

**Universiti Teknologi Mara**

**The Development of The Indexing Prototype  
Considering Tags Into The Inverted File: Case Study  
on Ftmsk's Official Letter**

**MOHD SHARIZAN MOHD SHARIFF**

Thesis submitted in fulfillment of requirement of  
**Bachelor of Science (Hons) In Information Technology**  
**Faculty of Information Technology and Quantitative Science**

**APRIL 2005**

## **DECLARATION**

I hereby declare that this thesis is based on the result found by myself. Materials of work found by the researcher are mentioned in references. This thesis has neither in whole nor in part been previously submitted for any degree

1 APRIL 2005

MOHD SHARIZAN MOHD SHARIFF

2002610014

## **ACKNOWLEDGEMENT**

The preparation of this project paper gave me lot opportunities to be exposed to such working in project development environment. Without the help and endless support from various people, the completion of this project paper would not be possible.

First, I would like to thank the Great One for giving me an opportunity and strength to complete this thesis. In particular, I would like to thank my supervisor, Cik Hayati Abd Rahman for all personal advice, valuable guidance, views and support from the initial process of this prototype development until the completion of this project. Thank you very much

My deepest thanks are also dedicated to my beloved family and friends who never failed to give me their love and constant support during the ups and downs in this research completion.

Finally, I would like to express my profound to all parties and individuals that had rendered their help directly or indirectly during the preparation and completion of this thesis

## **ABSTRACT**

The combination of IR and structure form (XML) make the retrieval process become more powerful than before. As far as the effectiveness of document retrieval is concerned, each segment (part) in the letter (document) has its own meaning or usage. Thus, term weight must be taken into consideration in order to make each segment (part) of the document more meaningful and to make the retrieval process produce more relevant output to the user. This idea is the basis for the prototype development. The prototype has been built using Visual Basic platform with MS Access as the data storage and structure. Inverted files technique had been chosen as the basis for the data structure in this prototype. The retrieval effectiveness is measured using redefined recall (R) and precision (P) that used to measure structured document. The evaluation will be done between the CAS (the prototype) and CO (benchmark) retrieval. The result of evaluation been done shows that the term weighting assist in production of more relevant output to user query rather than ignorance of it in structured document. Each part of the segment in the structured form of the document become more identical in query process with the used of term weighting inserted in the tags.

## **TABLE OF CONTENT**

TITLE	i
DECLARATION	ii
ACKNOWLEDGEMENT	iii
ABSTRACT	iv
TABLE OF CONTENT	v

## **CHAPTER 1 INTRODUCTION**

1.1 RESEARCH BACKGROUND	1
1.2 PROBLEM DESCRIPTION	2
1.3 PROJECT OBJECTIVE	2
1.4 PROJECT SCOPE	3
1.5 PROJECT SIGNIFICANCE	3
1.6 PROJECT LIMITED	3