

WEIGHTING IN INDEXING PROCESS FOR DOCUMENT IN BAHASA INDONESIA USING INDRI FRAMEWORK

Hendrex Herdi, Julio Adisantoso

Department of Computer Science

Faculty of Mathematics and Natural Sciences

Bogor Agricultural University

julioipb@gmail.com

2010

Abstract

A very large amount of information has stimulated the development of information search engine to help users in finding information they need. To retrieve the information according to the user's needs, information search engine should be able to work well. One of the factors that can affect the performance of search engines is indexing. The purpose of this research is to implement automatic indexing process using Indri framework with TF-IDF and BM25 term weighting. This testing used 30 queries and 2000 documents. The testing result showed that the performance of information search engine is better when we use the BM25 term weighting than TF-IDF term weighting. However, the performance of information search engine with BM25 term weighting and TF-IDF term weighting gave good results with around 64% average precision. The number of indexed documents for indexing will affect the indexing time. Increasing of the number of indexed documents will increase the indexing time.