

Universiti Teknologi MARA

**2D Visualization of Terms and Documents
in Malay Language**

Nur Hamizah binti Mat Saad

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

JULY 2012

ACKNOWLEDMENT

Alhamdulillah, praise and thank to Allah because of His Almighty and His utmost blessings, I was able to finish this research within the duration given. Firstly, my special thanks go to my supervisor, Dr Normaly Kamal bin Ismail. He guided me to finish this thesis with patient. Without him, I would not have been able to proceed in developing the prototype smoothly. For my CSP650 (Project Formulation) lecturers, Dr. Sharifalillah Binti Nordin and Dr. Noor Elaiza Binti Abdul Khalid, thank you for the lesson and support given along this two semesters.

Special appreciation also goes to my beloved parents Mat Saad bin Bakar and
With the courage they gave me, I was able to finish
this thesis. I want to thanks to my sister who also send dua to Allah for me to
complete this project.

Last but not least, I would like to give my gratitude to all the lectures involved and my dearest friend who helps me doing research and discussion together. Not to forget, for those who participated in the survey for usability test, I would like to say thank you as my appreciating towards your cooperation.

ABSTRACT

In the technology era, information is just at our fingertip. Many information nowadays can be search in digitize way. The output of the data is still in the listed form. This linear (one dimension) form makes user hardly to find information related to the data requested. There are many way that the search data can be display. In addition, Malay document are still displayed in textual listed form. Thus, prototype of 2 dimension is develop by using Malay document. With this prototype, users are able to see the graphical relationship between relevant Hadith documents based on selection of term. TF*IDF method used to develop vector space model for term-document relationship. While Cosine Similarity technique use for document-document relationship. After the prototype has develop and able to run smoothly, the functionality and usability test is done. The discussion is done base on the survey form given to the user. Most of the users agree that functions in the prototype give many benefits. Moreover, they find that display information in 2 dimensional is more interesting as compare to linear form.

TABLE OF CONTENT

CONTENTS	PAGE
SUPERVISOR'S APPROVAL	iii
DECLARATION	iv
ACKNOWLEDGEMENT	v
ABSTRACT	vi
TABLE OF CONTENTS	vii
LIST OF FIGURES	xi
LIST OF TABLES	xiv
CHAPTER ONE: INTRODUCTION	
1.1 Project Background	1
1.2 Problem Statement	1
1.3 Objective	3
1.4 Scope	3
1.5 Significant	3
1.6 Summary	4
CHAPTER TWO: LITERATURE REVIEW	
2.1 Introduction	5
2.2 Documents and Terms in Malay Language	5
2.3 2-Dimensional Visualization	5

CHAPTER 1

INTRODUCTION

1.1 Project Background

Nowadays, the internet is used for a variety of things. About anything that users want to know, just type it in the internet browser, then several results will be display.

There are over 300 million peoples who converse in Malay language and considered as the fourth largest language group. Malay documents in the internet are still in limited range to be searched.

As computers have become widely used and powerful tool to produce pictures. There are many kinds of data sets can be used and effective visualization scheme depend on the characteristics of the data. Data sets can be distributed over a two-dimensional (2D) region of space, three-dimensional (3D) region or higher-dimensional space.

Definition of visualization is to form a picture of someone or something (Oxford dictionary, 2006). Visualization can be made through images, diagram or animation. 2D and 3D are included in visualization. In this project, 2D will be used to visualize the relationship between Malay documents.

1.2 Problem Statement

Malay languages is also known as Bahasa Malaysia. Documents based on Malay language can be searched on the web nowadays. But, there is the limitation whereby we are not guaranteed to find what we want. This shows that there was lacks somewhere. One of the problems stated by Amato, Lecce and Piuri (2007), is the user interface. In addition, as stated by Amato *et al.* (2007), somehow, the representation of the documents is still in textual sorted list. They consist of wordy in the