

## ABSTRAK

Penilaian uraian singkat otomatis atau disebut juga *Automatic Short Answer Scoring* (ASAS) adalah salah satu media *assessment* pada *e-learning*. Pemanfaatan ASAS dalam proses *assessment* pada *e-learning* masih kurang dibandingkan dengan penilaian otomatis pertanyaan pilihan ganda. Hal ini disebabkan metode-metode yang telah dikembangkan oleh peneliti ASAS belum mampu memberikan skor yang sepadan dengan skor yang dihasilkan oleh guru. Maka dari itu, masih terbuka banyak ruang penelitian untuk mengembangkan metode-metode yang diterapkan pada ASAS.

ASAS pada prinsipnya adalah membandingkan antara kalimat jawaban siswa dan kalimat jawaban kunci dengan menggunakan metode terkomputerisasi. Hasil dari proses perbandingan ini kemudian digunakan sebagai dasar untuk menentukan skor siswa. Terdapat dua hal yang krusial dalam proses perbandingan tersebut, pertama adalah penyediaan jawaban kunci yang memadai sehingga mampu menangkap keberagaman dari kalimat jawaban siswa. Kedua adalah tersedianya sebuah metode yang mampu digunakan untuk mengukur kesamaan antara kalimat jawaban siswa yang beragam dengan kalimat jawaban kunci. Penyediaan kalimat jawaban kunci dapat dilakukan dengan dua cara yaitu manual dan otomatis. Penyediaan secara manual sangat tergantung dari kemampuan guru dalam menciptakan variasi kalimat dan membutuhkan waktu yang lama. Di sisi lain penyediaan kalimat jawaban kunci dengan cara otomatisasi masih membutuhkan proses yang kompleks, salah satunya harus menyediakan korpus sesuai dengan materi yang diujikan. Sebuah mekanisme penyediaan jawaban kunci yang mampu menghasilkan variasi jawaban kunci yang memadai serta tidak membutuhkan proses yang kompleks masih perlu mendapatkan perhatian. Sementara itu, beberapa metode pengukuran kesamaan kalimat pada ASAS yang telah dikembangkan juga masih mempunyai celah, yaitu tidak dapat secara langsung diterapkan pada domain/topik dan bahasa yang berbeda. Salah satu metode pengukuran kesamaan kalimat yang dapat diterapkan secara langsung tanpa melalui pembelajaran adalah metode yang berbasis pada *string similarity*. Metode ini masih mempunyai kekurangan yaitu pertama, masih belum akurat untuk mengukur kalimat yang bermakna sama dengan menggunakan kosakata berbeda. Kedua, belum akurat untuk mengukur kalimat yang bermakna sama tetapi mempunyai panjang yang sangat berbeda. Permasalahan yang kedua pada *string similarity* ini yang menjadi fokus penelitian pada disertasi ini.

Pada disertasi ini diusulkan dua metode untuk menyelesaikan permasalahan yang masih terdapat pada ASAS. Pertama adalah mengusulkan penerapan metode *Maximum Marginal Relevance* (MMR) untuk membangkitkan variasi jawaban kunci secara otomatis. Kedua adalah mengusulkan metode baru yaitu *Geometry Average Normalized-Longest Common Subsequence* (GAN-LCS) yang mampu digunakan untuk mengukur kesamaan kalimat yang mempunyai makna sama akan tetapi mempunyai perbedaan panjang yang besar. Dari hasil penelitian didapatkan bahwa metode MMR mampu membangkitkan jawaban kunci yang berasal dari jawaban siswa dengan sebesar 91,95. Sedangkan Kinerja GAN-LCS menghasilkan nilai *Root Mean Square Error* (RMSE) sebesar 0,884 dan nilai korelasi sebesar 0,468.

**Kata kunci** – *Short answer scoring*, MMR, GAN-LCS, *String similarity*

## ABSTRACT

Automatic Short Answer Scoring (ASAS) is one of the assessment media in e-learning. The use of ASAS in the assessment process in e-learning is still lacking compared to the automatic assessment of multiple choice questions. This is because the methods that have been developed by ASAS researchers have not been able to give a score commensurate with the scores generated by the teacher. Therefore, there is still plenty of room for research to develop methods applied to ASAS.

ASAS in principle is to compare between student answer sentences and key answer sentences using computerized methods. The results of the comparison process are then used as a basis for determining student scores. There are two crucial things in the comparison process, the first is the provision of adequate key answers so that they are able to capture the diversity of the students' response sentences. Second is the availability of a method that is able to be used to measure the similarity between diverse student answer sentences and key answer sentences. Provision of key answer sentences can be done in two ways namely manual and automatic. Provision manually depends on the ability of teachers to create variations of sentences and requires a long time. On the other hand the provision of key answer sentences through automation still requires a complex process, one of which must provide the corpus in accordance with the material being tested. A mechanism for providing key answers that is able to produce an adequate variety of key answers and does not require complex processes still needed. Meanwhile, several methods of measuring the similarity of sentences in ASAS that have been developed also still have a gap, which cannot be directly applied to different topics and languages. One method of measuring the similarity of sentences that can be applied directly without going through learning is a method based on string similarity. This method still has shortcomings, The first, it is still not accurate to measure sentences that have the same meaning but use different vocabulary. The secondly, it is not accurate to measure sentences that have the same meaning but have very different lengths. The second problem in string similarity is the focus of research in this dissertation.

In this dissertation two methods are proposed to solve the problems that still exist in ASAS. The first is to propose the application of the Maximum Marginal Relevance (MMR) method to automatically generate variations of key answers. The second is to propose a new method, namely Geometry Average Normalized-Longest Common Subsequence (GAN-LCS) which is able to be used to measure the similarity of sentences that have the same meaning but have large differences in length. From the results of the study it was found that the MMR method was able to generate key answers derived from student answers by 91.95. Whereas GAN-LCS Performance produces a Root Mean Square Error (RMSE) value of 0.884 and a correlation value of 0.468.

**Keywords :**Short answer scoring, MMR, GAN-LCS, String similarity