

INTISARI

APLIKASI REPOSITORI DATA ILMIAH MENGGUNAKAN INFORMATION RETRIEVAL DENGAN ALGORITMA LATENT SEMANTIC INDEXING

Bondan Bhaskara

10/297610/PA/13030

Sumber daya informasi digital ilmiah (*scientific data*) meningkat secara signifikan dengan banyaknya penelitian yang menghasilkan data ilmiah dalam jumlah besar. Data ilmiah biasanya berhubungan satu sama lain dalam hal bidang yang sama, maupun obyek ilmiah yang sama. Karena banyaknya data yang tersedia terkadang membuat proses pencarian data yang berhubungan membutuhkan waktu yang lama dan sulit. Penelitian ini mengimplementasikan metode *information retrieval* yang efektif dan memberikan hasil pencarian yang sesuai dengan keinginan user aplikasi pengumpulan data ilmiah

Pada penelitian ini diimplementasikan metode *Latent Semantic Indexing* pada repositori data ilmiah. Pencarian dokumen dengan metode ini dapat mencari hasil berdasarkan basis konseptual dari dokumen. Proses dekomposisi matriks yang terjadi di metode ini dapat mencari hubungan kata dengan topik. Hasil dari dekomposisi matriks adalah nilai vektor dari semua dokumen dan *query* dari pengguna yang digunakan untuk mencari nilai kesamaannya.

Berdasarkan hasil pengujian, kinerja sistem repositori data ilmiah yang dinilai dengan *precision* dan *recall* mampu memberikan hasil yang relevan, dan jumlah dokumen berpengaruh terhadap kecepatan sistem pencarian.

Kata kunci: Information Retrieval, Latent Semantic Indexing, Singular Value Decomposition, Scientific Data, Data Repository

ABSTRACT

IMPLEMENTATION OF LATENT SEMANTIC INDEXING ALGORITHM IN SCIENTIFIC DATA REPOSITORY APPLICATION

Bondan Bhaskara

10/297610/PA/13030

Scientific based digital information is increasing significantly equal to increasingly research with digital data. Scientific data is usually related one to another in terms of same field or scientific object. Because a lot of data is available, search process can take long and difficult. In this research, it is implement an effective method in information retrieval, providing search results based on user query in scientific repository.

In this research, Latent Semantic Algorithm is implemented in scientific data repository. This searching method can find results from conceptual basis of basis. Matrix decomposition process from this method can be used to extract topic from document collection. Result from matrix decomposition is vector value from all documents and search query to calculate the similarity score.

Based on test result, scientific data repository performance which is assessed with precision and recall, can give relevant result. Number of documents in database affect the performance of system.

Keyword: Information Retrieval, Latent Semantic Indexing, Singular Value Decomposition, Scientific Data, Data Repository