

Analisis efek modifikasi algoritma latent semantic analysis terhadap akurasi deteksi plagiarisme paper bahasa Indonesia dan Inggris = Analysis of latent semantic analysis algorithm modification effect against the accuracy of plagiarism detection between Indonesian and English paper / Emily Lomempow

Emily Lomempow, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20411041&lokasi=lokal>

Abstrak

ABSTRAK

Latent Semantic Analysis (LSA) adalah metode yang dapat digunakan untuk membandingkan kesamaan antar teks dengan memanfaatkan representasi kata ke dalam komputasi statistik berdasarkan konteksnya dalam teks tersebut. Pada skripsi ini dirancang sistem yang dapat mendeteksi plagiarisme antara paper bahasa Indonesia dengan paper bahasa Inggris. Sistem dirancang berdasar pada metode LSA, tetapi dengan beberapa modifikasi untuk meningkatkan kecepatan komputasi serta keakuratan program. Metode LSA yang digunakan adalah hasil penelitian yang berasal dari program Simple-O. Dimana, metode ini memiliki keunggulan di waktu proses yang lebih cepat karena mengurangi vector space dalam proses SVD. Beberapa modifikasi dirancang untuk memperoleh hasil yang paling akurat, antara lain menghilangkan stopwords sebelum pemrosesan dan pembentukan matriks term-document dengan keywords dari paragraf referensi saja. Gabungan dari kedua modifikasi tersebut memberikan hasil yang paling akurat dengan persentase 81,82% sampai dengan 90,91%. Kemudian nilai dari pengujian akan dicek berdasarkan batas mutlak, system ranking, maupun perhitungan distribusi normal untuk menentukan adanya indikasi plagiarisme. Hasil pengecekan plagiarisme paling akurat diperoleh menggunakan perhitungan distribusi normal, dengan persentasi 79,49% sampai dengan 87,81%.

<hr>

ABSTRACT

Latent Semantic Analysis (LSA) is a method to find the similarity between two texts using the statistical representation of the words based by its contextual means in each text. The system in this thesis is designed to be able to detect plagiarism between two paper written in different languages, which are Indonesian and English. The system is designed using modified version of LSA which is first developed for Simple-O program. This modified version of LSA excel in fast computation as the effect of vector space reduction in SVD process. Several forms of modification are developed to bring forward the most accurate algorithm for the program, for instances are excluding stopwords from LSA processing and creating term-document matrix using words from reference paragraph only. The algorithm composed using the two modifications produces the best result with 81,82% to 90,91% accuracy. The value obtained from the test will be used to decide if there is an indication of plagiarism between two paragraphs using an absolute threshold, ranking system, or based by normal distribution calculation. The most accurate results are obtained from normal distribution calculation based detection with 79,49% to 87,81% rate of success.