

**RETRIEVING MALAY TRANSLATED AL-QURAN
BY USING N-GRAMS WITH RELEVANCE
FEEDBACK**

NURAFZAN BINTI MOHAMMAD

**THESIS SUBMITTED IN FULFILMENT OF THE REQUIREMENTS FOR
BACHELOR OF SCIENCE (Hons) INTELLIGENT SYSTEMS**

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

OCTOBER 2004

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 and the Programme Head of Bachelor Sc. (Hons.) Intelligent Systems, Encik Mohd Zaki Zakaria who is giving good management services throughout my degree program.

To my friends, 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, Haji Mohammad bin Abdullah, 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 عِلْمٌ بِالصَّوَابِ
أَمِين

ABSTRACT

Information Retrieval is one of the fields in information technology that perform a task to search effectively and efficiency. This project is broad the study on the information retrieval. To ensure retrieving the information works effectively, relevance feedback technique implement in this project. Relevance feedback technique help to reduce uncertainty of documents retrieve by incorporating user's feedback in the system. Where the result is retrieved based on user judgment instead of computer generated result. This project also employs bigram algorithm as a retrieval method that is used to retrieve the relevant al-Quran documents based on user's query.

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	ii
LIST OF TABLE	vi
LIST OF FIGURES	ix
ABSTRACT	x
CHAPTER	
1 INTRODUCTION	1
1.0 Problem Background	1
1.1 Problem Description	2
1.2 Project Scope	3
1.3 Project Objective	4
1.4 Project Methodology	4
1.5 Hardware Requirement	6
1.6 Software Required	6
1.6.1 C Language	6
1.7 Project Benefits	7
1.8 Summary	7
2 LITERATURE REVIEW	9
2.0 Introduction	9
2.1 An Information Retrieval System	11
2.2 Conflation Method	13
2.3 N-grams	15
2.4 Relevance Feedback	21
2.5 Al-Quran Retrieval System	27
2.5.1 The Humanities Text Initiatives, Uni Of Michigan, USA	28
2.5.2 Brown University's Scholarly Technology Group, UK	29
2.5.3 Bandung Institute Of technology, Indonesia	30

CHAPTER 1

INTRODUCTION

1.0 Problem Background

Information retrieval (IR) is one of the fields in Information Technology that involves in finding some desired information in a stored of information on a database. Information retrieval actually is a communication process. It was means by which used of an information system can found not only documents and records, also graphics images or sound that meet users' need or interest.

Due to the rapid growth of the information technology in this age, information retrieval technique has been applied to generate a few number of Al-Quran Retrieval System, where is Al-Quran translated verses can be retrieved and viewed based on users' queries. Day by day, the number of Al –Quran Retrieval System increased and got an improvement to ensure the effective and relevance information retrieved. Some of Al-Quran Retrieval System has been developed as a web based application. Despite it, there is a limited number of web based application system on retrieving Malay Translated Al-Quran.

Most of the existing Al-Quran Retrieval System employs exact match as a conflation technique and was deal with uncertainty result, where the result of the retrieval is did not meet the user satisfaction. Due to this problem, this project appends bigram, one of the conflation techniques. In order to improve the information retrieval result, the concept of relevance feedback employ in this project,