

**Web-Based Directory for Malay Hadis Retrieval System  
Using Inverted Indexed Files Technique**

**BY**

**SITIFARIZA BINTIMOHD YUSAK  
BACHELOR OF COMPUTER SCIENCE (Hons)**

**THESIS SUBMITTED IN PARTIAL FULFILLMENT OF  
THE REQUIREMENT FOR THE DEGREE OF  
BACHELOR OF COMPUTER SCIENCE**

**FACULTY OF COMPUTER AND MATHEMATICAL  
SCIENCES  
UNIVERSITITEKNOLOGI MARA**

**NOV 2010**

## **ACKNOWLEDGEMENT**

First of all, Alhamdulillah for giving me a strength and bless in completing this thesis.

This research also would not be possible and successful without the help and support, advice, guidance from the course co-ordinator, Dr. Noor Elaiza Abdul Khalid and En. Fakhru Hazman Yusoff and my supervisor, Assoc. Prof. Nurazzah Abdul Rahman for the idea, suggestion and comment in helping me to complete this project.

Many thanks to my beloved parent for giving me full support, understanding and courage throughout the research without hassle.

Finally, to my colleagues, thanks for the moral support that they given me and other for their opinion, support, suggestion and cooperation to do this research and finished up this prototype.

## ABSTRACT

Information retrieval is to find all documents relevant for a user query in a collection of documents. Information retrieval (IR) is fast becoming the dominant form of information access over taking traditional database style searching. An inverted file typically includes the vocabulary, a vector containing all distinct words in the text collection and a list of all document numbers in which each distinct word occurs. The lists of document numbers constitute the largest part of an index (N.Ziviani, E.Silva de Moura, CNavarro, R.A. Baeza-Yates, 2000). Inverted files require less space and provide greater functionality. Web directories are taxonomies for the classification of Web documents. This kind of IR systems presents a specific type of search where the document collection is restricted to one area of the category graph. Generally, existing Mutiara Hadith provide a search engine to search hadith using stemming and thesaurus but the main problem is it will searching by non-specific hadith. Thus mean, it will retrieve all word from the word entry. In order to improve the retrieval of relevant document, web-based directory which search hadith by main topic will be created. The objective of this project is to develop web-based directory hadith by topics for Hadith Shahih Bukhari and Hadith Shahih Muslim text document by using inberted indexed files technique. It is also to evaluate the retrieval effectiveness by using Recall and Precision formula. Evaluation from this result shows that retrieval effectiveness is well again usability in the query and inverted index using web-based directory.

Keyword: Information retrieval, Web directory, Hadith, retrieval effectiveness.

# TABLE OF CONTENTS

	Page
<b>DECLARATION</b>	ii
<b>ACKNOWLEDGEMENT</b>	in
<b>ABSTRACT</b>	iv
<b>TABLE OF CONTENT</b>	v
<b>LIST OF TABLE</b>	viii
<b>LIST OF FIGURE</b>	ix
<b>1. Chapter 1: Introduction</b>	
1.1 Introduction	1
1.2 Problem statement	2
1.3 Objective	3
1.4 Scope of Project	4
1.5 Significance of Project	4
<b>2. Chapter 2: Literature review</b>	
2.1 Introduction	5
2.2 Definition of Web directory	6
2.3 Definition of Inverted Indexed files	7
2.4 Retrieval Process	9
2.5 Related Research on Inverted Indexed files	10

2.5.1	Inverted files and dynamic signature files for optimization of Web directories	10
2.5.2	Inverted Files versus Signature Files for Text Indexing	10
2.5.3	Automatic Construction of Web Directory using Hyperlink and Anchor Text	11
2.6	Approaches of Inverted Indexed files	11
2.6.1	Full-Text database	11
2.6.2	Technique of Inverted Indexed files	12
2.6.2.1	Extended Inverted Indexed Files versus Compression Inverted Indexed Files	12
2.6.3	Type of inverted index	13
2.6.3.1	Frequency index	13
2.6.3.2	Scheme-independent index	14
2.7	Comparison of Traditional Information Retrieval	14
2.7.1	Vector space model	14
2.7.2	Boolean Model	15
2.8	Mathematical modeling for inverted index files	16
2.8.1	Vector space model	16
2.8.2	Boolean model	18
3.	<b>Chapter 3: Research Methodology</b>	
3.1	Introduction	20
3.2	Research / Project Formulation Framework	20