

Universiti Teknologi MARA

**Indigenous Language Translation
Using Example-Based Approach**

Neriza Binti Nazar

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

July 2013

ACKNOWLEDGEMENT

In the name of Allah, the Most Gracious and the Most Merciful and Him alone is worthy of praise. Alhamdulillah, praise and thank to Allah because of His Almighty and His utmost blessings, I was able to finish this project within the time duration given.

Firstly, my special thanks and gratitude go to my supervisor, Associate Professor Dr. Nursuriati binti Jamil for her full support, patience, motivation and helpful knowledge. Her guidance helped me to think and perform better throughout this amazing journey. I am also indebted to my former supervisor Abdul Rahman bin Gobil for guiding me as well. Their cooperation in sharing information and guidance are the key to completing this thesis. My thanks are extended to my coordinator, Dr Suriyani Ariffin and Dr Noor Atiqah Sia Abdullah for their consultant and for their willingness in sharing their experience working with thesis.

Special appreciation also goes to my beloved parents and family that become my backbones who have supported me up to this moment, teaching me the meaning of life, and giving me all the love in the world. Thanks to them, I feel strong when I was feeling down and useless. They also are the motivation that triggers the idea of this project title.

Not to be forgotten my dearest friends and course mate for the simultaneous discussions, all the sleepless nights working together and for all the wild and fun vacations we had to release each other stress. It has been an honour to spend my once in a life time journey with all of you.

Thank you.

JULY 2013

NERIZA BINTI NAZAR

ABSTRACT

Language has been a very crucial communication medium since the existence of human. Thousands of language created ever since. However, not all of them survive the human history. Language is a great asset to a culture and it represents the identity of a community. Yunani and Greek are some example of dead language. Of course there is available human translator however the number is decreasing day by day. In fear of Suluk becoming a dead language, this research is conducted. The aim is to create a translation application to preserve this indigenous language as well as a learning tool for beginner to learn Suluk language. This can be achieved by identifying methods that can be used in translating Malay to Suluk language, developing the prototype based on selected method and testing the prototype. After conducting the preliminary study, the approaches considered for this project is statistical approach, example-based approach and direct translation approach. To produce more reliable result, example-based approach is chosen considering both languages (Suluk and Malay) have different word order. In fact, deep linguistic understanding is not required thus making the example-based approach the ideal method to be implemented. The end product of this project is a prototype that can be used to translate Malay to Suluk language. Translation is limited to direction commands only due to time constraint. This is a small step in helping the nation to preserve the heritage of our country.

TABLE OF CONTENTS

CONTENTS	PAGE
SUPERVISOR'S APPROVAL	ii
DECLARATION	iii
ACKNOWLEDGEMENT	iv
ABSTRACT	v
TABLE OF CONTENTS	vi
LIST OF FIGURES	ix
LIST OF TABLES	xi

CHAPTER ONE: INTRODUCTION

1.1	Background Of Study	1
1.2	Problem Statement/Motivation	2
1.3	Objectives	4
1.4	Scope Of Project	4
1.5	Research Significance	5
1.6	Overview Of The Thesis	5

CHAPTER TWO: LITERATURE REVIEW

2.1	Natural Language	6
2.2	Malay And Suluk Language Characteristic	7
2.3	Application Significant To Natural Language Processing (NLP)	8
2.4	Machine Translation	8
2.5	Machine Translation Methods/Approaches	9
2.5.1	Example-Based Approach	10
2.5.2	Statistical Approach	12
2.5.3	Direct Approach	14

2.6	Comparison Of Approaches	16
2.7	Summary	17

CHAPTER THREE: RESEARCH METHODOLOGY

3.1	Research Methods	18
3.2	Preliminary Study	20
3.3	Data Collection	20
3.4	Research Design	23
	3.4.1 System Development Design	25
3.5	Research Planning	27
3.6	Summary	29

CHAPTER FOUR: SYSTEM DESIGN AND IMPLEMENTATION

4.1	Example-Based Algorithm Design	30
4.2	Example-Based Algorithm Phase	33
	4.2.1 Exact Match Phase	33
	4.2.2 Segmentation Phase	34
	4.2.3 Combination Phase	35
4.3	Interface Design	35
4.4	Summary	36

CHAPTER FIVE: RESULT AND FINDING

5.1	Translation Results	37
	5.1.1 Exact Match Translation	38
	5.1.2 Variable+Phrase Translation	40
	5.1.3 Phrase+Variable Translation	41
5.2	Result Analysis	42
5.3	Summary	45