

Implementasi pembobotan inverse document frequency - normalized document vektor pada SIMPLE-O = Implementation of inverse document frequency- normalized document vector weighting to SIMPLE-O

Vivi Zulvi Fauziah, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20249069&lokasi=lokal>

Abstrak

Metode Latent Semantic Analysis (LSA) adalah salah satu metode yang digunakan untuk sistem penilaian essay secara otomatis (essay grading). Metode ini merepresentasikan suatu tulisan essay kedalam bentuk matriks dimana pada matriks tersebut dihitung frekuensi kemunculan suatu kata. Matriks tersebut kemudian diolah dengan cara penganalisisan bernama SVD (Singular Value Decomposition) yang akan mengkompres matriks besar menjadi matriks yang lebih kecil namun matriks yang lebih kecil tersebut mampu mewakili korelasi antar kata yang lebih baik dibandingkan dengan matriks sebelumnya.

Pada skripsi ini akan dibahas mengenai aplikasi SIMPLE-O yang dikembangkan dengan cara memberi pembobotan pada kata-kata kunci pada matriks awal. Pembobotan yang digunakan adalah IDF-NDV (Inverse Document Frequency - Normalized Document Vector) yang telah diujikan pada penelitian sebelumnya. Setelah diujikan, pembobotan IDF-NDV masih perlu perbaikan untuk mendekati human raters, dan dibandingkan dengan SIMPLE-O lama, terdapat perbedaan sebesar 11,78%.

Method of Latent Semantic Analysis (LSA) is one of the methods used for automatic essay grading system (grading essays). This method represents an essay written in the matrix form where the matrix is calculate the frequency of occurrence of a word. Matrix is then processed by analyzing method called SVD (Singular Value decomposition), which compresses the large matrix into smaller matrices, but that smaller matrix can represent the correlation between a better word than the previous matrix.

In this final project will discuss about the application of SIMPLE-O who developed a way to give weighting to the key words at the matrix. The Weighting method was used is IDF-NDV (Inverse Document Frequency - Normalized Document Vector) that have been tested in previous research. After tested, the IDF-NDV Weighting still need improvement to approach the human raters, and compared with Original SIMPLE-O, there is a difference of 11,78%.