

Universiti Teknologi MARA

**KEYWORD SEARCH FOR DOA
RETRIEVAL USING N-GRAM
METHOD**

ANIS SAFURAA BINTI AHMAD TAJUDIN

Thesis submitted in fulfillment of the
requirements for Bachelor of Computer Science
(Hons) Computer Science
Faculty of Computer and Mathematical Sciences

JULY 2015

ACKNOWLEDGEMENT

Alhamdulillah, praises and thanks to Allah because His Almighty and His utmost blessings, I was able to finish this research within time duration given. Firstly, my special thanks to my supervisor, PM Zaidah Binti Ibrahim. She gives many advices, criticism, guidance and creative ideas during completed this final year project. The special thanks also dedicated to my final year project lecturer, Dr Norhaslinda Binti Kamaruddin gives the support and criticism and also guidance to achieve of this final year project. Special appreciation also goes to my beloved mother for her encouragement, patience, support, financial support and sacrifice that she have given me during the course of this thesis. Last but not least, I would like to give my gratitude to my dearest friends that they help directly or indirectly to completed my final year project

ABSTRACT

The rapid growth of the Internet has increased the number of stored and available data tremendously for online documents. One of the documents is Doa which is an act of supplication. Doa is derived from Arabic word which means that one asks for the fulfillment of a need or that some harm be removed from him/her. One way to retrieve the appropriate Doa is to construct a prototype to search and retrieve the appropriate Doa. There are some applications that require using the exact match of the keyword search with the words stored in the database. This approach leads to the retrieval of insignificant results. Thus, this project utilizes N-gram method for the search and retrieval process. Various words may have similar meaning and to increase the accuracy of the retrieved result, this project applies the dice and overlap coefficient algorithm to find the synonyms of the searched word. A relational database for the Malay words has been constructed in two table which is title of doa and verse and the prototype to retrieve the Doa in Arabic, its' meaning in Malay and the use of it has been developed. The result is overlap coefficient percentage result better than dice coefficient.

Keywords: Relational Database, Keyword Search, Dice and Overlap Coefficient, N-Gram Algorithm, Malay keyword.

TABLE OF CONTENTS

CONTENT	PAGE
SUPERVISOR'S APPROVAL	1
STUDENT'S DECLARATION	ii
ACKNOWLEDGEMENT	iii
ABSTRACT	iv
TABLE OF CONTENT	vi
LIST OF FIGURES	viii
LIST OF TABLES	ix
CHAPTER 1: INTRODUCTION	1
1.0 Background	1
1.1 Problem Statement	2
1.2 Objectives	2
1.3 Possible Solution	2
1.4 Scope and Limitation	2
1.5 Significance of Project	2
CHAPTER2: LITERATURE REVIEW	4
2.0 Introduction	4
2.1 Doa Retrieval	5
2.2 Introduction of keyword Search	5
2.3 Method	5
2.3.1 Edit Distance	6
2.3.2 Symbol-Based Tree Traverse Search Scheme	6
2.3.3 N-Gram Based	7

2.3.4 Jaro	8
2.4 Related Work	10
2.4.1 Improving Arabic Information Retrieval System	
Using N-Gram Method.	10
2.4.2 Effectiveness of Stemming and n-grams String Similarity	
Matching on Malay Documents	12
2.4.3 N-gram Statistics.in English and Chinese:	
Similarities and Differences	13
2.5 Proposed Solution	14
CHAPTER 3: RESEARCH APPROACH AND METHODOLOGY	15
3.0 Introduction	15
3.1 Methodology Overview	15
3.1.1 Requirement Planning	16
3.1.2 Design Activities	18
3.1.3 Development	22
3.1.4 Testing and Outcomes	23
3.2 Software and Hardware Requirement	23
CHAPTER 4: RESULT AND FINDINGS	25
4.0 Introduction	23
4.1 Data Collection	23
4.2 N-Gram	24
4.3 Searching Function	28
4.4 Prototype Interface	30
4.5 Experiment Results	32
CHAPTER 5: CONCLUSIONS AND FUTURE RESEARCH	33
5.0 Introduction	33