ASPECT BASED OPINION MINING OF HOTEL REVIEWS USING APRIORI ALGORITHM FOR FREQUENT ASPECT FINDING

By

Mochammad Athariq Kanz Hutama
11302019

BACHELOR'S DEGREE
in

INFORMATION TECHNOLOGY
FACULTY OF ENGINEERING AND INFORMATION TECHNOLOGY

SWISS GERMAN UNIVERSITY

SWISS GERMAN UNIVERSITY
The Prominence Tower
Jalan Jalur Sutera Barat No. 15, Alam Sutera
Tangerang, Banten 15143 - Indonesia

July 2017

Revision after Thesis Defence on 20 July 2017

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.

Mochammad Athariq Kanz Hutama

Student _____ Date _____

Approved by:

Alva Erwin, ST., M.Sc., MTI

Thesis Advisor _____ Date _____

Dipl. –Inf. Kho I Eng

Thesis Co-Advisor _____ Date _____

Dr. Ir. Gembong Baskoro, M.Sc.

Dean _____ Date _____

SWISS GERMAN UNIVERSITY

ABSTRACT

ASPECT BASED OPINION MINING
ON HOTEL REVIEW
USING APRIORI ALGORITHM FOR FREQUENT ASPECT FINDING

By

Mochammad Athariq Kanz Hutama
Alva Erwin, ST., M.Sc., MTI, Advisor
Dipl. –Inf. Kho I Eng, Co-Advisor

SWISS GERMAN UNIVERSITY

The emergence of Web 2.0 leads to the development of online review, such as product and hotel review, where the customer can freely write their opinion towards the hotel. Consumer opinion is the important source of information for people when deciding to book the hotel. However, with the increasing amount of review, it is making impossible for people to read all of the review. Additionally, it is also hard for hotel management to monitor the consumer opinion.

Aspect Based Opinion Mining is a study of people's opinion and the aspect towards the entities. This thesis is a research for finding the product aspect and the opinion polarity from the hotel review. There are 3 steps to perform the research: (1) Identifying frequent product aspect using apriori algorithm; (2) Identifying polarity of the review using SentiWordNet and CoreNLP; (3). The Precision and recall for predicting aspect and opinion pair is averaging 0.57 and 0.64 respectively, with the average F-Measure of 0.6, and can generate average of 75% of relevant frequent aspects. Using SentiWordNet and CoreNLP has high accuracy for predicting the polarity of the sentiment, with the average of 0.86 and 0.8 respectively.

Keywords: Text Mining, Sentiment Analysis, Natural Language Processing, Opinion Mining.

© Copyright 2017
by Mochammad Athariq K.H
All rights reserved

SWISS GERMAN UNIVERSITY

DEDICATION

"Dedicated to my beloved parents for their effort
To make me able to achieve this far"

SGU®
SWISS GERMAN UNIVERSITY

ACKNOWLEDGEMENTS

I would like to give my huge appreciation to my family for their support fulfilling and giving me everything to complete this Bachelor Thesis.

Huge appreciation for the Advisor and Co-Advisor, Mr. Alva Erwin and Mr. Kho I Eng for their time and advice during the process of this research.

Lovely thanks to my brother Nara, and Yudhistira for their effort and big help on fulfilling this thesis research. Also for accompany me during the making of this thesis.

Huge thanks to all my friends, for your support and help. Even though I can't write all the name, the appreciation is for all of you.

SGU
SWISS GERMAN UNIVERSITY

Regards,

Mochammad Athariq Kanz Hutama

TABLE OF CONTENTS

Page

STATEMENT BY THE AUTHOR 2

ABSTRACT 3

DEDICATION 5

ACKNOWLEDGEMENTS 6

TABLE OF CONTENTS 7

LIST OF FIGURES 10

LIST OF TABLES 11

CHAPTER 1 - INTRODUCTION 13

1.1. Research Background 13

1.2. Research Problem 15

1.3. Research Objective 15

1.4. Significance of Study 15

1.5. Research Question 16

1.6. Research Hypothesis 17

1.7. Research Limitation 17

CHAPTER 2 – LITERATURE REVIEW 18

2.1. Opinion Mining 18

2.2. Aspect Based Opinion Mining 19

2.2.1. Core Task in Aspect Based Opinion Mining 19

2.2.2. Opinion Word Finding 20

2.2.3. Aspect Extraction 20

2.3. Pre Processing 21

2.3.1. Tokenization 21

2.3.2. Lemmatization 21

2.3.3. Part-Of-Speech Tagging 22

2.4. Apriori Algorithm 24

SWISS GERMAN UNIVERSITY

2.5. Evaluation Metrics 25
2.6. Related Works 27
2.6.1. Feature Extraction depending on frequency of noun and noun phrases 27
2.6.2. Using Most Relevant Frequent Cluster of Terms for Aspect Opinion Mining in Microblogs 27
2.6.3. Unsupervised Opinion Mining from Unstructured Review 28
2.6.4. Sentiment Classification 28
CHAPTER 3 – RESEARCH METHODS 31
3.1. Data Sets 32
3.2. Pre-Processing and POS Tagging 32
3.2.1. Tokenization 32
3.2.2. Lemmatization 33
3.2.3. Part-Of-Speech Tagging 34
3.3. Opinionated Sentence Finding 35
3.4. Frequent Aspect Extraction 35
3.5. Sentiment Polarity Classification 37
3.5.1. SentiWordNet Approach 37
3.5.2. CoreNLP Recursive Deep Model 38
3.6. Aspect Opinion Summarization 38
3.7. View Web Interfaces 39
CHAPTER 4 – EXPERIMENTAL RESULTS AND DISCUSSIONS 40
4.1. Experiment Setup 40
4.1.1. Hardware Environment 40
4.1.2. Software Environment 40
4.1.3. Data Sets 41
4.2. Experiment Result 42
4.2.1. Experiment #1 Grand Hyatt 42
4.2.2. Experiment #2 Fairmont Jakarta 43
4.2.3. Experiment #3 Tentrem Yogyakarta 44
4.2.4. Experiment #4 W Bali 46
4.2.5. Experiment #5 Eco Stay Bali 47
4.3. Result Performance and Discussion 48

SWISS GERMAN UNIVERSITY

4.3.1. Evaluation Metrics **49**
4.3.2. Generated Aspect – Opinion Evaluation **50**
4.3.3. Sentiment Polarity Evaluation **51**
4.3.4. Result Discussion **53**
CHAPTER 5 – CONCLUSION AND RECOMMENDATION **57**
5.1. Conclusion **57**
5.2. Future Works **58**
GLOSSARY **59**
REFERENCES **60**
APPENDICE **63**
Curriculum Vitae **68**

SGU
SWISS GERMAN UNIVERSITY