

**IMAGE AND TRANSLATED AL-QURAN VERSES RETRIEVAL
SYSTEM USING THESAURUS APPROACH BASED
ON MALAY QUERY WORDS**

MOHD ZAINI MAT ABAS

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

**FACULTY OF INFORMATION TECHNOLOGY
AND QUANTITATIVE SCIENCES
UNIVERSITI TEKNOLOGI MARA
SHAH ALAM**

2001

ACKNOWLEDGEMENTS

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ
 الْحَمْدُ لِلَّهِ رَبِّ الْعَالَمِينَ
 الصَّلَاةُ وَالسَّلَامُ عَلَى أَشْرَفِ الْأَنْبِيَاءِ وَالْمُرْسَلِينَ

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 Science (Hons.) Information Technology, Assoc. Prof. Kalsom Nasir who are 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, Mohd Rosmadi Mokhtar, Nazean Jomhari 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, Mat Abas Mat Yusof, Jariah Abdul Ghaffar, Fatimah Haron 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.

وَاللَّهُ عَالِمُ بِالصَّوَابِ
 أَمِين

ABSTRACT

Information can be in the form of texts, images or sounds. Current technology features allow Internet users to retrieve any documents that available in the online database from remote hosts or locally. The main issue in this information age was the efficiency and effectiveness of the retrieval system that used by the information provider. A good retrieval system should provide tools to perform searching accurately based on user requirement. The study of information retrieval is how to determine and retrieve from a mass of prepared information; the part that is relevant to particular information needs. There are many methods available to achieve this objective. Conflation is one of it. In English and French language conflation methods have been successful to identify words variants from present and historical database. As well as Malay language, conflation methods such as string-similarity method and stemming algorithms have been applied to Malay documents retrieval system. Another way of conflating related term is a thesaurus, which lists synonymous terms, and sometimes the relationship between them. In thesaurus approach, the behaviors of the language and the subject areas are important because it was totally language dependent. This study has evaluated and identified the effectiveness of thesaurus in Malay document retrieval system using Al-Quran test collections which consists Al-Quran documents, Al-Quran images, relevant judgments and one set of queries. As the result of this study, a retrieval system that used thesaurus will be obtained more retrieved documents compare to exact match searching. However, it will reduce the efficiency and the effectiveness of the retrieval system. Although, it can be better if a proper study on this area can be perform together with a specialist who have good Islamic knowledge.

TABLE OF CONTENTS

| | | Page |
|------------------------|--|-------------|
| DECLARATION | | ii |
| ACKNOWLEDGEMENT | | iii |
| ABSTRACT | | iv |
| CONTENTS | | v |
| LIST OF TABLES | | viii |
| LIST OF FIGURES | | x |
| | | |
| CHAPTER I | INTRODUCTION | |
| 1.1 | Background | 1 |
| 1.2 | Objectives | 3 |
| 1.3 | Significant of Project | 3 |
| 1.4 | Scope | 3 |
| 1.5 | Summary | 4 |
| | | |
| CHAPTER II | LITERATURE REVIEW | |
| 2.1 | Introduction | 5 |
| 2.2 | Conflation Methods | 6 |
| 2.3 | Stemming | 7 |
| 2.4 | Thesaurus | 8 |
| 2.5 | Type Of Images | 9 |
| | 2.5.1 Joint Photographic Expert Group (JPEG) | 10 |
| | 2.5.2 Portable Network Graphic (PNG) | 10 |
| | 2.5.3 Graphical Interchange Format (GIF) | 11 |
| 2.6 | Al-Quran Online | 11 |

CHAPTER I

INTRODUCTION

1.1 Background

The string searching problem is to find all occurrences of a query word in a translated Al-Quran documents. The problem is a fundamental concern of computer science. String searching methods are included in every text editor, word processor, search engine and many other applications. Besides that, people may want to know the originality of the words of their query. In this case images of Al-Quran documents can be proof for the query result. All those images will retrieve based on Malay query words that entered by user.

The main function of information retrieval system is to provide the user with tools to perform searching effectively and efficiently. In order to achieve these goals a study on the behaviors of the language also must be firstly done. Technique such as edit distance, stemming, thesaurus and many others will be used to compare which technique is the best. In this paper, a thesaurus approach will study on to retrieve related documents from Al-Quran base on the user query. Al-Quran has 6,236 verses and every each verse describes the Islamic rule, tauhid and the past and future stories as guidance to mankind. Every Muslim must refer to Al-Quran in whatever activity in their life. As the technology developed, an effort to accommodate Muslim needs should be done. With this online Al-Quran Retrieval System, it can provide Malay Muslim to have information about Al-Quran at their convenient time.