

**UNIVERSITI TEKNOLOGI MARA**

**WEB INFORMATION RETRIEVAL (IR)  
INTERACTION**

**AMIROHAIME MAHAIDIN**

**THESIS SUBMITTED IN PARTIAL FULFILMENT FOR THE  
BACHELOR OF COMPUTER SCIENCE (HONS)**

**MAY 2008**

## ACKNOWLEDGEMENTS

Praise to Allah the Almighty for giving me superior knowledge throughout this period of study and for all that has bestowed on me. It is with His power the study is completed.

I would like to express a very special gratitude and thankfulness to my supervisor Assoc.Prof Nurazzah Abd Rahman and coordinator subject, Dr. Siti Salwa Salleh who giving me reliance and shrewd guidance that have profited me very much. They whose gave a lot of support in this knowledge quest. The special thankful for Principle of Sekolah Kebangsaan Karak and Sekolah Kebangsaan Sri Bentong, Karak Pahang, to teachers especially Mrs Noraini Jaafar and all student involved.

My special appreciation is due to the Dean of the Faculty Information and Technology and Quantitative Science, Prof. Dr Zainab Abu Bakar and the Programmed Head of Bachelor (Hons.) Computer Science, Dr. Zaleha Zainal Abidin who giving good management services throughout my degree program. My credits are also due to the computer laboratory staff of the faculty for their most generous assistance.

To my friends Maszurah Othman, and my entire classmate in this faculty, who are simply too numerous to indicate, thanks for your helping with your time, understanding and buttress. May Allah will reciprocate you for all those things.

Last but not least, to my parents Mahaidin Abdullah, and my sibling, I would like to forward my obliged to them for their continuous support during my study period, their patience and benevolence. Thank you.

# TABLE OF CONTENT

<b>CONTENT</b>	<b>PAGE</b>
<b>APPROVAL</b>	
<b>DECLARATION</b>	
<b>ACKNOWLEDGEMENT</b>	ii
<b>TABLE OF CONTENT</b>	iii
<b>LIST OF TABLES</b>	vi
<b>LIST OF FIGURES</b>	vii
<b>ABSTRACT</b>	viii

## **CHAPTER 1: INTRODUCTION**

1.0 Background	1
1.1 Problem Statement	3
1.2 Objectives	3
1.3 Scope of Research / Project	4
1.4 Contribution of Research / Project	5

## **CHAPTER 2: LITERATURE REVIEW**

2.0 Introduction	6
2.1 Related Researches on Web Information Retrieval	
2.1.1 Problem secondary schools student are facing in retrieving relevant information.	7
2.1.2 Web-based Information Retrieve systems to help secondary school students.	9

## **ABSTRACT**

A good information retrieval (IR) interaction should provide tools to perform searching accurately based on user requirement. There are many website retrieve variant to the same relevant website. To overcome these website variants problems, conflation was introduced. Stemming method was proven to be successful in reducing words with the same stem to a common form and is evidenced by the work of many researchers. Web information interaction retrieves the relevant website to user and applied stemming algorithm. This website project consists of interactive guidance to retrieve the relevant URL. Searching from query given to URL document will retrieve the relevant URL suggestion by expert. The testing is for measuring the recall and precision in retrieve and relevant document. The project testing shown that most of documents retrieved, does not relevant to the queries. As a conclusion we must applied more than one algorithm in term to retrieve relevant document accurately based on user requirement.

# CHAPTER 1

## INTRODUCTION

### 1.0 Background

This chapter will explain briefly on the introduction to the topic “Web Information Retrieval (IR) Interaction for Secondary school Students Use of the Web to Solve Assigned Specific Task”. This chapter also consists the problem of statement for the research, objectives, scopes and contributions of research.

Information retrieval (IR) is a part of computer science, which studies the retrieval of information (not data) from a collection of written documents. The retrieved documents aim at satisfying a user information need usually expressed in natural language. Rijsbergen (1979), the of IR, defines IR as a study on how to determine and retrieve from a corpus of stored information, the part that is relevant to particular information needs.

IR is deals with the representation, storage, organization of, and access to information items (*Ricardo Baeza-Yates, 1999*). This research focus on web IR interaction with secondary school students for their task. Interaction is two ways communication that has request and feedback, this concept of interaction is apply in web IR. In two ways communication IR should be effectiveness and efficiency.

IR should be effectiveness because commonly IR measured to precision and recall but IR should not be efficiency because efficiency refer to performance time and difficult to measure a time. Precision is the ratio of the number of