

ABSTRAK

Skripsi merupakan tugas akhir mahasiswa sebagai syarat kelulusan. Setiap tahun skripsi akan terus bertambah banyak seiring dengan kelulusan mahasiswa. Sebagai akibatnya jumlah informasi yang didapat dari skripsi akan terus bertambah banyak. Skripsi ini bertujuan membuat suatu aplikasi yang dapat mengklasifikasikan skripsi berdasarkan dokumen – dokumen abstrak yang telah ada. Menggunakan teknik sistem temu kembali informasi .

Sistem temu kembali informasi (*information retrieval system*) digunakan untuk menemukan kembali (*retrieve*) informasi - informasi yang relevan terhadap kebutuhan pengguna dari suatu kumpulan informasi secara otomatis. Dengan sistem informasi temu kembali, terdapat dua proses penting yaitu proses *indexing* dan pencarian dokumen. Proses *indexing* adalah proses pemberian bobot kata berdasarkan frekuensi kemunculan kata pada satu dokumen dan menyimpannya kedalam koleksi dalam *database* yang disebut *index*. *Index* yang terbentuk dapat digunakan untuk pencarian dokumen yang sesuai dengan *query user*, dimana hasil pencarian disusun berdasarkan tingkat kemiripan dokumen. Proses pencarian kemiripan inilah yang digunakan untuk melakukan klasifikasi dokumen atau teks.

Untuk melakukan klasifikasi digunakan *vector space model* dengan *term* (kata) sebagai pembentuk ruang vektor. Setiap vektor *term* mengandung bobot yang digunakan untuk menghitung tingkat kemiripan dokumen. Tingkat kemiripan dokumen dihitung menggunakan *cosinus similarty*. Untuk klasifikasi dokumen menggunakan algoritma *rocchio*. Dalam algoritma *rocchio* terdapat 2 proses penting yaitu *training* dan *testing*. Proses *training* dilakukan secara manual, bertujuan untuk membentuk model dokumen yang merupakan *centroid* dari sebuah kategori dokumen. Kategori dokumen yang digunakan adalah basisdata, jaringan dan komputasi. Proses *testing* adalah proses pengujian dimana dokumen akan secara otomatis diklasifikasikan oleh sistem.

Skripsi ini menggunakan 60 dokumen abstrak skripsi sebagai model *training* dan 10 dokumen abstrak skripsi untuk pengujian. Untuk menghitung tingkat *precision* dilakukan dengan membandingkan hasil *quisioner* dan hasil klasifikasi sistem. Dari hasil perhitungan tingkat *precision* didapatkan *precision* terendah 80% dan *precision* tertinggi 100% dari kuisisioner yang dibagikan kepada responden sebanyak 30 orang. *Precision* rata – rata adalah 93.67% dan hasil ini baik.

ABSTRACT

Thesis is the final project, students as a condition of graduation. Each year the paper will continue to grow much in line with the graduation of students. As a result of information obtained from theses will continue to increase. This thesis aims to create an application that can classify the thesis based on the document - the document has no abstract. The methodology used is the information retrieval system.

Information retrieval system (information retrieval system) used to find the back (retrieve) information - information that is relevant to the needs of users from a collection of information automatically. With information retrieval systems, there are two important processes: process indexing and searching documents. Indexing process is the process of weighting by frequency of occurrence of words in a word document and save it into a collection in a database called the index. Index that is formed can be used to search according to user queries, where search results have been prepared based on the similarity of the relevant documents to irrelevant documents based on similarities in the training documents with user queries. Similarity search process is used to perform the classification of documents or texts.

This thesis documents menggunakan 60 abstract theses as a model training and 10 thesis abstracts documents for testing. Precision is done by comparing the results of questionnaires and the results of the classification system. Number of is respondents 30 people and the lowest precision is 80% and the highest precision is 100%. Precision - average is 93.67% and show the good result.