JURNAL

SISTEM KOREKSI JAWABAN *ESSAY* MENGGUNAKAN METODE *COSINE SIMILARITY*

Oleh:

IKA PRAMISWARI

12.1.03.02.0067

Dibimbing oleh :

1. **M.Rizal Arief, S.T., M.Kom.**
2. **Daniel Swanjaya, M.Kom.**

PROGRAM STUDI TEKNIK INFRMATIKA

FAKULTAS TEKNIK

UNIVERSITAS NUSANTARA PGRI KEDIRI

2017

Artikel Skripsi
Universitas Nusantara PGRI Kediri

SURAT PERNYATAAN
ARTIKEL SKRIPSI TAHUN 2017

Yang bertanda tangan di bawah ini:

Nama Lengkap	: IKA PRAMISWARI
NPM	: 12.1.03.02.0067
Telepun/HP	: 085745509222
Alamat Surel (Email)	: ikapramiswari@gmail.com
Judul Artikel	: Siatem Koreksi Jawaban *Essay* Menggunakan Metode Cosine Similarity
Fakultas – Program Studi	: Fakultas Teknik – Teknik Informatika
Nama Perguruan Tinggi	: Universitas Nusantara PGRI kediri
Alamat Perguruan Tinggi	: Jl. K.H. Ahmad Dahlan No. 76, Mojoroto, Kediri, Jawa Timur 64112

Dengan ini menyatakan bahwa :

a. artikel yang saya tulis merupakan karya saya pribadi (bersama tim penulis) dan bebas plagiarisme;
b. artikel telah diteliti dan disetujui untuk diterbitkan oleh Dosen Pembimbing I dan II.

Demikian surat pernyataan ini saya buat dengan sesungguhnya. Apabila di kemudian hari ditemukan ketidaksesuaian data dengan pernyataan ini dan atau ada tuntutan dari pihak lain, saya bersedia bertanggungjawab dan diproses sesuai dengan ketentuan yang berlaku.

Mengetahui		Kediri, 28 Januari 2017
Pembimbing I	Pembimbing II	Penulis,
M.Rizal Arief,ST.,M.Kom. NIDN. 0716027504	Daniel Swanjaya, M.Kom. NIDN. 0723098303	Ika Pramiswari NPM.12.1.03.02.0067

SISTEM KOREKSI JAWABAN *ESSAY* MENGGUNAKAN METODE *COSINE SIMILARITY*

Ika Pramiswari
12.1.03.02.0067
Fakultas Teknik –Teknik Informatika
ikapramiswari@gmail.com
M.Rizal Arief, ST., M.Kom dan Daniel Swanjaya, M.Kom
UNIVERSITAS NUSANTARA PGRI KEDIRI

ABSTRAK

Ika Pramiswari: Sistem Koreksi Jawaban Essay Menggunakan Metode *Cosine Similarity*, Skripsi, Teknik Informatika, Fakultas Teknik, UN PGRI Kediri, 2016.

Evaluasi hasil belajar menjadi komponen yang sangat penting dalam proses pembelajaran, karena hasil evaluasi merupakan indikator dari pemahaman siswa terhadap materi ajar yang diberikan. Ujian dengan jawaban dalam bentuk *essay*, umumnya digunakan sebagai salah satu bentuk evaluasi untuk menentukan kemampuan siswa, dimana pilihan jawaban tidak diberikan. Siswa harus menjawab dengan kalimat, sehingga jawaban tersebut dapat bervariasi berdasarkan pemikiran mereka. Salah satu kelemahan dari jawaban dalam bentuk *essay* adalah sulitnya menilai jawaban, dan cenderung memakan waktu.

Pembuatan *essay* dalam penilaian hasil kegiatan belajar dinilai sebagai metode yang tepat, karena melibatkan kemampuan siswa untuk mengingat, mengorganisasikan, mengekspresikan dan mengintegrasikan gagasan yang dimiliki siswa tersebut. Tetapi, pemeriksaan *essay* membutuhkan waktu yang lebih lama karena sebagian besar dilakukan secara manual yaitu dengan cara membaca *essay* satu persatu, dari pada tes-tes yang bersifat objektif seperti pilihan ganda dan jawaban singkat. Para pengajar perlu menghabiskan banyak waktu untuk menilai jawaban ujian siswa mereka.

Metode TF-IDF merupakan metode untuk menghitung bobot setiap kata yang paling umum digunakan pada *information retrieval*. Metode ini juga terkenal efisien, mudah dan memiliki hasil yang akurat. Metode ini menggabungkan dua konsep untuk perhitungan bobot yaitu, frekuensi kemunculan sebuah kata di dalam sebuah dokumen tertentu dan *inverse frekuensi* dokumen yang mengandung kata tersebut.

Kesimpulan dari skripsi yang berjudul Sistem Koreksi Jawaban Essay Menggunakan *Metode Cosine Similarity* ini adalah: (1). Telah dihasilkan program sistem koreksi jawaban essay yang bertujuan untuk membantu pengajar mengkoreksi jawaban essay secara mudah. (2). Dengan adanya aplikasi ini pengajar dapat mengoreksi tanpa harus membawa kertas-kertas jawaban dari siswa. (3). Dengan adanya sistem koreksi jawaban essay komputerisasi, maka diharapkan dapat mempermudah pihak sekolahan maupun guru dalam koreksi jawaban essay, sehingga pihak sekolah maupun guru dapat menyingkat waktu serta mempermudah penilaian.

KATA KUNCI : *Essay, Metode Cosine Similarity, Metode TF-IDF*

I. LATAR BELAKANG

Menurut Bagio Rahardjo (2007), Evaluasi hasil belajar menjadi komponen yang sangat penting dalam proses pembelajaran, karena hasil evaluasi merupakan indikator dari pemahaman siswa terhadap materi ajar yang diberikan. Ujian dengan jawaban dalam bentuk *essay*, umumnya digunakan sebagai salah satu bentuk evaluasi untuk menentukan kemampuan siswa, dimana pilihan jawaban tidak diberikan. Siswa harus menjawab dengan kalimat, sehingga jawaban tersebut dapat bervariasi berdasarkan pemikiran mereka. Salah satu kelemahan dari jawaban dalam bentuk *essay* adalah sulitnya menilai jawaban, dan cenderung memakan waktu.

Pembuatan *essay* dalam penilaian hasil kegiatan belajar dinilai sebagai metode yang tepat, karena melibatkan kemampuan siswa untuk mengingat, mengorganisasikan, mengekspresikan dan mengintegrasikan gagasan yang dimiliki siswa tersebut. Tetapi, pemeriksaan *essay* membutuhkan waktu yang lebih lama karena sebagian besar dilakukan secara manual yaitu dengan cara membaca *essay* satu persatu, dari pada tes-tes yang bersifat objektif seperti pilihan ganda dan jawaban singkat. Para pengajar perlu menghabiskan banyak waktu untuk menilai jawaban ujian siswa mereka.

Untuk mengurangi kesalahan kesalahan dalam mengoreksi jawaban *essay*,beberapa penieliti telah membuat penelitian.(Ratna A., Budiardjo B., dan Hartanto. 2007). Sistem Penilaian OtomatisUntuk Menilai Ujian Dalam Bahasa Indonesia. Dipetik Desember 21, 2012, dariDepartemen Elektro, Fakultas Teknik, Universitas Indonesia, Depok 16424,
Indonesia:
http://journal.ui.ac.id/upload/artikel/02_AAP-Ratna_SIMPLE.pdf)

Pada peneliti ini suatu cara yang dapat yang dapat mempercepat penilaian *essay*, sedang banyak dikembangkan. Metode yang digunakan beraneka ragam, salah satunya adalah metode *Cosine Similarity*. Metode *Cosine Similiraty* adalah metode untuk menghitung kesamaan dari dua dokumen. Untuk menyamakan frekuensi jangka setiap kata pada kalimat yang ada digunakan persamaan Tf atau *Term Frequency*, *Term Frequency* mengekstrak dokumen menjadi proses yang terdiri dari kumpulan kata perkalimat. Tujuannya adalah menyamakan kedua kalimat pada suatu dokumen yang nantinya akan dibandingkan, sehingga kita dapat melangkah ke tahap selanjutnya yaitu tahapan *similarity*. Pada tahapan ini nantinya kedua kalimat yang sudah disamakan menggunakan Tf, akan dicari

nilai kesamaan pada kedua dokumen atau jawaban. Sehingga yang mendekati prosentase kesamaan bisa dikatakan bahwa jawaban terhadap soal yang ada adalah benar. Tingkatan prosentase akan mempengaruhi penilaian pada akhirnya. Dengan menggunakan metode tf dan *cosine Similarity* ini kita bisa lebih mudah untuk menentukkan hasil jawaban.

Karena beberapa alasan inilah mendorong penulis untuk membuat SISTEM KOREKSI JAWABAN ESSAY MENGGUNAKAN METODE COSINE SIMILARITY

II. METODE

III. Proses *Stemming* Bahasa Indonesia

Berdasarkan analisa *morfologi* yang telah dibahas sebelumnya, maka terdapat 5 aturan tahapan pada proses *stemming* dalam bahasa Indonesia (Tala., 2003). Aturan tersebut adalah:

1. Pada tahap awal yaitu menangani partikel *infleksional.*

Tabel 2.3. Aturan Pembentukan Partikel *Infleksional* (Tala, 2003).

2. Aturan tahap kedua menangani kata ganti milik *infleksional*.

Sufiks	Pengganti	Kondisi Ukuran	Kondisi Tambahan	Contoh
Kah	NULL	2	NULL	Bukukah→ buku
Lah	NULL	2	NULL	Adalah → ada
tah*	NULL	2	NULL	Apatah → apa
pun**	NULL	2	NULL	Bukupun → buku

Tabel 2.4. Aturan Pembentukan Kata Ganti Milik *Infleksional* (Tala, 2003).

Sufiks	Pengganti	Kondisi Ukuran	Kondisi Tambahan	Contoh
Ku	NULL	2	NULL	Bukuku → buku
Mu	NULL	2	NULL	bukumu → buku
Nya	NULL	2	NULL	bukunya → buku

3. Aturan tahap ketiga menangani urutan *prefiks derivasional* pertama.

Tabel 2.5. Aturan Pembentukan *Prefiks Derivasional* Pertama(Tala., 2003).

Prefiks	Pengganti	Kondisi Ukuran	Kondisi	Contoh
meng	NULL	2	NULL	Mengukur → ukur
meny	S	2	V…*	Menyapu → sapu
men	NULL	2	NULL	Menduga → duga
men	t	2	V…*	Menuduh → tuduh
mem	p	2	V…*	Memilah → pilah
mem	NULL	2	NULL	Membaca → baca
me	NULL	2	NULL	Merusak → rusak
Prefiks	**Pengganti**	**Kondisi Ukuran**	**Kondisi**	**Contoh**
peng	NULL	2	NULL	Pengukur → ukur
peny	s	2	V…*	Penyelam → selam
pen	NULL	2	NULL	Pendaki → daki
pen	t	2	V…*	Penari → tari
pem	p	2	V…*	Pemilah → pilah
pem	NULL	2	NULL	Pembaca → baca
di	NULL	2	NULL	Diukur → ukur
ter	NULL	2	NULL	Tersipu → sipu
ke	NULL	2	NULL	Kekasih → kasih

*kata dasar dimulai huruf *vocal*

4. Aturan tahap keempat menangani urutan *prefiks Derivasional* kedua.

Tabel 2.6. Aturan Pembentukan *Prefiks Derivasional* Kedua (Tala., 2003).

Prefiks	Pengganti	Kondisi Ukuran	Kondisi Tambahan	Contoh
Ber	NULL	2	NULL	berlari → lari
Bel	NULL	2	ajar	belajar → ajar
Be	NULL	2	kerja	bekerja → kerja
Per	NULL	2	NULL	penjelas → jelas
Pel	NULL	2	ajar	pelajaran → ajar
Pe	NULL	2	NULL	pekerja→kerja

5. Aturan tahap kelima menangani *sufiks Derivasional*.

Tabel 2.7. Aturan Pembentukan *Sufiks Derivasional* (Tala., 2003).

Sufiks	Pengganti	Kondisi Ukuran	Kondisi Tambahan	Contoh
Kan	NULL	2	Prefix є {ke,peng}	tarikkan → tarik (meng)ambilkan →ambil
An	NULL	2	Prefix є {di,meng,ter }	makanan→makan (per)janjian→janji
I	NULL	2	Prefix є {ber,ke,peng}	tandai→tanda (men)dapati→dapat

Kondisi ukuran adalah jumlah minimum suku kata dalam sebuah kata. Karena dalam bahasa Indonesia, kata dasar setidaknya mempunyai 2 suku kata. Maka kondisi ukuran dalam proses *stemming* bahasa Indonesia adalah dua. Adapun suku kata didefinisikan memiliki satu *vocal*. (Aziz Mustofa.,2009)

1. Algoritma *TF/IDF* (*Term Frequency – Inversed Document Frequency*)

Metode TF-IDF merupakan metode untuk menghitung bobot setiap kata yang paling umum digunakan pada *information retrieval*. Metode ini juga terkenal efisien, mudah dan memiliki hasil yang akurat (Abdul Aziz Ma'arif., 2015). Metode ini menggabungkan dua konsep untuk perhitungan bobot yaitu, frekuensi kemunculan sebuah kata di dalam sebuah dokumen tertentu dan *inverse frekuensi* dokumen yang mengandung kata tersebut(Ari Andika., 2015).

Frekuensi kemunculan kata di dalam dokumen yang diberikan menunjukkan seberapa penting kata tersebut di dalam dokumen tersebut. Frekuensi dokumen yang mengandung kata tersebut menunjukkan seberapa umum kata tersebut. Sehingga bobot hubungan antar sebuah kata dan sebuah dokumen akan tinggi apabila frekuensi kata tersebut tinggi di dalam dokumen dan frekuensi keseluruhan dokumen yang ada mengandung kata tersebut yang rendah pada kumpulan dokumen (*database*) (Ari Andika., 2015).

a. Perhitungan *TF/IDF*

Pada algoritma *TF/IDF* digunakan rumus untuk menghitung bobot (W) masing-masing dokumen terhadap kata kunci dengan rumus yaitu : (Adhit Herwansyah., 2012).

Dimana:

$$W_{dt} = tf_{dt} * IDF_t$$

d = dokumen ke-d

t = kata ke-t dari kata kunci

W = bobot dokumen ke-d terhadap kata ke-t

tf = banyaknya kata yang dicari pada sebuah dokumen

IDF =*Inversed Document Frequency*

IDF = log2 (D/df)

D = total dokumen

df = banyak dokumen yang mengandung kata yang dicari

Setelah bobot (W) masing-masing dokumen diketahui, maka dilakukan proses pengurutan dimana semakin besar nilai *W*, semakin besar tingkat similaritas dokumen tersebut terhadap kata kunci, demikian sebaliknya. (Adhit Herwansyah., 2012).

C.Metode *Cosine Similarity*

Cosine similarity adalah ukuran kesamaan yang lebih umum digunakan dalam *information retrieval* dan merupakan ukuran sudut antara vektor dokumen Da(titik (ax.bx)) dan Db (titik (ay by)). Tiap vektor tersebut merepresentasikan setiap kata dalam setiap dokumen (teks) yang dibandingkan dan membentuk sebuah segitiga, sehingga dapat diterapkan hukum kosinus untuk menyatakan bahwa

$$cos(C) = a^2 + b^2 - c2/2a\,Ь$$

Dimana

$$a^2 = \alpha_x^2 + \alpha_y^2 \,,\; b^2 = b_x^2 + b_y^2$$

Dan

$$C^2 = (b_x - \alpha_x)^2 + (\alpha_y - b_y)^2$$

Gantikan nilai-nilai tersebut untuk α , Ь , dan C , Sehingga didapatkan

$$Cox\ C = \frac{\alpha_x b_x + \alpha_y b_y}{\alpha_x^2 + \alpha_y^2 \;\; x \;\; Ь_{x^2} + Ь_{y^2}}$$

IV. HASIL DAN KESIMPULAN

A. Tampilan Program

1. Tampilan *User*

Halaman pertama atau home adalah tampilan saat pertama kali program dibuka.

Gambar 5.1.Tampila nawal program

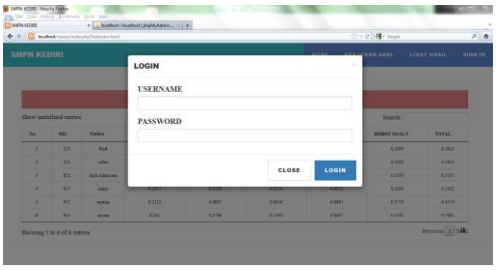
Pada tampilan awal terdapat 4 menu yaitu *Home,* Kerjakan soal, Lihat hasil dan *Sign in.* Menu *Home* adalah menu yang akan menampilkan tampilan awal. Menu kerjakan Soal menampilkan nama, nisn, dan soal yang akan dijawab oleh peserta didik telah disediakan admin. Pada menu Lihat Hasil dalah menu yang menampilkan hasil nilai siswa dari perhitungan *cosine similarity*. untukmenu *Login* hanya bisa diakses oleh admin untuk keperluan *insert, update,* soal dan kunci jawaban, kemudian perhitungan jawaban siswa untuk menentukan nilai hasil jawaban siswa.

Gambar 5.2.Tampilan halaman admin

Setelah siswa mengisi jawaban ,maka data akan disimpan kedalam *database*. Data jawaban terdapat pada halaman admin dan hanya dapat dilihat setelah *login* aplikasi.

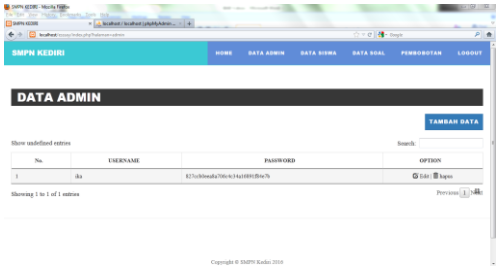
Terdapat 4 menu yang tersedia selain menu *home* padahalaman utama admin yaitu data *admin*, data siswa, data soal, pembobotan dan *logout*.

2. Halaman Admin



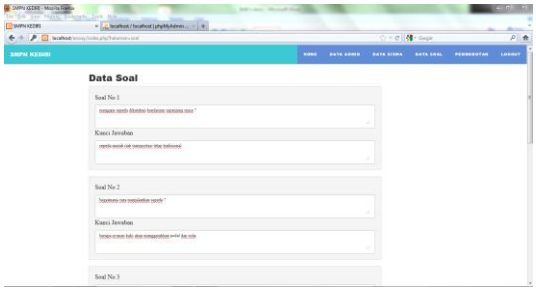
Gambar 5.3.Halaman login *admin*

Pada halaman utama admin akan menampilkan *id* dan *pasword login*.Menu *login* digunakan untuk login admin kedalam aplikasi. Dalam halaman *login, admin* harus memasuk kan *id* dan *password* yang sudah disimpan didatabase. Jika *id* dan *password* admin benar, maka admin akan masuk kedalam sistem, tetapi jika salah maka sistem tidak akan mengizin kan untuk masuk kedalam aplikasi.



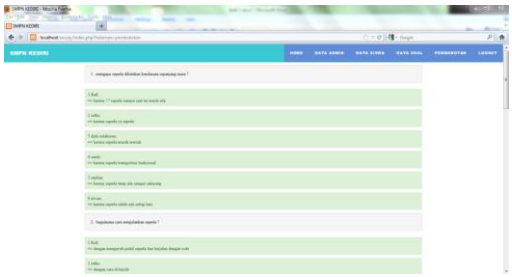
Gambar 5.4.Halaman utama admin

Menu data admin merupakan halaman yang dimanfaatkan untuk merubah *login id* dan *pasword*.



Gambar5.5.Halaman *detail* lihat Data Soal

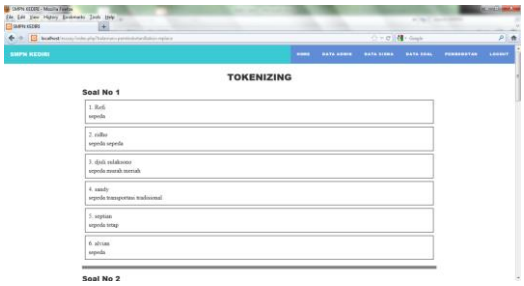
Pada halaman ini dimanfaatkan bagi admin untuk *insert*, *update* soal dan kunci jawaban.



Gambar 5.6.Tampilan *detail* data jawaban dari masing masing siswa

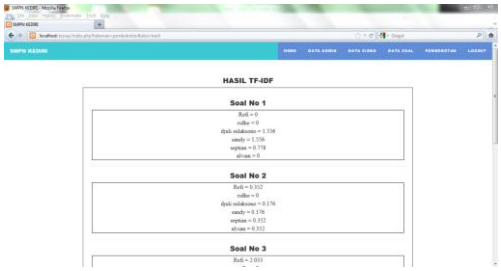
Pada tampilan ini adalah jawaban seluruh siswa, admin dapat mengetahui nama dan jawaban yang telah di *input kan* oleh

siswa.



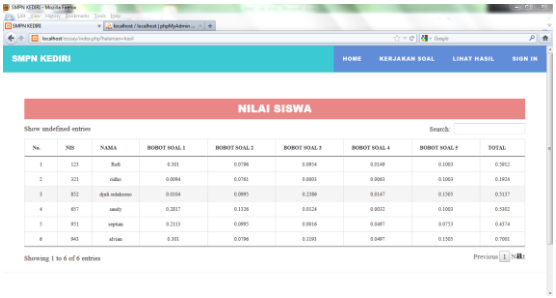
Gambar 5.7.Tampilan *detail tokenizing*

Setelah di*submit* menuju proses selanjutnya, dari masing- masing jawaban siswa admin dapat langsung melakukan pembobotan selanjutnya adalah melakukan proses *tokenizing* yaitu, menghitung jumlah kata masing-masing jawaban berdasarkan pada kunci jawaban. satuan token dan melakukan *stopword* pembuangan kata yang tidak penting dan tanda baca.



Gambar 5.8.Tampilan detail *TF-IDF*

Padatahapanstemmingakandijadikan query untukpembobotanpada tahapan *TF-IDF*. Halaman ini menampilkan seluruh nama siswa beserta hasil perhitungan .Semakin tinggi hasilnya.



Gambar 5.9.Tampilan Hasil Penilaian

Hasil akhir perhitungan dan program telah selesai.

V. DAFTAR PUSTAKA

Alfina, Tahta, Budi Santosa dan Ali Ridho Barakbah. 20212 . *Analisa Perbandingan Metode Hirarchical Clustering, K-means dan Gabungan Keduanya dalam Cluster Data.*Surabaya: Institut Teknologi Sepuuh Nopember (ITS).

Gie, The Liang. 1995. *Pengertian Skripsi*.Cina

Handoyo, Rendi, R. Rumani M dan Surya Micrandi Nasution. 2014.*Perbandingan Metode Cludtering Menggunakan Metode Single Linkage dan K-mens pada Pengelompokan Dokumen*. Bandung : Universitas Telkom.

Milan,Mc Schumacher. 2001. *Pengertian Penelitian*. San Francisco.

Pradnyana , Gede Aditra dan Ngurag Agus Sanjaya ER.2012.Prancangan dan *Implementasi Automated Documen Iintegration* dengan *Menggunakan Algoritma Complete Linkage Agglomerative Hierarchical Clustering.*

Santoso. 2007. *Pengertian Data Mining*. Jakarta. Mediakita.

Soekanto, Soerjno. 1998. *Penjelasan Penelitian*.Bandung

Swanja, Daniel, Chastine, Fatichah dan Diana Purwitasari. 2015. *Hypergraph – partitioning pada Co – Aurhorship Graph untuk Pengolompokan Penulis Berdasarkan Topik Penelitian*. Surabaya : Institut Teknologi Sepuluh Nopember(ITS).

Ridho Refi Septian. 2016. Perancangan Aplikasi Pengukuran Similaritas Pada Artikel Dengan Algoritma TF-IDF Di CV DIGIMEDIA CORPORINDO. Kediri.

Utomo, Mardi Siswo. 2013. *Implemetase Steamer Tala pada Aplikasi Berbasis Web*. Program Study Teknik Informatika, Universitas Stikubank.

Witten. 2005. *Data Mining*. New York

Wijaya, Ramadani. 2007. *Pengertian Penelitian Rekayasa*. Jakata. Mediakita.