

Pengembangan sistem penilaian esai otomatis dengan metode LSA dengan koreksi tingkat kata = Automatic essay grading system development using LSA with word level correction

Randy Sanjaya, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20456730&lokasi=lokal>

Abstrak

Pada skripsi ini dikembangkan sistem dengan basis bahasa pemrograman Java untuk menilai esai dalam bahasa Indonesia menggunakan algoritma yang lebih efisien dan optimal. Algoritma ini terdiri dari 4 tahap. Pertama adalah Latent Semantic Analysis (LSA) yang digunakan untuk memperoleh dan menyimpulkan hubungan kontekstual dari arti kata suatu teks. Kedua, Single Value Decomposition SVD untuk memperoleh variasi penyebaran dari hubungan tersebut. SVD mengidentifikasi dimana variasi muncul paling banyak, sehingga memungkinkan untuk mencari pendekatan yang terbaik pada data asli menggunakan dimensi yang lebih kecil. Ketiga, Latent Semantic Indexing LSI yaitu metode pengindeksan dan pengambilan untuk mengidentifikasi pola didalam hubungan antara term dan konsep yang dimiliki didalam koleksi teks yang tidak terstruktur sehingga memperoleh vektor yang merepresentasi teks tersebut. Terakhir, Cosine Similarity Measurement CSM untuk memperoleh nilai kemiripan antara teks dengan dokumen referensi.

Untuk mengatasi permasalahan tata bahasa dan kosa kata pada esai, dalam karya ini diajukan teknik koreksi otomatis untuk memeriksa kata dalam pustaka kata untuk penyetaraan kata dengan arti yang serupa ataupun kata yang tidak memiliki arti spesifik. Kemudian, algoritma jarak Jaro-Winkler digunakan untuk memeriksa kesalahan kata yang disebabkan secara tidak sengaja. Dengan jarak Jaro-Winkler, kita dapat menentukan apakah 2 buah kata dapat dikatakan serupa. Hal ini sangat penting saat memeriksa dokumen yang berisi kesalahan penulisan, karena dapat mempengaruhi hasil LSA. Dengan sistem ini, nilai yang diperoleh serupa dengan nilai berdasarkan human-rater. Dengan pustaka kata yang terdiri dari 116 kata sinonim dan 2014 kata tugas, akurasi yang dihasilkan adalah 85.082 13.423.

In this thesis, a Java based system for grading essays in Indonesian language using a more efficient and optimal algorithm is developed. This algorithm consisted of 4 stage. The first stage is Latent Semantic Analysis LSA , which is used to obtain and conclude the contextual relation of words meaning in a text. The second stage uses Single Value Decomposition SVD to obtain scatter variance from the relations. SVD identifies where variances appear at most, therefore is enabled to find the best approach to the original data using reduced dimensions. The third stage is Latent Semantic Indexing LSI which is an indexing and retrieval method to identifies patterns in relation between terms and concepts contained in unstructured text collection and results with a vector representing the text. The last stage is Cosine Similarity Measurement CSM to obtain similarity value from the text and answer document.

To resolve problems stemmed from grammar and vocabulary, in this work we propose an auto correction technique to check a word from word library for equalization of word with same or no specific meaning. Then, Jaro Winkler distance algorithm is used to check word errors caused by accident when typing. With the distance, we can determine whether two strings of word are similar. This is extremely important when scanning text with typos, as it will affect the result from LSA. Using this system, the value obtained is similar to the value obtained from human rater. With word library consisting of 116 words for synonym check and 204 function words, the resulting accuracy is 85.082 13.423.