

ABSTRACT

During the past decades, researches about automatic grading have become interesting issue. These studies focuses on how to make machines are able to help human on assessing students' learning outcomes. Automatic grading enables teachers to assess student's answers with more objective, consistent, and faster. Automatic learning outcomes' assesment can be done with various types of questions and assessment models, such as multiple choice and essay. Especially for essay model, it has two different types, i.e. long essay and short essay (short answer). Almost of the previous researches merely developed automatic essay grading (AEG) instead of automatic short answer grading (ASAG). Afterward, the majority of these studies used private data, furthermore most of their data used English as their language.

This study aims to assess the lexical similarity of short answer to the questions and answers in Indonesian language without any language semantic's tool. This research uses pre-processing steps consisting of case folding, tokenization, stemming, and stopword removal. The proposed approach is String-Based Similarity methods with adding keyword extraction features and using a number of alternative reference answers provided by teachers. The dataset used by this research is private data. It consists of 7 questions, 34 alternative reference answers and 224 student's answers.

The experiment results show that the proposed approach is able to achieve a correlation value between 0.65419 up to 0.66383 at Pearson's correlation, with Mean Absolute Error (MAE) value about 0.94994 until 1.24295. The proposed approach also leverages the correlation value and decreases the error value in each method.

Keywords : short answer, automatic scoring, natural language processing, string-based similarity, keyword matching.

INTISARI

Selama beberapa dekade terakhir, penelitian tentang penilaian otomatis menjadi topik yang menarik. Penelitian tersebut berfokus pada bagaimana membuat mesin memungkinkan untuk menilai hasil belajar siswa. Penelitian otomatis memungkinkan guru untuk menilai jawaban siswa dengan lebih objektif, konsisten, dan cepat. Model penilaian esai memiliki dua jenis, yaitu esai panjang dan esai pendek (jawaban pendek). Sebagian besar penelitian sebelumnya hanya mengembangkan *Automatic Essay Grading* (AEG) daripada *Automatic Short Answer Grading* (ASAG). Selanjutnya, mayoritas penelitian tersebut menggunakan data privat dan sebagian besar data tersebut menggunakan bahasa Inggris.

Penelitian ini bertujuan untuk menilai kemiripan leksikal jawaban pendek pada pertanyaan dan jawaban berbahasa Indonesia tanpa menggunakan bantuan jaringan semantik bahasa apapun. Penelitian ini menggunakan tahapan pra-pemrosesan yang terdiri dari *case folding*, tokenisasi, *stemming*, dan *stopword removal*. Pendekatan yang diusulkan adalah metode *String-Based Similarity* dengan menambahkan fitur pencocokan kata kunci dan menggunakan sejumlah alternatif jawaban referensi yang disediakan oleh guru. Data yang digunakan pada penelitian ini adalah data privat yang terdiri dari 7 pertanyaan, 34 alternatif jawaban referensi dan 224 jawaban siswa.

Hasil eksperimen menunjukkan pendekatan yang diusulkan mampu meraih nilai korelasi antara 0,65419 sampai dengan 0,66383 pada korelasi pearson, dengan nilai *Mean Absolute Error* (MAE) antara 0,94994 sampai dengan 1,24295. Pendekatan yang diusulkan juga mampu menaikkan nilai korelasi dan menurunkan nilai *error* pada masing-masing metode.

Kata kunci – jawaban pendek, penilaian otomatis, pengolahan bahasa alami, *string-based similarity*, pencocokan kata kunci.