persen dan mencapai 20 juta orang ditahun 2014. Kementerian Teknologi dan Informatika (KOMINFO) mengeluarkan hasil riset sebuah lembaga *e-Marketer* yang mengatakan populasi *netter* tanah air mencapai



83,7 juta orang [1]. Pada tahun 2017 pengguna internet di Indonesia meningkat 8% menjadi 143,26 juta pengguna. Jika dibandingkan dengan data populasi penduduk Indonesia yang sebanyak 262 juta jiwa, maka pengguna internet di Indonesia sudah mencapai 54,68% dari total populasi di Indonesia. Perkembangan ini membawa pengaruh terhadap berbagai aspek peradaban, mulai dari sosial, pendidikan, kesehatan, administrasi, dan masih banyak lagi [2].

Pendidikan di tanah air juga tidak luput dari pengaruh internet, hal ini dilihat dari banyaknya keikutsertaan internet dalam aspek-aspek pendidikan, mulai dari administrasi-administrasi pendidikan, sampai dengan model pembelajaran yang telah *ter-mordenisasi* menggunakan internet. Adanya mata pelajaran Teknologi Informasi dan Komunikasi (TIK), juga membuktikan betapa pentingnya mempelajari kemajuan teknologi khususnya internet itu sendiri. Salah satu bentuk pemanfaatan internet dalam dunia pendidikan adalah dengan hadirnya model pembelajaran *e-learning*. Pada awal mulanya penerapan *e-learning* yaitu pada saat memanfaatkan *disk* untuk menyampaikan materi melalui video, banyak orang berpendapat bahwa *e-learning* mirip dengan televisi [3].

Saat ini mulai banyak yang tertarik untuk mengimplementasikan sebuah kelas konvensional kedalam bentuk *virtual classroom* (VC). Ujian online merupakan bentuk *virtual classroom*. Dalam pemanfaatannya ujian online sering digunakan untuk menghilangkan keterbatasan berupa ruang kelas untuk ujian dan dengan memberikan kemudahan akses oleh peserta ujian. Namun disisi lain, pemanfaatan internet dalam *virtual classroom* memberikan dampak kerentanan terhadap kecurangan saat ujian online dilaksanakan, tindakan curang ini dilakukan oleh peserta ujian seperti dalam ujian *essay*. Penggunaan *virtual classroom* saat ujian online akan memberikan kesempatan kepada peserta ujian untuk menjiplak/mencontek pekerjaan *essay* peserta lainnya karena ketiadaan pengawasan pelaksanaan ujian secara fisik. Tentu sebagai sebuah lembaga pendidikan yang menjunjung tinggi kejujuran, kerentanan dari ujian online ini akan membawa dampak kemerosotan akhlak dan kemampuan intelektual peserta ujian.

Dalam pengembangan sebuah *virtual classroom* dapat dilakukan pendeteksian untuk

mengindikasikan tindak kecurangan yang terjadi saat ujian online, tentu hal ini berkaitan dengan implementasi metode yang sesuai terhadap sistem *e-learning*. Salah satu metode yang dapat digunakan dalam mendeteksi kemiripan jawaban *essay* adalah *Latent Semantic Analysis Method (LSA)*. Metode ini merupakan sebuah teori dan metode untuk menggali dan mempresentasikan konteks yang digunakan sebagai sebuah arti kata dengan memanfaatkan komputasi statistik untuk sejumlah *corpus* yang besar [4].

2. METODE PENELITIAN

2.1 E-Learning

E-learning (electronik learning) atau dalam bahasa Indonesia yang berarti “pembelajaran elektronik”, merupakan sebuah fasilitas elektronik yang dewasa ini sudah menjadi bagian dari gaya pembelajaran hampir disemua lembaga atau instansi pendidikan. *E-learning* saat ini sering dikaitkan dengan internet dan *website*, padahal sebenarnya *e-learning* tidak hanya sebatas penggunaan internet dan *website* semata, namun sejatinya *e-learning* adalah penggunaan alat elektronik dalam proses belajar mengajar.

Para ahli memberikan pernyataan yang berbeda-beda mengenai pengertian *e-learning*, namun mengarah pada pokok yang sama. Menurut Rusman (2012) *e-learning* merupakan segala aktivitas belajar yang menggunakan bantuan teknologi elektronik. Melalui *e-learning*, pemahaman siswa tentang sebuah materi tidak tergantung pada guru/instruktur tetapi dapat diperoleh dari media elektronik. Teknologi elektronik yang banyak digunakan misalnya internet, tape video atau audio, penyiaran melalui satelit, televisi interaktif, *virtual reality* serta CD-ROM. Menurut Daryanto (2010) *e-learning* merupakan sistem pembelajaran yang dapat membantu kegiatan pembelajaran dengan memanfaatkan media elektronik. Dari beberapa pengertian dari para ahli tersebut memusatkan pengertian *e-learning* pada sistem pembelajaran yang memanfaatkan penggunaan media elektronik [5].

2.2 HTML

HTML yang merupakan singkatan dari *Hypertext Markup Language* adalah serangkaian kode program yang merupakan dasar dari representasi visual sebuah halaman

web. Didalamnya berisi kumpulan informasi yang disimpan dalam *tag-tag* tertentu, dimana *tag-tag* tersebut digunakan untuk melakukan format terhadap informasi yang dimaksud. Berbagai pengembangan telah dilakukan terhadap kode HTML dan telah melahirkan teknologi-teknologi baru di dalam dunia pemrograman web. Kendati demikian, sampai sekarang HTML tetap berdiri kokoh sebagai dasar dari bahasa web seperti PHP, ASP, JSP dan lainnya. Bahkan secara umum, mayoritas situs web yang ada di internet pun masih tetap menggunakan HTML sebagai teknologi utama [6].

2.3 Python

Python merupakan salah satu bahasa pemrograman yang populer di dunia kerja Indonesia. Python juga merupakan bahasa pemrograman yang relatif mudah untuk pemula. Selain itu di akademik pun banyak akademisi yang menggunakan python untuk menyelesaikan penelitiannya di bidang robotika, *data science*, ekonomi, antariksa dan berbagai macam bidang lainnya. Python secara *default* telah terpasang di beberapa sistem operasi berbasis Linux seperti Ubuntu, Linux Mint, dan Fedora. Untuk sistem operasi lain, sudah tersedia *installer* yang disediakan untuk sistem operasi tersebut [6].

2.4 Algoritma Latent Semantic Analysis

Latent Semantic Analysis (LSA) adalah suatu teori atau algoritma untuk mengekstrak dan merepresentasikan kesamaan makna dari kata-kata menggunakan perhitungan statistika yang diterapkan pada keseluruhan teks. LSA menggunakan metode aljabar linear yaitu *Singular Value Decomposition* (SVD). LSA akan membentuk matriks yang merepresentasikan asosiasi antara term-dokumen yang merupakan *semantic space*, yakni kata-kata dan dokumen yang berasosiasi dekat akan ditempatkan dekat satu sama lain [7].

Metode *Latent Semantic Analysis* (LSA) memiliki beberapa langkah dalam mendeteksi kemiripan teks pada beberapa dokumen string antara lain:

1. Parsing Text

Parsing adalah sebuah proses yang dilakukan seseorang untuk menjadikan sebuah kalimat menjadi lebih bermakna atau berarti dengan cara memecah kalimat tersebut menjadi kata-kata atau frase. *Parsing* di dalam

pembuatan aplikasi dokumen yang semula berupa kalimat-kalimat berisi kata-kata dan tanda pemisah antar kata seperti titik (.), koma (,), spasi dan tanda pemisah lainnya menjadi kata-kata saja baik itu berupa kata-kata penting maupun tidak penting. *Parsing text* dibagi menjadi tiga bagian seperti:

a. Tokenizing

Tokenizing merupakan proses mengidentifikasi unit terkecil (*token*) dari suatu struktur kalimat. Tujuan dilakukannya *tokenizing* ini adalah untuk mendapatkan *term-term* yang nantinya akan diindeks. Pengklasifikasian *token* dilakukan untuk teks yang dipisahkan dengan spasi atau enter adalah suatu dokumen.

b. Filtering

Filtering merupakan proses dimana *token-token* yang didapat dari proses *tokenizing* akan diseleksi dari *token-token* yang dianggap tidak penting (*stoplist*). *Stoplist* merupakan kata yang sering muncul dan bisa diabaikan pada proses *filtering*.

c. Stemming

Stemming adalah suatu proses yang bertujuan untuk mengambil kata dasar dari kata yang berimbuhan atau kata tunggal dari kata bentukan. Hal itu mengurangi jumlah *term* yang berbeda dalam koleksi.

2. Singular Value Decomposition (SVD)

Singular Value Decomposition (SVD) adalah satu metode untuk memecahkan masalah-masalah matematik linier. Pada SVD, matriks akan didekomposisi menjadi tiga komponen matriks. Komponen matriks pertama mendeskripsikan entitas baris sebagai vektor orthogonal matriks. Komponen matriks kedua mendeskripsikan matriks diagonal yang berisi nilai skalar dan yang ketiga adalah matriks entitas kolom sebagai vektor orthogonal matriks.

Metode SVD berdasarkan pada teori aljabar linier yang menyatakan bahwa suatu matriks A yang berukuran $m \times n$, mempunyai nilai singular yang merupakan akar pangkat dua *eigenvalue* $A^T A$.

Adapun rumus yang digunakan sebagai berikut:

$$A_{mn} = U_{mm} \times S_{mn} \times V_{nn}^T \quad (1)$$

Dimana:

A : matriks yang didekomposisi

U : matriks ortogonal U (matriks vektor singular kiri)
 S : matriks diagonal S (matriks nilai singular)
 V : *transpose* matriks ortogonal V (matriks vektor singular kanan)
 m : jumlah baris matriks
 n : jumlah kolom matriks

Setelah memperoleh 3 matriks dari proses SVD, proses berikutnya yang perlu dilakukan adalah mereduksi dimensi dari matriks dengan cara mengurangi dimensi dari matriks kedua yang berupa matriks diagonal. Pengurangan dimensi matriks diagonal ini dilakukan dengan cara mengeset semua nilai diagonal matriks ke dua menjadi nol kecuali diagonal dimensi yang dipilih. Dan jika ketiga komponen matriks tersebut dikalikan maka menghasilkan matriks rekonstruksi yang lain dengan tujuan untuk nilai korelasi yang diinginkan.

Jika nilai-nilai singular dari matriks S diurutkan berdasarkan nilainya, maka k nilai terbesar dapat tetap disimpan dalam matriks tersebut, dan nilai-nilai lain yang lebih kecil dapat diset menjadi nol. Jika baris dan kolom yang berkaitan pada matriks U dan V juga diset menjadi nol, maka hasil kali dari ketiga matriks ini akan membentuk matriks baru yang merupakan matriks *least square approximation*.

Dengan menghapus elemen yang tidak menunjukkan arti, berarti menghapus *noise* yang berada pada vektor pada matriks. Sehingga vektor menjadi lebih pendek dan mengandung elemen yang sangat signifikan dengan data awal saja. Karena tujuan dari LSA bukanlah mendeskripsikan vektor secara verbal tetapi mampu untuk merepresentasikan *term*, dokumen dan menghasilkan *query* yang tidak ambigu dan beredundansi dengan sesama *term*, dimensi nilai k untuk mereduksi matriks bisa menggunakan dua atau tiga vektor saja.

3. Pengukuran kemiripan (*similarity*)

Salah satu pengukuran kemiripan (*similarity*) adalah dengan menghitung sudut antara dua vektor yang bersangkutan, disini antara vektor dari permintaan (*query*) dengan vektor dokumen yang akan dinilai (*document*). Adapun rumus yang digunakan yaitu rumus *cosine similarity*:

$$SV = \cos \theta = \frac{AB}{\|A\| \|B\|} \quad (2)$$

Dimana:

SV : *Similarity Value*

A : vektor A

B : vektor B

$\|A\|$: panjang vektor A

$\|B\|$: panjang vektor B

Antara vektor A dan B membentuk sudut α dimana \cos dari sudut tersebut menunjukkan kemiripan (*similarity*) dari kedua sudut tersebut, bukan menunjukkan jarak (*distance*) antara vektor yang satu dengan yang lainnya.

2.5 Singular Value Decomposition

Pada string kunci jawaban dan jawaban siswa telah melalui tahap *preprocessing*. Selanjutnya dilakukan pencocokan *string* untuk menghitung nilai kemiripan antara kunci jawaban dan jawaban siswa.

Tabel 1 Perhitungan bobot kemunculan kata

	Q	D1	D2	D3
Mampu	1	1	1	1
Mata	2	1	2	1
Cembung	1	1	0	0
Pipih	1	1	1	1
Lensa	1	1	1	1
Lihat	1	0	1	0
Objek	1	0	1	0
Daya	0	0	1	0
Tebal	0	0	1	1

Dibutuhkan nilai *eigen* dan nilai vektor terlebih dahulu. Dimana untuk mencari nilai *eigen* dan nilai vektor, dimulai dengan *transpose* matriks awal.

$$A^T = \begin{bmatrix} 1 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \end{bmatrix}$$

Selanjutnya mencari nilai matriks awal yang dikalikan dengan matriks *transpose* kemudian sebaliknya.

$$A^T A = \begin{bmatrix} 1 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \end{bmatrix}$$

$$A^T A = \begin{bmatrix} 7 & 4 & 4 & 4 \\ 4 & 5 & 2 & 2 \\ 4 & 2 & 5 & 3 \\ 4 & 2 & 3 & 5 \end{bmatrix}$$

$$AA^T = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 & 1 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \end{bmatrix}$$

$$AA^T = \begin{bmatrix} 3 & 1 & 2 & 2 & 2 & 1 & 3 & 1 & 2 \\ 1 & 2 & 1 & 0 & 2 & 1 & 2 & 0 & 1 \\ 2 & 1 & 3 & 2 & 2 & 2 & 3 & 1 & 1 \\ 2 & 0 & 2 & 2 & 1 & 1 & 2 & 1 & 1 \\ 2 & 2 & 2 & 1 & 3 & 2 & 3 & 0 & 2 \\ 1 & 1 & 2 & 1 & 2 & 2 & 2 & 0 & 1 \\ 3 & 2 & 3 & 2 & 3 & 2 & 4 & 1 & 2 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 \\ 2 & 1 & 1 & 1 & 2 & 1 & 2 & 0 & 2 \end{bmatrix}$$

cari nilai *eigen* dengan cara hasil perkalian matriks $A^t A$ dikurangi dengan λ

$$\text{Eigen} = (A^T \times A) - \lambda I$$

$$\begin{bmatrix} 7 & 4 & 4 & 4 \\ 4 & 5 & 2 & 2 \\ 4 & 2 & 5 & 3 \\ 4 & 2 & 3 & 5 \end{bmatrix} - \begin{bmatrix} \lambda & 0 & 0 & 0 \\ 0 & \lambda & 0 & 0 \\ 0 & 0 & \lambda & 0 \\ 0 & 0 & 0 & \lambda \end{bmatrix} = \begin{bmatrix} 7-\lambda & 4 & 4 & 4 \\ 4 & 5-\lambda & 2 & 2 \\ 4 & 2 & 5-\lambda & 3 \\ 4 & 2 & 3 & 5-\lambda \end{bmatrix}$$

Buatlah persamaan *polynomial* dari matriks tersebut dengan mengalikan setiap elemen dan dikurangi elemen dari sisi berlawanan. Dari persamaan dibawah akan didapatkan nilai *eigen* masing-masing dokumen.

$$\lambda^4 - 21\lambda^3 + 300\lambda^2 + 858\lambda + 1375 = 0$$

$$\lambda_1 = 15,400 = Q$$

$$\lambda_2 = 1,235 = D1$$

$$\lambda_3 = 3,366 = D2$$

$$\lambda_4 = 2,000 = D3$$

Urutkan nilai *eigen* dari urutan besar ke urutan kecil

$$Q = 15,400$$

$$D2 = 3,366$$

$$D3 = 2,000$$

$$D1 = 1,235$$

Selanjutnya pembentukan matriks V dengan menggunakan hasil dari kalkulasi normalisasi *eigen* vektor kemudian *transpose* matriks V.

$$V = \begin{bmatrix} Q & D1 & D2 & D3 \\ -0,636 & -0,761 & -0,129 & 0 \\ -0,421 & 0,482 & -0,769 & 0 \\ -0,457 & 0,307 & 0,443 & -0,707 \\ 0,457 & 0,307 & 0,443 & 0,707 \end{bmatrix} \rightarrow V = \begin{bmatrix} Q & D2 & D3 & D1 \\ -0,636 & -0,761 & -0,129 & 0 \\ -0,421 & 0,482 & -0,769 & 0 \\ -0,457 & 0,307 & 0,443 & -0,707 \\ 0,457 & 0,307 & 0,443 & 0,707 \end{bmatrix}$$

$$V^T = \begin{bmatrix} -0,636 & -0,421 & -0,457 & 0,457 \\ -0,129 & 0,769 & 0,443 & 0,443 \\ 0 & 0 & -0,707 & 0,707 \\ -0,761 & 0,482 & 0,307 & 0,307 \end{bmatrix}$$

Matriks S didapatkan dari diagonal akar nilai *eigen* yang sudah diurutkan dari yang terbesar.

$$Q = 15,400 \mid D2 = 3,366 \mid D3 = 2,000 \mid D1 = 1,235$$

$$\begin{bmatrix} \sqrt{\lambda_1} & 0 & 0 & 0 \\ 0 & \sqrt{\lambda_2} & 0 & 0 \\ 0 & 0 & \sqrt{\lambda_3} & 0 \\ 0 & 0 & 0 & \sqrt{\lambda_4} \end{bmatrix} \Rightarrow \begin{bmatrix} 3,924 & 0 & 0 & 0 \\ 0 & 1,835 & 0 & 0 \\ 0 & 0 & 1,414 & 0 \\ 0 & 0 & 0 & 1,111 \end{bmatrix}$$

Selanjutnya pembentukan matriks U dengan mengkalikan $A.V.S^{-1}$. Matriks A dan V diurutkan terlebih dahulu dari urutan yang

paling besar ke urutan paling kecil. Dengan persamaan

$$U = \begin{bmatrix} 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} -0,636 & -0,129 & 0 & -0,761 \\ -0,421 & -0,769 & 0 & 0,482 \\ -0,457 & 0,443 & -0,707 & 0,307 \\ -0,457 & 0,443 & 0,707 & 0,307 \end{bmatrix} \begin{bmatrix} 2,882 & 0 & 0 & 0 \\ 0 & 6,164 & 0 & 0 \\ 0 & 0 & 7,999 & 0 \\ 0 & 0 & 0 & 10,181 \end{bmatrix}$$

$$U = \begin{bmatrix} 0,386 & -0,248 & -0,5 & -0,025 \\ 0,233 & 0,483 & 0 & -0,533 \\ 0,386 & -0,248 & 0,5 & -0,025 \\ 0,269 & -0,489 & 0 & 0,251 \\ 0,395 & 0,413 & 0 & 0,131 \\ 0,279 & 0,171 & 0,5 & 0,408 \\ 0,502 & -0,006 & 0 & -0,302 \\ 0,107 & -0,419 & 0 & -0,302 \\ 0,279 & 0,171 & -0,5 & 0,408 \end{bmatrix}$$

Penyederhanaan sesuai dengan nilai dimensi(k). Penyederhanaan pada matriks U dengan mengurangi jumlah kolom. Pada matriks S akan dilakukan pengurangan pada jumlah baris dan kolom. Jika k bernilai 2 maka matriks U hanya memiliki 2 kolom, sedangkan matriks S memiliki 2 kolom dan 2 baris.

$$U = \begin{bmatrix} 0,386 & -0,248 \\ 0,233 & 0,483 \\ 0,386 & -0,248 \\ 0,269 & -0,489 \\ 0,395 & 0,413 \\ 0,279 & 0,171 \\ 0,502 & -0,006 \\ 0,107 & -0,419 \\ 0,279 & 0,171 \end{bmatrix} \quad S = \begin{bmatrix} 3,924 & 0 \\ 0 & 1,835 \end{bmatrix}$$

Melakukan pencarian matriks vektor Q pada vektor D sebelum masuk ke proses perhitungan *cosine similarity* pada tahap selanjutnya.

Dengan membentuk vektor *query* q^t yang dibentuk dengan cara yang sama dengan membentuk matriks A.

$$\bar{q} = [1 \ 0 \ 1 \ 1 \ 1 \ 1 \ 1 \ 0 \ 1] \begin{bmatrix} 0,386 & -0,248 \\ 0,233 & 0,483 \\ 0,386 & -0,248 \\ 0,269 & -0,489 \\ 0,395 & 0,413 \\ 0,279 & 0,171 \\ 0,502 & -0,006 \\ 0,107 & -0,419 \\ 0,279 & 0,171 \end{bmatrix} \begin{bmatrix} 0,255 & 0 \\ 0 & 0,545 \end{bmatrix}$$

$$\bar{q} = [0,617 \ -0,813]$$

Dengan melakukan perhitungan yang sama pada tiap D, maka didapatkan nilai vektor sebagai berikut.

$$D1 = [0,407 \ -0,813]$$

$$D2 = [0,444 \ 0,470]$$

$$D3 = [0,445 \ 0,470]$$

2.6 Cosine Similarity

Sekarang dapat dihitung hasil kali titik antara vektor *query* yang sudah dipetakan dengan masing-masing vektor dokumen untuk mendapatkan nilai similaritas dari matriks vektor masing-masing D terhadap Q.

Menghitung nilai *cosine similarity* untuk membandingkan kemiripan antara dua vektor, yaitu kunci jawaban(Q) dengan jawaban siswa (D)

$$D1 = \frac{(Q_1 \cdot D_{11}) + (Q_2 \cdot D_{12})}{\sqrt{(Q_1)^2(D_{11})^2 + (Q_2)^2(D_{12})^2}}$$

$$= \frac{(0,617)(0,407) + (-0,135)(-0,813)}{\sqrt{(0,617)^2 + (0,407)^2} \cdot \sqrt{(-0,315)^2 + (-0,813)^2}} = 0,644 \times 100 = 64,4$$

$$D2 = \frac{(0,617)(0,444) + (-0,135)(0,470)}{\sqrt{(0,617)^2 + (0,444)^2} \cdot \sqrt{(-0,315)^2 + (0,470)^2}} = 0,566 \times 100 = 56,6$$

$$D3 = \frac{(0,617)(0,445) + (-0,135)(0,470)}{\sqrt{(0,617)^2 + (0,445)^2} \cdot \sqrt{(-0,315)^2 + (0,470)^2}} = 0,566 \times 100 = 56,6$$

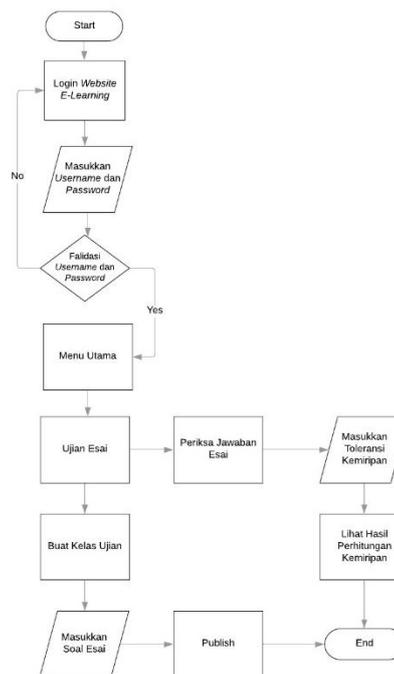
Dari perhitungan terakhir, didapatkan nilai masing-masing siswa yang terdapat pada D1, D2, dan D3.

3. HASIL DAN PEMBAHASAN

3.1 Flowchart Sistem

Setelah menganalisis sistem, maka didapatkan *flowchart diagram* untuk sistem *e-learning*, dan *flowchart* metode *Latent Semantic Analysis (LSA)*. Adapun alur kerja *flowchart diagram* sistem ditunjukkan oleh Gambar 1 sebagai berikut:

- User mengunjungi laman login situs *e-learning*
- User memasukkan *username* dan *password*.
- User masuk ke Menu Utama.
- User memilih menu Ujian.
- User memilih sub-menu Periksa Jawaban Esai.
- User memasukkan toleransi kemiripan jawaban esai.
- User melihat hasil perhitungan kemiripan jawaban esai.



Gambar 1 Flowchart sistem

3.2 Pengujian Sistem Similaritas

Pengujian dilakukan untuk memeriksa apakah sistem berhasil untuk melakukan deteksi terhadap jawaban yang terdapat kata yang mirip. Pengujian ini dilakukan pada *website* ujian online dengan kondisi tidak ada jawaban yang benar benar sama, namun hanya terdapat beberapa kata yang sama pada jawaban yang berbeda. Untuk mendukung hasil keakuratan hasil, maka ditambahkan metode pembobotan *LSA* yang berbeda, yaitu pembobotan *Tf-Idf*. Nantinya akan dilakukan pengecekan nilai similaritas tertinggi dan terendah dari kedua metode pembobotan. Pada perhitungan pembobotan, pembentukan *term document matrix* dilakukan dengan memasukan *query*. Untuk lebih jelasnya dapat dilihat pada tabel dibawah ini:

Tabel 2 Perhitungan similaritas percobaan 1

String	Matriks	SVD	Tf-Idf
Doc.1 Mampu cembung pipih daya	[1, 1, 1, 0, 0, 0, 0, 1, 0]	64.08	53.25
Doc.2 Mampu mata lensa ojek tebal	[1, 0, 0, 1, 0, 1, 1, 0, 1]	63.19	63.00
Doc.3 Mata lensa lihat objek	[0, 0, 0, 1, 1, 1, 0, 0, 1]	32.45	53.41
Query		Matriks	
Mampu cembung pipih lensa lihat objek tebal		[1, 1, 1, 1, 1, 1, 1, 0, 0]	

	0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]	
<i>E-learning</i> sebagai sembarang pengajaran dan pembelajaran yang menggunakan rangkaian elektronik (LAN, WAN, atau internet) untuk menyampaikan isi pembelajaran, interaksi, atau bimbingan.	[0, 1, 1, 3, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]	Doc.3
<i>E-learning</i> adalah Proses pembelajaran jarak jauh dengan menggabungkan prinsip-prinsip dalam proses pembelajaran dengan teknologi	[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 2, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]	Doc.4
<i>E-learning</i> adalah suatu sistem atau konsep pendidikan yang memanfaatkan teknologi informasi dalam proses belajar mengajar	[0, 1, 0, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0]	Doc.5
<i>E-learning</i> sebagai sembarang pengajaran dan pembelajaran yang menggunakan rangkaian elektronik (LAN, WAN, atau internet) untuk menyampaikan isi pembelajaran, interaksi, atau bimbingan.	[0, 1, 1, 3, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]	Doc.6
<i>E-learning</i> adalah suatu sistem atau konsep pendidikan yang memanfaatkan teknologi informasi dalam proses belajar mengajar	[0, 1, 0, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0]	Doc.7
<i>E-learning</i> adalah Pembelajaran yang disusun dengan tujuan	[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,	Doc.8

menggunakan sistem elektronik atau komputer sehingga mampu mendukung proses pembelajaran	0, 0, 0, 0, 0, 1, 0, 2, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 1, 1, 1, 1, 0, 0]	
Query	Matriks	
<i>E-learning</i> adalah suatu sistem atau konsep pendidikan yang memanfaatkan teknologi informasi dalam proses belajar mengajar	[0, 1, 0, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0]	

Pada dokumen string Tabel 5 terdapat string atau kata yang sama pada beberapa dokumen dan ada pula dokumen yang sama sekali tidak memiliki kata yang sama pada dokumen manapun. Pada dokumen 1 dan 2 terdapat beberapa kata yang sama namun dari kedua dokumen tersebut, tidak terdapat pada dokumen lainnya. Total banyak kata pada *term document matrix* sebanyak 54 kata.

Tabel 6 Hasil perhitungan similaritas percobaan 4

Id	SVD	Tf-Idf	Human Rates
Doc.1	0.0	0.0	Tidak sama
Doc.2	0.0	0.0	Tidak sama
Doc.3	8.2749	26.7756	
Doc.4	98.7563	39.492	
Doc.5	100.0	99.9999	Sama
Doc.6	11.5055	24.6291	
Doc.7	100.0	99.9999	Sama
Doc.8	83.1824	32.5083	

Dapat dilihat dari hasil percobaan pada Tabel 6 ditemukan pola nilai similaritas tertinggi berada pada dokumen ke-4 dengan *similarity* sebesar 98.7563 dan nilai terendah pada dokumen ke-3 sebesar 8.2749. Dan telah dibandingkan dengan metode pembobotan *Tf-Idf* menghasilkan pola nilai similaritas yang sama untuk nilai terbesar yakni 39.492, sedangkan untuk nilai terendah terjadi perbedaan, dimana nilai terendah berada pada dokumen 6 dengan nilai similaritas 24.6291.

Pada percobaan Tabel 6 terdapat dokumen yang benar-benar mirip dan dokumen yang benar-benar berbeda. Dokumen tersebut telah terdeteksi sama dan telah terdeteksi berbeda oleh sistem, hal ini membuktikan bahwa sistem telah berhasil mendeteksi *text* yang benar-benar sama dan *text* yang benar-benar berbeda.

3.3 Pengujian Lama Kerja Sistem

Pengujian lama kerja sistem dilakukan dengan parameter, yaitu terhadap banyak peserta ujian dan banyak kata pada jawaban. Penggunaan parameter ini disebabkan karena ukuran matriks sangat dipengaruhi oleh banyak peserta dan banyak kata yang berbeda. Semakin banyak pesertakan kata yang berbeda, maka akan semakin besar pula ukuran matriks.

Pengujian pertama dilakukan dengan parameter terhadap banyak peserta ujian, hasil dari ujicoba ini adalah sebagai berikut:

Tabel 7 Hasil Perhitungan waktu percobaan 5

Ujicoba ke	Jumlah Peserta	Lama Kerja Sistem
Ke-1	3 orang	0.325- sec.
Ke-2	6 orang	0.613- sec.
Ke-3	9 orang	0.892- sec.
Ke-4	12 orang	0.995- sec.
Ke-5	15 orang	1.113- sec.
Rata Rata waktu dalam penambahan peserta tiap 3 peserta		0.197- sec.

Ujicoba pada Tabel 7 dilakukan dengan kondisi jawaban setiap peserta adalah sama, dan dilakukan perhitungan berulang ulang sengan kondisi tiap kali perhitungan selanjutnya peserta ujian ditambahkan sebanyak 3. Dari total 5 kali perhitungan didapatkan bahwa rata-rata penambahan waktu untuk penambahan tiap 3 peserta adalah 0.197-sec.

Pengujian selanjutnya dilakukan dengan kondisi dengan parameter terhadap banyak kata yang berbeda dalam satukali perhitungan, hasil dari ujicoba ini adalah sebagai berikut:

Tabel 8 Hasil perhitungan waktu percobaan 6

Ujicoba ke	Jumlah Kata	Lama Kerja Sistem
Ke-1	3 kata	0.335- sec.
Ke-2	6 kata	0.353- sec.
Ke-3	9 kata	0.362- sec.
Ke-4	12 kata	0.797- sec.
Ke-5	15 kata	1.809- sec.
Rata Rata waktu dalam penambahan peserta tiap 3 peserta		0.1184- sec.

Ujicoba tersebut dilakukan dengan kondisi jumlah peserta adalah sama, dan dilakukan

perhitungan berulang ulang sengan kondisi tiap kali perhitungan selanjutnya jawaban peserta ujian seluruhnya ditambahkan sebanyak 3 kata yang berbeda. Dari total 5 kali perhitungan didapatkan bahwa rata rata penambahan waktu untuk penambahan tiap 3 kata adalah 0.1184-sec.

4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan mengenai implementasi *Latent Semantic Analysis Method (LSA)* untuk mendeteksi kemiripan jawabn *essay*, maka diperoleh kesimpulan yaitu:

1. Dalam percobaan pertama dan ke 3 telah membuktikan bahwa sistem mampu mendeteksi jawaban *essay* yang benar-benar sama denga nilai kemiripan antara 90 -100 %
2. Semkain besar ukuran matriks, maka akan semakin lama pula waktu yang dibutuhkan sistem untuk menghitung similaritas jawaban. Adapun rata-rata peningkatan waktu kerja sistem terhadap penambahan 3 peserta dan 3 kata berbeda berkisar antara 0.118 - 0.197 *second*.

5. SARAN

Pembentukan nilai similaritas dokumen berpengaruh pada penggunaan *query* dalam pembentukan *term document matrix* dan proses *parsing text*. Dalam *parsing text* terdapat proses eliminasi kata yang tidak berguna dan pengembalian kata ke bentuk awal, serta terdapat proses penyamaan makna kata yang berbeda. Pada penelitian ini tidak digunakan proses peyamaan kata untuk makna yang sama sehingga sistem masih mendeteksi bahwa kata yg berbeda namun satu makna masih diletakkan pada kordinat matriks yang berbeda

Dalam pembentukan indeks kemiripan dibutuhkan referensi untuk mengukur akurasi secara ilmiah. Serta penggunaan dokumen dengan *human rates* dapat dijadikan acuan dalam proses perhitungan similaritas.

DAFTAR PUSTAKA

- [1] Hidayat, W. "Penggunaan Internet Indonesia Nomor Enam Dunia", [online]. Available: <https://kominfo.go.id/content/detail/4286>

- /pengguna-internet-indonesia-nomor-enam-dunia/0/sorotan_media. [Accessed: 11 Nov 2014].
- [2] Soemartono H.K., "Pengguna dan Perilaku Internet Indonesia", *APJII*, No. 1, Vol. 23 Hal. 1, 2018.
- [3] Suranto, B, "Virtual Classroom : Strategi Pembelajaran Berbasis Synchronous E-Learning", *Seminar Nasional Aplikasi Teknologi Informasi*. 1-1, 2009.
- [4] Nova, K., Dadang, S., & Ardiyanto, W. "Aplikasi Pendeteksi Plagiat Dengan menggunakan Metode Latent Semantic Analysis (Studi Kasus : Laporan TA PCR)", 2012.
- [5] Jhon, D, "Teori E-Learnig Menurut Beberapa Cendikiawan", [online]. Available: <https://www.silabus.web.id/e-learning/>, [Accessed: 2019]
- [6] Pressman, Roger. *Rekayasa Perangkat Lunak Pendekatan Praktisi (Buku I)*, Yogyakarta: Penerbit Andi & McGraw-Hill Book Co, 2002.
- [7] Asih, S., & Hendry, "Perancangan Aplikasi Tes Esai Online Dengan Memanfaatkan Algoritma Latent Semantic Analysis", *Dokumen Artikel Ilmiah*, 1-14. 2014.
-