

**2D Visualization on Malay Articles**

**BY**

**SIDI BUKHARI BIN SIDI OMAR**

**BACHELOR OF COMPUTER SCIENCE (HONS)**

**THESIS SUBMITTED IN PARTIAL FULFILLMENT OF  
THE REQUIREMENT FOR THE DEGREE OF  
BACHELOR OF COMPUTER SCIENCE (HONS)**

**FACULTY OF COMPUTER AND MATHEMATICAL  
SCIENCES  
UNIVERSITI TEKNOLOGI MARA**

**MAY 2011**

# Acknowledgement

First of all, I would like to say Alhamdulillah, praise to Allah who has given me the strength and good health to do this project until it is done. Not forgotten, my family who has support me all the time. Then, I would like to thanks my friends who have given encouragement and ideas for me to complete this project. Not to mention, the community at sourceForge.net who have help me solving many problems that related to this project. Last but not least, I would like to thanks to my supervisor, Mr. Normaly Kamal Ismail. Without you, this project would have never been finished. You have given many support, ideas, and encouragement for me to complete this project.

# Abstract

Currently, browsing through web sites can only be done in a text form. Although there are some research has been made to browse in graphical form, it is not widely used. Hence, not many people know about it. If users browse through web sites that contains many Malay documents, then it would be in text form. Because of that, users cannot gain the overview of the relationship of those Malay documents. Hence, a prototype is built to overcome this problem. The Malay documents is collected and processed with cosine similarity technique to obtain the relationship between Malay documents. Then, the relationship is visualized where users can see the overview of the relationship of those Malay documents. The prototype was evaluated with usability questionnaire. Each question is then discussed and users agreed on some of it. Then it is concluded that users agreed that the prototype is much better than the linear system, the system that we compared to.

# Table of Contents

<b>DECLARATION</b>	<b>ii</b>
<b>ACKNOWLEDGEMENT</b>	<b>iii</b>
<b>ABSTRACT</b>	<b>iv</b>
<b>TABLE OF CONTENTS</b>	<b>v</b>
<b>LIST OF FIGURES</b>	<b>vii</b>
<b>LIST OF TABLES</b>	<b>ix</b>

<b>1. Chapter 1 – Introduction</b>	<b>1</b>
1.1 Problem Statement	2
1.2 Objective	3
1.3 Scope	3
1.4 Significant	3
<b>2. Chapter 2 – Literature Review</b>	<b>4</b>
2.1 Introduction	4
2.2 Related Research on Visualizing Data Relationship	4
2.2.1 Geomorphologic Metaphor	4
2.2.2 Multidimensional Scaling	5
2.2.3 Embedding Technique	6
2.2.4 Conclusion	6
2.3 Related Research on Visualizing Clustered Web	7
2.3.1 Interface Design	7
2.3.2 Agglomerative Approach	7
2.3.3 Topology-Driven Approach	8
2.3.4 Conclusion	9
2.4 Related Research on Information Visualization Toolkit	9
2.4.1 prefuse Toolkit	10
2.4.2 Conclusion	10
2.5 Related Research on Document Clustering	11
2.5.1 Document Clustering in Malay Document	11
2.5.1.1 Singular Value Decomposition	12
2.5.2 Document Clustering by Similarity Measures	12
2.5.2.1 Metric	13
2.5.2.2 Euclidean Distance	13
2.5.2.3 Cosine Similarity	14
2.5.2.4 Jaccard Coefficient	14
2.5.2.5 Other Similarity Measure Techniques	14
2.6 Related Research on Information Retrieval	15
2.6.1 Document Analysis Techniques	15
2.6.2 Information Retrieval in Web Search Engines	16
2.6.3 Information Retrieval Problems	16

# CHAPTER 1

## Introduction

### 1.1 Background

As technology has advanced, so do internet. This is because people can access more information on the internet now than the time when internet was started. When we speak about information, information can be obtained from any web source and it can be in any form. For example, information can be a picture, a sound, a video, or even a digital document. A digital document can be journal, article, newspaper, magazine, e-book, etc (Norzaidah *et al.*, 2009). Since there are a billion of information on the internet, it is impossible for people to seek what they intend to find. Therefore, search engines like Google and yahoo have become a very useful tool for people to use to find what they intend to find (Li *et al.*, 2007).

Over 300 million peoples spoke using Malay language and Malay language is number four in ranking on the most used language (Norzaidah *et al.*, 2009). Furthermore, there are about 118 institutions all over the world that teach and doing research on Malay language (Asmah, 1968). Moreover, Malay is a context-dependent language (Azhar, 1988). Hence, a Malay document is a document that most words it used is in Malay language (Norzaidah *et al.*, 2009).

Visualize is forming mental picture into something (Oxford English Mini Dictionary & Thesaurus, 2008, p.740) or forming a geometric or graphic representation from collection of data (“Visualize,”). From the definition above, we can say that visualize is a projection of an object to various form of graphic representation. For