

**INFORMATION EXTRACTION FROM SOCIAL MEDIA
TWITTER USING NATURAL LANGUAGE PROCESSING FOR
ANDROID MOBILE APPLICATION**

By

Sri Krisna Endarnoto

A Thesis submitted to the Faculty of

INFORMATION TECHNOLOGY

Department of
INFORMATION TECHNOLOGY

In Partial Fulfillment of the Requirements for

BACHELOR'S DEGREE

IN

INFORMATION TECHNOLOGY

SWISS GERMAN UNIVERSITY

Swiss German University
EduTown BSDCity
Tangerang 15339
INDONESIA

Telp. +62 21 3045 0045
Fax. +62 21 3045 0001
E-mail: info@sgu.ac.id
www.sgu.ac.id

July 2011

Revision after the Thesis Defense on July 26th 2011

STATEMENT BY THE AUTHOR

I hereby declare that this submission is my own work and to the best of my knowledge, contains no material previously published or written by another person, nor material which to a substantial extent has been accepted for the award of any other degree or diploma at any educational institution, except where due acknowledgement is made in the thesis.

Sri Krisna Endarnoto

Date

Approved by:

Anto Satriyo Nugroho, Dr. Eng.

Date

James Purnama, M. Sc

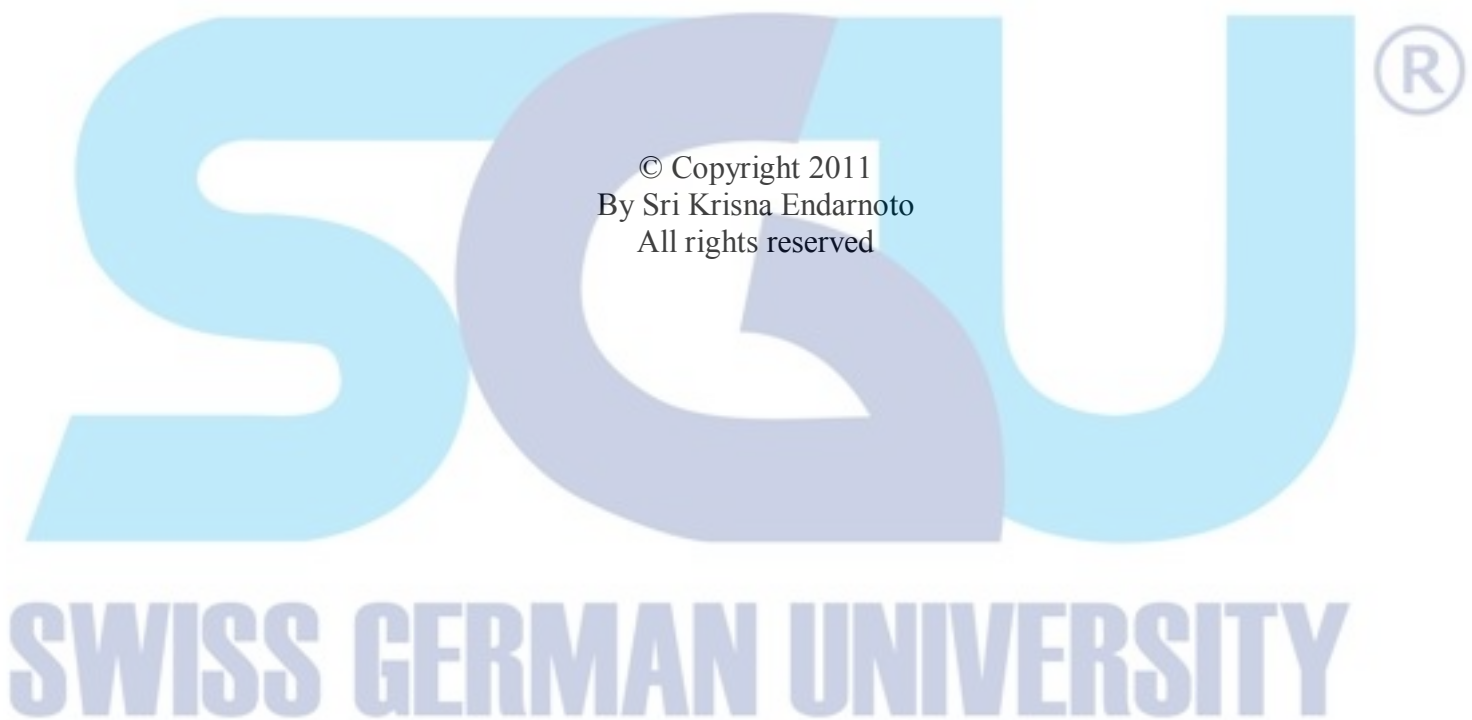
Date

Chairman of the Examination Steering Committee

Date

Sri Krisna Endarnoto

COPYRIGHT



ABSTRACT

INFORMATION EXTRACTION FROM SOCIAL MEDIA TWITTER USING NATURAL LANGUAGE PROCESSING FOR ANDROID MOBILE APPLICATION

By

Sri Krisna Endarnoto

SWISS GERMAN UNIVERISTY

Bumi Serpong Damai

Anto Satriyo Nugroho, Dr. Eng, Major Lecturer

Traffic jam in Jakarta has become a crucial problem for society. A Traffic Management Center has been built by Polda Metro Jaya, police unit in Jakarta, to help people to get the latest information of traffic. Twitter has been used to spread the news of traffic by them. With its limitation, Twitter doesn't provide good user interface in the case of traffic condition report. The main objective of this project is to develop a system that can extract information from TMC's Twitter to be presented in a map view by using Google Map and implement it in Android-based mobile application. Natural Language Processing can be used for information extraction. A tweet will be tokenized, each token will be assigned to a particular part-of-speech tag while analyzing the sentence by using rule based approach. Based on the rules, information of traffic can be extracted in the form of template which consist of time, origin, destination and condition. Google Map will present those information in 3 different colors for 3 different levels of traffic. Thus, providing the society an alternative to get traffic information from a reliable source with good user interface. Early experiment with limited vocabulary and rules has showed promising result.

Keywords : Natural Language Processing, information extraction, mobile application

DEDICATION

I dedicate this thesis to all the victims of traffic jam.



ACKNOWLEDGMENTS

The author wishes to thank my one and only God, Allah SWT for His blessing and guidance that are the source of my health, strength, idea, patience and other good things that led me to finish this chapter of my life called Bachelor Thesis.

Secondly, for the best role model for human to follow, Muhammad SAW. His way of life has inspired me so I can do my best for life, especially for this thesis.

Many great appreciation is presented to Anto Satriyo Nugroho. He is maybe the most brilliant man I've ever met. Both academically and personally. I don't know what to say but thank you. For every sacrifice that he made, for every laugh and pain, for everything. Also to James Purnama who has guided me right from the beginning with his great ideas and friendliness.

Thanks to Franciscus Chandra Pawitra, Rinto Priambodo and Rindang that have helped me and Sonny in Android development.

The thesis will not be completed without full support from my parents, Hartono and Srie Roosdarin and my sister, Sri Rosalina. Also to Sonny Pradipta who has been a great partner so far.

Prasetyo Eko Sulaksono, Nasrul Effendi, Pujas Leksono and Arandityo Narutomo for being very helpful in the difficult time, and being such a great friends.

And all of my friends that have been supporting me the whole time. To all of you, please accept my sincerest thanks.

TABLE OF CONTENTS

STATEMENT BY THE AUTHOR	2
COPYRIGHT	3
ABSTRACT	4
DEDICATION	5
ACKNOWLEDGMENTS	6
CHAPTER 1 – INTRODUCTION.....	11
1.1 Background.....	11
1.2 Objective	11
1.3 Research Problem	12
1.4 Scope of Analysis.....	12
1.5 Limitations	12
1.6 Methodology	14
CHAPTER 2 – LITERATURE REVIEW	15
2.1 Natural Language Processing.....	15
2.2 Android	19
2.3 Social Networking	22
2.4 Twitter.....	23
CHAPTER 3 – METHODOLOGY.....	24
3.1 Application Overview.....	24
3.2 Design.....	25
3.3 Agni Information Extraction Engine	43
3.4 Visualization	50
3.5 Testing System.....	53
CHAPTER 4 – RESULT & DISCUSSION.....	54
4.1 List of Infrastructure.....	54
4.2 Test Result	54
CHAPTER 5 – CONCLUSION AND FUTURE WORK.....	63
5.1 Conclusion	63
5.2 Future Work.....	63

GLOSSARY.....	64
REFERENCES.....	66
LIST OF PUBLICATIONS	68
APPENDIX I – PUBLISHED WORK.....	69
APPENDIX II – COMPONENT TEST CASE.....	74
APPENDIX III – SURVEY RESULT	76
APPENDIX IV – SAMPLE TWEETS.....	77
APPENDIX V – CONDITION CLASSIFICATION.....	79
CURRICULUM VITAE.....	80

