

PENGEMBANGAN APLIKASI ASSESSMENT TOOL MENGGUNAKAN METODE COSINE SEMANTIC SIMILARITY UNTUK AUTOMATIC SCORING JAWABAN TES URAIAN PADA MATA PELAJARAN BASIS DATA DI SMKN 1 SURABAYA

Muhammad Walid Arbiantono

Pendidikan Teknologi Informasi, Fakultas Teknik, Universitas Negeri Surabaya

Email : muhammad.17050974005@mhs.unesa.ac.id

Ekohariadi

Pendidikan Teknologi Informasi, Fakultas Teknik, Universitas Negeri Surabaya

Email : ekohariadi@unesa.ac.id

Abstrak

Teknologi informasi dan komputer (TIK) telah mengalami perkembangan yang sangat signifikan, begitupun dengan dunia pendidikan yang juga telah mengalami proses perubahan dalam sistem pendidikan dan pengajaran. Hal ini menuntut para pendidik untuk terus mengikuti perubahan dan perkembangan teknologi untuk mendukung proses pembelajaran. Aplikasi *assessment tool* merupakan aplikasi untuk merancang dan mengaplikasikan penilaian pada sistem ujian tes uraian secara otomatis. Aplikasi tersebut dikembangkan menggunakan metode *cosine semantic similarity* untuk menerapkan *automatic scoring* pada jawaban tes uraian. Aplikasi *assessment tool* dikembangkan menggunakan metode *research and development* oleh Plomp. Berdasarkan hasil validasi media oleh para ahli, aplikasi *assessment tool* mendapatkan presentase sebanyak 90% dengan tingkat sangat layak untuk diterapkan pada proses pembelajaran. Hasil akhir penilaian menggunakan metode *cosine semantic similarity* tidak ada perbedaan yang signifikan dibandingkan dengan penilaian secara manual dikarenakan penilaian manual memperoleh skor rata – rata 85,14 dan penilaian sistem memperoleh skor rata rata 84,91. Aplikasi tersebut juga menerapkan *automatic scoring* dan memiliki tingkat kesamaan 0,410 dengan sistem manual berdasarkan uji realibilitas dengan metode Kappa. Dapat dilihat dari nilai kesamaan tersebut, menunjukkan bahwa *automatic scoring* dan penilaian jawaban tes uraian secara manual memiliki kesamaan yang cukup baik sehingga sistem dapat digunakan untuk menilai jawaban *essay* secara otomatis.

Kata Kunci : Evaluasi, Tes Uraian, Cosine Semantic Similarity.

Abstract

Information and computer technology (ICT) has experienced a very significant development, as well as the world of education which has also undergone a process of change in the education and teaching system. This requires educators to keep abreast of changes and technological developments to support the learning process. The assessment tool application is an application to design and apply assessments to the test essay test system automatically. The application was developed using the method cosine semantic similarity to apply automatic scoring to the answers to the description test. The application was assessment tool developed using the method research and development by Plomp. Based on the results of media validation by experts, the application assessment tool gets a percentage of 90% with a very feasible level to be applied to the learning process. The final result of the assessment using the method cosine semantic similarity there is no significant difference compared to the manual assessment because the manual assessment obtained an average score of 85.14 and the system assessment obtained an average score of 84.91. The application also applies automatic scoring and has a similarity level of 0.410 with the manual system based on the reliability test using the Kappa method. It can be seen from the similarity value, indicating that automatic scoring and manual assessment of test answers have a fairly good similarity so that the system can be used to assess answers essay automatically.

Keywords: Evaluation, Essay, Cosine Semantic Similarity.

PENDAHULUAN

Teknologi informasi dan komputer (TIK) telah mengalami perkembangan yang sangat signifikan, begitupun dengan dunia pendidikan yang juga telah mengalami proses perubahan dalam sistem pendidikan dan pengajaran. Proses pengajaran beberapa intitusi pendidikan untuk proses pengajarannya juga sudah menerapkan untuk mengembangkan sistem *e-learning*.

Penelitian yang mengadopsi *Codeigniter* sebagai platform dalam pembuatan soal tes online, berhasil diimplementasikan di dalam kelas (Wibawa, et al., 2018). Pernyataan di atas menjadi inspirasi dalam membuat tes berbasis komputer menggunakan platform sistem manajemen pembelajaran (LMS). Pengembangan LMS membutuhkan perencanaan yang baik yaitu terorganisir sehingga pengembangan dapat disampaikan tujuan

pembelajaran secara keseluruhan (Sujatmiko, Wibawa, Dermawan, & Ongge, 2021).

Media *e-learning* dalam proses pembelajaran termasuk bentuk media yang dilakukan dengan memanfaatkan peralatan elektronik yang bisa terhubung internet dalam penyampaian, pelatihan, sampai proses penilaian materi pembelajaran (Ratna, Budiardjo, & Hartanto, 2007).

Dalam proses evaluasi media *e-learning* bisa disimpulkan bahwasanya hasil belajar menjadi sangat penting, dikarenakan indikator – indikator pemahaman siswa terhadap materi yang diberikan oleh pengajar merupakan hasil dari evaluasi atau penilaian. Evaluasi yang dilaksanakan dengan penilaian dilakukan secara online dapat diberikan soal dalam jenis apapun, salah satunya adalah dalam bentuk sistem tes uraian.

Sistem tes uraian dalam penilaian merupakan evaluasi dimana siswa harus menjawab dengan kalimat sesuai pemahamannya, dikarenakan pilihan jawaban tidak disediakan (Ratna, Budiardjo, & Hartanto, 2007), hal tersebut siswa bisa terlatih dalam menyampaikan informasi yang didapatkan dari pengajar secara verbal. Penilaian tes uraian juga diharapkan mampu untuk membantu pemahaman serta berpikir siswa menjadi lebih mendalam akan suatu ilmu (Hamza, 2013). Dengan hal ini para pengajar tetap melakukan sistem tes uraian ini untuk melaksanakan proses penilaian dalam evaluasi tingkat kemampuan pemahaman siswa, meskipun pada kenyataan memberikan penilaian yang objektif pada jawaban siswa sangatlah tidak mudah.

Berdasarkan hasil observasi dari guru pengajar mata pelajaran di SMK Negeri 1 Surabaya untuk memeriksa jawaban tes uraian pengajar memerlukan waktu yang sangat banyak, semakin soal tes uraian berjumlah banyak dan banyaknya jumlah siswa yang mengerjakan soal tes uraian, maka pengajar akan semakin banyak untuk mengoreksi jumlah soal tes uraian. Dalam hal ini kualitas penilaian menyebabkan adanya penurunan dan bahkan penilaian sudah tidak objektif dimana objektifitas merupakan suatu hal yang dapat diukur dan berada diluar pikiran atau persepsi manusia, sementara subjektifitas merupakan fakta yang berada didalam pikiran manusia sebagai perasaan, persepsi, dan keyakinan. Kegiatan penilaian tes uraian merupakan kegiatan yang memakan waktu dikarenakan beberapa peneliti menyatakan bahwa banyak perbedaan versi evaluasi di kelas yang diberikan oleh masing – masing pendidik terhadap sifat subjektifitas dari penilai tes uraian sehingga membuat para siswa merasakannya sebagai sumber ketidakadilan (Valenti, Neri, & Cucchiarelli, 2003).

Beberapa metode sudah di implementasikan untuk memecahkan permasalahan tersebut seperti, *e-Rater*, dan *Project Essay Grade* (Valenti, Neri, & Cucchiarelli, 2003), selain metode tersebut terdapat metode *cosine similarity*

untuk menghitung atau membobotkan setiap kata yang dapat memecahkan pada permasalahan penilaian tes uraian secara otomatis. Dan perlu adanya penambahan *semantic similarity* untuk menangani sinonim atau kesamaan kata yang mempunyai kesamaan makna.

Metode *cosine similarity* merupakan salah satu metode untuk memperhitungkan kemiripan beberapa dokumen yang sama. Setiap kata dalam dokumen akan disamakan frekuensinya atau bisa disebut *Term Frequency* (tf). *Term Frequency* (tf) adalah penentuan dalam pembobotan sebuah dokumen di setiap kata. Setiap terdapat kata (tf) nilai dapat diperhitungkan pada dokumen terhadap bobot kata, bisa disimpulkan jika didalam dokumen bobot semakin besar, maka akan menghasilkan nilai kesamaan yang semakin besar.

Semantic similarity adalah metode untuk merepresentasikan istilah (*term*) berbasis vector yang mampu mendeteksi inti (*semantic*) kalimat. *Semantic* memiliki fungsi utama untuk menghitung tingkat kemiripan (*similarity*) dari dokumen dengan membandingkan representasi vektor setiap dokumen tersebut. Identifikasi kalimat merupakan tahap pertama yang dilakukan dalam proses perhitungan kemiripan kalimat. Tahap ini yakni memilah kalimat sesuai dengan jenis atau kriteria katanya. Proses yang dilakukan untuk menentukan jenis kata pada kalimat yaitu menggunakan basis data *wordnet*.

Wordnet adalah sebuah basis data leksikal yang diorganisasikan berdasarkan hubungan semantik. *Synonym sets* (*synset*) mengorganisasikan kata benda, kata kerja, dan kata sifat yang menggantikan konsep leksikal dasar. Hubungan semantik yang terbentuk antara lain berdasarkan pada sinonim, antonim, hiponim dan meronym. Penelitian ini serupa dengan yang dilakukan oleh (Nudin, Siahan, & Yuhana, 2012), yang menggunakan *wordnet* sebagai media untuk mencari kemiripan kata berdasarkan atribut emosi pada konten lagu. Terdapat contoh kasus misalnya, yaitu apabila seseorang ingin mencari lagu berdasarkan emosi yang ia alami saat ini, yakni "*I am euphoric*". Maka dari query tersebut dilakukan pemrosesan menggunakan metode *rule-based* dan pendekatan semantik yang kemudian sistem mendapatkan hasil kontek yang dibutuhkan, yaitu "*euphoric*". Akan tetapi, dalam klasifikasi musik tidak terdapat "*euphoric*" sehingga diperlukan kamus untuk mencari kata yang mempunyai kemiripan. Hasil yang didapat dalam pencarian tersebut yang memiliki kemiripan dengan "*euphoric*" adalah "*happy*".

Oleh karena itu, dapat memanfaatkan Teknik reduksi untuk efektifitas pemberian nilai terhadap suatu kata dan sebagai contoh penelitian yang dilakukan oleh Fawaz S. Al-Anzi pada tahun 2017 yang berjudul "*Toward an Enhanced Arabic Text Classification Using Cosine*

Similarity and Latent Semantic Indexing” melaporkan bahwa *cosine similarity* digunakan dalam pola pengenalan dan klasifikasi teks sebagai metode pembelajaran dimana dapat ditentukan dengan sudut cosinus antara dua vector. Metode *cosine similarity* bertujuan dimana hasil perbandingan dari kesamaan kedua kalimat yang dibandingkan dan nilai similaritynya akan digunakan sebagai nilai akhir.

Pada intitusi pendidikan bisa diterapkan untuk pengembangan aplikasi penilaian tes uraian secara otomatis, dalam hal ini mampu untuk memecahkan permasalahan subjektifitas pendidik, sehingga permasalahan – permasalahan saat proses terjadinya evaluasi penilaian jawaban tes uraian yang terjadi pada saat ini secara manual dapat teratasi.

Berdasarkan gagasan diatas, maka penulis ingin melaksanakan penelitian “Pengembangan Aplikasi *Assessment Tool* Menggunakan Metode *Cosine Semantic Similarity* untuk *Automatic Scoring* Jawaban Tes Uraian pada Mata Pelajaran Basis Data di SMKN 1 Surabaya.” Bahwa penelitian ini mengembangkan media aplikasi *assessment tool* dengan menggunakan metode *cosine semantic similarity* untuk merancang dan mengaplikasikan penilaian pada sistem ujian tes uraian secara otomatis.

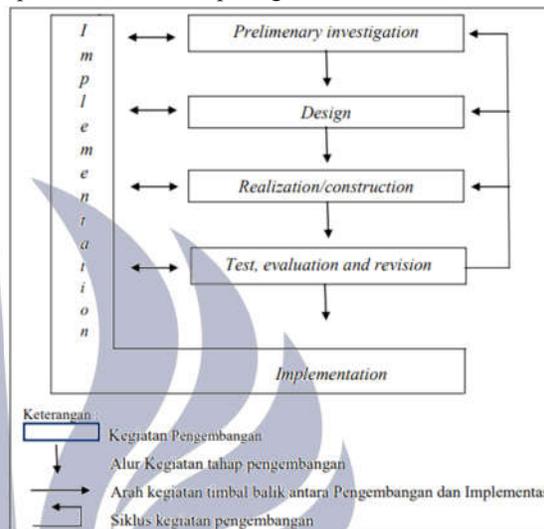
Berdasarkan gagasan diatas peneliti dapat merumuskan dan menyelesaikan permasalahan penelitian ini diantaranya : (1). Bagaimana mengembangkan media aplikasi *assessment tool* menggunakan metode *cosine semantic similarity* untuk *automatic scoring* pada jawaban tes uraian ?. (2). Bagaimana penerapan metode *cosine semantic similarity* dalam aplikasi *assessment tool automatic scoring* pada jawaban tes uraian?. (3). Bagaimana reliabilitas antar penilai (*Inter-rater reliability*) skor yang diperoleh dari *assessment tool* dan manual (*human rater*) ?

Tujuan penelitian ini berdasarkan rumusan masalah diatas memiliki tujuan antara lain : (1). Untuk mengetahui cara rancangan sistem aplikasi *assessment tool* menggunakan metode *cosine semantic similarity* untuk *automatic scoring* pada jawaban test uraian. (2). Untuk mengetahui hasil aplikasi *assessment tool* menggunakan metode *cosine semantic similarity* untuk *automatic scoring* pada jawaban tes uraian. (3). Untuk mengetahui konsistensi penilaian media aplikasi *assessment tool* menggunakan metode *cosine semantic similarity* untuk *automatic scoring* pada jawaban tes uraian.

METODE

Penelitian dan pengembangan (R&D) merupakan metode yang digunakan dalam penelitian ini dengan model penelitian dan pengembangan dari Plomp. Menurut Plomp (1997) penelitian ini diperlukan desain yang mengutarakan “*we characterized educational design in short as method*

within which one is working in systematic way toward the solving of a ‘make’ problem”. Yang mengartikan secara singkat bahwasanya mengkarakteristikan model dalam bidang Pendidikan merupakan metode dengan bekerja secara sistematis untuk memecahkan dari permasalahan yang dibuat. Pada umumnya bidang pendidikan dalam pemecahan masalah yang dipaparkan oleh Plomp (1997:5) dapat divisualisasikan pada gambar 1.



Gambar 1. Desain penelitian dan pengembangan Plomp 1997 (Sumber Plomp : 1997)

Tahapan penelitian adalah proses dalam melaksanakan penelitian agar bisa dilakukanya sesuai kebutuhanya sehingga dapat diproses pelaksanaanya akan lebih sistematis, terarah, dan terstruktur. Berikut adalah tahapan serta penjelasan model penelitian pengembangan menurut Plomp.

Investigasi (*Preliminary Investigation*)

Investigasi merupakan fase yang melakukan pengumpulan informasi dengan cara mengidentifikasi masalah yang terkait dengan aplikasi *assessment tool* tes uraian. Dalam tahapan pengumpulan informasi ini bertujuan untuk memperkuat latar belakang masalah, rumusan masalah dalam penelitian. Pengumpulan informasi atau data ini dilakukan dengan observasi dan wawancara salah satu guru di SMK Negeri 1 Surabaya.

Desain (*Design*)

Desain (*Design*) merupakan fase dalam merancang sebuah produk yang berupa aplikasi *assessment tool* tes uraian. Tahap desain dalam sebuah aplikasi yang pertama yaitu proses mendesain dengan mengambil beberapa data yang sebelumnya sudah dikumpulkan. Kemudian merancang desain alur program aplikasi yang akan dijalankan.

Realisasi / Kontruksi (Realization/Costruction)

Realisasi/Konstruksi (Realization/Costruction) merupakan fase yang menghasilkan bentuk dasar produk aplikasi *assessment tool* tes uraian sebagai hasil dari realisasi fase desain. Pada proses ini aplikasi *assessment tool* tes uraian mulai dikembangkan sesuai apa yang telah dirancang oleh peneliti pada tahapan desain.

Tes, Evaluasi, dan Revisi (Test, Evaluation, and Revision)

Dalam tahapan ini dilakukanya kegiatan validasi instrumen. Validasi ini merupakan suatu proses kegiatan untuk menilai validitas intrumen produk dari aplikasi *assessment tool* tes uraian. Kegiatan validasi intrumen ini dilakukan uji validitas dari beberapa dosen yang ahli di bidang pendidikan dan media untuk menilai atau memberikan *judgment*. Intrumen yang divalidasi antara lain RPP, Materi pembelajaran, dan Soal sebagai bahan penelitian.

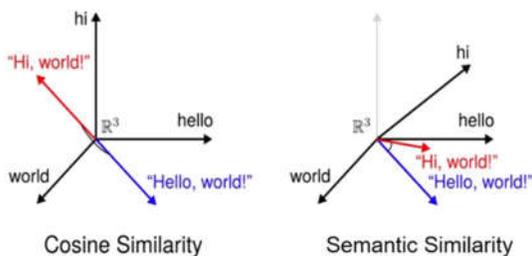
Implementasi (Implementation)

Dalam fase ini setelah proses validasi terhadap aplikasi *assessment tool* mendapatkan hasil valid, maka selanjutnya aplikasi *assesment tool* tes uraian tersebut diimplementasikan dalam intitusi pendidikan yang telah direncanakan. Dalam proses implementasi aplikasi *assessment tool* tes uraian tetap harus dalam tahapan evaluasi agar dapat terlihat kekurangan serta hambatan untuk perbaikan selanjutnya.

Rancangan Sistem

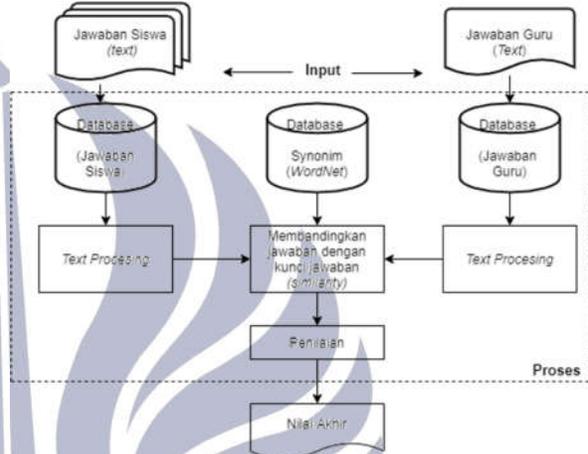
Tahapan dari hasil pengembangan yaitu membuat aplikasi *assesment tool* tes uraian. Pada apikasi ini menggunakan metode atau algoritma *cosine similariy* untuk mendeteksi kesamaan kata (tf) dan *semantic similarity* untuk mencari persamaan kata yang memiliki pengertian yang sama (*synonim*) pada jawaban siswa dengan kunci jawaban guru.

Perbedaan proses dalam algoritma *cosine similarity* dan *semantic similarity* dapat digambarkan seperti gambar 2.



Gambar 2. Perbedaan proses cosine similarity dan semantic similarity

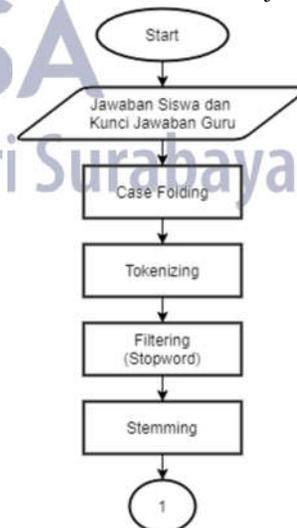
Pada Gambar 2 proses *cosine similarity* dan *semantic similariy* diatas dijelaskan bahwasanya pada tahap cosine similariy kata / kalimat “Hi, world!” dengan “Hello, world!” sangat bertentangan. Sedangkan dalam tahapan semantic similarity kata / kalimat “Hi, world!” dengan “Hello, world!” harus dianggap mempunyai arti atau pengertian yang sama. Oleh sebab itu penelitian ini diperlukan penambahan metode *semantic similarity* untuk membantu mencari kesamaan makna dalam kata / kalimat. Rancangan sistem aplikasi *assessment tool* divisualisasikan seperti gambar 3.



sistem aplikasi *assessment tool* mempunyai beberapa tahapan yang dikerjakan sistem yaitu :

- a. *Input Data*
 1. Pertanyaan soal uraian
 2. Kunci jawaban guru
 3. Jawaban siswa
- b. *Text Procesing*

Pada tahap pemrosesan teks pada aplikasi assesment tool ini akan dijelaskan pada gambar 3.



Gambar 4 . Desain proses *text processing*

Pada gambar 4. desain pemrosesan teks diatas dijelaskan tahap-tahapan pemrosesan teks antara lain :

1) *Case Folding*

Pada tahap case folding ini menjalankan pemrosesan teks, hal ini seluruh jawaban siswa dan kunci jawaban guru yang telah masuk kedalam sistem akan dilakukan pemrosesan *case folding* dengan tujuan mengubah semua huruf dalam dokumen menjadi huruf kecil.

2) *Tokenisasi*

Tahapan selanjutnya adalah proses tokenisasi yaitu dengan cara memisahkan kata pada kalimat berdasarkan spasi, enter, tabulasi, titik (.), maupun koma (,).

3) *Stopword*

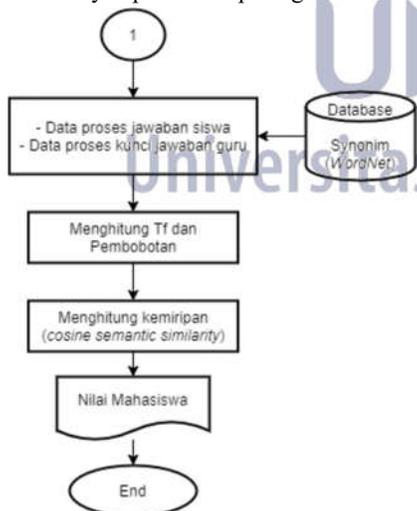
Pada tahapan ini setelah seluruh kata pada jawaban berhasil dipisahkan, selanjutnya adalah daftar *stopword*. Pada tahapan ini kata pada jawaban siswa banyak ditemukan kata yang frekuensinya tinggi tetapi tidak mempunyai makna atau nilai skor yang tinggi. Kata tersebut akan dimasukkan dalam daftar kata *stopword* dengan contoh seperti kata 'dan, di, ke' kata ini termasuk kata daftar *stopword*.

4) *Stemming*

Tahap pemrosesan yang terakhir yaitu *stemming*, dalam proses ini mencari kata dasar dari setiap kata. Hal ini bermanfaat untuk mencari kata dasar dari setiap kata yang ditemukan.

c. Pengukuran *Text Similarity*

Selanjutnya untuk proses pengukuran text similarity dapat dilihat pada gambar 5.



Gambar 5 . Desain proses pengukuran text similarity

1) *Term Weighting*

Proses selanjutnya adalah proses menghitung tf pembobotan yang menghitung frekuensi dari setiap kata. Frekuensi kata berdasarkan pada jumlah kata pada jawaban siswa. Setiap ditemukan kata dari jawaban siswa yang sama dengan kata kunci jawaban guru akan menghitung pembobotan dengan menggunakan persamaan. Konsep pembobotan ini dihitung berdasarkan frekuensi konsep dibagi dengan total frekuensi yang ditemukan

a) Input : term

b) Proses : menghitung bobot kata (menggunakan persamaan berikut)

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

Dimana :

w_{if} = Bobot Kata terhadap dokumen

Tf_{if} = Jumlah kemunculan kata/term T_j

N = Jumlah dokumen dalam database

n_i = Jumlah dokumen yang mengandung/term

T_j = (Minimal terdapat satu kata yaitu term T_j)

2) Menghitung Kemiripan

Pada proses ini dilakukan perhitungan kemiripan dari jawaban essay siswa dengan kunci jawaban guru, dengan menggunakan rumus persamaan seperti dibawah ini :

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

Dimana :

w_{qk} = Bobot Query

w_{dk} = Bobot Dokumen.

Dalam tahapan proses menghitung kemiripan hanya membandingkan antara jawaban siswa dengan kunci jawaban guru pada aplikasi *assessment tool*. Proses ini menghasilkan *output* nilai kemiripan antara jawaban siswa dengan kunci jawaban guru yang selanjutnya akan dikonversikan untuk dijadikan hasil nilai siswa.

d. Konversi *Score Similarity* menjadi *Score Tes Uraian*.

Score similarity yang dihasilkan akan dikonversi menjadi *score tes uraian* siswa berdasarkan rentang nilai versi penilaian manusia (*human rater*). Berikut rentang skor yang ditunjukkan oleh tabel (1).

Tabel 1. Rentang *score* jawaban siswa

<i>Rage Score</i>	
<i>Score Similarity</i>	<i>Score Human Rater</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

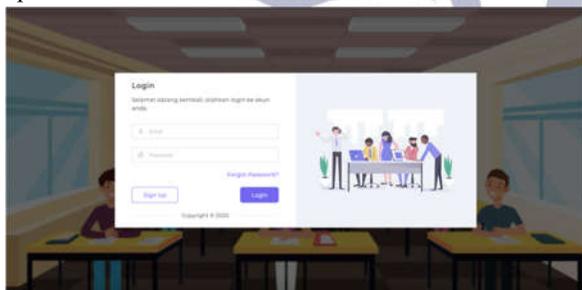
(Fuat, 2010)

Hasil penilaian siswa pada tabel 1 tersebut dibagi dengan jumlah soal tes uraian yang sesuai diberikan oleh guru. Misalnya terdapat 5 soal tes uraian, jika siswa berhasil menjawab dengan 1 soal jawaban benar dan nilai similarity atau kemiripannya 1 maka siswa tersebut memperoleh skor $(100/5) = 20$.

HASIL DAN PEMBAHASAN

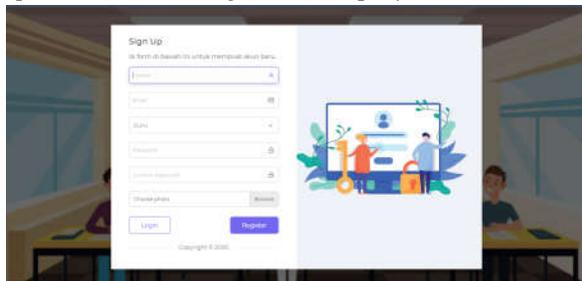
Hasil Media Pembelajaran

Hasil dari penelitian ini adalah aplikasi *assessment tool* tes uraian untuk menilai secara otomatis. Kegiatan proses penilaian ini dapat dilakukan dengan menggunakan bantuan koneksi internet. Dibawah ini merupakan tampilan dari aplikasi *assessment tool* tes uraian.



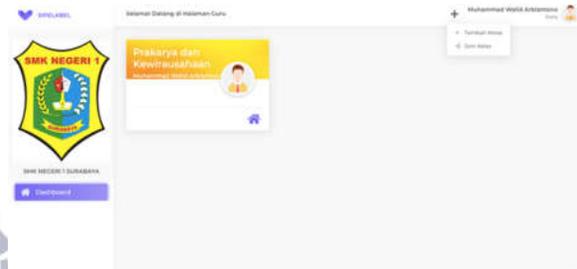
Gambar 6 . Tampilan login

Pada tgmbar 6 merupakan tampilan login ini *user* mengisi email dan password kemudian klik 'login'. Apabila *user* belum mempunyai akun maka bisa klik 'sign up' untuk mendaftar agar bisa mempunyai akun.



Gambar 7. Tampilan *sign up*

Pada gambar 7 yaitu tampilan *sign up* ini untuk *user* baru yang belum memiliki akun, pada tampilan tersebut *user* bisa mendaftarkan sebagai 'guru' atau sebagai siswa. Dengan mengisi form yang telah sistem sediakan untuk proses pembuatan akun.



Gambar 8. Tampilan dashboard

Pada gambar 8 merupakan tampilan dashboard ini berisikan informasi mengenai data kelas atau mata pelajaran yang sudah di ambil / diprogram. Kemudian pada tampilan dashboard guru terdapat tambahan menu tambah kelas dan *join* kelas, sedangkan pada tampilan siswa hanya terdapat *join* kelas.



Gambar 9. Tampilan Tambah Kelas

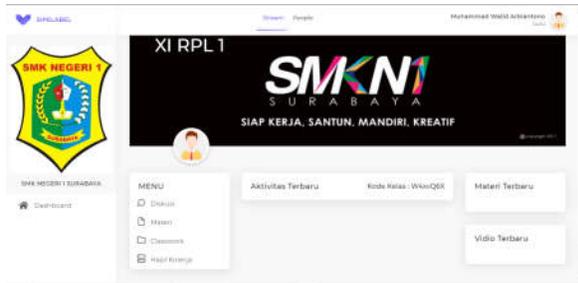
Pada gambar 9 merupakan tampilan tambah kelas merupakan fitur yang dimiliki oleh guru. Pada tampilan tersebut guru bisa memogram atau membuat kelas. Terdapat nama kelas, *section*, *subjek*, dan *room* yang harus guru inputkan. Setelah penambahan kelas berhasil maka akan muncul kode untuk digunakan siswa bergabung dikelas tersebut.



Gambar 10. Tampilan gabung kelas

Pada gambar 10 merupakan tampilan gabung kelas tersebut, siswa hanya memasukkan kode agar bisa bergabung dikelas tersebut. Kode tersebut siswa dapatkan dari guru yang telah membuat *room* / kelas.

Pengembangan Aplikasi *Assessment Tool* Menggunakan Metode *Cosine Semantic Similarity* untuk *Automatic Scoring* Jawaban Tes Uraian pada Mata Pelajaran Basis Data di SMKN 1 Surabaya



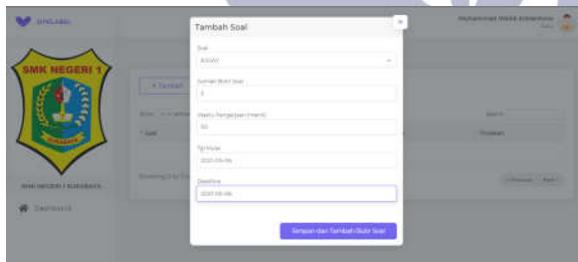
Gambar 11. Tampilan kelas

Pada gambar 11 merupakan tampilan kelas berisikan fitur diskusi, materi, *classwork*, dan hasil kinerja. Pada tampilan kelas ini hanya bisa ditambahkan oleh guru.



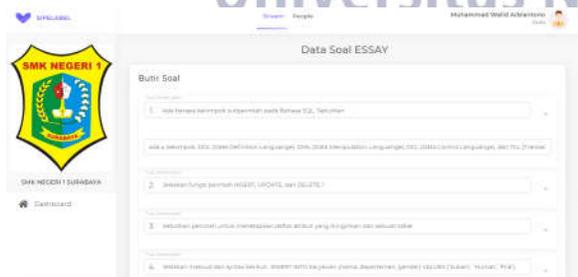
Gambar 12. Tampilan tambah materi

Pada gambar 12 yaitu tampilan tambah materi untuk memberikan materi kepada siswa, bisa berupa pdf, ppt, maupun video.



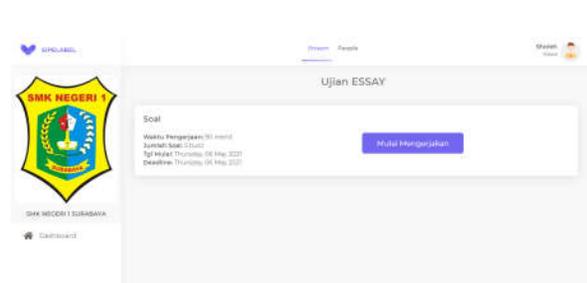
Gambar 13. Tampilan tambah soal

Pada gambar 13 merupakan tampilan tambah soal tersebut merupakan soal essay, guru bisa memprogram jenis soal, jumlah butir soal, waktu pengerjaan, tanggal mulai serta tanggal selesai pengerjaan soal.



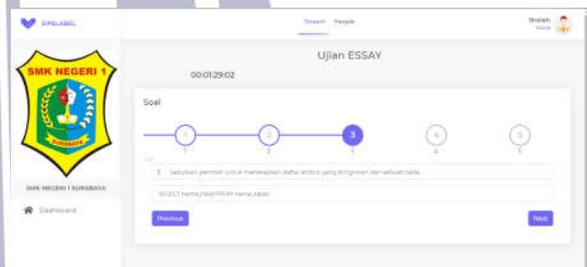
Gambar 14. Tampilan data soal

Pada gambar 14 merupakan tampilan data soal adalah guru memasukkan soal dan kunci jawaban pada aplikasi *assessment tool* ini.



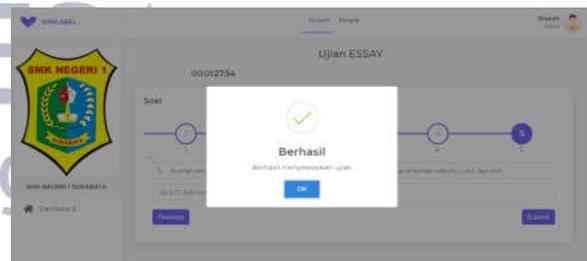
Gambar 15. Tampilan memulai tes uraian

Pada gambar 15 merupakan tampilan memulai tes uraian merupakan tampilan halaman tes uraian siswa, pada halaman ini siswa bisa langsung menekan tombol / *button* 'mulai mengerjakan' agar bisa mengerjakan soal atau tes uraian. Pada halaman memulai tes terdapat keterangan waktu mengerjakan soal, jumlah soal, tanggal mulai, dan deadline saat mengerjakan.



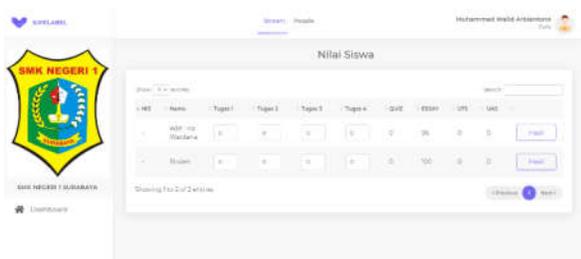
Gambar 16. Tampilan tes uraian

Pada gambar 16 merupakan tampilan tes uraian ini adalah proses mengerjakan soal atau tes uraian. Siswa menjawab langsung dibagian bawah soal yang sudah tersedia. Siswa bisa memilih tombol *previous*, *next*, dan setelah semua jawaban sudah terisi maka akan ada tombol *submit* untuk mengakhiri atau menyelesaikan tes uraian. Sistem akan otomatis *submit* jika batas waktu pengerjaan telah habis.



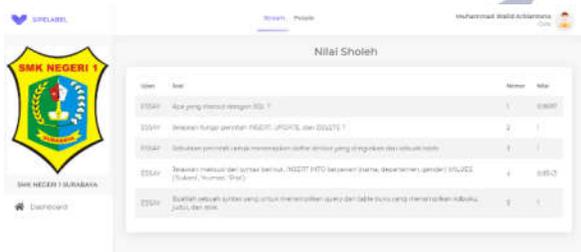
Gambar 17. Tampilan selesai menyelesaikan tes uraian

Pada gambar 17 merupakan tampilan selesai menyelesaikan tes uraian ditas setelah kita menekan tombol *submit*, maka akan muncul notifikasi 'Berhasil menyelesaikan ujian', hal ini menandakan tes uraian yang siswa kerjakan sudah selesai atau berakhir. Untuk melihat nilai akhir tes uraian siswa maupun guru bisa melihat dibagian "hasil kinerja" yang secara otomatis nilai tes uraian akan keluar seperti pada gambar 18.



Gambar 18. Tampilan nilai siswa

Pada gambar 18 tampilan nilai siswa merupakan hasil tes uraian yang telah siswa kerjakan, yang secara otomatis keluar nilai akhirnya. Pada button “hasil” jika diklik maka hasil akan seperti gambar 19.



Gambar 19. Tampilan hasil nilai siswa

Pada gambar 19 tampilan hasil nilai siswa nomor soal 2,3, dan 5 mendapatkan nilai similarity 1 yang mempunyai kesamaan dengan kunci jawaban. Sedangkan soal nomor 1 mendapatkan nilai similarity 0,96 dan soal nomor 4 mendapatkan 0,95 yang menandakan hampir memiliki kesamaan dengan kunci jawaban. Tetapi pada hasil akhir mendapatkan nilai skor 100 dikarenakan jawaban soal nomor 1 dan 4 masuk dalam kategori *score similarity* 0,91 – 1 yang mendapatkan nilai 100.

Pembahasan Hasil Validasi

Pada hasil validasi instrumen penelitian ini terdapat 3 validator, diantaranya 2 Dosen Jurusan Teknik Informatika Unesa, dan 1 Guru SMKN 1 Surabaya. Tahap validasi ini menggunakan skala likert dengan hasil presentase yang diterapkan ke dalam kriteria skor yang disajikan dalam tabel 2.

Tabel 2. Kriteria presentase kevalidan

Presentase (%)	Skor
0-20	Sangat Kurang
21-40	Kurang
41-60	Cukup
61-80	Valid
81-100	Sangat Valid

(Riduwan, 2015)

Lembar validasi RPP berisi kriteria tujuan pembelajaran, dan langkah – langkah pembelajaran. Lembar validasi Materi pembelajaran berisi kriteria format, bahasa, dan isi materi. Lembar validasi Soal berisi

validasi isi, bahasa dan penulisan soal. Lembar validasi Media berisi kriteria kepuasan pengguna (*user satisfaction*), kualitas sistem (*system quality*), dan kualitas informasi (*information quality*).

Hasil validasi ini memperoleh hasil presentase dan keterangan sebagai berikut :

Tabel 3. Hasil validasi instrumen

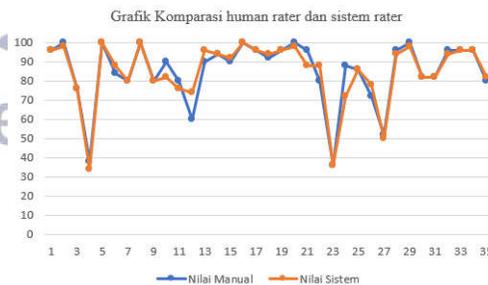
No	Validasi	Presentase Rata – rata(%)	Kategori
1	RPP	89,78	Sangat Valid
2	Materi	93,04	Sangat Valid
3	Soal	95,00	Sangat Valid
4	Media	90,00	Sangat Valid

Berdasarkan tabel 3 menunjukkan hasil validasi RPP memperoleh hasil presentase 89,78% dengan kategori sangat valid. Hasil validasi materi pembelajaran memperoleh hasil presentase 93,04% dengan kategori sangat valid. Hasil validasi soal memperoleh hasil presentase 95,00% dengan kategori sangat valid. Dan hasil Media pembelajaran *assessment tool* memperoleh hasil presentase 90,00% dengan kategori sangat valid. Berdasarkan hasil tersebut hal ini menunjukkan bahwa perangkat dan media pembelajaran ini sangat valid untuk digunakan dalam proses penelitian.

Pembahasan Hasil Penelitian

Penelitian ini dilaksanakan di SMKN 1 Surabaya yang berlokasi di Jl Smea No. 4 Surabaya. Populasi dari penelitian ini merupakan siswa kelas XI Jurusan Rekayasa Perangkat Lunak. Sampel yang digunakan dalam penelitian ini merupakan kelas XI RPL 1 dengan jumlah siswa 35 orang.

Berikut hasil komparasi hasil nilai penelitian dari kedua rater sistem dan manual.



Gambar 19. Grafik komparasi nilai dari *human rater* dan *sistem rater*

Berdasarkan gambar 19 menunjukkan komparasi nilai siswa dari *rater sistem* dengan *rater manual*. Hasil akhir penilaian menggunakan metode *cosine semantic similarity* tidak ada perbedaan yang signifikan dibandingkan dengan penilaian secara manual dikarenakan penilaian manual

memperoleh skor rata – rata 85,14 dan penilaian sistem memperoleh skor rata rata 84,91. Tetapi perlu dicermati dalam pemilihan jawaban tes uraian yang akan dijadikan kunci jawaban dikarenakan *semantic similarity* atau *wordnet* memiliki keterbatasan pada dataset, hal tersebut agar tidak menyebabkan kata – kata yang memiliki makna (bias) tidak terdeteksi.

Dalam penelitian kuantitatif, teknik analisis data yang digunakan yaitu analisis hasil akhir dari penilaian atau *rater* aplikasi *automatic scoring* dan penilaian manual (*human rater*) yang dilakukan oleh guru mata pelajaran basis data di SMKN 1 Surabaya. Kemudian untuk mengetahui tingkat konsistensi *scoring* perlu adanya uji reliabilitas antar penilai (*inter-rater reliability*) skor yang diperoleh dari aplikasi *assessment tool* dan manual (*human rater*).

Uji reliabilitas antar *rater* ini menggunakan koefisien Kappa dikarenakan hanya terdapat 2 *rater*. Koefisien merupakan satu sarana untuk menentukan tingkat konsistensi atau kesetaraan antar *rater* untuk memberikan rating penilaian tes uraian. Adapun teknik uji reliabilitas antar *rater* menggunakan bantuan *software* SPSS 25.0. Terdapat ahli yang mengklasifikasikan nilai koefisien Kappa yaitu Fleiss (1975). Menurut Fleiss (1975) klasifikasi nilai Kappa adalah sebagai berikut:

Tabel 4. Kategori nilai Kappa

Presentase	Predikat
$k < 0.4$	<i>poor agreement</i>
$0.40 < k < 0.75$	<i>good</i>
$k > 0.75$	<i>excellent agreement</i>

Kemudian hasil tes uraian nilai siswa dari kedua *rater* diolah dengan menggunakan *software* SPSS 25.0 dengan memperoleh hasil sebagai berikut :

Tabel 5. Hasil uji reliabilitas inter rater koefisien Kappa

Symmetric Measures					
		value	Asymptotic Standard Error	Approximate T ^b	Approximate Significance
Measure of Agreement	Kappa	.410	.084	8.967	.000
N of Valid Cases		35			

Berdasarkan tabel 5 dapat disimpulkan bahwasanya hasil uji reliabilitas *inter rater* dengan menggunakan koefisien Kappa memperoleh hasil 0,41 dengan predikat *good* ataupun baik menurut Fleiss (1975). Hal ini dapat disimpulkan bahwasanya kedua *rater* yaitu *rater* sistem (aplikasi *assessment tool*) dengan *rater* manual (*human*) memiliki tingkat konsistensi yang baik sehingga aplikasi *assessment tool* ini dapat digunakan untuk menilai jawaban tes uraian secara otomatis dengan akurat.

Penelitian ini sebanding dengan peneliti yang dilakukan oleh Kaja Zupanc pada tahun 2017 berjudul “*Automated Essay Evaluation with Semantic Analysis*” yang

menyatakan bahwa sistem yang menerapkan metode *semantic* memberikan akurasi penilaian yang lebih tinggi dari pada 9 sistem evaluasi tes uraian yang lebih canggih, keuntungan dari umpan balik penilaian secara otomatis adalah anonimitas, instan, dan dorongan untuk perbaikan berulang dengan memberikan siswa lebih banyak latihan dalam menulis tes uraian. Sementara penelitian yang dilakukan oleh Cokorda Gde Abimayu pada tahun 2020 yang berjudul “*Automatic Essay Answer Rating Using the Text Similarity Method*” menyatakan bahwa penelitian ini direpresentasikan dalam bentuk *vector* menggunakan *word2vec* atau *wordnet* untuk menunjukkan keakuratan.

PENUTUP

Simpulan

Berdasarkan hasil penelitian yang telah dilakukan, simpulan ini dapat ditarik sebagai berikut :

1. Hasil pengembangan aplikasi *assessment tool* menggunakan algoritma *cosine semantic similarity* untuk *automatic scoring* jawaban tes uraian yang diterapkan pada siswa kelas XI RPL mata pelajaran basis data di SMKN 1 Surabaya telah dilaksanakan sesuai prosedur dan pengembangan penelitian ini menggunakan metode riset dan development oleh Plomp.
2. Hasil uji validitas media pembelajaran atau aplikasi *assessment tool* untuk *automatic scoring* tes uraian ini oleh 3 validator ahli memperoleh presentase 90,00% yang dikategorikan sangat valid untuk diimplementasikan dalam proses pembelajaran. Hasil uji intrumen lainya yaitu RPP mendapatkan presentase 89,79% yang dikategorikan sangat valid, instrument materi pembelajaran mendapatkan presentase 93,04% yang dikategorikan sangat valid, dan intrumen soal mendapatkan presentase 95,00% yang dikategorikan sangat valid.
3. Hasil akhir penilaian menggunakan metode *cosine semantic similarity* tidak ada perbedaan yang signifikan dibandingkan dengan penilaian secara manual dikarenakan penilaian manual memperoleh skor rata – rata 85,14 dan penilaian sistem memperoleh skor rata rata 84,91 . Hasil penelitian menggunakan uji reliabilitas *inter rater* dengan menggunakan koefisien Kappa yang mengolah nilai siswa dari kedua *rater* yaitu *rater* sistem aplikasi *assessment tool* dan *rater* manual (*human rater*) dengan memperoleh hasil 0,410. Berdasarkan uji realibilitas *inter rater* dengan metode Kappa menunjukkan bahwa penilaian *automatic scoring* dan tes uraian secara manual memiliki konsistensi koefisien yang baik sehingga sistem dapat digunakan untuk menilai jawaban tes uraian secara otomatis dengan akurat.

Saran

Berikut saran untuk peneliti atau pengembang selanjutnya sebagai berikut:

1. Hasil uji reliabilitas *inter rater* dengan menggunakan koefisien Kappa memperoleh hasil skor 0,410 dengan kategori baik. Diharapkan untuk peneliti atau pengembang selanjutnya peneliti lebih banyak dalam mengampil responden atau sampel penelitian, agar hasil uji reliabilitas memperoleh hasil yang semakin bervariasi dan baik.
2. Untuk peneliti atau pengembang selanjutnya dalam tahap pembuatan media *assessment tool* ini diharapkan menambahkan metode ataupun algoritma dalam menilai tes uraian yang lebih akurat.
3. Penelitian ini hanya dilakukan pada mata pelajaran Basis Data, untuk itu penulis mengharapkan agar implementasi aplikasi *assessment tool* menggunakan metode *cosine semantic similarity* untuk *automatic scoring* jawaban tes uraian dapat diterapkan pada mata pelajaran yang lainnya.

UCAPAN TERIMA KASIH

Penulis dalam hal ini mengucapkan terima kasih kepada :

1. Keluarga yang mendo'akan, mendukung dan berjuang apapun itu demi anaknya berhasil.
2. Dosen pembimbing skripsi Bapak Prof. Dr. Ekohariadi, M.Pd. yang telah membimbing, meluangkan banyak waktunya, dan bersedia mengarahkan hingga artikel ilmiah ini selesai. Bapak Salamun Rohman Nudin, S.Kom., M.Kom. dan Bapak Setya Chendra Wibawa, S.Pd., M.T. selaku dosen penguji yang telah memberikan masukan dan arahan sehingga artikel ilmiah dapat menjadi lebih baik.
3. SMK Negeri 1 Surabaya yang telah memberikan izin dalam pelaksanaan penelitian ini dengan lancar, khususnya kepada Ibu Dra. Retno Pinasti, S.ST., dan Ibu Dra. Siti Chodidjah, S.ST., MM.

DAFTAR PUSTAKA

- Abimanyu, C. G., & Sanjaya, N. A. (2020, May). Automatic Essay Answer Rating Using the Text Similarity Method. *Jurnal Elektronik Ilmu Komputer Udayana, Volume 8, No 4.*, 463-468. doi:10.24843/JLK.2020.v08.i04.p12
- Al-Anzi, F. S. (2017). Toward an Enhanced Arabic Text Classification Using Cosine Similarity and Latent Semantic Indexing. *Journal of King Saud University*, 189-195. doi:10.1016/j.jksuci.2016.04.001
- Fitri, R., & Asyikin, A. N. (2015). Aplikasi Penilaian Ujian Essay Otomatis Menggunakan Metode Cosine Similarity. *Jurnal POROS TEKNIK, Volume 7 No. 2, Desember 2015 : 54-105*, 88-94. doi:10.31961/porosteknik.v7i2.218

- Fleiss, J. L. (1975). Measuring Agreement between Two Judges on the Presence or Absence. *Biometrics*, 651 - 659.
- Fuat, R. (2010). Sistem Penilaian Esai Otomatis Pada Elearning menggunakan Metode Cosine Similarity.
- Gomaa, W. H. (2012). Short Answer Grading Using String Similarity and Corpus-Based Similarity. *International Journal of Advanced Computer Science and Applications*, 115 - 121.
- Hamza, S. S. (2013, Desember). Sistem Koreksi Soal Essay Otomatis Dengan Menggunakan Metode Rabin Karp. *Jurnal EECCIS Vol. 7, No. 2*, 153-158.
- Nudin, S. R., Siahian, D. O., & Yuhana, U. L. (2012). Query Bahasa Alami Pada Mesin Pencari Dengan Menggunakan Metode Rule-Based dan Pendekatan Semantik Natural Language Query on Search Engine Using Rule-Based Method and Semantic Approach. C-13-1 - C-13-8.
- Plomp, T. (1997). *Educational Design: Introduction. From Tjeerd Plomp (eds)*. Educational & Training System Design: Introduction. Design of Education and Training (in Dutch). Utrecht (the Netherlands): Lemma.: Netherland. Faculty of Educational Science.
- Putri, R. A., Hartanto, B. B., & Djoko, d. (2007, April). SIMPLE: Sistem Penilai Esai Otomatis Untuk Menilai Ujian Dalam Bahasa Indonesia. *Makara, Teknologi,, Vol, 11, No.1*, 5-11. doi:10.7454/mst.v11i1.435
- Riduwan. (2015). *Dasar - dasar Statistika*. In Alfabeta.
- Sujatmiko, B., Wibawa, S. C., Dermawan, D. A., & Ongge, H. A. (2021). Computer Based Test (CBT) using Moodle-based High Order Thinking Skills in network design subjects. *IOP Conference Series: Materials Science and Engineering*, 1098(5), 052112. doi:10.1088/1757-899x/1098/5/052112
- Sulistyo, M. E., Saptono, R., & Asshidiq, A. (2015, Juli). Penilaian Ujian Bertipe Essay Menggunakan Metode Text Similarity. *Telematika, Vol. 12, No. 02*, 146-158. doi:10.31315/telematika.v12i2.1422
- Valenti, S., Neri, F., & Cucchiarelli, A. (2003). An Overview of Current Research on Automated Essay Grading. *Journal of Information Technology Education, Volume 2.*, 319 - 330. doi:doi.org/10.28945/331
- Wibawa, S. C., Wahyuningsih, Y., Sulistyowati, R., Abidin, R., Lestari, Y., Noviyanti, & Maulana, D. A. (2018). Online test application development using framework CodeIgniter. *IOP Conference Series: Materials Science and Engineering*, 296. doi:10.1088/1757-899X/296/1/012041
- Zupanca, K., & Bosni, Z. (2017). Automated essay evaluation with semantic analysis. *Knowledge Based Systems*, 1-24. doi:10.1016/j.knosys.2017.01.006