

**THE STUDY OF EXISTING MALAY ALGORITHM PERFORMED ON
WORDS BEGINNING WITH 'D'**

ELLY JOHANA JOHAN

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

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

2000

ACKNOWLEDGEMENTS

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

Praise be to Allah (swt). His loves and peace be upon Nabi Muhammad (saw) and his family and companions. Thank to Allah (swt) for giving me the time and strength to finish writing this thesis.

I would like to express my sincere gratefulness and gratitude to my supportive supervisor, PM. Dr. Zainab Abu Bakar for her invaluable guidance, encouragement, and advice during the course of this thesis.

My thanks also go to all my friends for their understanding and co-operation where I spent most of my time, for sharing happy and unforgettable moments together.

Last but not least, I would like to express my gratitude to my beloved family for their encouragement, patience, support, financial support and sacrifice they have given me during the course of this thesis.

ABSTRACT

This thesis concerns a Malay language documents retrieval system. Stemming algorithm, database Quran translated documents and electronic root dictionaries are used in order to complete this study. The performance of a Malay stemming algorithm is tested based on words that beginning with 'd', using two experiments. First, use the original set of data collections. Second, the data that have been modified in order to correct the error that exists in database Quran translated documents and in electronic root dictionary. The results of these experiments are based on the 24 order of the rules that consist of prefix, suffix, prefix-suffix pair and infix. The main objective is to minimize the unstemming, understemming, overstemming and other problems that occurred when 'd' words stemmed. It is achieved the objective when the best order of rule to used to stem the words that beginning with 'd' is met. The best rule combinations are 15, 17 and 18. These experiments can serves as a benchmark for future research in Malay language. Furthermore it can help those who are interested to know about certain subject matters from the Al-Quran where the document retrieval system will automatically retrieve all relevant documents in response to the users' queries.

TABLE OF CONTENTS

	Page
DECLARATION	ii
ACKNOWLEDGEMENTS	iii
ABSTRACT	iv
TABLE OF CONTENTS	v
LIST OF TABLES	ix
LIST OF FIGURE	x
CHAPTER I INTRODUCTION	
1.1 Background	1
1.2 Problem Description	4
1.3 Scope Of The Research	4
1.4 Objective Of The Research	5
1.5 Significance Of The Research	5
1.6 Limitation Of The Research	6
1.7 Outline Of The Thesis	6

CHAPTER I

INTRODUCTION

1.1 BACKGROUND

Information Retrieval (IR) can be defined broadly as the study of how to determine and retrieve from a corpus of stored information the portions, which are responsive to the particular information, needs (Sembok 1989). IR is also concerned with text representation, text storage, text organisation and the retrieval of stored information items that are similar in some sense to information requests received from users. The term IR covers a wide range of disciplines and has some similarities with many other areas of information processing, e.g., management information systems database management systems, decision support systems, question-answering systems, natural language processing, as well as document retrieval systems.