

Implementasi latent semantic indexing dengan singular value decomposition pada aplikasi information retrieval untuk dokumen berbahasa Indonesia = Implementation of latent semantic indexing with singular value decomposition

Laskito Harmantyo, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20242712&lokasi=lokal>

Abstrak

Aplikasi information retrieval dalam wujud mesin pencari atau search engine sudah dikenal luas oleh pengguna internet. Informasi bisa didapatkan dengan melakukan pencocokan istilah dalam dokumen dengan istilah yang ingin dicari. Kata yang dimasukkan dalam query dicari kehadirannya dalam sejumlah dokumen. Pencocokan ini merupakan pencocokan secara harfiah atau yang dikenal dengan lexical matching. Metode pencocokan secara harfiah ini dapat memberikan hasil yang kurang akurat karena pencocokan harfiah hanya melihat kesamaan bentuk kata atau istilah saja tanpa melihat makna dan korelasi istilah tersebut. Latent Semantic Indexing (LSI) mengatasi masalah tersebut dengan menggunakan pendaftaran dan penyusunan istilah secara statistik. Struktur laten dituangkan dan digambarkan secara matematis dalam elemen-elemen matriks yang terukur. Teknik Singular Value Decomposition (SVD) digunakan untuk melakukan estimasi struktur penggunaan kata dalam dokumen-dokumen. SVD juga dipakai untuk meminimalisasi perhitungan dan meningkatkan performa information retrieval. Pembobotan pada dokumen dan query ditambahkan untuk meningkatkan performa relevansi perolehan dokumen. Implementasi LSI dengan SVD dilakukan secara web-based, dengan koleksi dokumen berupa judul dan abstrak dari sejumlah skripsi mahasiswa teknik elektro. Hasil ujicoba memberikan kesimpulan bahwa dari seluruh rangkaian proses perolehan informasi, 95% waktunya dihabiskan untuk kalkulasi SVD. Selain itu, pemakaian fitur klasifikasi dapat mempercepat proses sampai 64 kali waktu tanpa klasifikasi. Berbagai metode pembobotan juga terbukti memberikan hasil perolehan dokumen yang berbeda pada query yang sama, dengan tingkat relevansi yang berbeda. Berdasarkan perhitungan Non Interpolated Average Precision, skema pembobotan frekuensi kata untuk dokumen dan query merupakan skema yang paling baik dalam hal presisi.

.....Information retrieval in form like search engine has already known by a lot of internet users. Information can be obtained by doing terms matching. Terms in documents are matched dan compared by terms given by users in query. Such method like this known as lexical matching. This method can give inaccurate results because the mechanism is only matching and seeing the similarities of words without giving any concern of meaning or relevancy. Latent Semantic Indexing (LSI) try to compensate this problem by doing indexing and arranging terms in statistical manner. Latent structures are developed in mathematical way as values in matrices elements. Singular Value Decomposition (SVD) is used to estimating the structure of words in documents. In addition, SVD minimalize the matrices calculation and improves the performance of information retrieval application. Weighting scheme to documents and query added to improves the performance of relevancy retrieval. LSI with SVD implemented in web based way, with title and abstract from student of electrical engineering papers as document corpus. The experiment gives a fact that from all LSI process time, for about 95% is spent for SVD calculation. Classification feature of this application gives an acceleration up to 64 times of common process time (without classification). The number of user agent accessing the application gives a slow effect of processing time in linear manner. The great number of user,

the longer process time. Various kind of weighting scheme makes a different documents retrieval result at the same queries. Calculation of Non-Interpolated Average Precision stated that word frequency weighting scheme for both document and query is the best in precision performance.