

**Universiti Teknologi MARA**

**The Development of XML Document  
Retrieval Using Query Weighting**

**Fauziah Binti Mat Sa'at @ Puteh**

Thesis submitted in fulfillment of the requirements for  
**Bachelor of Science (Hons) Business Computing**  
**Faculty of Information Technology And**  
**Quantitative Science**

November 2005

## **DECLARATION**

I hereby declare that the work in this thesis is my own except for quotations and summaries, which have been duly acknowledged.

21<sup>st</sup> November 2005

FAUZIAH BINTI MAT SA'AT @ PUTEH  
2003283312

## **ACKNOWLEDGEMENT**

In the name of Allah, the Compassionate, the Merciful, Praise be to Allah, Lord of the Universe, and Peace and Prayers be upon His final Prophet and Messenger.

I would like to express my sincere gratefulness and gratitude to my supervisor Cik Hayati Abd Rahman for her guidance, encouragement and advices during the course of this thesis. She is the one who encouraged me the freedom of individual endeavor essential to the fulfillment of my work and gave very possible assistance in the pursuit of this study.

A pleasure to acknowledge to the lecture, Puan Marina Yusoff and also my friends for their cooperation, encouragement, contribution and spending time together to discuss about the research.

Last but not least, in collection of many kindness, continuous encouragements for perseverance and patience, my grateful thanks to my family especially my parents, Mat Sa'at @ Puteh bin Abdullah and Chik binti Mohd Nong for their supported. Thanks to all.

## **ABSTRACT**

Query is one of the processes involved in information retrieval in order to obtain the retrieved documents. Query is importance in information retrieval because it will affect the retrieved documents. The poor formulated queries will lead to the poor document retrieval such as the retrieved document might be too large or too small and no provision for ranking documents. There are many techniques that we can use to process the user query and these different kinds of techniques will give a different result for documents retrieval. One of the techniques is using query weighting algorithm which is either the system will calculate the weight of the importance words or the user itself will give a weight to the importance words. This research is focus on developing an XML documents retrieval using a query weighting algorithm in order to retrieve FTMSK official letter. This technique has been successfully tested and it proves that query weighting technique is very effective in XML document retrieval.

**Keywords:** Query Weighting, XML Documents Retrieval, Query.

# TABLE CONTENTS

CONTENTS	PAGE
TITLE PAGE	
APPROVAL	
DECLARATION	iii
ACKNOWLEDGEMENTS	iv
ABSTRACT	v
CONTENTS	vi- viii
LIST OF TABLES	ix
LIST OF FIGURES	x
<b>CHAPTER ONE : INTRODUCTION</b>	
1.0 Research Background	1
1.0 Problem Description	2
1.2 Project Objectives	2
1.3 Project Scope	2
1.4 Project Significance	3
1.5 Project Limitations	3
1.6 Project Overview	3
<b>CHAPTER TWO : LITERATURE REVIEW</b>	
2.0 Introduction to Information Retrieval (IR)	5
2.1 Introduction to Query In Information Retrieval	5
2.2 About Term Weighting	6
2.3 Introduction to Query Weighting	7
2.4 How Does Ranking Works With Query Weighting	8
2.5 Advantages of Query Weighting and Term Weighting	8
2.6 About Vector Space Model	9
<b>CHAPTER THREE : RESEARCH METHODOLOGY</b>	
3.0 Introduction	10
3.1 Project Methodology Overview	10