

## INTISARI

# ANALISIS PENGGUNAAN MODEL KOMBINASI VSM-LSA DALAM SISTEM *INFORMATION RETRIEVAL* PADA DATA KARYA ILMIAH UNIVERSITAS GADJAH MADA

Oleh

M. AL AMIN

22/510437/PPA/06841

Jenis *query* dapat mengakibatkan pencarian informasi yang tidak optimal terutama pada data karya ilmiah. Misal, pada *query* spesifik dengan menggunakan *query* tersebut kebanyakan model sistem temu balik informasi mengalami kesulitan dalam mengembalikan dokumen relevan salah satu contoh model yang sering digunakan dalam sistem *Information Retrieval* (IR) adalah *Vector Space Model* (VSM) yang telah menjadi pendekatan utama untuk mengatasi masalah ini dengan menggunakan metode perhitungan jarak antara kata kunci dan dokumen untuk menghasilkan *similarity*. Namun, VSM mengalami masalah ketika dokumen mempunyai hubungan semantik dengan *query*, yang mengakibatkan ketidakakuratan dalam pencarian informasi.

Penggunaan model kombinasi VSM-LSA dan model LSA terpisah, dalam proses pengembalian informasi memberikan solusi dengan menggunakan penjumlahan nilai kemiripan untuk kedua model dan menormalisasikan nilai kemiripan khusus untuk model LSA dari hasil awal -1 sampai 1 menjadi 0-1. Metode ini menganalisis struktur hubungan kata-dokumen dalam bentuk matriks. Hal ini dapat mengoptimalkan hasil pengembalian informasi relevan sedangkan model VSM merepresentasikan dokumen dan *query* sebagai vektor di ruang multidimensi dengan pembobotan menggunakan TF-IDF (*Term Frequency-Inverse Document Frequency*).

Penelitian ini menerapkan model VSM, LSA, dan kombinasi VSM+LSA dalam sistem *Information Retrieval* (IR) dengan menggunakan data karya ilmiah Universitas Gadjah Mada. Hasil model dievaluasi efektivitas kinerjanya dengan menggunakan nilai *Mean Average Precision* (MAP) dan rata-rata presisi untuk setiap *rank* hasil *retrieved*. Efisiensi diukur dengan waktu yang diperoleh dalam pengembalian informasi dengan jenis *query* yaitu sederhana maupun spesifik yang digunakan untuk setiap model. Metode yang diajukan tidak dapat secara optimal mengembalikan informasi relevan berdasarkan *query* spesifik maupun sederhana dengan hasil hampir sama dengan model LSA baik secara efisiensi waktu yang diperoleh maupun efektivitas kinerjanya. Namun, untuk *query* sederhana model VSM mendapatkan efektivitas dan efisiensi kinerja optimal dibandingkan kedua model lainnya dengan nilai MAP 0.62870 dengan waktu yang diperoleh 0.01562 s untuk pengembalian dokumen dengan *query* sederhana.

**Kata Kunci:** *Latent Semantic Analysis, Vector Space Model, Information Retrieval, Topic Modelling*

## ABSTRACT

### ***ANALYSIS OF VSM-LSA COMBINED MODEL IN INFORMATION RETRIEVAL SYSTEMS FOR SCIENTIFIC PAPERS AT GADJAH MADA UNIVERSITY***

Oleh

M. AL AMIN

22/510437/PPA/06841

This type of query can result in a search for information that is not optimal, especially in scientific work data. For example, in a specific query using this query, most information retrieval system models have difficulty returning relevant documents. One example of a model that is often used in Information Retrieval (IR) systems is the Vector Space Model (VSM) which has become the main approach to solving problems. This is done by using a distance calculation method between keywords and documents to produce *similarity*. However, VSM experiences problems when documents have a semantic relationship with the query, which results in inaccuracies in information retrieval.

The use of a combined VSM-LSA model and a separate LSA model, in the process of returning information provides a solution by using the sum of the *similarity* values for both models and normalizing the special *similarity* value for the LSA model from the initial result of -1 to 1 to 0-1. This method analyzes the structure of word-document relationships in matrix form. This can optimize the results of returning relevant information while the VSM model represents documents and queries as vectors in multidimensional space with weighting using TF-IDF (Term Frequency-Inverse Document Frequency).

This research applies VSM, LSA, and a combination of VSM+LSA models in the Information Retrieval (IR) system using scientific work data from Gadjah Mada University. The model results are evaluated for their performance effectiveness using the Mean Average Precision (MAP) value and the average precision for each rank of retrieved results. Efficiency is measured by the time obtained in returning information with the type of query, namely simple or specific, used for each model. The proposed method cannot optimally return relevant information based on specific or simple queries with almost the same results as the LSA model both in terms of time efficiency and performance effectiveness. However, for simple queries the VSM model obtains optimal performance effectiveness and efficiency compared to the other two models with a MAP value of 0.62870 with a time obtained of 0.01562 s for document retrieval with simple queries.

**Keywords:** *Latent Semantic Analysis, Vector Space Model, Information Retrieval, Topic Modelling*