

IMPLEMENTASI *INFORMATION RETRIEVAL SYSTEM* MENGUNAKAN TEKNIK *VECTOR SPACE MODELS(VSM)* UNTUK SISTEM PENGUJIAN KOMPUTER BERBASIS TEKS

Elin Panca Saputra¹, Sugiono², Supriatiningsih³

Program Studi Sistem Informasi, Universitas Bina Sarana Informatika, Indonesia^{1,2,3})

Jl. Kamal Raya No.18, RT.6/RW.3, Cengkareng Bar., Kecamatan Cengkareng, Kota Jakarta Barat, Daerah
Khusus Ibukota Jakarta 11730, Indonesia

elin.epa@bsi.ac.id¹; sugiono.sgx@bsi.ac.id²; supriatiningsih.stq@bsi.ac.id³

Abstract : *Computer-based exams are a system of examinations that utilize information technology in managing answers and evaluating the results of tests that have been done by students. In general, computer-based examinations are still multiple choice or multiple choice, this causes students to think less developed patterns. Indeed essay-based exams must be applied to develop students' thinking patterns in understanding a scientific field. "Information retrieval" is study of procedures and methods used to rediscover stored information from various relevant resources. With the act of indexing, calling, calling data back (penarikkaning). VSM (Vector Space Model) is one of the IR methods by looking for similarity (similarity) of a document based on the query entered. In this study the VSM method will be applied to the problem of essay-based computer exams to get the level of similarity of an answer with the query determined by the teacher. Case studies that have been done with data in the form of 1 essay question answered by 4 students with each different answer can display similarity levels of 0.14, 0.0, 0.10 and 0.20. So that it can be concluded that the sequence of ranking similarity answers to queries is the 4th, 1st, 3rd and 2nd students.*

Key Word : *Text Based Computer Exam*

1. PENDAHULUAN

Implementasi dengan Information Retrieval bertujuan untuk meningkatkan proses pencarian informasi, dan untuk meningkatkan efisiensi pencarian informasi yang optimal untuk memenuhi kebutuhan pengguna (Qasim & Dkk, 2015) Beragam manfaat media komputer dalam dunia pendidikan diantaranya untuk keperluan ujian baik bersifat *online* maupun *offline*. Akan tetapi kebanyakan ujian dengan menggunakan komputer masih berupa pilihan ganda atau *multiple choice* sehingga peserta ujian hanya memilih salah satu jawaban yang dianggap benar saja dan tidak perlu banyak berfikir untuk menyelesaikannya. Hal tersebut dapat menyebabkan kurang maksimalnya pola berfikir siswa. Dengan sistem ujian komputer yang berbasis *essay* para siswa diharapkan mampu mengembangkan pola pemikirannya terhadap suatu mata pelajaran. Sistem ujian berbasis *essay* pada umumnya

sudah diterapkan namun dalam pelaksanaannya masih menggunakan metode manual dalam proses pengkoreksiannya.

Dari pencarian informasi saat ini menghasilkan *penarikan* yang tinggi dan tingkat keakuratan yang rendah. *penarikan* yang tinggi dapat diartikan bahwa dokumen yang didapatkan dalam penelusuran dokumen adalah banyak, sedangkan tingkat akurasinya rendah dapat diartikan bahwa dokumen yang diharapkan dapat ditemukan sedikit atau rendah. Untuk memecahkan Permasalahan untuk mengatasi masalah di atas adalah dengan membuat sistem Information Retrieval yaitu menggunakan metode *Vector Space Model (VSM)*. Metode VSM dipilih karena cara kerja model yang efisien, mudah dalam representasi dan dapat diimplementasikan pada text mining. Algoritme tersebut berbeda dalam mengambil pendekatan, yang berbeda dalam menganalisis kesamaan serta menghitung

skor. Salah satu pendekatan yang berperingkat tinggi dan digunakan adalah *Vector Space Model* (A. Jain & Dkk, 2017).

Vector Space Model (VSM) saat ini sudah diterapkan dalam bidang keilmuan seperti *Computational Linguistic* (K. Erk & Dkk, 2010). Penggunaan VSM juga dapat diterapkan dalam sistem temu kembali informasi atau *Information Retrieval (IR)*. Penerapan sistem IR dalam ujian berbasis esay dapat mempermudah sistem dalam memberikan nilai kemiripan suatu jawaban terhadap kunci jawaban atau *query* yang telah ditentukan oleh pengajar.

Vector Space Model digunakan untuk merepresentasikan teks abstrak menjadi vektor yang digunakan untuk mengidentifikasi keterkaitan dokumen (C. Slamet & Dkk, 2018). Dari metode ini, hasil ringkasan dapat dihasilkan dari dokumen.

Pada penelitian ini akan dibahas mengenai penerapan *Information retrieval* pada sistem ujian komputer berbasis esay menggunakan metode *Vector Space Model*. Dengan menggunakan metode VSM kami akan menghitung penarikan dan *precision* antara jawaban esay siswa dengan jawaban esay yang telah diinputkan oleh pengajar sehingga didapatkan tingkat kemiripan antara jawaban dengan kunci jawabannya. metode menggunakan model ruang vektor (VSM) untuk mewakili dokumen untuk vektor, di mana masing-masing dimensi sesuai bobot, dan kemudian menghitung kesamaan antara dokumen dengan kumpulan dokumen yang memiliki kesamaan lebih tinggi (C. Qimin & Dkk, 2015).

Kami mengusulkan metode untuk mengotomatiskan proses peringkasan dokumen *teks* berdasarkan frekuensi istilah dalam dokumen pada tingkat yang berbeda - paragraf dan kalimat. Untuk meringkas dokumen, kesamaan antara paragraf dan kalimat dalam paragraf dianggap menggunakan *Vector Space Model (VSM)* (N. Gupta & Dkk, 2013). di samping itu *Vector Space Model (VSM)* metode kami

sangat efektif dan efektif dapat meningkatkan akurasi pengelompokan yang signifikan ditunjukkan dalam hasil percobaan (C. Qimin & Dkk, 2015).

2. LANDASAN TEORI

Text Mining

Text mining merupakan salah satu bidang khusus dari data mining. Secara garis besar *text mining* sebagai proses menemukan informasi yang dilakukan oleh user yang berinteraksi dengan sekumpulan dokumen menggunakan tool analisis yang merupakan komponen dalam data mining (D. W. Brata and A. Hetami, 2015).

Beda halnya dengan Data mining terdiri dari algoritma inti yang memungkinkan seseorang akan dapat untuk mendapatkan sebuah dasar dasar wawasan dan pengetahuan dari data besar (E. P. Saputra, 2017).

Seperti ekstraksi data selain itu *text mining* sebuah pemrosesan bahasa alami yang lebih luas dan telah menjadi bidang aktif dalam sains selama beberapa dekade ini. Namun, penggunaannya dalam bisnis telah menjadi semakin penting ketika menjadi web (C. Fay, 2018).

Text mining diperlukan sebelum melakukan proses pencarian *similarity* suatu dokumen. Dalam tahapan *text mining* terdapat tahapan *Preprocessing text*, yakni tindakan menghilangkan karakter-karakter tertentu yang terkandung dalam dokumen, seperti koma, tanda petik dan lain-lain serta mengubah semua huruf kapital menjadi huruf kecil. Selain itu, dalam tahap *text preprocessing* ini dilakukan *tokenization*. *Tokenization* merupakan proses pengolahan token yang terdapat dalam rangkaian teks, sehingga dokumen akan dipecah-pecah menjadi term (A. Aziz & Dkk, 2016).

Berikut tahapan-tahapan *text mining*:

- a. *Casefolding*: mengubah semua huruf dalam dokumen menjadi huruf kecil (*lowercase*).

- b. Tokenizing: memotong tiap kata dalam kalimat atau parsing dengan menggunakan spasi sebagai delimitter yang akan menghasilkan token berupa kata.
- c. Filtering: menyaring kata yang didapat dari tokenizing yang dianggap tidak penting atau tidak memiliki makna dalam proses text mining yang disebut stopword. Stopword berisi kata-kata umum yang sering muncul dalam sebuah dokumen dalam jumlah banyak namun tidak memiliki kaitan dengan tema tertentu.
- d. Stemming: mengembalikan kata-kata yang diperoleh dari hasil filtering ke bentuk dasarnya, menghilangkan imbuhan awal (prefix) dan imbuhan akhir (suffix) sehingga didapat kata dasar.

Metode Term Frequency-Inversed Document Frequency (TF/IDF)

Basis pembobotan TF-IDF merupakan jenis pembobotan yang melibatkan pengukuran statistik untuk mengukur seberapa penting sebuah kata dalam kumpulan dokumen. Tingkat kepentingan meningkat ketika sebuah kata muncul beberapa kali dalam sebuah dokumen tetapi diimbangi dengan frekuensi kemunculan kata tersebut dalam kumpulan dokumen (D. W. Brata and A. Hetami, 2015).

TF merupakan pembobotan yang sederhana dimana penting tidaknya sebuah kata diasumsikan sebanding dengan jumlah kemunculan kata tersebut dalam dokumen, sementara IDF merupakan pembobotan yang mengukur seberapa penting sebuah kata dalam dokumen apabila dilihat secara global pada seluruh dokumen (T. M. Isa and F. Abidin, 2013).

Pada metode TF-IDF yang dapat memberikan nilai untuk masing-masing - setiap data dan *Vector Space Model (VSM)* menghitung nilai kesamaan antara kata kunci serta masing-masing pada data besar. Hasil perhitungan diwakili oleh tingkat

kesamaan data terhadap kata kunci (A. Rokhim and A. A. Yaqin, 2017).

Pada dasarnya TF dan IDF merupakan proses pembobotan suatu kata dalam suatu dokumen dihitung berdasarkan banyaknya kata tersebut muncul. Berikut persamaan yang digunakan dalam proses TF-IDF:

$$IDF_{(t)} = \log(D/df_{(t)})$$

Dimana :

IDF = rasio frekuensi dokumen pada kata ke-t dari query

D = jumlah dokumen yang digunakan

$Df_{(t)}$ = jumlah dokumen yang mengandung kata ke-t dari query

$$TF-IDF_{(d,t)} = TF_{(d,t)} * IDF_{(t)}$$

Dimana :

D = dokumen ke-d

T = kata ke-t dari kata kunci

Tf = frekuensi banyaknya kata ke-t dari query pada dokumen ke-d

TF-IDF = bobot dokumen ke-d terhadap query ke-t

IDF = rasio frekuensi dokumen pada kata ke-t dari kata kunci

Vector Space Model

Kemiripan antara vektor dokumen dan vektor query akan dihitung dengan pendekatan cosine similarity[10]. Perhitungan similarity menggunakan VSM dapat ditulis menggunakan persamaan :

$$Sim(di,dj) = \frac{D_i D_j}{D_1 |D_2|} = \frac{W_i q . W_j}{\sqrt{W_i q^2 + W_j j^2}}$$

Dimana $Sim(di,dj)$ = similaritas antara query dengan dokumen

D = panjang vektor pada dokumen

W_{ij} = bobot term pada dokumen

W_{iq} = bobot term pada query

3. HASIL DAN PEMBAHASAN

Pada penelitian ini dilakukan pengujian terhadap similaritas antara jawaban guru dengan jawaban murid terhadap suatu soal yang sama. Langkah awal guru menginput pertanyaan dan jawaban dari pertanyaan tersebut yang nantinya akan digunakan sebagai query. Selanjutnya siswa menjawab

pertanyaan dari guru yang nantinya jawaban tersebut akan digunakan sebagai dokumen.

Soal : Aplikasi Komputer Adalah.?

Jawaban(Guru) : Sebuah perangkat lunak (software) program komputer yang ditulis dalam bahasa pemrograman dan berfungsi melakukan perintah sesuai dengan keinginan dari pembuat aplikasi.

Jawaban(Siswa 1) : program pada komputer yang dibuat untuk menjalankan perintah programmer

Jawaban(Siswa 2) : windows, microsoft word, winamp, dan lain-lain

Jawaban(Siswa 3) : software yang ada pada komputer yang digunakan sebagai alat untuk menjalankan perintah pengguna komputer

Jawaban(Siswa 4) : sekumpulan instruksi yang ditulis dalam bahasa pemrograman yang berfungsi untuk memerintah komputer agar dapat menjalankan perintah dari penggunanya.

1) Pengolahan dengan Microsoft Excel

Pada tahap awal, terlebih dahulu dilakukan preprocessing text yakni proses casefolding, tokenizing, filtering dan streaming.

Pada tahap *casefolding*, suatu dokumen disamakan bentuk hurufnya menjadi kecil semua atau lower case dan dihilangkan semua tanda baca yang ada. Selanjutnya pada tahapan *tokenizing* dokumen dipecah per kata menggunakan operator spasi.

Tabel 1. Dokumen yang sudah dilakukan Casefolding dan Tokenizing

T	S1	S2	S3	S4
sebuah	program	windows	software	sekumpulan
perangkat	pada	microsoft	yang	instruksi
lunak	komputer	word	ada	yang
software	yang	office	pada	ditulis
program	dibuat	winamp	komputer	dalam
komputer	untuk	dan	yang	bahasa
yang	menjalankan	lain	digunakan	pemrograman
ditulis	perintah	lain	sebagai	yang

	ah			
dalam	programer		alat	berfungsi
bahasa			untuk	untuk
permr ogran man			menjalank an	memerinta h
dan			perintah	komputer
berfu ngsi			pengguna	agar
melak ukan			komputer	dapat
perint ah				menjalank an
sesuai				perintah
denga n				dari

Setelah dilakukan proses *Casefolding* dan *Tokenizing* selanjutnya dilakukan proses *Filtering* yaitu dengan menghilangkan kata-kata *stop word* atau kata sambung serta menghilangkan kata-kata yang sama pada setiap dokumen.

Tabel 2. Dokumen yang sudah dilakukan Filtering

Q	D1	D2	D3	D4
sebuah	program	windows	software	sekumpul an
perangkat	kompute r	micr osof t	komputer	instruksi
lunak	dibuat	wor d	digunakan	ditulis
software	menjala nkan	offic e	sebagai	bahasa
program	perintah	wina mp	alat	pemprogra man
komputer	program er		menjalank an	berfungsi
ditulis			perintah	memerint ah
dalam			pengguna	komputer
bahasa			komputer	menjalan kan
permr ogran aman				perintah
berfungsi				pengguna nya
melakuka n				
perintah				

sesuai				
keinginan				
pembuat				
aplikasi				

Setelah proses Filtering selanjutnya dilakukan proses Streaming yakni proses pengembalian kosakata menjadi kata dasar dengan menghilangkan imbuhan-imbuhan kata yang ada.

Tabel 3. Dokumen yang sudah dilakukan Streaming

Q	D1	D2	D3	D4
aplikasi	buat	microsoft	alat	bahasa
bahasa	jalan	office	guna	fungsi
buat	komputer	winamp	guna	guna
fungsi	perintah	windows	jalan	instruksi

ingin	program	word	komputer	jalan
komputer			perintah	komputer
laku			software	kumpul
lunak				perintah
perangkat				program
perintah				tulis
program				
sesuai				
software				
tulis				

Setelah tahapan text mining selesai selanjutnya dilakukan perhitungan pembobotan dari setiap term menggunakan persamaan TF-IDF.

Tabel 4. Tabel perhitungan TF dan IDF

TABEL PERHITUNGAN TF-IDF													
Term	TF					DF	D/D F	IDF	TF*IDF				
	Q	D1	D2	D3	D4				Q	D1	D2	D3	D4
alat	0,00	0,00	0,00	1,00	0,00	1,00	4,00	0,60	0,00	0,00	0,00	0,60	0,00
aplikasi	1,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
bahasa	1,00	0,00	0,00	0,00	1,00	1,00	4,00	0,60	0,60	0,00	0,00	0,00	0,60
buat	1,00	1,00	0,00	0,00	0,00	1,00	4,00	0,60	0,60	0,60	0,00	0,00	0,00
fungsi	1,00	0,00	0,00	0,00	1,00	1,00	4,00	0,60	0,60	0,00	0,00	0,00	0,60
guna	0,00	0,00	0,00	1,00	1,00	2,00	2,00	0,30	0,00	0,00	0,00	0,30	0,30
ingin	1,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
instruksi	0,00	0,00	0,00	0,00	1,00	1,00	4,00	0,60	0,00	0,00	0,00	0,00	0,60
jalan	0,00	1,00	0,00	1,00	1,00	3,00	1,33	0,12	0,00	0,12	0,00	0,12	0,12
komputer	1,00	1,00	0,00	1,00	1,00	3,00	1,33	0,12	0,12	0,12	0,00	0,12	0,12
kumpul	0,00	0,00	0,00	0,00	1,00	1,00	4,00	0,60	0,00	0,00	0,00	0,00	0,60

laku	1,00	0,0 0	0,0 0	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
lunak	1,00	0,0 0	0,0 0	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
microsoft	0,00	0,0 0	1,0 0	0,00	0,00	1,00	4,00	0,60	0,00	0,00	0,60	0,00	0,00
office	0,00	0,0 0	1,0 0	0,00	0,00	1,00	4,00	0,60	0,00	0,00	0,60	0,00	0,00
perangkat	1,00	0,0 0	0,0 0	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
perintah	1,00	1,0 0	0,0 0	1,00	1,00	3,00	1,33	0,12	0,12	0,12	0,00	0,12	0,12
program	1,00	1,0 0	0,0 0	0,00	1,00	2,00	2,00	0,30	0,30	0,30	0,00	0,00	0,30
sesuai	1,00	0,0 0	0,0 0	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
software	1,00	0,0 0	0,0 0	1,00	0,00	1,00	4,00	0,60	0,60	0,00	0,00	0,60	0,00
tulis	1,00	0,0 0	0,0 0	0,00	1,00	1,00	4,00	0,60	0,60	0,00	0,00	0,00	0,60
winamp	0,00	0,0 0	1,0 0	0,00	0,00	1,00	4,00	0,60	0,00	0,00	0,60	0,00	0,00
windows	0,00	0,0 0	1,0 0	0,00	0,00	1,00	4,00	0,60	0,00	0,00	0,60	0,00	0,00
word	0,00	0,0 0	1,0 0	0,00	0,00	1,00	4,00	0,60	0,00	0,00	0,60	0,00	0,00

Setelah melakukan perhitungan terhadap TF dan IDF selanjutnya kita lakukan tahapan berikutnya yakni

menghitung jarak antar vektor dengan persamaan VSM.

Tabel 5. Tabel Perhitungan VSM

TABEL PERHITUNGAN VSM									
Term	TF-IDF					VSM			
	Q	D1	D2	D3	D4	Q*D1	Q*D2	Q*D3	Q*D4
alat	0,00	0,00	0,00	0,36	0,00	0,00	0,00	0,00	0,00
aplikasi	1,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
bahasa	1,00	0,00	0,00	0,00	0,36	0,00	0,00	0,00	0,36
buat	1,00	0,36	0,00	0,00	0,00	0,36	0,00	0,00	0,00
fungsi	1,00	0,00	0,00	0,00	0,36	0,00	0,00	0,00	0,36
guna	0,00	0,00	0,00	0,09	0,09	0,00	0,00	0,00	0,00
ingin	1,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
instruksi	0,00	0,00	0,00	0,00	0,36	0,00	0,00	0,00	0,00
jalan	0,00	0,02	0,00	0,02	0,02	0,00	0,00	0,00	0,00
komputer	1,00	0,02	0,00	0,02	0,02	0,02	0,00	0,02	0,02
kumpul	0,00	0,00	0,00	0,00	0,36	0,00	0,00	0,00	0,00
laku	1,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
lunak	1,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
microsoft	0,00	0,00	0,36	0,00	0,00	0,00	0,00	0,00	0,00
office	0,00	0,00	0,36	0,00	0,00	0,00	0,00	0,00	0,00
perangkat	1,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
perintah	1,00	0,02	0,00	0,02	0,02	0,02	0,00	0,02	0,02
program	1,00	0,09	0,00	0,00	0,09	0,09	0,00	0,00	0,09
sesuai	1,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
software	1,00	0,00	0,00	0,36	0,00	0,00	0,00	0,36	0,00
tulis	1,00	0,00	0,00	0,00	0,36	0,00	0,00	0,00	0,36
winamp	0,00	0,00	0,36	0,00	0,00	0,00	0,00	0,00	0,00
windows	0,00	0,00	0,36	0,00	0,00	0,00	0,00	0,00	0,00
word	0,00	0,00	0,36	0,00	0,00	0,00	0,00	0,00	0,00
SQRT	3,74	0,71	1,35	0,93	1,43	0,70	0,00	0,63	1,10

Setelah mengkonversi tiap-tiap term menjadi vektor dan mencari bobot antar vektor atau tahapan VSM langkah

selanjutnya kita cari jarak antara tiap-tiap kata pada query dengan masing-masing dokumen.

Tabel 6. Tabel perhitungan Jarak dan Dot Produk

Q	TC-S				DOT PRODUK			
	D1	D2	D3	D4	Q*D1	Q*D2	Q*D3	Q*D4
0,00	0,00	0,00	0,36	0,00	0,00	0,00	0,00	0,00
0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
0,36	0,00	0,00	0,00	0,36	0,00	0,00	0,00	0,13
0,36	0,36	0,00	0,00	0,00	0,13	0,00	0,00	0,00
0,36	0,00	0,00	0,00	0,36	0,00	0,00	0,00	0,13
0,00	0,00	0,00	0,09	0,09	0,00	0,00	0,00	0,00
0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
0,00	0,00	0,00	0,00	0,36	0,00	0,00	0,00	0,00
0,00	0,02	0,00	0,02	0,02	0,00	0,00	0,00	0,00
0,02	0,02	0,00	0,02	0,02	0,00	0,00	0,00	0,00
0,00	0,00	0,00	0,00	0,36	0,00	0,00	0,00	0,00
0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
0,00	0,00	0,36	0,00	0,00	0,00	0,00	0,00	0,00
0,00	0,00	0,36	0,00	0,00	0,00	0,00	0,00	0,00
0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
0,02	0,02	0,00	0,02	0,02	0,00	0,00	0,00	0,00
0,09	0,09	0,00	0,00	0,09	0,01	0,00	0,00	0,01
0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
0,36	0,00	0,00	0,36	0,00	0,00	0,00	0,13	0,00
0,36	0,00	0,00	0,00	0,36	0,00	0,00	0,00	0,13
0,00	0,00	0,36	0,00	0,00	0,00	0,00	0,00	0,00
0,00	0,00	0,36	0,00	0,00	0,00	0,00	0,00	0,00
0,00	0,00	0,36	0,00	0,00	0,00	0,00	0,00	0,00
1,93	0,50	1,81	0,86	2,04	0,14	0,00	0,13	0,40

Setelah langkah-langkah diatas selanjutnya kita cari tingkat similarity duatu dokumen dengan query dengan menghitung jaraknya

Tabel 7. Hasil pencarian Similarity dari tiap-tiap dokumen terhadap query

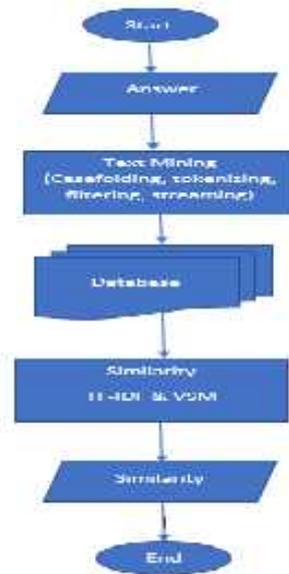
0,14	0,00	0,10	0,20
------	------	------	------

Dari tabel diatas dapat disimpulkan bahwa jawaban dari siswa ke-1 memiliki tingkat similarity sebesar 0,14 siswa ke-2 sebesar 0,0 siswa ke-3 sebesar 0,10 dan siswa ke-4 sebesar 0,20 terhadap jawaban yang diberikan oleh guru.

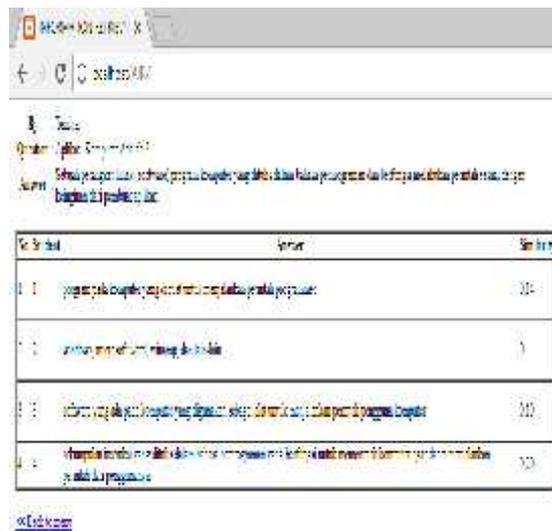
2) Implementasi Aplikasi

Aplikasi ini merupakan bentuk implementasi penerapan Information Retrieval dengan menggunakan metode Vector Space Model(VSM). Aplikasi ini berfungsi untuk mencari tingkat kemiripan (similarity) antara jawaban yang telah diinputkan guru(query) dengan jawaban yang telah diinputkan oleh siswa(dokumen). Aplikasi ini masih dalam bentuk modul sederhana yang nantinya diperlukan implementasi terhadap aplikasi induk seperti e-learning maupun sistem ujian ainnya.

Sistem berjalan aplikasi ini yaitu guru terlebih dahulu menginput soal beserta jawabannya yang nantinya dianggap sebagai query. Selanjutnya setiap siswa aja memberikan jawaban yang nantinya dianggap sebagai dokumen yang akan diproses tingkat similaritasnya terhadap query.



Gambar 1. Flowchart Sistem



Gambar 2. Hasil perhitungan VSM menggunakan Aplikasi

4. KESIMPULAN

Berdasarkan penelitian diatas penulis dapat mengambil kesimpulan bahwa metode VSM dapat mempermudah guru dalam memberikan penilaian terhadap tingkat kemiripan jawaban siswa pada kasus ujian berbasis esay menggunakan komputer.

Hasil pengujian terhadap suatu data berupa satu pertanyaan esay yang dijawab

oleh 4 orang siswa dapat diperoleh tingkat similaritas tertinggi dari siswa ke-4 yakni

Daftar Pustaka

- [1] L. M. Qasim Abualigah and E. S.Hanandeh, "Applying Genetic Algorithms to Information Retrieval Using Vector Space Model," *Int. J. Comput. Sci. Eng. Appl.*, vol. 5, no. 1, pp. 19–28, 2015.
- [2] A. Jain, A. Jain, N. Chauhan, V. Singh, and N. Thakur, "Information Retrieval using Cosine and Jaccard Similarity Measures in Vector Space Model," *Int. J. Comput. Appl.*, vol. 164, no. 6, pp. 28–30, 2017.
- [3] K. Erk, S. Padó, and U. Padó, "A flexible, corpus-driven model of regular and inverse selectional preferences," *Comput. Linguist.*, vol. 36, no. 4, pp. 723–763, 2010.
- [4] C. Slamet, A. R. Atmadja, D. S. Maylawati, R. S. Lestari, W. Darmalaksana, and M. A. Ramdhani, "Automated Text Summarization for Indonesian Article Using Vector Space Model," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 288, no. 1, 2018.
- [5] C. Qimin, G. Qiao, W. Yongliang, and W. Xianghua, "Text clustering using VSM with feature clusters," *Neural Comput. Appl.*, vol. 26, no. 4, pp. 995–1003, 2015.
- [6] N. Gupta, P. C. Saxena, and J. P. Gupta, "Document summarisation based on sentence ranking using vector space model," *Int. J. Data Mining, Model. Manag.*, vol. 5, no. 4, p. 380, 2013.
- [7] D. W. Brata and A. Hetami, "Perancangan Information Retrieval (IR) Untuk Pencarian Ide Pokok Teks Artikel Berbahasa Inggris Dengan Pembobotan Vector Space Model," *J. Ilm. Teknol. Inf. Asia*, vol. 9, no. 1, pp. 53–59, 2015.
- [8] E. P. Saputra, "Prediksi Keberhasilan Telemarketing Bank Untuk," *J. Ilmu Pengetah. Dan Teknol. Komput.*, vol. 2, no. 2, pp. 66–72, 2017.
- [9] C. Fay, "Text Mining with R : A Tidy Approach ," *J. Stat. Softw.*, vol. 83, no. Book Review 1, pp. 1–3, 2018.
- [10] A. Aziz, R. Saptono, and K. P. Suryajaya, "Implementasi Vector Space Model dalam Pembangkitan Frequently Asked Questions Otomatis dan Solusi yang Relevan untuk Keluhan Pelanggan," *Sci. J. Informatics*, vol. 2, no. 2, p. 111, 2016.
- [11] T. M. Isa and F. Abidin, "Mengukur Tingkat Kesamaan Paragraf Menggunakan Vector Space Model untuk Mendeteksi Plagiarisme," *Semin. Nas. dan ExpoTeknik Elektro*, pp. 229–234, 2013.
- [12] A. Rokhim and A. A. Yaqin, "Implementasi Metode Term Frequency Inversed Document Frequence (Tf-Idf) Dan Vector Space Model," *J. SPIRIT*, vol. 9, no. 1, pp. 34–38, 2017.