

APLIKASI PENILAIAN UJIAN ESSAY OTOMATIS MENGGUNAKAN METODE COSINE SIMILARITY

Rahimi Fitri¹, Arifin Noor Asyikin²

mahaluddin@yahoo.com

(¹) Staf Pengajar Jurusan Teknik Elektro Politeknik Negeri Banjarmasin

Ringkasan

Ujian dengan sistem esai merupakan bentuk evaluasi dimana pilihan jawaban tidak disediakan, dan siswa harus menjawab dengan kalimat, sehingga dapat melatih siswa dalam menyampaikan sesuatu informasi secara verbal, selain itu ujian esai juga menuntut pemahaman yang lebih baik akan suatu ilmu dan dapat digunakan untuk mengukur tingkat pemahaman manusia akan suatu ilmu secara lebih mendalam. Ujian dengan system ini tetap menjadi pilihan pengajar untuk mengevaluasi tingkat kemampuan pemahaman siswa walaupun kenyataannya tidak mudah untuk memberikan penilaian yang objektif pada jawaban siswa. Pengajar memerlukan waktu yang banyak untuk memeriksa jawaban essay, semakin banyak jumlah ujian dan banyaknya jumlah pelajar yang mengikuti ujian, maka semakin banyak jumlah ujian yang dikoreksi oleh pengajar. Hal ini menyebabkan kualitas penilaian menurun dan terkadang penilaian tidak bersifat objektif lagi.

Beragam metode yang dapat digunakan untuk melakukan penilaian jawaban ujian esai secara otomatis salah satunya yang digunakan adalah metode cosine similarity. Metode Cosine Similarity adalah metode untuk menghitung kesamaan atau kemiripan dari dua buah dokumen, pada penelitian ini digunakan dokumen jawaban ujian siswa dan dokumen kunci jawaban pengajar. Untuk menyamakan frekuensi setiap kata yang terdapat didalam kalimat yang ada digunakan persamaan Tf atau Term Frequency, Term Frequency merupakan suatu faktor yang menentukan bobot kata yang didasarkan pada jumlah frekuensi kata dalam sebuah dokumen.

Berdasarkan hasil penelitian bahwa sistem penilaian esai otomatis menggunakan metode cosine similarity telah berjalan dengan baik untuk ujian esai dalam bahasa Inggris Hasil uji coba menunjukkan kesesuaian nilai sistem dengan nilai yang diberikan oleh pengajar adalah rata-rata 89,48%

Kata Kunci : *Ujian Esai, Cosine similarity, Term, Term Frequency, Bobot Kata*

1. PENDAHULUAN

Seiring dengan berkembangnya teknologi informasi dan komputer (TIK), dunia pendidikan mengalami perubahan sistem pendidikan dan pengajaran yang cukup signifikan. Beberapa institusi pendidikan mulai mengembangkan sistem *e-learning* atau *elektronik learning* dalam proses pengajarannya.

E-Learning merupakan salah satu metode pembelajaran dimana penyampaian materi pembelajaran, pelatihan, perkuliahan bahkan proses penilaian dilakukan dengan menggunakan peralatan elektronik yang terhubung dengan jaringan akses jarak jauh seperti internet atau intranet (Putri Ratna, et al., 2007; Aji, et al., 2011).

Dalam *e-learning*, proses evaluasi hasil belajar menjadi komponen yang sangat penting, karena hasil evaluasi merupakan indikator dari pemahaman siswa terhadap

materi yang ajar yang diberikan. Evaluasi dilakukan dengan mengadakan ujian yang dilaksanakan secara online dengan salah satu jenis soal yang diberikan adalah dalam bentuk sistem esai.

Ujian dengan sistem esai merupakan bentuk evaluasi dimana pilihan jawaban tidak disediakan, dan siswa harus menjawab dengan kalimat (Putri Ratna, et al., 2007), sehingga dapat melatih siswa dalam menyampaikan sesuatu informasi secara verbal, selain itu ujian esai juga menuntut pemahaman yang lebih baik akan suatu ilmu dan dapat digunakan untuk mengukur tingkat pemahaman manusia akan suatu ilmu secara lebih mendalam (Hamza, et al., 2013). Ujian dengan system ini tetap menjadi pilihan pengajar untuk mengevaluasi tingkat kemampuan pemahaman siswa walaupun kenyataannya tidak mudah untuk memberikan penilaian yang objektif pada jawaban siswa.

Pengajar memerlukan waktu yang banyak untuk memeriksa jawaban *essay*, semakin banyak jumlah ujian dan banyaknya jumlah pelajar yang mengikuti ujian, maka semakin banyak jumlah ujian yang dikoreksi oleh pengajar. Hal ini menyebabkan kualitas penilaian menurun dan terkadang penilaian tidak bersifat objektif lagi. Objektivitas adalah hal-hal yang bisa diukur yang ada di luar pikiran atau persepsi manusia dan subjektivitas adalah fakta yang ada di dalam pikiran manusia sebagai persepsi, keyakinan dan perasaan (Assadi, 2008). Salah satu kesulitan penilaian *essay* adalah subjektivitas. Banyak peneliti menyatakan bahwa sifat subjektif dari penilaian *essay* menyebabkan variasi penilaian di kelas yang diberikan oleh penilai manusia yang berbeda, yang dirasakan oleh siswa sebagai sumber ketidakadilan. Selanjutnya penilaian *essay* adalah kegiatan memakan waktu (Valenti, et al., 2003).

Ada beberapa metode yang dapat digunakan untuk mengatasi permasalahan tersebut seperti : *Project Essay Grade* (PEG), *Intelligent Essay Assessor*, *e-Rater*, *Bayesian Essay Test Scoring System* (Betsy) (Valenti, et al., 2003). Metode lainnya seperti *Latent Semantic Analysis* (LSA) yang pada awalnya digunakan untuk melakukan *indexing* pada sistem temu balik informasi (Deerwester, et al., 1990; Aditya, et al., 2009), kemudian metode ini dikembangkan lagi dalam sebuah system yang bernama SIMPLE atau system penilaian esai otomatis untuk menilai ujian dalam bahasa Indonesia dimana setiap kalimat yang terdapat dalam setiap jawaban diekstraksi dan direpresentasikan kedalam bentuk matematik. Nilai dari jawaban esei diperhitungkan, dengan mencocokkan ada atau tidak adanya kata yang dianggap penting, pada kelompok kata dalam matriks yang telah dipersiapkan oleh *human rater* (penilai manusia). (Putri Ratna, et al., 2007).

Selain metode yang disebutkan diatas terdapat metode lain yang dapat digunakan untuk melakukan penilaian jawaban esai otomatis yaitu dengan metode *cosine similarity*, (Fuat, 2010).

Metode Cosine Similarity adalah metode untuk menghitung kesamaan atau kemiripan dari dua buah dokumen. Untuk menyamakan frekuensi setiap kata yang terdapat didalam kalimat yang ada digunakan persamaan Tf atau *Term Frequency*, *Term Frequency* merupakan suatu faktor yang menentukan bobot kata yang didasarkan pada jumlah frekuensi kata dalam sebuah dokumen. Nilai jumlah kemunculan suatu kata (tf) diperhitungkan dalam pemberian bobot

terhadap suatu kata. Semakin besar jumlah kemunculan suatu kata (tf tinggi) dalam dokumen, semakin besar pula bobotnya dalam dokumen atau akan memberikan nilai kesesuaian yang semakin besar (Yates, et al., 1999). Metode ini bertujuan adalah menyamakan kedua kalimat pada suatu dokumen yang nantinya akan dibandingkan, kemudian hasil perbandingan akan digunakan sebagai nilai *similarity*.

Pengembangan aplikasi penilaian jawaban ujian esai secara otomatis ini diterapkan pada soal *essay* bahasa Inggris dan diharapkan mampu menyelesaikan permasalahan subjektivitas pengajar yang terjadi pada saat proses penilaian jawaban ujian esai secara manual.

Sesuai dengan latar belakang, maka tujuan dalam penelitian ini adalah menganalisa mekanisme penilaian esai yang selama ini telah dilaksanakan pengajar, serta merancang dan mengaplikasikan penilaian ujian esai otomatis menggunakan metode *cosine similarity*

2. METODE PENELITIAN

Soal *esei* merupakan salah satu bentuk evaluasi dimana pilihan jawaban dari soal tidak disediakan, dan siswa harus menjawab dengan kalimat, sehingga jawaban dapat sangat bervariasi sesuai dengan pemikiran masing-masing peserta ujian (Putri Ratna, et al., 2007).

Penilaian adalah salah satu proses untuk mengambil keputusan dengan menggunakan informasi yang didapat melalui pengukuran hasil belajar baik yang menggunakan instrumen tes maupun yang tidak. Penilaian dengan metode esai menjadi pilihan dalam mengevaluasi tingkat kemampuan dari siswa walaupun faktanya tidak mudah dalam memberikan penilaian yang objektif pada setiap siswa pada saat mengevaluasi jawaban dari soal esai. Evaluasi pembelajaran dengan menggunakan bentuk esai ini oleh banyak peneliti dianggap alat yang sangat sesuai untuk mengukur hasil pembelajaran, begitu juga untuk mengamati kemahiran dan pemahaman berpikir tingkat tinggi seperti sintesis dan analisis (Fuat, 2010; Putri Ratna, et al., 2007)

Rancangan Penelitian

Rancangan penelitian yang akan digunakan dalam penelitian ini adalah sebagaimana ditunjukkan pada gambar 1. Berdasarkan rancangan system diatas maka terdapat beberapa hal yang harus dilakukan dalam system yaitu :

a. Input Data

Data yang dimasukkan kedalam sistem adalah data yang dimasukkan oleh pengajar dan siswa yaitu :

1. Pertanyaan soal.
2. Kunci jawaban.
3. Jawaban siswa.

b. Pemrosesan Teks

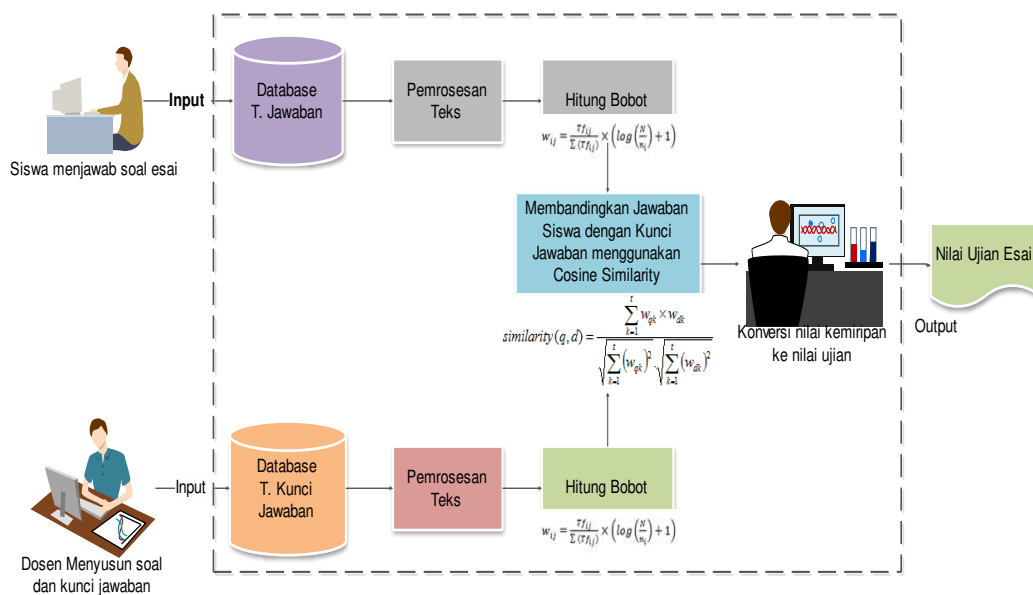
Dalam tahapan ini seluruh dokumen yang dimasukkan kedalam sistem akan dilakukan pemrosesan awal. Tahapan dalam pemrosesan teks adalah sebagaimana ditunjukkan oleh gambar 3.2. Seluruh dokumen yang dimasukkan harus dalam bentuk file *.txt, jika dokumen yang dimasukkan dalam format *.pdf maka diperlukan suatu *library* yang berfungsi untuk mengekstraksi pdf. Tahapan selanjutnya adalah memisahkan setiap kata didalam kalimat berdasarkan spasi, enter, tabulasi, titik (.) dan koma (,), teknik ini sering disebut tokenisasi..

Setelah seluruh kata didalam dokumen berhasil dipisahkan, selanjutnya adalah

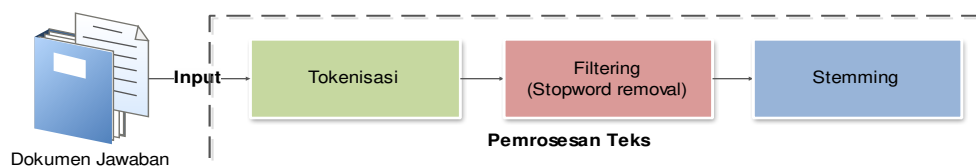
membuang kata yang termasuk kedalam daftar *stopword*. Pada saat pemisahan kata dalam dokumen sering kali ditemukan banyak kata yang dengan frekuensi yang tinggi tetapi tidak memiliki makna penting untuk dijadikan konsep, kata-kata tersebut biasanya masuk kedalam daftar kata *stopword* contoh nya adalah kata '*dan*' kata ini merupakan kata yang termasuk dalam daftar *stopword*.

Selanjutnya adalah stemming, dalam tahapan ini sistem akan mencari kata dasar dari setiap kata yang ditemukan. Dalam penelitian ini, pada tahapan stemming, berguna untuk mencari kata dasar dari setiap kata yang ditemukan. Berikut ini adalah tahapan praproses berdasarkan IPO (Input, Proses, Output).

- Input : dokumen/*corpus* dari kunci jawaban dan dari jawaban siswa.
- Proses : tokenisasi, *stopword* dan *stemming*.
- Output : daftar kata



Gambar 1. Rancangan Penelitian (Sumber : diolah 2015)



Gambar 2. Pemrosesan Teks (Sumber : diolah 2015)

- c. Menghitung tf dan pembobotan kata
Tahapan selanjutnya adalah menghitung frekuensi dari setiap kata yang terbentuk. Frekuensi kata berdasarkan jumlah kata didalam dokumen (Yates, et al., 1999).selanjutnya setiap kata yang ditemukan dihitung bobotnya menggunakan persamaan 9. Pembobotan konsep dihitung berdasarkan frekuensi konsep dibagi total frekuensi yang ditemukan.
- Input : Term
 - Proses : menghitung bobot kata (menggunakan persamaan berikut)

$$w_{ij} = \frac{Tf_{ij}}{\sum(Tf_{ij})} \times \left(\log \left(\frac{N}{n_i} \right) + 1 \right)$$

Dimana:

- w_{ij} = bobot kata/term t_j terhadap dokumen d_i
- tf_{ij} = jumlah kemunculan kata/term t_j dalam d_i
- N = jumlah semua dokumen yang ada dalam database
- n_i = jumlah dokumen yang mengandung kata/term t_j
- t_j = (minimal ada satu kata yaitu term t_j)

- d. Perhitungan kemiripan menggunakan *cosine similarity*
Tahapan ini dilakukan perhitungan kemiripan dari jawaban esai siswa dengan jawaban kunci yang dimiliki dosen, dengan menggunakan persamaan berikut ini.

$$similarity(q, d) = \frac{\sum_{k=1}^t w_{qk} \times w_{dk}}{\sqrt{\sum_{k=1}^t (w_{qk})^2} \cdot \sqrt{\sum_{k=1}^t (w_{dk})^2}}$$

Dimana:

- w_{qk} : Bobot query
- w_{dk} : Bobot dokumen

Jadi pada tahapan ini hanya membandingkan dua buah dokumen. Pada tahapan ini menghasilkan output berupa nilai kemiripan antara kedua dokumen tersebut yang selanjutnya akan dikonversi menjadi nilai siswa.

- e. Konversi Nilai Kemiripan menjadi Nilai Ujian Esai.
Nilai kemiripan yang dihasilkan sebelumnya dikonversi menjadi nilai jawaban ujian esai siswa berdasarkan rentang nilai versi penilaian manusia (*human rates*), rentang nilai ini mengacu pada penelitian yang dilakukan oleh Fuat tahun 2010. Adapun rentang nilai tersebut

adalah seperti yang ditunjukkan oleh tabel (1).

Tabel 1. Rentang Nilai Jawaban Siswa

Perbandingan Rentang Nilai	
Nilai Kemiripan	Nilai <i>human rates</i>
0.01 - 0.10	10
0.11 - 0.20	20
0.21 - 0.30	30
0.31 - 0.40	40
0.41 - 0.50	50
0.51 - 0.60	60
0.61 - 0.70	70
0.71 - 0.80	80
0.81 - 0.90	90
0.91 - 1	100

Sumber : (Fuat, 2010)

Hasil penilaian nilai siswa seperti tabel 1 akan dibagi sesuai dengan jumlah soal yang ada. Misal jika terdapat 4 soal essay maka jika siswa menjawab dengan benar dan nilai kemiripannya adalah 1 maka nilai yang diperoleh siswa adalah $(100/4)=25$

Metode Pengembangan Aplikasi

Metode yang digunakan dalam pengembangan aplikasi penilaian ujian esai otomatis menggunakan algoritma *cosine similarity* ini adalah menggunakan metode *software development live cycle* (SDLC) dengan tahapan sebagai berikut :

1. Analisis Kebutuhan Sistem

Tahapan awal yang harus dilakukan dalam pengembangan Aplikasi Penilaian ujian Online adalah menganalisa kebutuhan, pada tahapan ini terdapat beberapa hal yang harus dilakukan yaitu :

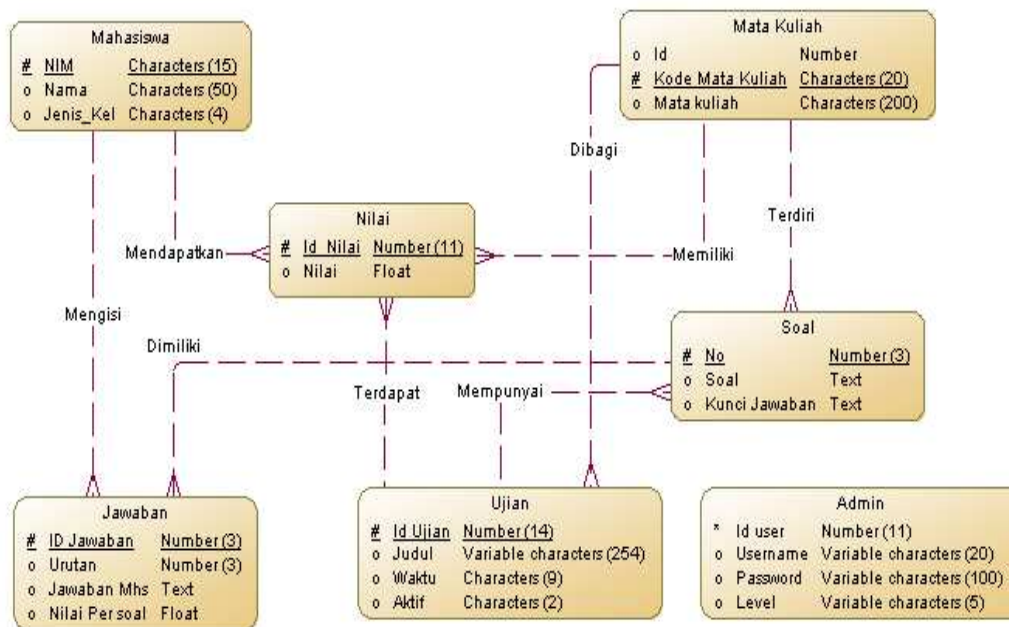
- a. Pengumpulan Data Awal
Pada tahapan ini Pengumpulan data awal yang berhubungan dengan pelaksanaan ujian yang dilaksanakan didalam kelas belajar mengajar. Untuk pengumpulan data awal maka perlu dilakukan analisa terhadap sistem penilaian ujian yang dilaksanakan secara essay.
- b. Menentukan Kebutuhan Fungsional Sistem
Kebutuhan fungsional dari system ini adalah sebagai berikut :
 1. Terdapat form untuk registrasi peserta
 2. Terdapat form *login* untuk peserta yang akan mengikuti ujian berdasarkan nim dan *password* masing-masing.
 3. Peserta dapat memilih mata kuliah yang akan diujikan berdasarkan tipe ujian yang telah aktif pada saat itu.

4. Sistem harus mampu menampilkan Soal ujian berbentuk essay.
5. Sistem Harus mampu menampilkan kunci jawaban yang tersedia pada sistem nantinya akan dicocokkan pada jawaban masing-masing peserta.
6. Soal ujian diambil secara acak dari bank soal. Tiap tipe ujian terdiri dari lima soal.
7. Terdapat Waktu yang disediakan oleh sistem diatur oleh admin.
8. Peserta tidak dapat mengikuti tipe ujian yang sama pada setiap mata kuliah lebih dari satu kali.
9. Sistem mampu menampilkan nilai per soal
10. Sistem mampu menampilkan total nilai dari seluruh soal yang sudah

dijawab peserta setelah selesai menjawab.

Pada bagian administrator

1. Administrator memiliki akses khusus untuk *input*, *edit* dan penghapusan soal ujian dari *bank* soal.
2. Administrator harus memeriksa siapa saja peserta yang berhak mengikuti ujian.
3. Administrator memiliki hak akses penuh untuk mengubah data peserta ujian.
4. Administrator menentukan tipe ujian yang akan diaktifkan agar bisa dikerjakan oleh peserta



Gambar 4. Conceptual Data Model

2. Kebutuhan Sistem

Pada subbab ini akan dijelaskan kebutuhan implementasi aplikasi penilaian ujian online yang meliputi kebutuhan perangkat keras dan kebutuhan perangkat lunak.

A. Kebutuhan Perangkat Keras

Kebutuhan perangkat keras dengan spesifikasi yang disarankan untuk menjalankan sistem dalam performa terbaik adalah sebagai berikut :

- a. PC atau Notebook dengan spesifikasi minimal:

- ✓ Processor dual core
- ✓ Memory 1 Gb
- ✓ Hardisk 500Gb

- b. Printer
- c. Modem

B. Kebutuhan Perangkat Lunak

Kebutuhan perangkat lunak yang diperlukan untuk menjalankan sistem ini agar berjalan dengan baik adalah sebagai berikut :

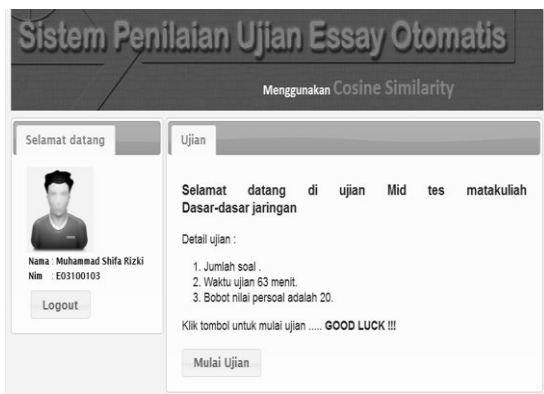
- a. WampServer Version
- b. Apache Version 2.2.8
- c. MySQL Version 5.0.5.1b

3. HASIL DAN PEMBAHASAN

Berdasarkan kebutuhan fungsional dari sistem yang dibangun, maka penelitian ini dikatakan berhasil jika :

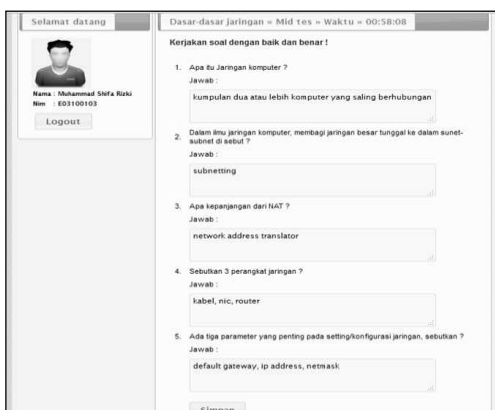
a. Sistem mampu menampilkan halaman ujian untuk siswa.

Gambar 5 berikut ini adalah halaman awal sebelum memulai ujian yang dapat diikuti siswa. Matakuliah yang diujikan adalah hanya mata kuliah yang tersedia. Sebelum masuk kehalaman ujian mahasiswa yang telah teregistrasi sebelumnya melakukan login kedalam sistem.

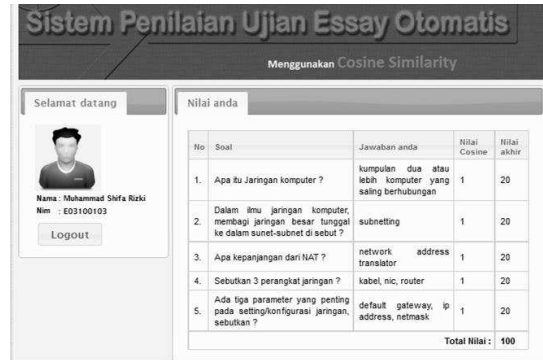


Gambar 5. Halaman Awal Sebelum Memulai Ujian

Selanjutnya untuk memulai ujian siswa menekan tombol Ujian seperti pada gambar 6 berikut ini. Pada halaman ujian siswa akan mengisi jawaban yang pada kotak jawaban yang telah disediakan. Soal yang diberikan akan diacak yang diambil di bank soal yang tersimpan didalam database. Pada halaman ujian sistem menyediakan waktu yang terhitung mundur dan apabila waktu habis maka sistem akan langsung menilai jawaban dari mahasiswa

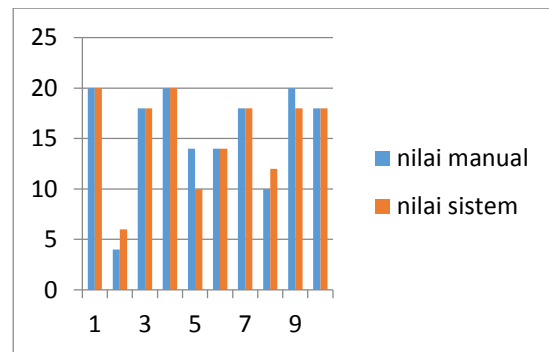


Gambar 6. Halaman Ujian

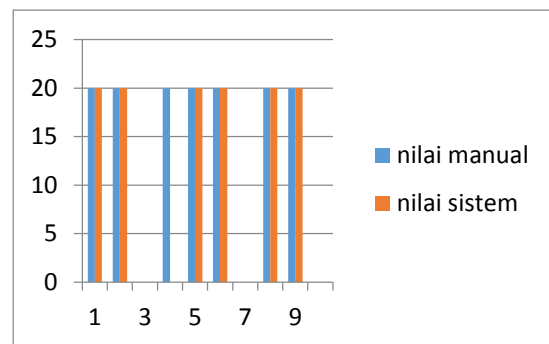


Gambar 7. Halaman Nilai Hasil Ujian

Setelah selesai ujian maka mahasiswa dapat langsung mengetahui hasil ujian seperti pada gambar 7. Pada gambar 7 diasumsikan bahwa mahasiswa menjawab soal dengan jawaban yang sama dengan kunci jawaban. Sehingga diperoleh nilai kemiripan sama dengan 1. Berdasarkan pengujian pada beberapa mahasiswa maka diperoleh hasil perbandingan nilai seperti pada gambar grafik berikut ini :



Gambar 8. Grafik Perbandingan hasil penilaian manual dengan hasil dari sistem pada soal nomor 1



Gambar 9. Grafik Perbandingan hasil penilaian manual dengan hasil dari sistem pada soal nomor 2

Berdasarkan jawaban siswa pada soal nomor 1 (Gambar 8) diperoleh hasil kesesuaian antara hasil penilaian yang dilakukan secara manual oleh manusia dengan penilaian yang dilakukan oleh sistem yaitu rata-rata 89,48 % dengan nilai kemiripan terkecil yaitu 67% , dan untuk jawaban untuk soal 2 (gambar 9) diperoleh hasil kesesuaian 100% jika jawaban tepat dan 0% jika jawaban salah, hal ini dikarenakan jawaban pada kunci jawaban hanya terdapat 1 kata saja yaitu 'subnetting', sehingga apabila mahasiswa menuliskan 'subneting' maka sistem akan menganggap salah karena secara susunan huruf berbeda. Sedangkan bagi pengajar jawaban siswa tersebut tetap dianggap benar.

4. SIMPULAN DAN IMPLIKASI

Berdasarkan penelitian yang telah dilaksanakan maka dapat diambil kesimpulan:

1. Penelitian ini telah mampu menganalisa sistem penilaian ujian essay secara manual sehingga dapat dijadikan acuan untuk membangun sistem penilaian otomatis
2. Metode penilaian kemiripan jawaban menggunakan cosine similarity telah berhasil di implementasikan dalam mengembangkan aplikasi penilaian ujian essay secara otomatis
3. Nilai kemiripan yang dihasilkan telah dikonversi kedalam nilai mahasiswa berdasarkan range yang telah ditentukan dibagi dengan jumlah soal yang disajikan
4. Berdasarkan hasil ujicoba yang telah dilakukan dapat disimpulkan bahwa kesesuaian antara nilai manual yang dilakukan oleh pengajar yang dibandingkan dengan nilai sistem adalah dengan rata-rata 89,48%
5. Sistem ini memiliki kekurangan dan kelebihan, untuk jawaban yang sama persis sistem mampu menilai 100% sedangkan untuk jawaban untuk tipe tertentu tidak dapat bisa menilai dengan baik.

5. DAFTAR PUSTAKA

- [1] Aditya, M., & Witanti, N. (2009). Penilaian Esai Jawaban Bahasa Indonesia Menggunakan Metode SVM - LSA. *Journal of Information Systems, Volume 5, Issues 1* , p 33-41.
- [2] Aji, R. B., Baizal, A., & Firdaus, Y. (2011). Automatic Essay Grading System

menggunakan Metode Latent Semantic Analysis. *Seminar Nasional Aplikasi Teknologi Informasi (SNATI 2011)*, ISSN : 1907-5022. Yogyakarta.

- [3] Deerwester, s., Harshman, R., Dumais, S., Furnas, G. W., & Landauer, T. (1990). Indexing by Latent Semantic Analysis.
- [4] Fuat, R. (2010). *Sistem Penilaian Esai Otomatis Pada Elearning menggunakan Metode Cosine Similarity*. Surabaya: Buku Tugas Akhir Mahasiswa Jurusan Sistem Informasi Fakultas Teknologi Informasi Institut Teknologi Sepuluh November .
- [5] Hamza, S., Sarosa, M., & Santoso, P. B. (2013). Sistem Koreksi Soal Essay Otomatis Dengan Menggunakan Metode Rabin Karp. 7.
- [6] Putri Ratna, A. A., Budiardjo, B., & Hartanto, D. (2007). SIMPLE : Sistem Penilaian Esei Otomatis Untuk Menilai Ujian Dalam Bahasa Indonesia. *Makara, Teknologi, Vol, 11, No.1* , 5-11.
- [7] Valenti, S., Neri, F., & Cucchiarelli, A. (2003). An Overview of Current Research on Automated Essay Grading. 2.
- [8] Yates, R., & Neto, B. (1999). *Modern Information Retrieval*. Addison Wesley Longman Limited.