Database Development And Online Digital Archive Retrieval System For Bahagian Arkib Dan Muzium (BAM) UiTM

BY

NORASHIDA BINTI SABARI

BACHELOR OF COMPUTER SCIENCE (HONS)

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE DEGREE OF BACHELOR OF COMPUTER SCIENCE (HONS)

FACULTY OF COMPUTER AND MATHEMATICAL SCIENCES

UNIVERSITI TEKNOLOGI MARA

MAY 2011

ACKNOWLEDGEMENT

First and foremost, all praises and thanks to Allah, Lord of al-Mighty, for His Guidance and will, for the revelation of some of His knowledge for me in the successful to complete this project.

Secondly, many thanks to my beloved family, team-mates, classmates and friends for never quitting in giving me full support, understanding and courage throughout the research without hassle.

This research would also not be possible and successful without the help and support from my supervisor, Assoc. Prof. Dr. NursuriatiJamil and course coordinator, Dr.SharifalillahNordin. Many thanks and appreciation goes to them for giving instructions, advices, motivation, support and guide the research in obtaining a good research. I would also like to En. Mohd Noor, persons from BahagianArkibdanMuziumUiTM (BAM) that help me in getting the data and some information about their department.

Finally, a deepest gratitude goes to my course colleagues of CS230 for their help and others who have, in one way or others, given me invaluable help, assistance and advice. And to the respondents for the cooperation they gave. Last but not least, to the seniors who have shared their knowledge. Thank you very much.

ABSTRACT

Information retrieval is finding material (usually documents) ofan unstructured nature (usually text) that satisfies an informationneed from within large collections (usually stored on computers). Information retrieval has become widely used and becoming the dominant form of information access instead of the traditional database style searching especially to search in the large databases. BahagianArkibdanMuzium (BAM) UiTM holds large numbers of data that will be referenced by users. The objectives of this project are to design and develop database for BAM and to develop a document (newspaper cutting) retrieval prototype based on description using Boolean Model. The scope of this project is newspaper cutting containing news on UiTM. In the future, thisproject needs enhancement to produce better result with implementing advance stemmingand retrieval techniques.

Keywords: Information Retrieval, Boolean Model, Stemming

TABLE OF CONTENTS

			Page
Approval			ii
Declaration			iii
Acknowledgements			iv
Abstract			v
List of Tables			ix
List of Figures			x
Chapter 1:		Introduction	1
	1.1	Research Background	1
	1.2	Problem Statement	2
	1.3	Objective	2
	1.4	Project Scope	3
	1.5	Project Significance	3
Chapter 2:		Literature Review	4
	2.1	Introduction	4
		2.1.1 Database Design and Development	4
		2.1.2 Information Retrieval	6
		2.1.3 Digital Archive	8
		2.1.4 BahagianArkibdanMuziumUiTM	10
	2.2	Related Research on Document Retrieval	12
		2.2.1 Basic Document Retrieval System	12
		2.2.1.1 Document processing	13
		2.2.1.2 Query processing	14
		2.2.1.3 Matching of Query to Documents	15
		2.2.2 Theoretical Models	16
		2.2.2.1 The Boolean Retrieval Model	16

2.2.2.2 The Vector Space Model 17
2.2.2.3 The Probabilistic Model 17
2.2.2.4 The Language Model 18

Chapter 3: Methodology 21
3.1 Introduction 21
3.2 Research Framework 22
3.2.1 SDLC Diagram 22
3.2.2 Research Flow 23
3.3 Research Activities 26
3.3.1 Requirement Analysis 26
3.3.2 Data Collection 26
3.3.3 System Design and Development 28
3.3.3.1 Document Processing 31
3.3.3.2 Query Processing 32
3.3.3.3 Matching Process 32
3.3.4 System Testing 35
3.3.5 Documentation 35
3.4 Hardware and Software Requirements 35
3.5 Detail Discussion on the Proposed Technique 37
3.6 Conclusion 38

Chapter 4: Results and Finding 39
4.1 Introduction 39
4.2 To design and develop database for BAM 39
4.3 To develop a document (newspaper cutting) retrieval prototype based on description using Boolean Model. 41
4.4 To test the functionality of the prototype. 44