UNIVERSITI TEKNOLOGI MARA

GENDER IDENTIFICATION FROM FACEBOOK STATUS USING TERM CLASSIFICATION AND OCCURRENCE

Danna Majdiah Binti Anafiah

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

July 2013

ACKNOWLEDGEMENT

Alhamdulillah, praise to Almighty Allah S.W.T with His willing giving me the opportunity to complete my bachelor degree with this final year project which entitled Gender Identification From Facebook Status Using Term Classification and Occurrence. This thesis and project submitted as a partial fulfilment of the requirements for the Bachelor of Computer Science (Hons) under the Faculty of Computer and Mathematical Sciences, Universiti Teknologi Mara.

First of all, I would like to express my deepest gratitude to my current supervisor Prof. Madya Dr. Nursuriati Binti Jamil, my former supervisor Mr. Abdul Rahman Bin Gobil, my coordinator Dr. Suriyani Ariffin and my former coordinator Dr. Nur Atiqah Sia Abdullah for their full support throughout my research and giving their cooperation, valuable information and suggestions during the completion of my final year project and thesis.

I also owe my deepest thanks to my fellow friends who have been directly or indirectly contributed throughout this academic exploration, supporting my work and helped me in my project progress until it is fully completed.

Last but not least, my appreciation goes to my parents, my boyfriend, my siblings and other family members for their unconditional continuous love and encouragement during the last two years until my final year of study.

Thank You.

ABSTRACT

Computer technology has caused a big impact in our daily life. People nowadays mostly socialize through social networks since it is efficient and the fastest way to communicate regardless geographical boundaries. However, social networks have brought some drawbacks and disadvantages to the society. Some irresponsible people use social networks to do crimes and violation to other social network's users by faking their true identities including their name, age and gender. Therefore, this research will focus on gender identification from Facebook status posted and how gender can be classified through the occurrence and frequency of language style used by male and female users. This research focused on Facebook status in Malay language only due to time constraint and scope limitation of the project. The preliminary result of the research showed that different gender have different language and writing style where women tend to use more emotion and question words while men tend to use more opinion and insult words. This research will use the term frequency to identify whether the status posted on the Facebook is written by male or female user. As a conclusion, the product of this research most probably could help to reduce the identify fraud in social networks.

TABLE OF CONTENTS

CHAPTER 3: METHODOLOGY

CHAPTER 4 : SYSTEM DESIGN AND IMPLEMENTATION

CHAPTER 5 : RESULTS AND FINDINGS