AUTOMATED DOCUMENT CLASSIFICATION FOR NEWS ARTICLE IN BAHASA INDONESIA BASED ON TERM FREQUENCY INVERSE DOCUMENT FREQUENCY (TF-IDF) APPROACH

By

Ari Aulia Hakim
12110005

A thesis submitted to the Faculty of

ENGINEERING AND INFORMATION TECHNOLOGY

in partial fulfillment of the requirements
for the
BACHELOR'S DEGREE
in

INFORMATION TECHNOLOGY

SWISS GERMAN UNIVERSITY
EduTown BSD City
Tangerang 15339
Indonesia

Revision after the Thesis Defense on 15 July 2014

STATEMENT BY THE AUTHOR

I hereby declare that this submission is my own work and to the best of my knowledge, it 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.

Ari Aulia Hakim
_____ _____
Student Date

Approved by:

Mr. Alva Erwin, MSc
_____ _____
Thesis Advisor Date

Dipl.-inf. Kho I Eng
_____ _____
Thesis Co-Advisor Date

Dr. Ir. Gembong Baskoro, M.Sc

_____ _____

Dean Date

ABSTRACT

AUTOMATED DOCUMENT CLASSIFICATION FOR NEWS ARTICLE IN BAHASA INDONESIA BASED ON TERM FREQUENCY INVERSE DOCUMENT FREQUENCY (TF-IDF) APPROACH

By

Ari Aulia Hakim
Alva Erwin, Advisor
Kho I Eng, Co-Advisor

SWISS GERMAN UNIVERSITY

The exponential growth of the data both in the digital or printed media may lead us to the information explosion era, where most of the data cannot be maintained easily. The research in the text mining might prevent the world to enter that era. One of the text mining studies that can help in maintaining the data is automated text classification. This research can classify one or more articles based on predefined categories. Automated text classification can be considered important, due to the big number of the data exist, and text classification may not be handled manually, because it will consume a lot of time and human resource. Then, the classifier developed by implementing term frequency inverse document frequency (TF-IDF).

Keywords: Text Mining; Automated Document Classification; TF-IDF;

DEDICATION

I dedicate this works for the future of the country I loved: Indonesia

ACKNOWLEDGEMENTS

Thanks to Allah for giving me the power and help to accomplish this research. Without the grace of Allah, I was not able to accomplish this work.

I would like to thank my parents very much for their pray, patience, motivation, and continues support. I also extend my thanks to Disya Oktaviani, Avin Mohanza Kasim, Harmando Taufik Gemilang, Kafin, Muhammad Fadhil Mudjiono, Muhammad Ghaffar Adipridhana, Norman Jaya Subrata, Richard Kartiyanta, Sayid Ali Hadi and all of my family members for their motivation and support.

I am grateful to my supervisor, Mr. Alva Erwin, ST. MSc for his enormous support, valuable guide, and assistance throughout the work of this research.

Special thanks for Dipl.-inf. Kho I Eng for his guide and comments.

TABLE OF CONTENTS

TABLE OF FIGURES

CHAPTER - 2

LIST OF TABLES

CHAPTER - 2