

PERBANDINGAN ALGORITMA TF/IDF DAN BLEU UNTUK PENILAIAN JAWABAN ESAI OTOMATIS

Hesti Wahyu Nugroho✉, Feddy Setio Pribadi, Ulfah Mediaty Arief, Sri Sukamta

Jurusan Teknik Elektro, Fakultas Teknik, Universitas Negeri Semarang, Indonesia

Info Artikel

Sejarah Artikel:

Diterima September 2014
Disetujui Oktober 2014
Dipublikasikan Oktober 2014

Keywords:

TF/IDF, BLEU, perluasan kunci jawaban, penyisipan huruf, Automatic Essay Scoring, Frequency Term, Inverse Document Frequency, web

Abstrak

Setiap proses pembelajaran memerlukan suatu evaluasi berupa tes. Dalam perkembangannya tes dapat dilakukan secara online. Jenis tes antara lain tes benar salah, test pilihan ganda, tes mencocokkan, dan tes esai. Sudah banyak penelitian di luar negeri maupun di dalam negeri yang mengembangkan metode-metode sebagai penilai jawaban esai otomatis. Sebagai contoh Eksperimen LSA untuk esai GMAT (Graduation Management Achievement Test) menghasilkan persetujuan dengan manusia sebesar 85%-91% (Valenti, Neri, & Cucchiarelli, 2003), Electronic Essay Rater (E-Rater) menghasilkan 87%-94% (Valenti, Neri, & Cucchiarelli, 2003), untuk penelitian di dalam negeri hasil uji coba SIMPLE (Ratna, Budiharjo, & Hartanto, 2007) menghasilkan 69.80%-94.64% untuk lima mahasiswa dan 77.18%-98.42% untuk sepuluh mahasiswa. Hasil yang didapatkan masih jauh dari yang diharapkan. Hal ini yang mendasari dilakukannya penelitian mengenai pembobotan yang baik dalam penilaian jawaban esai otomatis. Penelitian ini membandingkan Algoritma pembobotan TF/IDF dan BLEU dengan alat bantu Automatic Essay Scoring yang berbasis web untuk membobotkan jawaban siswa terhadap kunci jawaban esai serta meneliti pengaruh perluasan kunci jawaban serta penyisipan huruf pada kata kurang huruf (toleransi) untuk memaksimalkan hasil penilaian. Algoritma TF-IDF merupakan salah satu skema pembobotan istilah/term dalam pencarian dokumen yang terdiri pembobotan lokal Frequency Term (TF), pembobotan global Inverse Document Frequency (IDF), dan normalisasi. Sedangkan Algoritma BLEU merupakan salah satu sistem evaluasi otomatis yang dikenalkan oleh grup riset IBM. Dalam penelitian ini penulis menghitung nilai BLEU sampai 4-gram (4 kata) saja, sedangkan esai yang dinilai dibatasi pada esai yang merupakan jawaban dari pertanyaan pada tes esai (maksimal 250 kata per jawaban), bukan esai karangan yang panjang (lebih dari 250 kata) dan jawaban yang dinikahi berupa kalimat bukan berupa jawaban hasil perhitungan (matematika, fisika dan kimia). Uji coba dilakukan dengan 10 soal esai dengan 22 peserta ujian. Secara keseluruhan, rata-rata korelasi TF/IDF-penilaian guru mencapai 0.70 dengan nilai korelasi tertinggi mencapai 0.98, sedangkan rata-rata korelasi BLEU-penilaian guru mencapai 0.63 dengan nilai korelasi tertinggi mencapai 0.97, secara keseluruhan nilai yang dihasilkan algoritma TF/IDF lebih tinggi dari BLEU.

Abstract

Every learning process requires an evaluation of a test. In the development of the test can be done online. Types of tests include tests completely false, multiple choice tests, matching tests, and essay tests. There have been many studies abroad and within the country are developing methods as assessor automated essay answers. For example Experimental LSA for GMAT essay (Graduation Management Achievement Test) results in agreement with the human by 85% -91% (Valenti, Neri, & Cucchiarelli, 2003), the Electronic Essay Rater (E-Rater) resulted in 87% -94% (Valenti, Neri, & Cucchiarelli, 2003), for research on domestic trial results SIMPLE (Ratna, Budiharjo, & Hartanto, 2007) resulted in 69.80% -94.64% for the five students and 77.18% -98.42% for ten students. The results obtained are still far from the expected. It is for conducting research on a good weighting in the assessment of automated essay answers. This study compared the weighting algorithm TF / IDF and BLEU with Automatic Essay Scoring tools are web-based to membobotkan answers to the answer key student essay and examines the impact of the expansion of the answer key and the insertion of letters in the word less letters (tolerance) to maximize the results of the assessment. Algorithm TF-IDF weighting scheme is one term / terms in the search for documents consisting of local weighting Term Frequency (TF), Inverse Document Frequency weighting globally (IDF), and normalization. While the algorithm is one of the BLEU automatic evaluation system introduced by IBM research group. In this study the authors calculate the value to 4-gram BLEU (4 words) only, while the assessed essay is limited to the answers of the questions on the test essay (maximum of 250 words per answer), not an essay written by a long (more than 250 words) and answer dinikahi not be the answer in the form of sentence calculation results (mathematics, physics and chemistry). The test is done with 10 essays by 22 examinees. Overall, the average correlation of TF / IDF-teacher ratings reached 0.70 with the highest correlation value reaches 0.98, while the average correlation of BLEU-assessment guru reached 0.63 with the highest correlation value reaches 0.97. overall value of the resulting algorithm TF / IDF higher than BLEU.

PENDAHULUAN

Setiap proses pembelajaran memerlukan suatu evaluasi berupa tes, tidak terkecuali dengan proses pembelajaran secara *e-learning* (moodle dan edmodo). Tes adalah alat untuk memperoleh data tentang perilaku individu (Allen dan Yen, 1979: 1).

Jenis-jenis tes yang digunakan ada beranekaragam, tes benar salah (*True-False Test*), tes pilihan berganda (*Multiple Choice Test*), tes mencocokkan (*Matching Test*) dan tes uraian/esai (Sudijono, 2005). Jenis tes yang sering digunakan pada moodle dan edmodo adalah pilihan ganda dan isian singkat. Namun kedua jenis tes tersebut kurang memadai dalam proses belajar mengajar dan kurang dapat mengukur tingkat pemahaman akan ilmu yang didapat.

Tes esai memiliki beberapa kelebihan dibandingkan dengan tes objektif. Kelebihannya antara lain dapat mencegah timbulnya permainan spekulasi pada peserta tes dan melalui tes esai ini dapat diketahui tingkat kedalaman penguasaan peserta tes terhadap materi yang diujikan. Tes esai juga memiliki kekurangan, yaitu kesulitan dalam penilaiannya. Dibutuhkan banyak waktu, pikiran, dan tenaga dalam penilaiannya. Menurut Mason seperti yang disampaikan dalam (Valenti, Neri, & Cucchiarelli, 2003), hampir 30% dari waktu yang dimiliki seorang guru digunakan untuk menilai ujian atau evaluasi siswa. Selain itu, mungkin saja terdapat subjektivitas dari penilai ketika memeriksa dan menilai jawaban esai.

Kekurangan bentuk tes esai di atas dapat diatasi dengan penggunaan teknologi informasi dan komputasi, yaitu dengan penilaian esai secara otomatis menggunakan komputer. Hal ini mungkin dilakukan karena sekarang proses pembelajaran sudah mengarah menuju pembelajaran *e-learning* (moodle dan edmodo).

Sudah banyak penelitian di luar negeri maupun di dalam negeri yang mengembangkan metode-metode sebagai penilaian jawaban esai otomatis. Dari sekian penelitian ini didapat hasil yang belum sesuai dengan yang diharapkan tetapi masing-masing penelitian memiliki kelebihan

dan kekurangan. Karena itulah penilaian tes esai ini menarik untuk dikembangkan

Fokus utama permasalahan yang akan diteliti pada penelitian ini adalah:

1. Bagaimana hasil penilaian jawaban tes esai otomatis untuk jawaban berbahasa Indonesia antara algoritma pembobotan *Term Frequency/Inverse document Frequency* (TF/IDF) dengan algoritma pembobotan BLEU?
2. Bagaimana pengaruh perluasan kunci jawaban (*query expansion*) dan penggunaan algoritma penyisipan huruf dalam kata yang kurang huruf (toleransi) terhadap hasil penilaian, baik dengan algoritma pembobotan TF/IDF maupun algoritma pembobotan BLEU?

Tujuan penelitian ini adalah sebagai berikut:

1. Memanfaatkan/ mengimplementasikan sistem penilaian jawaban tes esai otomatis untuk jawaban berbahasa Indonesia menggunakan algoritma pembobotan *Term Frequency/Inverse document Frequency* (TF/IDF) dan algoritma pembobotan BLEU.
2. Mengetahui pengaruh perluasan kunci jawaban (*query expansion*) dan penggunaan algoritma penyisipan huruf dalam kata yang kurang huruf (toleransi) terhadap hasil penilaian, baik dengan algoritma pembobotan TF/IDF maupun algoritma pembobotan BLEU.

Manfaat yang diharapkan dari penelitian ini adalah ikut berkontribusi dalam perkembangan teknologi khususnya memberikan informasi mengenai hasil perbandingan penilaian tes esai otomatis dengan algoritma pembobotan TF/IDF dan algoritma pembobotan BLEU.

METODE

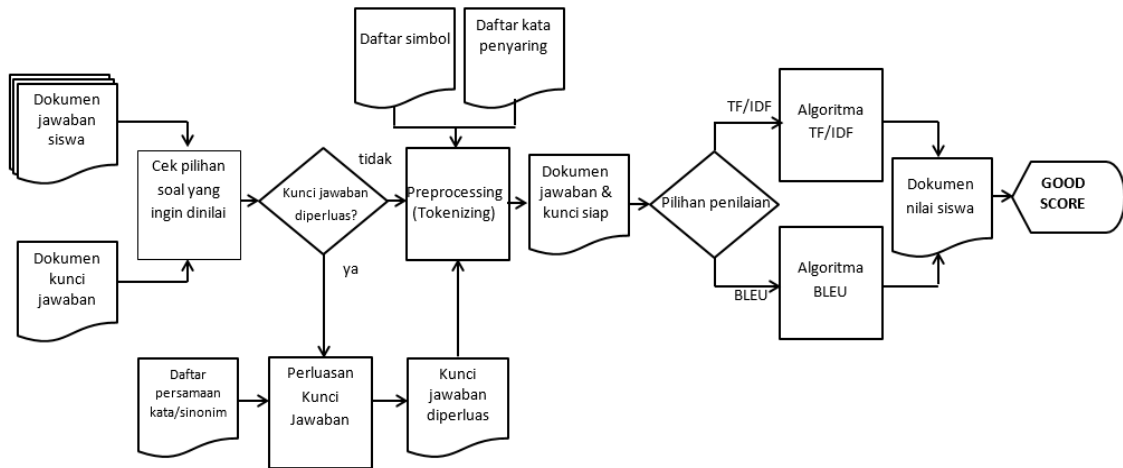
Metode pengumpulan data yang digunakan dalam penelitian ini yaitu metode observasi dan metode studi pustaka. Observasi

dilakukan di SMK Negeri 1 Bawen, hasil yang didapatkan data berupa soal yang terdiri dari 10 soal, kunci jawaban untuk masing-masing soal dan jawaban siswa dari 22 siswa yang masing-masing siswa mengerjakan 10 soal pada tes esai mata pelajaran Pendidikan Agama Islam kelas X-TB-A SMK Negeri 1 Bawen.

Penelitian ini membutuhkan alat bantu berupa aplikasi *Automatic Essay Scoring*. Dalam

penelitian ini alat bantu *Automatic Essay Scoring* dikembangkan dengan metode pengembangan perangkat lunak *Waterfall*. Tahapan pengembangan meliputi analisis sistem, desain, pengkodean dan *testing*.

Berikut rancangan alur proses penilaian jawaban siswa dalam tahapan desain:



Gambar 1. Rancangan Alur Proses Penilaian Jawaban Siswa

Proses penilaian terdiri atas tiga proses utama, yaitu perluasan kunci jawaban, *preprocessing*, dan pembobotan jawaban siswa terhadap kunci jawaban dengan menggunakan algoritma pembobotan TF/IDF dan BLEU.

1. Algoritma TF/IDF

Algoritma TF-IDF merupakan salah satu skema pembobotan istilah/*term* dalam pencarian dokumen. Ada beberapa tahapan pada proses pembobotan TF/IDF, yaitu:

Document Frequency (DF)

$$d \in D: t \in d$$

Keterangan:

d = dokumen ke - d t = kata ke -

t

Pembobotan Global Inverse Document Frequency (IDF)

$$IDF = \log_{10} \left(\frac{N}{df_t} \right)$$

Keterangan:

D atau N = jumlah dokumen

df_t = jumlah dokumen yang mengandung kata t

idf = log dari jumlah dokumen dibagi jumlah dokumen yang mengandung kata t

Pembobotan Lokal *Frequency Term (TF)*

$$TF = 1 + \log_{10}(tf_{t,d})$$

Keterangan:

$tf_{t,d}$ = jumlah kemunculan kata t pada dokumen d

Perhitungan TF/IDF

$$w_{t,d} = tf_{t,d} \times IDF_t$$

$$w_{t,d} = \left(1 + \log_{10}(tf_{t,d}) \right) \times \log_{10} \left(\frac{N}{df_t} \right)$$

Keterangan:

$w_{t,d}$ = bobot dokumen d terhadap kata t

2. Algoritma BLEU

BLEU (*Bilingual Evaluation Understudy*) merupakan salah satu sistem evaluasi otomatis

yang dikenalkan oleh grup riset IBM. BLEU mengukur modified *N-gram precision score* antara hasil terjemahan otomatis (jawaban siswa) dengan terjemahan rujukan (kunci jawaban) dan menggunakan konstanta yang dinamakan *brevity penalty*.

Rumus menghitung BLEU sebagai berikut (Papineni, Kishore; Salim Roukos; Todd Ward and Wei-Jing Zhu, 2001):

$$BP_{BLEU} = \begin{cases} 1 & \text{if } c > r \\ e^{(1-r/c)} & \text{if } c \leq r \end{cases}$$

$$P_n = \frac{\sum_{C \in corpus} \sum_{gram \in C} count_{clip}(n - gram)}{\sum_{C \in corpus} \sum_{gram \in C} count(n - gram)}$$

$$BLEU = BP_{BLEU} \cdot e^{\sum_{n=1}^N w_n \log p_n}$$

Keterangan:

BP = brevity penalty

c = jumlah kata jawaban siswa

r = jumlah kata kunci jawaban

P_n = modified precision score, jumlah pada jawaban siswa yang cocok dengan kunci jawaban dibagi dengan jumlah n-gram pada jawaban siswa

$W_n = 1/N$

N = jumlah n-gram (4)

3. Perluasan Kata Kunci Pencarian (*Query Expansion*)

Query Expansion adalah sebuah teknik yang digunakan pada mesin pencari di mana pengguna memberikan masukan kata-kata tambahan pada kata kunci pencarian (*query*). Hal ini dilakukan untuk meningkatkan *recall*, yaitu perbandingan antara jumlah dokumen relevan yang keluar dengan jumlah dokumen yang relevan (Manning, Raghavan, & Schutze, 2008).

Pertanyaan mendasar dalam melakukan query expansion adalah bagaimana cara membuat query baru yang sudah di-expand. Cara pertama adalah pengguna dapat memberikan tambahan kata-kata sendiri. Cara kedua adalah dengan analisis global menggunakan thesaurus, yaitu dengan menambahkan sinonim atau kata-kata yang berhubungan dengan kata-kata pada query. Thesaurus dapat diartikan sebagai

kumpulan term atau istilah pada domain tertentu dan relasi antara istilah-istilah tersebut. Terdapat thesaurus untuk domain kesehatan, matematika, ilmu computer, dan lain-lain (Chevallet, 2002). Relasi yang paling sering digunakan adalah sinonim.

4. Penggunaan Algoritma Penyisipan Huruf

Algoritma penyisipan huruf digunakan sebagai solusi dalam memperkecil kesalahan penulisan kata-kata oleh peserta tes yang masih bisa ditolerir manusia dan tidak bisa ditolerir sistem (Harisma, 2008). Algoritma ini biasanya digunakan dalam mesin pencari, contohnya pada google. Ketika kita memasukan query pada salah satu mesin pencari google, otomatis sistem mesin pencari google akan membenarkan query kita dengan pesan " *mungkin yang Anda maksud....*". Algoritma ini juga diterapkan dalam penelitian ini, dengan tujuan untuk memperkecil kesalahan penulisan kata-kata oleh peserta tes yang masih bisa ditolerir manusia sehingga hasil penilaian yang didapat bisa lebih maksimal.

Pada tahapan *testing/* pengujian menggunakan pengujian korelasi. Pengujian korelasi merupakan hasil perbandingan hubungan antara variable X (penilaian guru persoalan) dan variable Y (penilaian sistem).

HASIL DAN PEMBAHASAN

Uji coba dilakukan dengan menggunakan bantuan sistem penilai esai otomatis *Automatic Essay Scoring*. Berikut spesifikasi *software* yang digunakan:

Hosting:

Server: server47.idhostinger.com, IP server: 185.28.20.221, Versi Apache: Apache/2.2.14, Versi PHP: 5.3, Versi MYSQL: 5.1.61, Versi phpMyAdmin: 3.5.2.2

Browser

Browser : Google Chrome Version 34.0.1847.116 m

Berikut spesifikasi *hardware* yang digunakan:

Prosesor : Intel® Core™ i3-2350M CPU
@ 2.30GHz 2.30 GHz, RAM : 4.00 GB, *Operating system* : Windows 7 64-bit

Skenario uji coba yang dilakukan pada penelitian ini:

1. Uji coba 1, penilaian esai dengan menggunakan algoritma pembobotan *Term Frequency/Inverse document Frequency* (TF/IDF) dan algoritma pembobotan BLEU tidak menggunakan perluasan kunci jawaban/sinonim.

2. Uji coba 2, penilaian esai dengan menggunakan algoritma pembobotan *Term Frequency/Inverse document Frequency* (TF/IDF) dan algoritma pembobotan BLEU menggunakan perluasan kunci jawaban/sinonim.

Penilaian yang diperoleh baik dengan algoritma TF/IDF maupun algoritma BLEU merupakan hasil pencocokan dokumen jawaban siswa dengan dokumen kunci untuk tiap soal.

Tabel 1. Tabel Nilai Siswa

No	User	Penilaian Guru	Tanpa Sinonim		Sinonim	
			TF/IDF	BLEU	TF/IDF	BLEU
1	X-TB-A-2	66	39.9	44.7	45.7	46.5
2	X-TB-A-3	72	58.8	47	63.4	46.6
3	X-TB-A-4	55	40.1	42.3	43.05	40.7
4	X-TB-A-5	77	55.3	46.8	62	49.2
5	X-TB-A-7	59	53.6	33.8	58.3	33.9
6	X-TB-A-8	68	42.4	43.6	47	44.7
7	X-TB-A-9	83	60.7	56.3	63.2	55
8	X-TB-A-10	89	67.6	57.6	72.3	59.8
9	X-TB-A-11	91	71.6	60.4	76.4	60
10	X-TB-A-15	68	53.7	34.2	53.1	36
11	X-TB-A-16	77	62.7	46.6	65.4	45.8
12	X-TB-A-17	89	64.3	52.9	69.7	55.5
13	X-TB-A-18	89	65.3	57.5	73.5	61.5
14	X-TB-A-19	87	68.4	61.1	74.3	63.6
15	X-TB-A-21	74	35.65	35.95	41.1	38.65
16	X-TB-A-24	88	57.9	55.7	63.7	56
17	X-TB-A-26	89	63.1	57.7	70.2	61.2
18	X-TB-A-27	73	49	41	64.1	41.5
19	X-TB-A-29	78	63.8	55.2	69.4	56
20	X-TB-A-31	94	67.6	59.3	74.6	62.1
21	X-TB-A-33	72	66.6	53.6	69.2	52.9
22	X-TB-A-35	100	97	94.9	97	91.7
Rata-rata		79	59.32	51.73	64.39	52.68

Sumber: Data sekunder diolah, 2014

Tabel 1 menunjukkan nilai siswa dari penilaian guru, penilaian dengan algoritma TF/IDF dan penilaian dengan algoritma BLEU.

Tabel 2. Tabel Korelasi TF/IDF dan BLEU tanpa Perluasan Kunci Jawaban/Sinonim

Soal	Rata-Rata Nilai per Soal			Korelasi		Pengujian	
	Penilaian Guru	TF/IDF	BLEU	TF/IDF	BLEU	TF/IDF	BLEU
1	8.00	5.84	5.51	0.85	0.57	7.10	3.11
2	8.82	7.68	7.07	0.66	0.76	3.92	5.19
3	6.64	4.35	1.75	0.85	0.49	7.12	2.51
4	6.55	6.97	5.38	0.52	0.63	2.75	2.58
5	9.27	9.06	8.95	0.93	0.96	11.29	15.45
6	7.73	6.60	6.96	0.85	0.67	7.22	4.08
7	5.18	3.82	4.74	0.87	0.85	7.95	7.35
8	8.77	4.89	5.20	0.39	0.44	1.89	2.17
9	8.59	1.01	1.83	0.02	0.29	0.10	1.34
10	9.45	9.10	4.35	0.98	0.66	19.76	3.94
Rata	7.90	59.32	51.73	0.69	0.63	6.91	4.87

Sumber: Data sekunder diolah, 2014

Tabel 3. Tabel Korelasi TF/IDF dan BLEU dengan Perluasan Kunci Jawaban/Sinonim

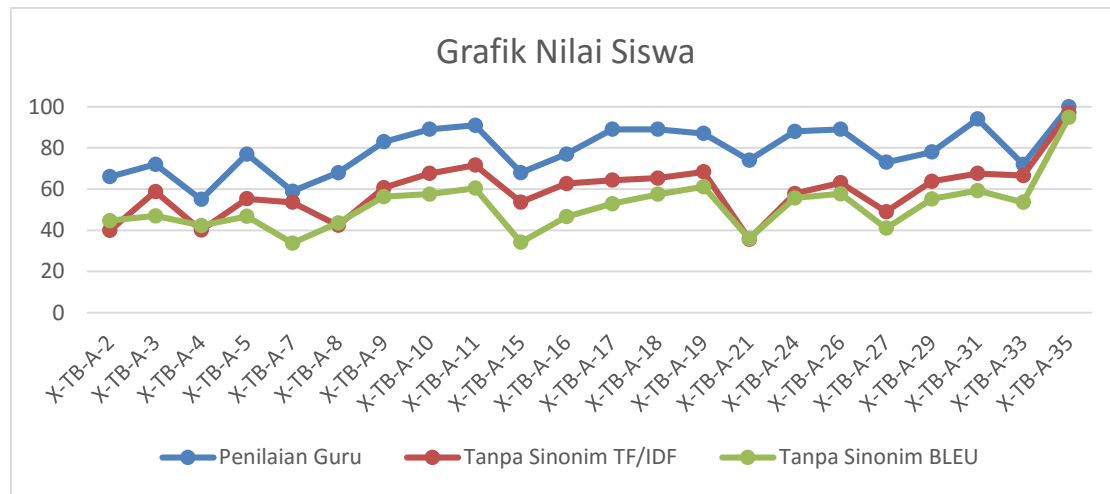
Soal	Rata-Rata Nilai per Soal			Korelasi		Pengujian	
	Penilaian Guru	TF/IDF	BLEU	TF/IDF	BLEU	TF/IDF	BLEU
1	8.00	6.15	5.57	0.84	0.59	6.95	3.24
2	8.82	9.09	6.99	0.81	0.82	6.23	6.47
3	6.64	4.33	1.55	0.77	0.48	5.36	2.47
4	6.55	7.30	5.51	0.43	0.60	2.13	3.33
5	9.27	9.07	8.93	0.93	0.97	10.90	16.80
6	7.73	7.65	7.25	0.89	0.55	8.91	2.92
7	5.18	4.66	4.89	0.93	0.87	11.68	7.97
8	8.77	5.57	5.47	0.38	0.41	1.83	2.04
9	8.59	1.02	1.83	0.02	0.29	0.11	1.34
10	9.45	9.54	4.69	0.98	0.74	19.76	4.96
Rata	7.90	64.49	52.68	0.70	0.63	7.39	5.15

Sumber: Data sekunder diolah, 2014

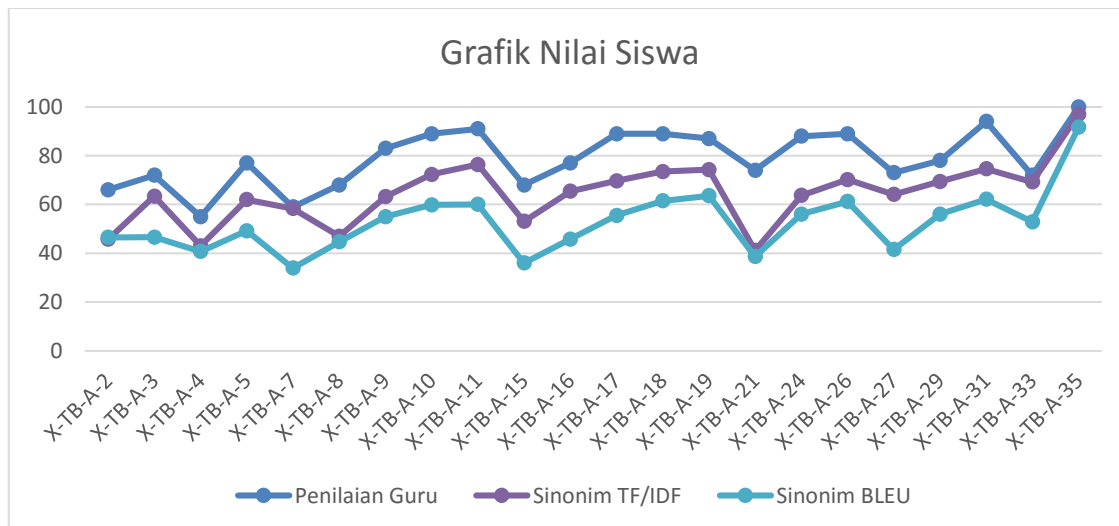
Keterangan :

 $\alpha = 5\%$ T table = $t(0.025, 20) = 2.423117$

	Ho	
	diterima	
	Ho	tidak diterima



Gambar 2. Grafik nilai siswa tanpa perluasan kunci jawaban/sinonim



Gambar 3. Grafik nilai siswa dengan perluasan kunci jawaban/sinonim

Tabel 2 dan 3 menunjukkan Korelasi TF/IDF dan BLEU dengan atau tanpa Perluasan Kunci Jawaban/Sinonim.

Gambar 2 dan 3 menunjukkan grafik nilai siswa dari penilaian guru, pembobotan TF/IDF dan pembobotan BLEU dengan atau tanpa perluasan kunci jawaban/sinonim.



Gambar 4. Contoh perluasan kunci jawaban/sinonim

Gambar 4 menunjukkan proses perluasan kunci jawaban/sinonim. Terlihat bahwa kata “QS.” Mengalami perluasan menjadi “surat” dan “surah”.

SIMPULAN

Sistem penilaian esai otomatis dengan algoritma pembobotan TF/IDF dan algoritma BLEU telah diimplementasikan dalam penelitian ini dan dilakukan uji coba untuk membandingkan hasil penilaian yang didapatkan. Simpulan yang diperoleh dari penelitian ini adalah sebagai berikut:

1. Berdasarkan rata-rata penilaian yang didapatkan antara hasil pembobotan TF/IDF dengan penilaian guru dan hasil pembobotan BLEU dengan penilaian guru, hasil penilaian pembobotan TF/IDF lebih tinggi dibandingkan pembobotan BLEU. Hal ini mungkin dikarenakan bentuk soal tes yang sebagian besar terdiri atas soal yang menyuruh siswa untuk menyebutkan sehingga jawaban dari siswa beragam.
2. Kedua algoritma pembobotan memiliki penilaian yang tinggi pada bentuk soal yang jawabannya berupa jawaban yang sudah pasti dan memiliki penilaian rendah pada soal uraian yang cara penyampaian jawabannya bisa beragam. Hal ini dikarenakan bentuk soal yang memiliki jawaban yang sudah pasti struktur jawaban, panjang jawaban serta kata-kata yang digunakan lebih mudah diproses/ dibobotkan.
3. Perluasan kunci jawaban atau sinonim membantu menaikkan hasil pembobotan baik pembobotan TF/IDF maupun pembobotan BLEU. Hal ini mungkin dikarenakan perluasan kunci jawaban pada sistem dapat menaikkan kemungkinan kecocokan kunci jawaban dengan variasi kata yang digunakan siswa pada jawabannya.
4. Penggunaan algoritma penyisipan huruf dalam pencocokan kata pada proses pembobotan membantu menaikkan korelasi hasil pembobotan baik pembobotan dengan algoritma

TF/IDF maupun BLEU, hal ini dikarenakan dalam penggunaannya algoritma penyisipan huruf memberikan toleransi pada kesalahan penulisan jawaban seperti halnya penilaian guru.

SARAN

Saran untuk penelitian sistem penilai esai otomatis untuk Bahasa Indonesia adalah sebagai berikut:

1. Agar lebih akurat, penelitian sistem penilai jawaban esai dapat dilakukan dengan algoritma pembobotan TF/IDF dengan memperhatikan susunan kata pada jawaban agar hasil yang lebih baik.
2. Penelitian lebih lanjut mengenai penilaian jawaban esai otomatis dengan algoritma pembobotan TF/IDF dapat menggunakan varian kunci jawaban/rubrik yang lebih banyak beserta skor untuk tiap kuncinya.

UCAPAN TERIMA KASIH

Ucapan terimakasih ditujukan kepada Prof. Dr. Fathur Rokhman M.Hum, Rektor Universitas Negeri Semarang (Unnes), Drs. Mohammad Harlanu M.Pd, Feddy Setio Pribadi, S. Pd., M.T., Drs. Suryono M.T., dan Drs. Agus Suryanto M.T., serta Dosen Fakultas Teknik Unnes.

DAFTAR PUSTAKA

- Arikunto, Suharsimi (2010). Prosedur Penelitian Suatu Pendekatan Praktik. Yogyakarta : Rineka Cipta.
- Connolly, Thomas M., Begg, Carolyn E. (2005). Database System : A Practical Approach to Design, Implementation, and Management, fourth edition. England: Addison-Wesley.
- Harisma, N. Z. (2008). Implementasi sistem penilaian esai otomatis metode LSA dengan tiga bobot kata kunci. Depok: Fakultas Teknik Universitas Indonesia.

- Hermawandi, D. (2008). Implementasi pembobotan SICBI pada aplikasi essay grading metode LSA. Depok: Fakultas Teknik Universitas Indonesia.
- Krisnanda, B. P. (2005). Sistem penilaian essay otomatis dengan menggunakan metode LSA. Depok: Fakultas Teknik Universitas Indonesia.
- Manning, C. D., Raghavan, P., Schutze, H. (2008). Introduction to information retrieval. New York: Cambridge University Press.
- Octaria, D. (2008). Implementasi skema pembobotan pada aplikasi penilaian esai otomatis metode LSA. Depok: Fakultas Teknik Universitas Indonesia.
- Pressman, Roger S. (2002). Rekayasa Perangkat Lunak :Pendekatan Praktisi (Buku 1). Yogyakarta : Andi
- Ratna, A.A.P., Budiardjo, B., Hartanto, D. (2007, April). SIMPLE: sistim penilai esai otomatis untuk menilai tes dalam bahasa Indonesia. Jurnal Makara Teknologi, 5-11.
- Septiantri, Heninggar (2009). Perbandingan Metode Latent Semantic Analysis dan Vector Space Model untuk Sistem Penilai Jawaban Esai Otomatis Bahasa Indonesia. Depok: Fakultas Teknik Universitas Indonesia.
- Subagyo, Pangestu (2002). Forecasting Konsep dan Aplikasi. Yogyakarta: BPFE Yogyakarta.
- Sudijono, Anas. (2005). Pengantar evaluasi pendidikan. Jakarta: PT Raja Grafindo Persada.
- Valenti, S., Neri, F., Cucchiearelli, A. (2003). An overview of current research on automated essay grading. Jurnal of Information Technology Education, Volume 2.
- Whitten, Jeffrey L (2007). System Analysis and Design Methods 7th. Mc Graw Hill Education. New York
- Williams, R. (2001). Automated essay grading: An evaluation of four conceptual models. In A. Herrmann and M. M. Kulski (Eds), Expanding Horizons in Teaching and Learning. Proceedings of the 10th Annual Teaching Learning Forum, 7-9 February 2001. Perth: Curtin University of Technology. <http://otl.curtin.edu.au/events/conferences/tlf/tlf2001/williams.html>