

**Universiti Teknologi MARA**

**Collection Of Malay Unverified Social  
Media Posts Using Crowdsourcing**

**AHMAD RUZAINI BIN AWANG AHMAD**

**Thesis submitted in fulfillment of the requirements  
for bachelor of Computer Science (Hons) CS230  
Faculty of Computer and Mathematical Science**

**February 2016**

## ACKNOWLEDGEMENT

Alhamdulillah, praises and thank to Allah S.W.T because of His Almighty and blessing, I was able to finish the report within the time. First of all, specials thanks to my supervisor, Miss Fadzlin Ahmadon as her brilliant idea and guidance make this project report complete smoothly. Besides, also thanks to CSP lecturer, Madam Syafaf Binti Ibrahim as her suggestion and idea also make this project more efficient.

Furthermore, appreciation to my parents, my father  
and my mother as giying me support in term of  
mental and physical.

Lastly, many thanks to the lecturers, friends, and my beloved classmate for sharing ideas and spending time for completing this report.

## ABSTRACT

Facebook is a social media platform that enables Facebook user to update status, pictures, videos and many more. As user can share and update their status or post on their Facebook newsfeed, therefore the information shared or posted is unreliable. The information shared or posted may be unverified and can lead to slander and misunderstanding. Therefore, this application will collect unverified Malay post using crowdsourcing. Crowds or Facebook users can send the unverified Malay post in their Facebook newsfeed through this application in the Facebook. Admin then will extract the keywords in the post using word repetition technique and save it into the database. Then script for the automated matching post to its categories will be run to store the new submitted posts to its type of category based on the list of keywords generated by the script. By that, this application may be able to help researchers, academicians or any other interested parties by providing a chart of what types of category most of users shared. Thus, the researchers or academicians are able to do more research or focus on the type of category that users shared and by having this information, the authorities and professionals in the related field can be involved to help solve the problem of unverified information being shared. This project is using Rapid Application Development (RAD) to design and develop the Facebook app and the web application.

## TABLE OF CONTENTS

<b>CONTENTS</b>	<b>PAGE</b>
<b>SUPERVISOR'S APPROVAL</b>	ii
<b>STUDENT'S DECLARATION</b>	iii
<b>ACKNOWLEDGEMENT</b>	iv
<b>ABSTRACT</b>	v
<b>TABLE OF CONTENTS</b>	vi
<b>LIST OF FIGURES</b>	x
<b>LIST OF TABLES</b>	xiv
<b>LIST OF ABBREVIATIONS</b>	xvii
<b>CHAPTER ONE: INTRODUCTION</b>	<b>1</b>
1.1 Background Study	1
1.2 Problem Statement	2
1.3 Objectives	2
1.4 Scope	3
1.5 Significance	3
1.6 Research Element	4
<b>CHAPTER TWO: LITERATURE REVIEW</b>	<b>5</b>
2.1 Unverified Information	5
2.2 Social Media	6
2.2.1 Facebook	7
2.2.2 Facebook Application Programming Interface	8
2.3 Sharing Information on Social Media	8
2.4 Collection of Information	12

2.5	Crowdsourcing	13
2.5.1	Approach of Crowdsourcing	15
2.5.2	Knowledge and Discovery Management (KDM)	17
2.6	Motivations for Crowdsourcing	19
2.7	Identifying Themes of Data	20
2.8	Repetitions	20
2.9	Visualization of Data	21
2.5	Summary	21
<b>CHAPTER THREE: METHODOLOGY</b>		<b>22</b>
3.1	Rapid Application Development (RAD)	22
3.1.1	Requirements Planning	25
3.1.2	Design	26
3.1.3	Construction	27
3.1.4	Testing and Implementation	28
3.2	Project Timeline	28
3.3	Summary	29
<b>CHAPTER FOUR: DESIGN</b>		<b>30</b>
4.1	Design Overview	30
4.2	Facebook Application Design Overview	35
4.2.1	Flowchart of Facebook Application	36
4.2.2	Use Case of Facebook Application	37
4.2.3	Design of Facebook Application	40
4.3	Keyword Extractor Program Overview	40
4.3.1	Flowchart of Keyword Extractor Program	42
4.3.2	Use Case of Keyword Extractor Program	43
4.3.3	Design of Keyword Extractor Program	50
4.4	Post Analyzer Program Overview	50
4.4.1	Flowchart of Post Analyzer Program	51