

37793

Universiti Teknologi MARA

**Development of Indexer Considering Tag
Weighting For XML Document**

Maslina Binti Wahid @ Buang

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

November 2005

DECLARATION

I certify that this thesis and the research to which it refers are the product of my own work and that any ideas or quotation from the work of other people, published or otherwise are fully acknowledged in accordance with the standard referring practices of the discipline

NOVEMBER 21, 2005

MASLIANA BINTI WAHID @ BUANG

2003283390

ACKNOWLEDGEMENTS

First and foremost, I would like to praise to God the Almighty for giving me the time and strength to finish writing this thesis and for all that has bestowed on me. It is with His ascendancy the study is completed.

I would like to express my sincere gratefulness and gratitude to my supervisor, Cik Hayati Bt Abdul Rahman and my thesis lecture, Puan Marina Yusoff for their invaluable guidance, encouragement and advice during the course of this thesis. They are 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.

My thanks also go to the lectures, supporting staff and special thanks for my senior, Mohd Sharizan Mohd Shariff for their cooperation and helpfulness in academic matter and their most generous assistance. I would like also to express my gratitude to my Fauziah, Edatul Muliana, Kamazzatul Arkmaz and Aminah for their cooperation and support on doing the similar project.

Last but not least in recollection of my kindness, continuous encouragement for perseverance and patience, my grateful thanks to all my family especially to my mother Puan Masri Bt Hj Anuar. Thank you.

ABSTRACT

Information retrieval makes the retrieval of document or collection of data in the storage media more efficient, effective, relevant, faster and more reliable than before. Such indexing techniques may influence the effectiveness of retrieval itself. The extension component within the indexing structure may also influence the performance of the retrieval process. This research is concerns to develop a prototype for indexing algorithm considering tag weighting for the XML document and also to test the indexer with the existing document. In order to perform efficient retrieval on documents, appropriate index structure or algorithm must be used which include the structural information. This research is used the inverted file method for the indexing techniques to develop the indexing algorithm of the FTMSK official letter. The relevancy of the document for the retrieval by using the algorithm has been successful achieved and it can prove that the prototype can increase the relevancy of document retrieval.

CONTENTS

		Page
APPROVAL		
DECLARATION		iii
ACKNOWLEDGEMENT		iv
ABSTRACT		v
CONTENTS		vi
LIST OF TABLES		ix
LIST OF FIGURE		x
CHAPTER 1	INTRODUCTION	
1.0	Research Background	1
1.1	Problem Statement	2
1.2	Project Objective	2
1.3	Project Scope	3
1.4	Project Significance	3
1.5	Summary	3
CHAPTER 2	LITERATURE REVIEW	
2.0	Introduction	4
2.1	Indexing	4
2.2	Indexing Techniques	5
2.3	Extensible Markup Language (XML)	6
2.4	Weighting	9
2.5	Summary	10
CHAPTER 3	RESEARCH METHODOLOGY	
3.0	Project Methodology Overview	11
3.1	Document Acquisition and Transformation	11