

Opinion Mining Approach to Improve Hotel Booking Process

By A.R.F Jawhara

149201C

**Faculty of Information Technology
University of Moratuwa
May 2017**

A.R.F Jawhara
149201C

Dissertation submitted to the Faculty of Information Technology, University of Moratuwa, Sri Lanka for the partial fulfillment of the requirements of the Master of Science in Information Technology

May 2017

Declaration

I declare that this thesis is my own work and has not submitted in any form for another degree or diploma at any university or other institution of tertiary education. Information derived from published or unpublished work of others has been acknowledged in the text and a list of list of references is given.

Name of Student

Signature of student

A.F Jawhara Rahuman

Date:

.....

Supervised by

Name of Supervisor

S.C. Premarathne

Senior Lecture

Faculty of Information Technology

University of Moratuwa

Signature of Supervisor

Date:

.....

Acknowledgements

I would like to express my gratitude to my supervisor Mr. S.C Premarathne, Senior Lecturer in University of Moratuwa, whose expertise, understanding added considerably to my research experience. I appreciate his vast knowledge and his skills in many areas and his guidance to writing report.

I would like to thank the other Lecture of University of Moratuwa, Prof. Asoka Karunananda for the knowledge and assistance provided by all level of the research project.

Not only that my thank should go to all the lectures of M.Sc. in Information technology degree program of Faculty of IT, who sharpens our knowledge and ideas throughout this M.Sc. Program

Moreover I would also like to thank all the batch mates of the M.Sc. in IT degree program who gave their valuable feedback to improve the result of research and my family for the support they provided me. I must acknowledge my husband and my best friend for the encouragement.

Abstract

Customer Opinions play a very crucial role in daily life. When we have to take a decision, others opinion also considered. The e-commerce field has developed to the point that more and more hotel companies provide online booking services to travellers as an integral part of their business model. Increasing numbers of hotel companies now provide such services as an integral part of their business model and their guests' experiences with their hotel. Some third-party services allow customers to add comments on each hotel at the affiliated website. The proposed tool features at hotel websites are based on fixed properties, allowing companies to take advantage of the huge number of available customer reviews to provide relevant information to consumers considering new services.

In this thesis we are going to see how Sentiment analysis tool is working used for mining reviews from online reviews those are posted by customers. How Apriori frequent item set mining algorithm can be used for find associate hotel features. Our main theme is to create a system for analyzing opinions which implies judgment of different consumer products.

This research should improve online hotel booking by building a customized tool that utilizes available customer reviews at the Trip Advisor website. In this thesis we are going to see how

Contents

Declaration.....	iii
Acknowledgements	iv
Abstract	v
List of Figures	x
List of Tables	xi
Chapter 1	1
Introduction	1
1.1 Prolegomena.....	1
1.2 Objective	2
1.3 Motivation	2
1.4 Proposed solution	3
1.5 Resource Required	4
1.6 Summary of chapters	4
1.7 Summary	5
Chapter 2	6
Development Challengers in Data Mining.....	6
2.1 Introduction.....	6
2.2 Early developments & Modern trends in Data mining	6
2.4 Future challenges.....	8
2.5 Summary	8
Chapter 3	9
Technology Adapted.....	9
3.1 Introduction.....	9
3.2 Tools & Technologies of Hotel Ranking system.....	9
3.3 Data Mining	9
3.3.1 Opinion Mining.....	10
3.3.2 Semantic Analysis.....	11
3.3.3 Sentiment Analysis or Sentiment Classification.....	11
3.4 Web Scraping.....	12

3.5. Xampp server.....	12
3.6 HTML.....	13
3.7 Hardware Requirements.....	14
3.8 Summary.....	14
Chapter 4.....	15
4.1 Introduction.....	15
4.1.1 Hypothesis.....	15
4.1.2 Users.....	15
4.1.3 Input.....	15
4.1.4 Output.....	16
4.1.6 Features.....	16
4.2 Methodology.....	16
4.2.1 Data Collection.....	17
4.2.2 Semantic Analysis.....	17
4.2.3 Aspect Oriented Customer Review Mining.....	18
4.2.4 Method of storing data to Database.....	20
4.2.5 Opinion Mining/Sentiment Analysis Tool.....	20
4.2.6 Natural Language Processing.....	21
4.2.7 Association Mining.....	21
4.3 System Architecture.....	22
4.4 Front End.....	22
4.5 Back End.....	23
4.9 Summary.....	23
Chapter 5.....	24
Design of IHBP.....	24
5.1 Introduction.....	24
5.2 General Architecture.....	24
5.2.1 Data Collection module.....	26
5.2.2 Semantic Analysis module.....	26
5.2.3 Review Data base module-Data set 1.....	27
5.2.4 Sentiment Analysis module.....	27

5.2.5 Data set-2 module.....	28
5.2.6 My Code module.....	28
5.2.7 Front End module	28
5.3 Summary	29
Chapter 6	30
Implementation of IHBP	30
6.1 Introduction.....	30
6.2 Overall Solution based	30
6.3 Data Collection	30
6.4 Feature Identification using Semantic Analysis.....	32
6.5 Implementation of Review Database	34
6.6 Sentiment Analysis Module.....	35
6.7 Implementation of MyCode Module	36
6.8 Implementation of the Frontend Module.....	38
6.9 Summary	39
Chapter 7	40
Evaluation	40
7.1 Introduction.....	40
7.2 Evaluation for Front End	40
7.3 Evaluation for Opinion Sentiment classifier - NLTK 2.0.4	41
7.4 Evaluate for Hotel Ranking.....	42
7.6 Summary	59
Chapter 8	60
Conclusion.....	60
8.1 Introduction.....	60
8.2 Achievement of Objective.....	60
8.3 Further work.....	61
8.4 Summary	61
References	62
Appendix A.....	63
Structure of Hotels Table in Database.....	64
Structure of Location Table in Database.....	64

Structure of Reviews Table in Database	65
After sentiment Analysis filtered only positive hotel features	65
Appendix B.....	66
Cording of Data base creation.....	67
Test script of efficiency of Sentiment Analysis.....	68
Apriori Test script	71
Hotel Ranking System/ MyCode PHP	73
Appendix C.....	76
Sample of User's survey feedback	

List of Figures

Figure 1: Aspect Identification	19
Figure 2: Top Level Architecture	25
Figure 3: Architecture of My SQL Database.....	27
Figure 4: Architecture of Front End.....	28
Figure 5: Sample of collected Reviews- Unstructured Data	32
Figure 6:QDA Miner lite interface	33
Figure 7: Analyzed Data-Structured Data	34
Figure 8: -Structure of Database	34
Figure 9:Front End Interface	38
Figure 10:Front End- list of location	40
Figure 11:Front End-Hotel Feature List	41
Figure 12:-Result of test script ID-122.....	42
Figure 13:Hotel Rank based on selected feature	43
Figure 14:Kandy Earl’s Regency hotel’s TripAdvisor user review screen	44
Figure 15:Kandy Earl’s Regency hotel’s HRS in Ranking	45
Figure 16: Colombo Hilton Hotels TripAdvisor user review screen	46
Figure 17:Colombo Hilton hotels HRS in Ranking	47
Figure 18:Survey Sample	48
Figure 19:HRS’s Overall performance chart.....	49
Figure 20: Apriori Test	58
Figure 21: Association Test	58
Figure 22: Test Result of Association.....	59
Figure 23: Structure of Hotels Table in Database	64
Figure 24: Structure of Location Table in Database	64
Figure 25: Structure of Reviews Table in Database.....	65
Figure 26: Structure of Reviews Table in Database.....	65
Figure 27: Survey of user 1	78
Figure 28: Survey of user 2	79
Figure 29: Survey of user 3	80
Figure 30: Survey of user 4	81
Figure 31: Survey of user 5	82
Figure 32: Survey of user 6	83
Figure 33: Survey of user 7	84
Figure 34: Survey of user 8	85

List of Tables

Table 1: User Feedback.....	49
Table 2: Transaction Table.....	51
Table 3: Table C1	51
Table 4: Table L1	52
Table 5: Table C2	53
Table 6: Table L2	54
Table 7: Table C3	55
Table 8: Table L3	56
Table 9: Table C4	56