

Analisis efek variasi nilai term-document matrix pada algoritma latent semantic analysis terhadap akurasi deteksi plagiarisme paper Bahasa Indonesia dan Inggris = Analysis of variation term document matrix value effect in latent semantic analysis algorithm against accuracy of plagiarism detection Indonesian and English paper

Mardiyah, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20431134&lokasi=lokal>

Abstrak

ABSTRAK

Pada skripsi ini telah dilakukan modifikasi metode untuk membangun sistem pendeteksian plagiarisme yang disebut dengan LSA atau Latent Semantic Analysis. Metode ini bekerja dengan mengekstrak dan merepresentasikan konteks yang digunakan sebagai sebuah arti kata dengan memanfaatkan komputasi statistik untuk sejumlah korpus yang besar dari teks. Modifikasi yang diterapkan yaitu mengubah panjang definisi dokumen pada Term-Document Matrix serta mengubah metode pengisian matriks pada Term-Document Matrix dari metode frekuensi ke metode biner. Hasil keluaran LSA yang dapat menentukan keakurasian sistem akan bervariasi sebagai akibat modifikasi sistem. Skripsi ini juga akan membahas metode dan alur yang digunakan untuk menganalisa perbedaan hasil keluaran LSA serta menampilkan hasil pengolahan data keluaran LSA yang kemudian akan memberikan nilai keakuratan masing-masing variasi sistem.

<hr>

ABSTRACT

This thesis has been modified method to build plagiarism detection system called LSA or Latent Semantic Analysis. This method works by extracting and representing context is used as a meaning of the word by using statistical computing to a large corpus of text. Modifications are applied by changing the length of the document definitions Term-Document Matrix and change the method of charging matrix in Term-Document Matrix of frequency to the binary method. The output of the LSA to determine the accuracy of the system will vary as a result of modifications to the system. This thesis will also discuss the methods and flow of used to analyze differences in the output of the LSA as well as displaying the data processing LSA output which will then provide the value of the accuracy of each of the various systems.