

PENGEMBANGAN SISTEM PENILAIAN OTOMATIS TERHADAP JAWABAN SOAL PENDEK DAN TERBUKA DALAM EVALUASI BELAJAR ONLINE BERBAHASA INDONESIA

Husni Thamrin¹

¹Program Studi Teknik Informatika,
Fakultas Komunikasi dan Informatika, Universitas Muhammadiyah Surakarta
Jl. A Yani Pabelan, Kartasura, Sukoharjo 57162
¹Email: husni.thamrin@ums.ac.id

Abstrak

Penggunaan sistem e-learning unggul dalam fleksibilitas jadwal dan fleksibilitas laju belajar. Keunggulan sistem dapat ditingkatkan jika terdapat fasilitas evaluasi belajar yang dapat melakukan penilaian otomatis. Pada dasarnya sistem penilaian otomatis tersedia dalam sistem e-learning namun hanya untuk jawaban tertutup. Penilaian otomatis untuk soal terbuka belum tersedia padahal soal terbuka mempunyai keunggulan dalam memotivasi kreatifitas siswa. Pada penelitian ini dilakukan modifikasi terhadap sistem e-learning agar soal jawaban pendek dapat dinilai secara otomatis dengan penilaian yang tidak kaku, artinya jika siswa menjawab soal dengan jawaban yang tidak sama namun mempunyai kemiripan dengan kunci jawaban, siswa masih mendapatkan nilai. Soal evaluasi menggunakan bahasa Indonesia sebagai pengantar. Nilai siswa ditentukan dari kesamaan atau kemiripan antara jawaban siswa dan kunci jawaban. Jika jawaban yang diketik siswa tidak ada dalam kamus, kemiripan jawaban dihitung dengan algoritma Levenshtein sedangkan jika jawaban terdapat dalam kamus, kemiripan jawaban dihitung dari sinonimitas kata bahasa Indonesia. Kecenderungan guru dalam menilai jawaban siswa menjadi acuan perhitungan nilai jawaban. Survei dilakukan kepada para guru untuk mengetahui nilai yang diberikan atas jawaban berbagai soal dengan berbagai nilai similaritas. Survei menunjukkan adanya dua pola pemberian nilai oleh guru. Guru cenderung lebih toleran terhadap kesalahan tipografi yang tidak menghasilkan kata baru. Guru cenderung mengurangi nilai siswa jika jawaban siswa tidak tepat sama, meskipun siswa memberikan jawaban yang bermakna sama (sinonim) dengan kunci jawaban. Algoritma yang diusulkan akan dapat memprediksi cara penilaian guru dalam menilai jawaban soal pendek dan terbuka.

Kata kunci: *penilaian otomatis, soal jawaban pendek, evaluasi belajar, bahasa Indonesia*

1. PENDAHULUAN

Pembelajaran elektronik (*e-learning*) memungkinkan proses pembelajaran dilakukan secara massal dengan jumlah peserta didik yang banyak. Pembelajaran elektronik memberi keuntungan pada penghematan ruang kelas dan tenaga pengajar, dan dapat dilangsungkan pada jarak jauh. Pada pengajaran klasikal, seorang pengajar secara efektif mampu mengelola kelas dengan jumlah peserta didik sekitar 25 orang (Figlio, 1998; Hillstock, 2005) sedangkan dengan pembelajaran elektronik, seorang pengajar diperkenankan mengelola kelas dengan kapasitas lebih besar tergantung sifat pembelajaran dan materi ajar (Wignal, 2011). Proses pembelajaran elektronik tidak memerlukan banyak ruang kelas. Ruang dibutuhkan pada saat dilakukan ujian atau tutorial khusus. Teknologi *e-learning* memungkinkan proses belajar mengajar dilakukan dari jarak jauh sehingga meningkatkan daya jangkau dan kesempatan pendidikan masyarakat luas (Nugraheni, 2009).

Keberhasilan pembelajaran elektronik dipengaruhi oleh keaktifan dan kemandirian peserta didik dalam mempelajari bahan ajar. Peserta didik yang aktif dan mandiri dapat memanfaatkan fasilitas pembelajaran dengan baik dan memperoleh sisi unggul *e-learning*. Keunggulan yang dimaksud adalah fleksibilitas jadwal dan fleksibilitas laju belajar. Jadwal dapat disesuaikan dengan aktivitas peserta didik dan laju belajar dapat diatur sesuai kemampuannya.

Keberhasilan pembelajaran elektronik dipengaruhi oleh desain pembelajaran dan materi ajar (Jona, 2000). Salah satu komponen penting dalam proses belajar mengajar adalah pemberian kuis sebagai upaya membandingkan apa yang telah dipahami oleh peserta didik dengan apa yang seharusnya dimengerti. Dalam sistem *e-learning*, soal kuis dapat diberikan sewaktu-waktu selama proses pembelajaran dan umpan balik terhadap jawaban diberikan sesaat setelah peserta didik

menjawab soal. Proses interaktif dalam kuis membuat pembelajaran elektronik menjadi menarik dan menantang.

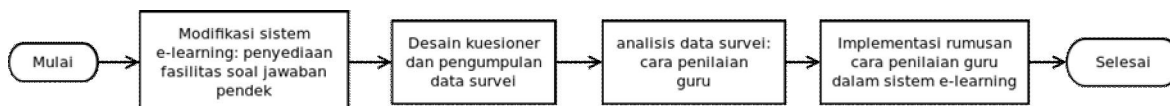
Terdapat berbagai jenis soal kuis dan ujian yang dapat diberikan dalam pembelajaran yang dapat dikategorikan sebagai soal terbuka dan soal tertutup. Soal tertutup meliputi soal pilihan ya-tidak, soal pilihan ganda dan soal pencocokan. Soal terbuka meliputi soal dengan jawaban pendek (yaitu terdiri atas kata, frase atau satu kalimat tunggal) dan soal uraian (yaitu terdiri atas kalimat majemuk atau banyak kalimat). Jawaban soal terbuka membutuhkan koreksi manual sedangkan soal jawaban soal tertutup dapat dikoreksi oleh program komputer. Jumlah peserta didik yang besar memberi konsekuensi lamanya waktu yang diperlukan bagi pengajar untuk melakukan koreksi manual. Oleh karena itu kebanyakan pertanyaan pada modul pembelajaran elektronik diberikan dalam bentuk soal tertutup sehingga tidak diperlukan koreksi manual. Pada kenyataannya, kebanyakan piranti lunak pembelajaran online telah memberikan fasilitas pembuatan soal dengan penilaian otomatis untuk jenis soal tertutup.

Soal jawaban pendek dan soal uraian merupakan soal terbuka yang menuntut peserta didik menuliskan secara kreatif jawaban yang soal yang diberikan. Peserta tidak sekedar memilih jawaban di antara jawaban yang disediakan. Soal terbuka mempunyai kelebihan dalam evaluasi belajar karena lebih menggambarkan kemampuan analitis peserta didik. Pemberian hanya soal tertutup dalam evaluasi dapat menyebabkan pikiran peserta didik dipenuhi pengetahuan yang lepas dari konteks. Pengetahuan yang lepas dari konteks sulit dicerna dan diingat, apalagi diaplikasikan di dunia nyata (Jona, 2000).

Tulisan ini mengungkapkan hasil penelitian tentang pengembangan sistem penilaian otomatis terhadap soal pendek dan terbuka pada pembelajaran elektronik. Skor otomatis diperoleh dari kemiripan antara jawaban siswa dan kunci jawaban yang dibobot dengan angka tertentu untuk mendekati cara penilaian guru. Jika jawaban siswa merupakan sebuah kata yang berbeda dari kunci, kemiripan ditentukan dari apakah jawaban tersebut merupakan sinonim dari kunci. Jika jawaban siswa bukan merupakan kata yang dikenal, kemiripan jawaban siswa terhadap kunci dihitung berdasarkan algoritma Levenshtein.

2. METODOLOGI

Penelitian ini diawali dengan proses modifikasi sistem *e-learning* agar terdapat fasilitas kuis atau soal dengan jawaban pendek (lihat Gambar 1). Beberapa sistem *e-learning* sudah memiliki fasilitas tersebut, misalnya MoodleTM atau ClarolineTM. Namun fasilitas yang tersedia umumnya mempunyai cara penilaian dengan melihat kesamaan atau ketidaksamaan jawaban siswa dengan kunci (Gonzales-Barbone & Llamas-Nistal, 2008). Modifikasi yang dilakukan dalam kegiatan penelitian ini berupa cara penilaian yang mempunyai rentang skor antara 0 dan 100. Nilai 0 diberikan pada jawaban yang siswa yang sama sekali berbeda dengan kunci, nilai 100 diberikan kepada jawaban yang tepat sama dengan kunci, dan nilai antara yang didasarkan pada kemiripan jawaban siswa dengan kunci. Kemiripan yang dimaksud dapat berupa kemiripan teks/string dan kemiripan makna (sinonimi).



Gambar 1. Alur kegiatan penelitian

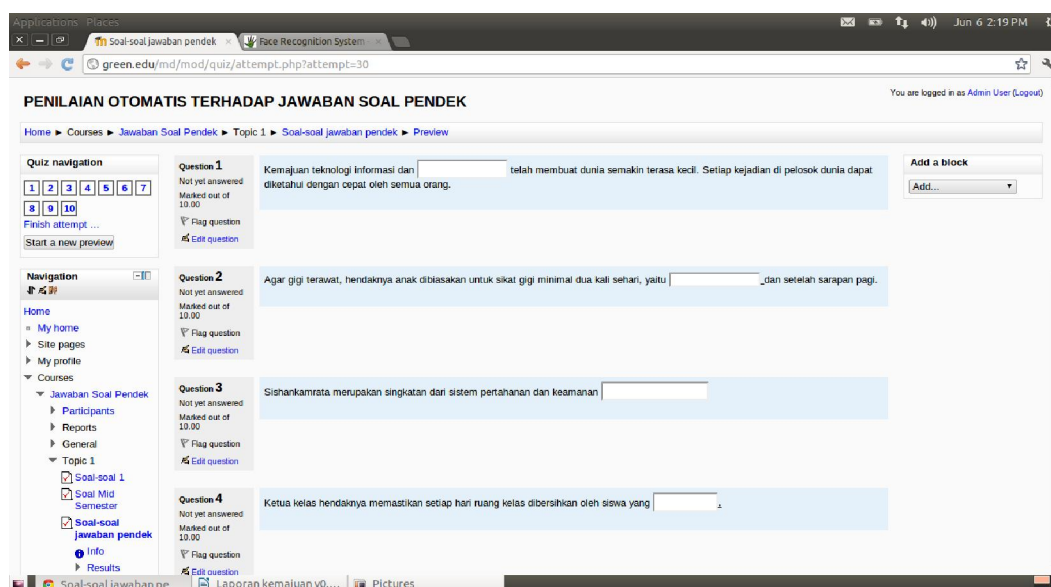
Penentuan nilai (skor) yang diberikan terhadap jawaban siswa didasarkan pada cara penilaian guru-guru sekolah. Pada penelitian ini dilakukan survei terhadap 124 guru SD, SMP dan SMA di Jawa Tengah. Para guru dihadapkan pada contoh soal berbahasa Indonesia yang telah dijawab siswa dan guru diminta memberi nilai (skor) terhadap jawaban siswa. Jawaban siswa bervariasi mulai dari kata yang tepat sama dengan kunci, sinonim, dan kata yang sama sekali tidak berhubungan makna. Beberapa jawaban siswa mengandung kesalahan tipografi (kesalahan ketik). Kesalahan ketik tersebut bisa membentuk kata baru ataupun tidak.

Berdasarkan data hasil survei dibuat prediksi nilai yang mungkin diberikan guru jika siswa memberi jawaban yang tepat sama dengan kunci, jawaban yang merupakan sinonim kunci dan

jawaban yang mengandung tipografi. Rumusan prediksi nilai akan diimplementasikan dalam sistem *e-learning* sehingga dapat dilakukan proses penilaian otomatis (*autoscoring*) terhadap jawaban siswa oleh sistem tersebut.

3. HASIL DAN PEMBAHASAN

Peneliti memilih Moodle sebagai sistem pembelajaran (*Learning Management System*, LMS) untuk dimodifikasi dengan penambahan pola pertanyaan baru. Moodle dipilih dengan beberapa alasan. Moodle merupakan LMS yang sangat dikenal dan memiliki fasilitas yang cukup lengkap. Moodle dibangun menggunakan bahasa pemrograman PHP yang merupakan bahasa pemrograman *open source*, dan Moodle sendiri bersifat *open source*. Moodle secara *default* menggunakan sistem manajemen basis data (*database management system*, DBMS) MySQL yang juga merupakan piranti lunak *open source*. DBMS yang lain, misalnya PostgreSQL dan Oracle, dapat pula digunakan jika data sudah berkembang semakin besar. Moodle dikembangkan berdasarkan lisensi GPL (*general public licence*) sehingga dapat dikembangkan dan dimodifikasi secara bebas tanpa perlu izin dan membayar biaya.



Gambar 2. Contoh soal jawaban pendek pada sistem *e-learning* hasil modifikasi

Bentuk antarmuka soal jawaban pendek hasil modifikasi tampak pada Gambar 2. Secara tampilan, tidak tampak perbedaan gaya (*style*) antara tipe soal hasil modifikasi dan tipe soal asli dan hanya tampak sebagai perbedaan nama. Namun algoritma yang digunakan dalam proses penilaian (skoring) telah mengalami perubahan.

Survei menyertakan 40 macam soal yang telah dijawab siswa. Dari 40 soal tersebut, terdapat 1 soal yang dijawab tepat sama dengan kunci, 2 soal dijawab dengan sinonim, dan 3 soal dijawab dengan kata-kata yang sama sekali tidak sama dengan kunci. Selain itu terdapat 4 soal dijawab dengan kesalahan ketik 1 huruf tanpa membentuk kata baru, 3 soal dijawab dengan kesalahan ketik 2 huruf tanpa membentuk kata baru, 1 soal dijawab dengan kesalahan ketik 1 huruf yang membentuk kata baru, dan 2 soal dijawab dengan kesalahan ketik 2 huruf yang membentuk kata baru.

Contoh soal yang dijawab dengan kesalahan ketik tanpa membentuk kata baru adalah:

Penduduk desa bergotong royong untuk membangun kembali _____ yang hanyut diterjang banjir.

Kunci jawaban untuk soal tersebut adalah “jembatan” tetapi siswa menjawab dengan “jembaran”. Adapun soal yang dijawab dengan kesalahan ketik dan membentuk kata baru adalah:

Setelah sarapan pagi, Nina _____ dengan memasukkan pakaian dan peralatan lain ke dalam ransel. Kepergiannya kali ini hanya untuk dua hari sehingga bawanya tidak banyak.

Kunci jawaban untuk soal tersebut adalah “berkemas” sedangkan siswa menjawab “berkeras”.

Tabel 1. Nilai yang diberikan guru terhadap jawaban siswa

Kaitan jawaban siswa dengan kunci jawaban	Rerata	Standar deviasi
Jawaban tepat sama	99,6	3
Sinonim	76	23
Tipografi 1 huruf, tidak membentuk kata	70	15
Tipografi 2 huruf, tidak membentuk kata	59	23
Tipografi 1 huruf, membentuk kata	43	35
Tipografi 2 huruf, membentuk kata	34	30
Tidak ada hubungan makna	35	30

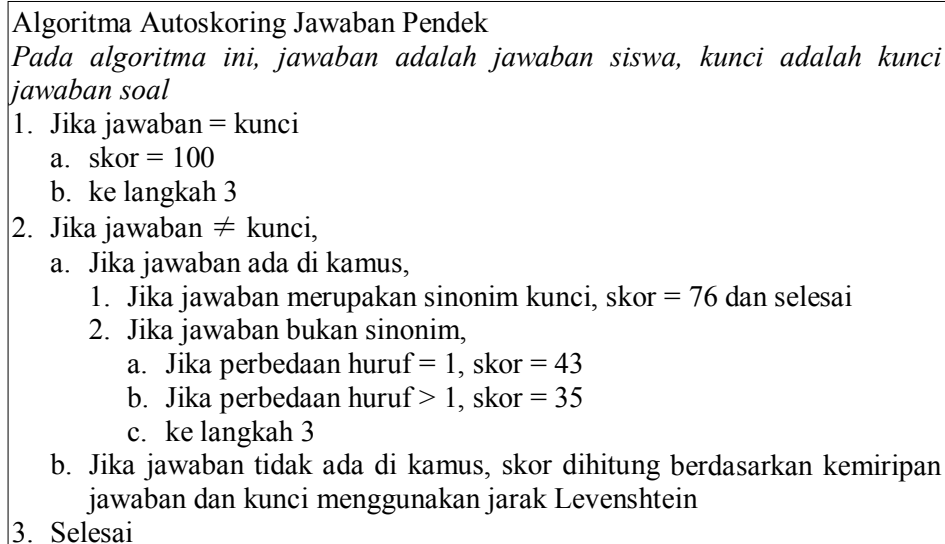
Setelah jawaban siswa dinilai oleh para guru, nilai tersebut dikumpulkan dan dianalisis. Tabel 1 memperlihatkan nilai yang diberikan oleh guru terhadap jawaban siswa yang dikelompokkan menurut keterkaitan antara jawaban siswa dan kunci. Tabel menampilkan rata-rata nilai yang diberikan guru dan standar deviasinya.

Penilaian guru terhadap jawaban yang tepat sama hampir mencapai nilai maksimum 100. Standar deviasi yang kecil, yaitu 3, menunjukkan bahwa hampir tidak ada perbedaan persepsi di antara para guru dalam memberikan nilai. Sementara itu, banyak guru tidak memberikan nilai nol pada jawaban siswa yang sama sekali salah atau tidak mempunyai hubungan makna dengan kunci jawaban. Secara rata-rata guru memberikan nilai 35 dengan standar deviasi 30 untuk jawaban yang sama sekali salah. Sebagian guru sepertinya cenderung mengangkat nilai siswa atau mencoba mengapresiasi upaya siswa dengan memberikan “upah menulis”.

Guru pada umumnya tidak memberi nilai maksimum terhadap jawaban yang merupakan sinonim dari kunci. Secara rata-rata nilai yang diberikan adalah 76. Untuk pola jawaban ini, variasi penilaian guru cukup besar terlihat dari standar deviasi yang mencapai 23. Jika ditilik dari soal dan jawaban yang diberikan, kemiripan makna jawaban dengan kunci cukup nyata, misalnya kemiripan kata “penat” dan “lelah” sebagai jawaban dan kunci dari pertanyaan: “Setelah bekerja seharian, pak Madi mencoba melepas _____ dengan berbaring di atas dipan.”. Kesan yang tampak adalah kurangnya toleransi terhadap jawaban yang tidak tepat sama.

Kesalahan tipografi yang lebih kecil secara umum lebih ditoleransi. Tabel 1 di atas menunjukkan bahwa jawaban dengan kesalahan ketik 1 huruf secara rata-rata diberi skor lebih besar daripada jawaban dengan kesalahan ketik 2 huruf. Akan tetapi kesalahan ketik yang membentuk kata baru lebih sulit ditoleransi. Skor yang diberikan terhadap jawaban dengan kesalahan ketik yang membentuk kata baru secara signifikan lebih kecil dibanding skor untuk jawaban dengan kesalahan ketik yang tidak membentuk kata baru.

Uraian beberapa paragraf di atas menunjukkan bahwa penentuan nilai (skor) terhadap jawaban siswa harus menempuh beberapa langkah. Algoritma pada Gambar 3 diusulkan untuk diterapkan pada sistem penilaian otomatis terhadap jawaban pendek dan terbuka. Penentuan nilai atas jawaban yang tidak terdapat pada kamus dilakukan berdasarkan rumusan yang diusulkan pada hasil penelitian oleh Thamrin dan Wantoro (2012), yaitu berdasarkan pada kemiripan jawaban dengan kunci yang proporsional dengan jarak Levenshtein.



Gambar 3. Algoritma penilaian otomatis atas jawaban pendek dan terbuka

Tindak lanjut yang masih diperlukan dalam penelitian ini adalah uji coba penerapan pada sistem penilaian otomatis, misalnya digunakan dalam evaluasi belajar, tes masuk sekolah/perguruan tinggi atau ujian kompetensi. Pengembangan lebih lanjut dari penelitian ini adalah menelaah penilaian otomatis pada teks jawaban yang lebih panjang. Penelitian ini masih membatasi telaah pada kemiripan makna kata (sinonim) dan kesalahan tipografi untuk jawaban pendek yang terdiri atas satu hingga dua kata. Sebagian penelitian telah bergerak lebih jauh dengan mencermati jawaban yang terdiri atas satu kalimat atau frase yang mempunyai lebih dari 3 kata, misalnya yang dilakukan Mohler dan Mihalcea (2009). Winarsono dkk (2009) juga telah meneliti penerapan metode syntactic-semantic similarity (SynSemSim) untuk penilaian otomatis jawaban teks singkat, namun sayang para peneliti dari Indonesia ini menggunakan jejaring kata bahasa Inggris, dalam penelitiannya sehingga belum dapat digunakan pada kalimat berbahasa Indonesia.

4. KESIMPULAN

Penelitian ini telah berupaya mengembangkan sistem penilaian otomatis terhadap jawaban pendek dan terbuka dengan memodifikasi sistem *e-learning* agar menyediakan model evaluasi belajar yang dikembangkan. Penilaian tidak hanya dilandaskan pada kesamaan atau ketidaksamaan jawaban dengan kunci yang mengarah kepada penilaian salah dan benar, tetapi menimbang pula faktor kemiripan makna (sinonim) dan kesalahan ketik (tipografi) sehingga mengarah pada skala nilai. Penelitian ini mengusulkan algoritma untuk memberi penilaian otomatis dengan beberapa langkah yang memperhatikan kesamaan jawaban dengan kunci, keberadaan kata dalam kamus, kemiripan makna kata dan jumlah kesalahan tipografi. Perhitungan pada penilaian otomatis diarahkan untuk menghasilkan nilai yang sama dengan nilai yang diberikan oleh para guru yang turut serta dalam survei.

Survei menunjukkan adanya dua pola pemberian nilai oleh guru. Guru cenderung lebih toleran terhadap kesalahan tipografi yang tidak menghasilkan kata baru. Guru cenderung mengurangi nilai siswa jika jawaban siswa tidak tepat sama, meskipun siswa memberikan jawaban yang bermakna sama (sinonim) dengan kunci jawaban. Namun kebanyakan guru tidak ingin memberi nilai nol pada siswa yang jawabannya tidak mempunyai hubungan makna dengan kunci.

DAFTAR PUSTAKA

- Figlio, D.N. (1999). Functional Form and the Estimated Effects of School Resources, *Economics of Education Review*, vol. 18, hal. 241-252.
- Gonzalez-Barbone, V., & Llamas-Nistal, M. (2008). eAssessment of open questions: An educator's perspective, dalam *Frontiers in Education Conference 2008*, hal. F2B-1. IEEE.
- Hillstock, L.G. (2005) A Few Common Misconceptions about Distance Learning, dalam *Proceedings of the 2005 ASCUE Conference*, Myrtle Beach, South Carolina, Juni 12-15.

- Jona, K. (2000) Rethinking the Design of Online Courses, dalam *Prosiding ASCILITE 2000*, New South Wales, Coffs Harbour, 9-14 Desember 2000.
- Nugraheni, E. (2009). Peranan Pendidikan Terbuka Dan Jarak Jauh Dalam Meningkatkan Daya Jangkau Pendidikan Tinggi Di Asia Tenggara, *Jurnal Pendidikan Terbuka dan Jarak Jauh*, vol. 10 no. 1, Maret 2009, hal. 1-9.
- Thamrin, H., Wantoro, J. (2012). Pengembangan Modul Penilaian Otomatis terhadap Jawaban Soal Pendek Terbuka dalam Sistem E-Learning, dalam *Prosiding SNFT Umsida 2012*, hal. B31-38.
- Wignal, E. (2011). Online Student-Teacher Ratios *On Teaching Online*, www.onteachingonline.com, diakses 24 Juni 2012.
- Winarsono, D., Siahaan D.D., Yuhana, U. (2009). Sistem Penilaian Otomatis Kemiripan Kalimat Menggunakan Syntactic-Semantic Similarity pada Sistem E-Learning, *Kursor*, vol. 5, no. 2, hal. 75-82.