

## ABSTRAK

Muta'abidin, Wildan. 2013. **Rancang Bangun Aplikasi *Search Engine* Ensiklopedia Hadits Menggunakan Teknik *Text Mining* dengan Algoritma VSM (*Vector Space Model*)**. Skripsi. Jurusan Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Sunan Gunung Djati Bandung. Pembimbing: (1) Drs. Opik Taupik Kurahman, M.Ag. (2) Wisnu Uriawan ST., M.Kom

---

Keywords : Encyclopedia, Hadits, Search, Text Mining, Vector Space Model

There are a lot of hadits scattered in holy books that looks complex in categorizing, therefore media is needed to collect it. One of the media is encyclopedia. Many encyclopedia is formed as book that contains the collection of hadits even in thousands. The thickness of the encyclopedia makes access process difficult. Nowadays with development of internet technology so the hadist encyclopedia based on web is built and then this application contains search engine. The development of web technology gives convenience in accessing and searching content that relate with encyclopedia content. And expected to help the user in find information with suitable the order.

This research uses text mining technique with vector space model algorithm to measuring similarity search result toward document. Text mining technique has three important phase that is : text preprocessing, text transformation, and pattern discovery. Text preprocessing phase existed from text cleaning and sentence fission to be words (tokenizing) so that be a stem. Then, the stem result will analyzed with counting word value and suitable keywords. Text transformation existed from filtering phase. The pattern discovery phase is valuing phase. The value is TF-IDF. From the value calculation result with TF – IDF, the calculate value document similarity will undertaken. In measuring level of similarity between keywords and document use vector space model algorithm with cosine similarity method.

The data used for testing is known that in the searching process to discover relevant results should enter the keyword which suitable Indonesian writing rules