

**CREATING DATABASE FOR THESIS ABSTRACT IN  
INFORMATION RETRIEVAL**

**RAZEDI BIN ABU**

**THESIS SUBMITTED IN PARTIAL FULFILMENT FOR THE  
DEGREE OF BACHELOR OF SCIENCE**

**FACULTY OF INFORMATION TECHNOLOGY  
AND QUANTITATIVE SCIENCES  
UNIVERSITITEKNOLOGI MARA  
SHAH ALAM**

**2002**

## ACKNOWLEDGEMENTS

الرحمن الرحيم  
t\_Jj Ji JULcJI  
الصلاة والسلام على اشرف الانبياء والمرسلين

Praise to Allah the Almighty for providing me a 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. Dr. Zainab Abu Bakar who is giving me reliance and shrewd guidance that have profited me very much. She who is also gave lots of support in this knowledge quest.

My special appreciation is due to the Dean of the Faculty of Information Technology and Quantitative Sciences, Assoc. Prof. Azizi Ngah Tasir and the Programme Head of Bachelor (Hons.) Information Technology, Puan Jamaliah who is 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, Khairul Helmy Naburi, Anuar b. Hj. Othman, Wan Iznif bin Wan Ibrahim and others in the faculty, who are simply too numerous to indicate, thanks for your time, understanding and buttress. May Allah will reciprocate you for all those things.

Last but not least, to my parents, Abu bin Daud , Ahmad bin Daud and my siblings, I would like to forward my obliged to them for their continuous support during my study period, their patience and benevolence. Thank you.

والله علم بالصواب  
أمين

## **ABSTRACT**

Information usually can exist like text, graphic, voice, movie and many more. But in the Internet features, to get the information we need to have some kind of web browsing search engine to retrieve the information and it is involve the database via online services. The main purpose is to retrieve the information efficient and effectively by using the best of search engine method. The best of search engine is getting the accurate of result that what is the users want without having any weakness. In order to make this search engine, exact match method is one of the popular in the information retrieval. In this study, it needs to evaluate, identify and make a conclusion or summary of the search method that has been use. In the exact match search engine method, it will give the result of thesis abstract document only for the exactly words that have in the thesis abstract collection. It is also consist the collection of thesis abstract, relevant judgment and a set of query that have in the thesis abstract. In order to prove this exact match method is giving the result like the theory, the experiment of exact match must be done. This experiment is involved using the query word (Experiment I and Experiment IV) and using the natural query words (Experiment II, Experiment III, Experiment V and Experiment VI). The result of this study show that a retrieval system that used exact matches will retrieved documents and looking the efficiencies and effectiveness of information retrieval.

## TABLE OF CONTENTS

	<b>Page</b>
<b>ACKNOWLEDGEMENT</b>	<b>ii</b>
<b>TABLE OF CONTENTS</b>	<b>iii</b>
<b>LIST OF TABLES</b>	<b>vi</b>
<b>LIST OF FIGURES</b>	<b>vii</b>
<b>ABSTRACT</b>	<b>viii</b>
 <b>CHAPTER</b>	
<b>1.0 INTRODUCTION</b>	
1.1 Background	1
1.2 Problem Description	4
1.3 Objectives	4
1.4 Scope	5
1.5 Summary	6
 <b>2.0 LITERATURE REVIEW</b>	
2.1 Introduction	7
2.2 String Similarity Method	8
2.3 N-gram String Similarity Method	10
2.4 Dynamic Programming Method	14
2.5 Thesaurus	23
2.6 Stemming	25
2.7 Summary	27
 <b>3.0 METHODOLOGY</b>	
3.1 Introduction	28
3.2 Thesis Abstract Collection	28
3.2.1 Abstract Document	29

# CHAPTER I

## INTRODUCTION

### 1.1 Background

Any information retrieval system, there must be have a sample of test collections that consist of document database and a set of queries for the database for which relevant judgment are available required to retrieve the information (Frakes 1992). Without the availability of these test collections, the information that cannot be retrieve in order to look their effective and efficiency.

Information retrieval (IR) is important in Information Technology. This project is concentrate on how to create database abstract so that IR is more effective and efficiency and at the same time to evaluate the searching method that will be use to retrieve the database test collection. The main function of this project is to collect and store all the collection of thesis abstracts and keep them systematically and to allow users to use this data effectively.

Nowadays, Information Technology or IT has enabled information to be captured, transmitted, stored, retrieved, manipulated or displayed (Altar 1996) via a computer that is linked to the Internet. Not just that, mobile computer and telephone also can retrieve the information as the technology becomes more sophisticated from time to time. Information can be in many forms such as text, image or sound. Advance in information technology has made it possible for a wide range of text-based