

## INTISARI

### Analisis Similaritas Jawaban Esai Dengan Kunci Jawaban Pada Algoritma *Cosine Similarity*

Oleh

Herlina

16/403684/PPA/05201

Ujian merupakan salah satu cara yang digunakan untuk mengukur kemampuan peserta didik dalam menerima pelajaran. Terdapat 2 jenis ujian, yaitu ujian dengan jawaban pilihan ganda dan dengan jawaban esai. Ujian dengan jawaban esai dinilai lebih mampu mengukur kemampuan peserta didik. Tetapi terdapat kendala dalam penilaian esai yaitu dibutuhkan waktu dan tenaga yang lebih dalam mengoreksi jawaban esai. Pengukuran kemiripan jawaban esai secara otomatis dapat menjadi solusi. Tetapi terdapat kendala dalam penilaian esai. Hal ini disebabkan bentuk jawaban esai yang tidak terstruktur, karena setiap peserta didik mempunyai kalimat tersendiri dalam menuliskan jawabannya. Oleh karena itu diperlukan metode yang tepat untuk diterapkan dalam mengukur kemiripan jawaban esai secara otomatis.

Penelitian ini menganalisis metode *cosine similarity* yang dikombinasikan dengan metode pembobotan Term-Frequency Inverse Document Frequency (TF-IDF). Untuk mengetahui performa *cosine similarity*, metode ini dibandingkan dengan 2 metode similaritas lainnya, yaitu metode *jaccard similarity* dan *dice similarity*. Penelitian ini juga menerapkan beberapa metode preprocessing seperti *case folding*, normalisasi, tokenisasi, n-gram, cek *typo* dengan algoritma Jaro-Winkler, dan menggabungkan kata negasi.

Pada tahap pengujian, penelitian ini menggunakan beberapa alternatif kunci jawaban. Pengujian juga dilakukan dengan menerapkan tahap preprocessing dan tanpa menerapkan tahap *preprocessing*. Hasil penelitian dengan menggunakan data yang ada menunjukkan bahwa metode *cosine* dan *dice similarity* mempunyai nilai yang sangat mirip dan dapat melakukan prediksi yang baik untuk mengukur tingkat kemiripan antara kunci jawaban dan jawaban untuk jawaban esai. Hal ini dapat dilihat dari hasil pengujian sistem dengan menerapkan *preprocessing* yang menghasilkan nilai F-measure dan akurasi yang sangat tinggi untuk *cosine* dan *dice similarity*.

**Kata Kunci:** *Cosine Similarity*, *Jaccard Similarity*, *Dice Similarity*, N-gram, Jaro-Winkler, Kemiripan Jawaban Esai

## ABSTRACT

### *Similarity Analysis Of Essay Answers To Answer Keys On Cosine Similarity Algorithm*

By

Herlina

16/403684/PPA/05201

An exam is a way that is used to measure the ability of students in accepting the lesson. There are two types of tests, a test with multiple choice answers and with essay answers. An exams with essay answers is considered more capable of measuring students' abilities. But there is a problem in the assessment of essays, which requires more time and energy in evaluating the answers. Measuring the similarity of essay answers automatically can be a solution. But this solution may not be implemented easily. This is due to the form of an unstructured essay answers because each student has their own sentence in writing the answers. For that, the right method is needed to be applied to measure the similarity of essay answers automatically.

This study analyzes the cosine similarity method combined with Term-Frequency Inverse Document Frequency (TF-IDF) weighting method. To find out cosine similarity performance, this method is compared to 2 other similarity methods, which are Jaccard similarity and Dice similarity method. This study also applies several preprocessing methods such as case folding, normalization, tokenization, n-gram, and typo checking with Jaro-Winkler algorithm.

In the testing phase, this study used several alternative answer keys. The testing was done with and without preprocessing phase. The results of the study using existing data showed that cosine and dice similarity had very similar values and have good predictions to measure the similarity between the answer key and the answer to the essay answer. This can be seen from the results of system testing by applying preprocessing which produced a very high F-measure and accuracy value for cosine and dice similarity.

**Keywords:** *Cosine Similarity, Jaccard Similarity, Dice Similarity, N-gram, Jaro-Winkler, Essay Answer Similarity*