

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

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

CHAPTER 2 LITERATURE REVIEW

2.1 INTRODUCTION	4
2.2 INDEXING	4
2.3 INVERTED FILES	6
2.4 EXTENSIBLE MARKUP LANGUAGE	7
2.5 WEIGHTING AND RANKING	9

CHAPTER 3 RESEARCH METHODOLOGY

3.1 PROJECT METHODOLOGY OVERVIEW	12
3.2 DATA ACQUISITION AND TRANSFORMATION	13
3.3 PROTOTYPE DESIGN AND ARCHITECTURE	14
3.4 SELECTING THE INDEXING TECHNIQUES	15
3.5 PROTOTYPE DEVELOPMENT	17
3.6 SYSTEM TESTING AND DEBUGGING	18
3.7 SYSTEM EVALUATION	19