

Analisis pengidentifikasian kata - kata yang muncul berulang menggunakan algoritma cosine similarity pada sistem penilaian esai otomatis simple o = Analysis identifying repetition of a word using cosine similarity in automated essay grading system simple o / Ria Yuliana

Ria Yuliana, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20411909&lokasi=lokal>

Abstrak

ABSTRAK

Pada skripsi ini akan membahas mengenai pengimplementasi pendeteksian pengulangan kata menggunakan Cosine Similarity, sedangkan untuk melihat makna antar kalimat akan dilakukan pembobotan kalimat yang sebelumnya di lakukan proses parsing, stemming, stopword dan melakukan pemisahan antar kalimat. Metode stemming yang akan digunakan adalah metode stemming Arifin-Setiono, proses stemming digunakan untuk mencari kata dasar dari setiap kata pada jawaban dan kemudian dibandingkan dengan kata yang ada pada database. Jika antar kalimat masing-masing dibandingkan dan memiliki nilai cosine similarity lebih dari 0.5 maka sistem tambahan akan menghapus satu kalimat tersebut.

Waktu proses antara sistem SIMPLE-O murni lebih cepat jika dibandingkan dengan SIMPLE-O dengan pendeteksian kata yang berulang antar kalimat dengan selisih waktu 0.22 detik. Korelasi nilai antara SIMPLE-O+Cosine dengan Human Raters adalah yaitu sebesar 0.38 untuk soal nomor satu dan untuk soal nomor dua memiliki nilai korelasi yang bernilai negatif 0.08 , untuk soal tiga 0.13, untuk soal empat 0.65 dan 0.022 untuk soal nomor lima.

<hr>

ABSTRACT

In this paper will discuss the use of word repetition detection implement Cosine Similarity, while to see the meaning of the sentence will be weighted between sentences previously performed process of parsing , stemming , stopword and separation between sentences . Stemming method to be used is a method of stemming Arifin - Setiono , stemming process used to find root of each word in student answer , and then compared with of the word database .

If the inter- sentence and has been compared and have value of cosine similarity more than 0.5 , the additional features on the system will remove a sentence .

The time between system processes SIMPLE - O more fast when compared with SIMPLE-O with additional features on the system that will repetitive detection word sentence with a time difference between 0.22 seconds. The correlation value of the Human Raters with the SIMPLE-O + Cosine is 0.383 for the first question, -0.08 for the second question, 0.13 for the third question, 0.65 for the fourth question and 0.02 for the fifth question.