

A Web-Based Library Search Engine Using Conflation Method

BY

ZULFAHANAN BINTI ABDUL GANI

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**

UNIVERSITI TEKNOLOGI MARA

NOV 2010

Acknowledgement

In the name of Allah the most Gracious and the most Merciful

May His blessing be upon the Prophet Muhammad s.a.w

I would like to express my profound gratitude to Allah S.W.T, for He has bestowed me with ideas, strength and opportunity to finish up this project.

My gratitude also goes especially to my supervisor, Prof. Madya Nurazzah bt Abdul Rahman, for her invaluable support, encouragement, supervision and useful suggestions throughout this research work. Her moral support and continuous guidance enabled me to complete my work. I am also thankful to Dr. Noor Elaiza Binti Abdul Khalid and Dr. Fakhrul Hazman bin Mohd Yusoff, my thesis coordinator, for all of his comments and valuable suggestions that he gave to me. My personal gratitude goes to my family, thanks for their love and supporting throughout my life. I also would like to thank to all of my friends, who shared their love and experiences with me. Finally thanked to all who directly or indirectly involved during completing this final project.

Abstract

Information retrieval (IR) is a field that helps a lot in searching process, so that is why it is used widely nowadays in many applications, systems and so on. With the development of the World Wide Web, the information search has grown to be a significant business sector of a global, competitive and commercial market. Many methods are used in order to make a powerful search engine including by using Stemming Algorithm which are in Conflation Method. Library system also needs to be grown with powerful search engine that can give an effective search ways to the users. As we know, library system is centre of information to the users and it must provide better search engine in order to ease user to get any information from it. Nowadays, there are many library search engine that are created, but usually for high institution like for university and just for some schools only. Therefore, first objective of this project is to develop a data collection for library search engine. Second, to design and develop library search engine for students using Conflation Method. In this project also, indexing technique where inverted index file is used in order to speed up searching. Actually, indexing is a data structure build on the text.

Keyword: Stemming Algorithm, inverted index file, library search engine.

Table of Contents

DECLARATION

ACKNOWLEDGEMENT

ABSTRACT

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

- 1. Chapter 1 - Introduction**
 - 1.1 Problem Statement
 - 1.2 Objective
 - 1.3 Scope
 - 1.4 Significant

- 2. Chapter 2 - Literature Review**
 - 2.1 Introduction to Information Retrieval
 - 2.1.1 Process in Information Retrieval
 - 2.1.2 Illustration of Information Retrieval System
 - 2.2 Overview of Search Engine
 - 2.2.1 How Do Search Engine Works
 - 2.2.2 Search Engine for Standalone System
 - 2.3 Library Search Engine
 - 2.4 Inverted Index File
 - 2.5 Related Work
 - 2.5.1 Conflation-based Comparison of Stemming Algorithms.
 - 2.5.2 Term Conflation Methods in Information Retrieval: Non Linguistic and Linguistic Approaches.
 - 2.5.3 Semantic Method of Conflation.
 - 2.6 Conclusion

3.	Chapter 3 - Methodology	
3.1	Research Methodology	13
	3.1.1 Planning	14
	3.1.2 Design	14
	3.1.3 Development	14
	3.1.4 Testing	14
	3.1.5 Documentation	14
3.2	Project Design Overview	15
3.3	Project Design Detail	17
	3.3.1 System Requirement/Information Gathering	17
	3.3.2 Test Collection	18
	3.3.3 System Design	18
	3.3.4 Result analysis	19
3.4	The Development Phase	20
3.5	Method That Will Be Used	21
	3.5.1 Exact Matching	21
	3.5.2 Stemming Algorithm	22
	3.5.3 Thesauri	23
3.6	Conclusion	25
4.	Chapter 4 - Result and Discussion	
4.1	Inverted Index File	26
	4.1.1 Book Inverted Index File	27
	4.1.2 Magazine Inverted Index File	29
	4.1.3 Newspaper Inverted Index File	31
4.2	Graphical Interface	33
4.3	Questionnaire Result	43
	4.3.1 Part A: Demographics	44
	4.3.2 Part B: Investigating on experience of using library search engine	45
	4.3.3 Part C: Knowing the difficulty level of this library search engine	47