

Sentiment Analysis of Hotel Service System

By

Kwang Chyun Yaw

A PROPOSAL

SUBMITTED TO

Universiti Tunku Abdul Rahman

in partial fulfillment of the requirements

for the degree of

BACHELOR OF BUSINESS INFORMATION SYSTEM (HONS)

Faculty of Information and Communication Technology

(Perak Campus)

May 2015

UNIVERSITI TUNKU ABDUL RAHMAN

REPORT STATUS DECLARATION FORM

Title: Sentiment Analysis of Hotel Service System

Academic Session: May 2015

I **KWANG CHYUN YAW** _____

declare that I allow this Final Year Project Report to be kept in

Universiti Tunku Abdul Rahman Library subject to the regulations as follows:

1. The dissertation is a property of the Library.
2. The Library is allowed to make copies of this dissertation for academic purposes.

Verified by,

(Author's signature)

(Supervisor's signature)

Address:

No 116, LRG CENDAWAN 3/4

TMN BAYAM, LUNAS

KEDAH

Supervisor's name

Date: _____

Date: _____

Sentiment Analysis of Hotel Service System

By

Kwang Chyun Yaw

A PROPOSAL

SUBMITTED TO

Universiti Tunku Abdul Rahman

in partial fulfillment of the requirements

for the degree of

BACHELOR OF INFORMATION TECHNOLOGY (HONS)

BUSINESS INFORMATION SYSTEM

Faculty of Information and Communication Technology

(Perak Campus)

May 2015

DECLARATION OF ORIGINALITY

I declare that this report entitled “Sentiment analysis of Hotel Service System” is my own work except as cited in the references. The report has not been accepted for any degree and is not being submitted concurrently in candidature for any degree or other award.

Signature: _____

Name : _____

Date : _____

Acknowledgement

To complete this final year project, I would like to sincere thanks to person who had helped me to solve the problems. Firstly, my supervisor Mr. Albert Yong Tien Fui had given many ideal and guided me in order to complete the project. Also, I would like sincere thanks to my friend who had gave suggestion and opinion to complete the project especially Loh Keh Hsin. Moreover I would like to thanks to teacher and lecture which are MS Zao May , Ms Aruna Raj , MS. Su Ling, MS. Bryce Kwang, MR. Mohd Asri, Ms. Ze Chian and MS. Wei Joo who participate to do the survey form for rating the 32question.

ABSTRACT

This project is developing a system which called sentiment analysis of hotel service system that used to analyze customers comment about hotel service on website. The objectives and scopes of this project are analyze how sentiment analysis detects positive or negative comment of the sentence and show sentiment result. In additional, this project is develop a system that able to access other websites to read customer comments and analyze sentiment sentences and show result to user. The existing methods to perform sentiment analysis are semantics analysis, Nasent algorithm, simple based-technology and text-parser method, pointwise mutual information, ES and NTCIR.

The problem of the sentiment analysis is not able to create word dictionary automatically which require user create manually. Moreover, most of sentiment analysis does not focus on analyze other hotel website's comment and accuracy of sentiment result may decreased when the sentence are too long. To solve these problems, there are few ways to solve it which are automate create the word dictionary, access certain website to get all comment and analyze it.

The methodology to perform sentiment analysis is break a long paragraph into sentence then spot the keyword such as "impressive" word to classify the word whether positive or negative and evaluate overall sentence is positive or negative comment about hotel service. Then only calculate for the result.

Evaluation part for testing the system accuracy is request few people who are good in English such as English teacher to evaluate few sentences in the prepared survey form that customer comment about the hotel service and also send the sentences to system to evaluate it.

Lastly, compare the analysis result which formula calculation is almost similar between human and the system result by using correlation method.

Table of Content

DECLARATION OF ORIGINALITY	i
ACKNOWLEDGEMENT	ii
ABSTRACT	iii
TABLE OF CONTENT	iv-vi
LIST OF FIGURES	ix-xi
LIST OF TABLE	xii
LIST OF ABBEVIATION	xiii
CHAPTER 1 INTRODUCTION	1
1-1 Project background	1-2
1-2 Problem Statement	3
1-3 Project Objectives	3
1-4 Expected Innovation	3
1-5 Detail of Expected Innovation	4
CHAPTER 2 LITERAL REVIEW	5
2-1 Literal Review	5
2-1-1 Semantics Analysis	5-7
2-1-2 Nasent Algorithm	7-8
2-1-3 Simple Based-keyword Technology	9-10
2-1-4 Text Parse	10

2-1-5 Naive Bayes Theorem	10-12
2-1-6 Pointwise Mutual Information	12-13
2-1-7 NTCIR	13-14
2-1-8 Enhance Sentiment	14-15
2-2 Fact Finding	15
2-2-1 Research	15
2-2-2 Data Collection	15-16
2-3 Evaluation Planning	16
CHAPTER 3 PROPOSED METHOD/ APPORACH	17
3-1 Methodology	17
3-1-1 Waterfall Model	17-18
3-1-2 V-Model	18-19
3-1-3 Incremental Model	20
3-2 Selecting Software Process Model	21
3-3 Sentiment Analysis Methodology	21
3-4 Tool to use	21-22
3-4-1 C#	22-23
3-4-2 Text File	23
3-5 Requirement Specifications	24
3-5-1 System Requirements	24
3-6-2 System Performance Definition	24
3-6 Timeline	25

3-6-1 Project I Schedule	25
3-6-2 Project II Schedule	25
CHAPTER 4 SYSTEM DESIGN	26
4-1 Block Diagram	26
4-2 Use Case Diagram	27
4-3 Activity Diagram	28
4-3-1 Analyze sentence activity diagram	28
4-3-2 Create word dictionary activity diagram	29
4-3-3 Analyze Hotel Comment	30
4-4 Class Diagram	31
4-5 Database Design	32
4-5-1 Sentiment Word Database	32
4-5-2 Hotel Comment Database	33-34
4-5-3 Hotel Name & Link Database	34-35
4-5-4 Staff ID & Password	35
4-6 System Graphic User Interface	36
4-6-1 Login Interface	36
4-6-2 Main Menu Interface	36-37
4-6-3 Word List Interface	37
4-6-4 Analyze Sentence Interface	38
4-6-5 Analyze Hotel Comment Interface	38-39

4-6-6 Edit User Interface	39
CHAPTER 5 IMPLEMENTATION AND TESTING	
5-1 Sentiment Value Calculation	40
5-2 Evaluation Result	41-46
5-3 Challenges	46
CHAPTER 6 Conclusions	47
6-1 Conclusion	47
6-2 Problem Encountered and Solution	48
6-2-1 System can't access network with proxy setting	48
6-2-2 Error may occur when network speed slow	48
6-3-3 Internet Explorer browser may pop out	48
6-3 Discussion	49
6-4 Future Work	49
REFERENCE	50-51
APPENDIX A TEXT PARSER TOOL	A-1
A-1 Sample Survey Form	A-1-A-6
APPENDIX B FINAL YEAR PROJECT WEEKLY REPORT	B-1
C-1 Final Year Project II Weekly Report	B-1-B-11
APPENDIX C SURVEY FORM RESULT	C-1

List of Figures

Figure Number	Title	Page
Figure 1.1	Example of system get the keyword by searching “<P> <CENTER>”	4
Figure 2.1	Example of Sentiment sentence	5
Figure 2.2	Example of after break sentiment sentence	6
Figure 2.3	Example of evaluate each sentiment sentence	6
Figure 2.4	Topics break down diagram	6
Figure 2.5	NaSent analyze sentence: “Unlike the surreal Leon, this movie is weird but likeable.”	7
Figure 2.6	NaSent analyze sentence: “Unlike the surreal but likeable Leon, this movie is weird.”	8
Figure 2.7	Example of simple keyword-based methodology scoring system	9
Figure 2.8	Parse Tree	10
Figure 3.1.1	Waterfall Model Diagram	17
Figure 3.1.2	V-Model Diagram	19
Figure 3.1.3	Incremental Model Diagram	20
Figure 3.1	Flow Diagram of sentiment analysis methodologies for sentiment analysis.	21
Figure 4.1	Block Diagram of Sentiment Analysis	26
Figure 4.2	Use case diagram of sentiment analysis	27

Figure 4.3.1	Activity diagram of analyze sentiment sentence	28
Figure 4.3.2	Activity diagram of create word sentiment dictionary	29
Figure 4.3.3	Activity diagram of analyze hotel comment	30
Figure 4.4	Class Diagram of System	31
Figure 4.5.1	shows PositiveWord.txt and Negative.txt data file	32
Figure 4.5.2	shows short length of sentence in Short.txt	33
Figure 4.5.4	shows long length of sentence in Long.txt	34
Figure 4.5.5	shows hotel name & address link in HotelName.txt	35
Figure 4.5.6	shows staff id & password store in StaffPass.txt	35
Figure 4.6.1	Login Interface	36
Figure 4.6.2	Main Menu Interface	37
Figure 4.6.3	WordList Interface	37
Figure 4.6.4	Analyze a Sentence Interface	38
Figure 4.6.5	Analyze Hotel Comment Interface	39
Figure 4.6.6	Edit user Interface	39
Figure 5.1	shows how to calculate sentiment value for the word.	40
Figure 5.2	shows assign value for each sentiment word.	41
Figure 5.3	shows human rate for 32 questions in survey form.	42
Figure 5.4	shows all formula and human average result.	43
Figure 5.5	shows correlation result between formula and human.	44

Figure 5.6	shows all formula and human average result after recognize intensifier grammars.	45
Figure 5.7	shows correlation result between formula and human after recognize intensifier grammars.	45

LIST OF TABLE

Table 3.1 Gantt Chart - Final Year Project 1	25
Table 3.2 Gantt Chart - Final Year Project 2	25

LIST OF ABBREVIATION

SA Sentiment Analysis

NTCIR National Institute of Informatics Test Collection Information Retrieval

PMI Pointwise Mutual Information

ES Enhance Sentiment

CHAPTER 1 INTRODUCTION

Chapter 1 Introduction

1.1 Project Background

A good hotel brand image is depending to how well will be the hotel services. To improve and consistent hotel service, sentiment analysis is very important for hotel management because they need the technology to evaluate their services. Without using the technology, it will cause hotel management waste time and resource to evaluate each survey form whether positive or negative comment because they need to evaluate sentiment result and root cause from the survey form manually. So, Sentiment Analysis is a very helpful technology because it able to help hotel management to evaluate many surveys in short time. Sentiment analysis also may cause hotel management make a wrong decision which can affect hotel image if the result of sentiment analysis is less accuracy. So, accuracy of sentiment analysis is the most concern in this project.

Sentiment Analysis also known as opinion mining which is a kind of process to indentify large number of posts, comments and online news whether the sentiment is positive, negative or neural through natural language processing. Besides that, it also can detect and measure sentiment in paper or report format. Sentiment analysis can automatically track, identify and understand customer feeling and attitude on blogs, forum, online survey form and paper survey form. It can track popularity of brand, perception of new product and anticipation and fame of company or rant detection.

Moreover, it can search opinions, posts and reviews around competition and predicting outcomes such as successful of lunching new product and service. So, it helps manager to make a better decision making on the business strategies faster than before. It also can help company easily to find out valuable information from survey form that can help to improve product and service. It allows company to understand root cause from the sentiment result and increase speed of detection company opportunities, threat and marketing targets.

Sentiment Analysis can result the positive or negative of the sentence by extracting exact words such as bad, good and not bad.

CHAPTER 1 INTRODUCTION

The following are advantages and disadvantages of using sentiment analysis:

Advantages

- Sentiment Analysis allows company to quickly identify and respond to positive and negative comment of the survey form, topic, product or service.
- Sentiment Analysis able to improve company customer services because company can study customer complaint and understand the root cause of the problem in short time.
- Sentiment Analysis allow company to figure out what are the customer preference, opinion and point of view on discussion and plan a better competitive strategies to increase profit.
- Sentiment Analysis able to standardize ways of quantifying the sentences.

Disadvantages

- Sentiment Analysis has trouble in understanding irony sentences because of there are not specific words to figure out. So, that is a most common mistake that sentiment analysis will make.
- The sentiment results do not really reflect the truth about how customers feel. For example, a customer does not like iPhone 6, but it doesn't mean that customer will not buy it.
- Different sentiment analysis tools may have different sentiment result which can confuse people.

CHAPTER 1 INTRODUCTION

1.2 Problem Statement

1. Sentiment Analysis not able to create the positive and negative word dictionary automatically which mean programmer need create their own word dictionary manually.
2. Most of Sentiment Analysis tools do not access hotel forums to extract the hotel service's comment and make comparison.
3. Accuracy of sentiment analysis may be reducing when the sentence is too long and complex.

1.3 Project Objective

1. To analyze how sentiment analysis detect positive or negative comment of the sentence and show the sentiment result.
2. To design a system that able to automate create word dictionary, access websites and calculate sentiment scoring for user compare.
3. To evaluate the system analyze sentence and increase accuracy of sentiment result.

1.4 Project Scope

1. To create word dictionary by accessing few websites to get the positive and negative word dictionary and store into database and allow user to user update in future which solve problem #1 and achieve objective #1 and #2.
2. To study few types of sentiment calculation and evaluate which calculation is more on human rating which solve problem #3 and achieve objective #3

CHAPTER 1 INTRODUCTION

1.5 Detail of Project Scope

1. To achieve project scope #1, the system will access to website which provide positive and negative words and get the words at page source of website, it will search the positive and negative word at source page of the website. After system searched the word and these words will store at database. So that the system don't need to access the website every time during analysis. The following is how the system gets the positive words. But before store in database, it allows user check the list of words which will display on screen and enable user to add and remove word manually.

```
72 awesome<BR>
73 <P><CENTER><font color=red><B>B</B></font></CENTER>
74
75 beaming<BR>
76 beautiful<BR>
77 believe<BR>
78 beneficial<BR>
79 bliss<BR>
80 bountiful<BR>
81 bounty<BR>
82 brave<BR>
83 bravo<BR>
84 brilliant<BR>
85 bubbly<BR>
86 <P><CENTER><font color=red><B>C</B></font></CENTER>
87
88 calm<BR>
89 celebrated<BR>
90 certain<BR>
91 champ<BR>
```

Figure 1.1 Example of system get the keyword by searching “<P><CENTER>” (Enhanced Learning 2010)

2. To achieve project scope #2, few types of sentiment calculation formulas will be research such as Naïve Bayes Theorem, Pointwise Mutual Information and NTCIR-6. Study this all calculation and prepare few set of survey form which contain sentiment sentences for human to rate, this survey form will use to evaluate which calculation is more closely with human rating.

CHAPTER 2 LITERATURE REVIEWRAL

Chapter 2 Literature Review

2.1 Literature Review

By using system of sentiment analysis for hotel services is a faster and easier way to evaluate sentiment from the survey form instead of manually. It normally use on social media to track company product and services. To perform sentiment analysis, there are several algorithms that able to analysis the sentence.

2.1.1 Semantics Analysis

Introduction

Semantics analyze the sentence which is relation of words and phase. It studies semantic, structures and meaning of the sentence. Semantics analysis do not spot the key words in the sentences to figure out the sentences is positive or negative such as meaningful or meaningless. It is because the word can be inverting the sentiment of word which it divides to 3 type of sentence. So, it will break a long sentence into few short sentences and defines positive or negative opinion for each sentence. Last, it only concludes overall sentiment result.

Example of inverting sentiment of word:

- Sentiment Shifters: “I think this book is **less meaningful** than other”.
- Connectives: “This TV program is boring **but meaningful**”.
- Models: “This gift **should be meaningful** to her”.



I've just registered for a trial with SimplyMeasured. I'm loving it so far. The reports are incredible. Their blog is so informative, too. However, my colleague thinks it's expensive. – date: December 16, 2013.

Figure 2.1 Example of Sentiment sentence (On Social Sentiment and Sentiment Analysis 2013).

CHAPTER 2 LITERATURE REVIEWRAL

(1) I've just registered for a trial with SimplyMeasured. (2) I'm loving it so far. (3) The reports are incredible. (4) Their blog is so informative, too. (5) However, my colleague thinks it's expensive.

Figure 2.2 Example of after break sentiment sentence (On Social Sentiment and Sentiment Analysis 2013).

- **Sentence 1** is neutral, as it doesn't offer any sentiment.
- **Sentence 2** expresses a positive opinion about the tool as a whole.
- **Sentence 3** expresses a positive opinion about the tool's reports.
- **Sentence 4** expresses a positive opinion about the company's blog.
- **Sentence 5** expresses a negative opinion about the tool's pricing, this time from the eyes of the writer's colleague.

Figure 2.3 Example of evaluate each sentiment sentence (On Social Sentiment and Sentiment Analysis 2013).

Doing sentiment analysis on a sentence, there are 5 main factors it will figure out to increase accuracy of result which are topics, aspects, sentiment, holder and time. Topic can break into aspects such as below

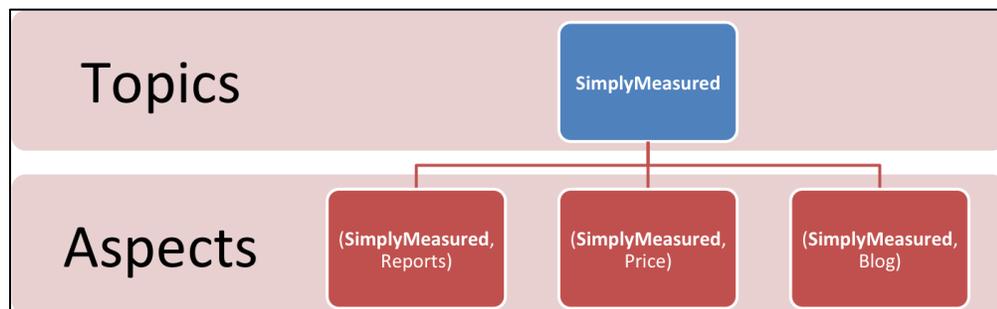


Figure 2.4 Topics break down diagram (Social Sentiment and Sentiment Analysis 2013)

Advantages:

- Can be done in document level, sentence level and entity and aspect.
- Break a long sentences in few parts and declare positive or negative opinion of each sentences

CHAPTER 2 LITERATURE REVIEWRAL

Disadvantage

- Does not able to study opposite connotation sarcasm and sentiment ambiguity.
- It able to breaks sentence but it not able to reduce number of words to increase accuracy of sentiment analysis.

2.1.2 NaSent Algorithm

Introduction

NaSent analysis is a recursive deep learning algorithm that enable computer to understand meaning of words in context which is similar to human. NaSent analysis normally use for evaluate the topic that related to politics, fashion or movie which is on situation and experience on the sliding rate of sentiment ranging.

Purpose of NaSent analysis allow computer to study sentiment sentence more like human way without constantly refer to dictionaries and rule which created by expert. NaSent Analysis does not much focus on single word but it is focus more on word to phases and sentences which are track word combination.

There are 2 grammatical structure trees that evaluating sentence:

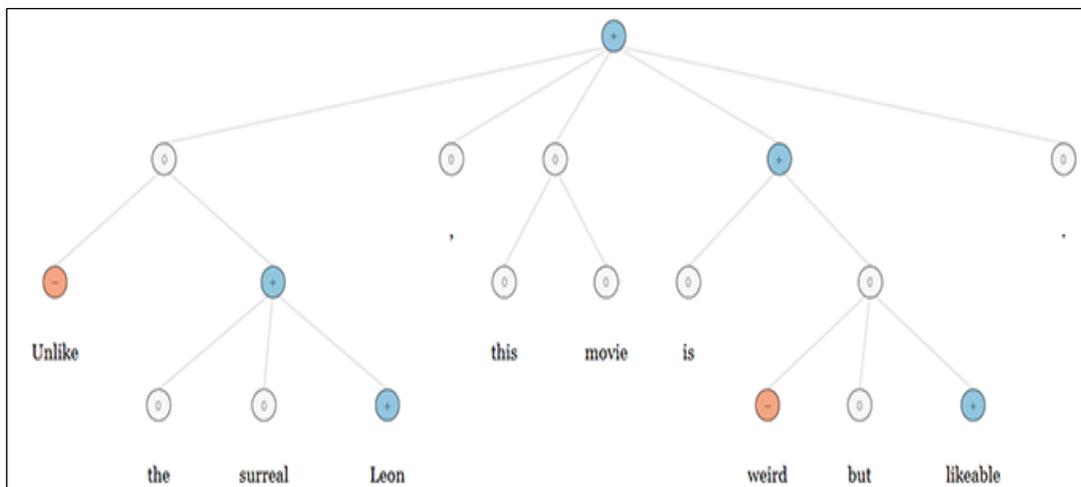


Figure 2.5 NaSent analyze sentence: “Unlike the surreal Leon, this movie is weird but likeable.” (Tom Abate 2013).

CHAPTER 2 LITERATURE REVIEWRAL

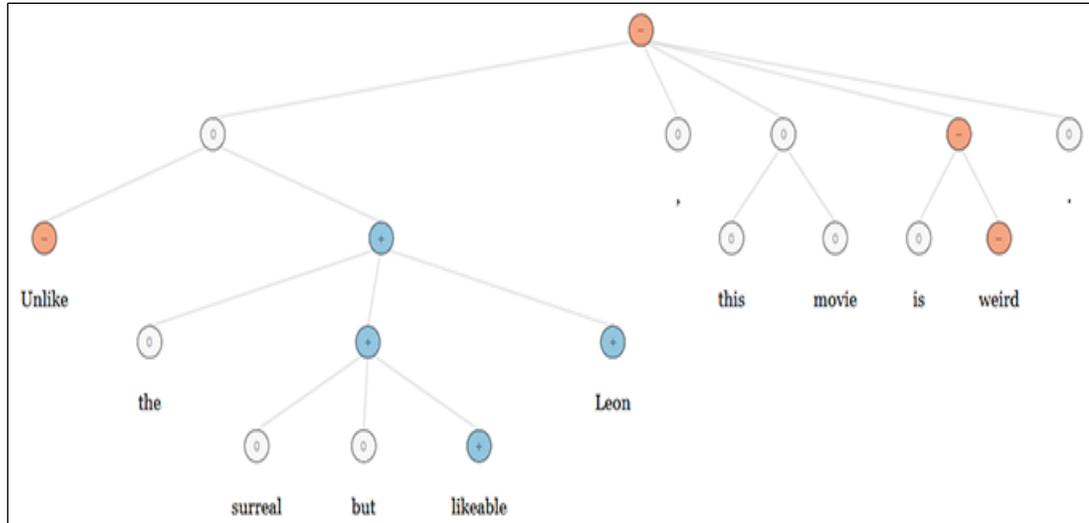


Figure 2.6 NaSent analyze sentence: “Unlike the surreal but likeable Leon, this movie is weird.” (Tom Abate 2013).

From the both figure show that red dot detected is refers negative to word and the blue dot detected is refer to positive word. For example, “this movie is weird”, it can rate as negative and “surreal but likeable” it can rate as positive.

Advantages:

- Doesn't much depend to human expert to create many rules and instructions to study sentiment sentence.
- Focus more on word to phases and sentences which are track word combination.
- Use parse group techniques to split the sentence into few phases into grammatical unit of meaning (Tom Abate 2013) and measure the sentiment sentence.

Disadvantages:

- Only useful for short phase because the long sentences can cause it not able to perform in principle way and result may less accuracy.

CHAPTER 2 LITERATURE REVIEWRAL

2.1.3 Simple keyword-based methodology

Introduction

Simple keyword-based methodology is one of the ways to analysis the sentence whether is positive or negative comment. It will identify the opinion of the word by indentify only the keywords. If the sentence contains more positive word then it will classify as positive comment.

It analyze the sentence through the terms presence and frequency and opinion word and phrases which is indentify the single word such as “good” and “bad” or “like” and “dislike”. It will differentiate the positive and negative word from word dictionary and give a binary weight score for the result. It will classify the sentiment sentence as very good, good, neural, bad and very bad. If the sentiment sentence is neural which mean either the sentence no have single charged word or positive and negative word is balance.

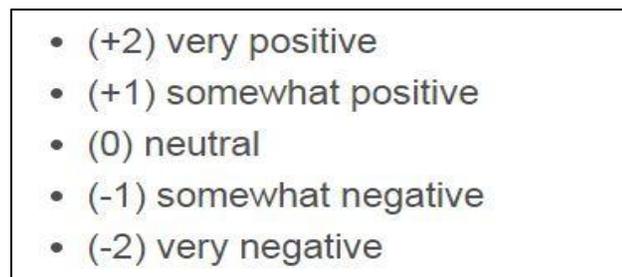
- 
- (+2) very positive
 - (+1) somewhat positive
 - (0) neutral
 - (-1) somewhat negative
 - (-2) very negative

Figure 2.7 Example of simple keyword-based methodology scoring system (Datapig 2014)

Simple keyword-based methodology is very simple and easily to apply but the result of accuracy will be less accurate when there is sentiment words are together such as “no go”, it can only spot single word and the result of sentiment will be good because it didn’t spot another word which is “no”. The accurate also will be less when the sentence does not contain charged word but the meaning is clearly negative comment such as “The movie ticket price is higher than other”.

Advantages:

- Easy and simple to implement

CHAPTER 2 LITERATURE REVIEWRAL

Disadvantages:

- Manually create and update positive or negative word dictionary.
- Only spot single word.

2.1.4 Text Parser

Introduction

Parser is a part of compiler which is use to breaks a string into individual parts. The parts can classify as noun, verb, adverb, adjective, interjection and conjunction. It receives a sentence and parses the sentence by determine the grammatical part and indentify relationships for each other. After parse, word will show in parse tree to describe how the sentence can built from grammar. It normally divides into 2 types which is lexical analysis and semantic parsing. Lexical analysis is use to divide sentence into few components while semantic parsing is use to determine meaning of the sentences.

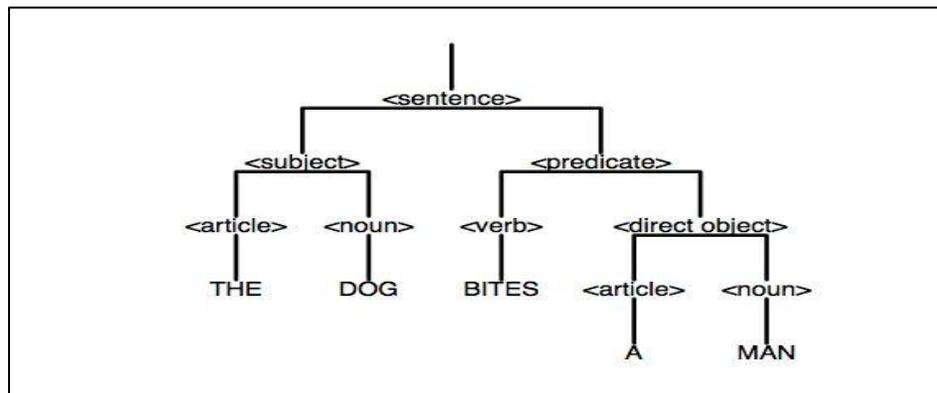


Figure 2.8 Parse Tree (Rober B Heckedorn, n.d.)

2.1.5 Naive Bayes Theorem

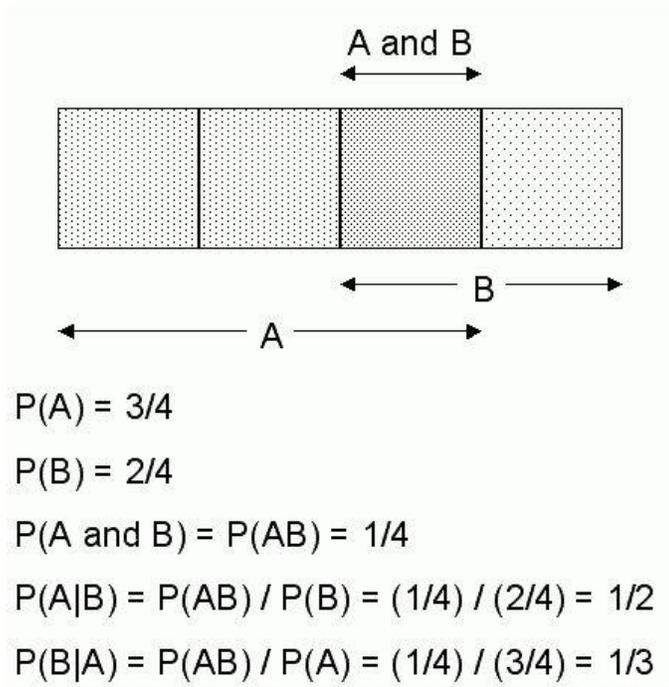
Naive Bayes Theorem is simple and basic text probability classifier because it can be trained very quickly and it also enables to determine conditional probability of event A conditioned by B and event B conditional by A. For example $P(A|B)$ and $P(B|A)$. It can be trained to estimate the document is positive or negative. Although it just a simple classification technique but it performs well in complex real world problem. It is useful when it deals with very large dataset and it is widely used because it performs better in

CHAPTER 2 LITERATURE REVIEWRAL

complicated classification method. It use in spam email detection, email sorting, document categorization language detection and sentiment detection. Naive Bayes Theorem is a better choice when programmer have limited source in terms of memory and CPU.

The following are Naive Bayes Theorem formula:

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{B}$$



A – Is represent number of positive or negative words detected.

B- Is represent total number of positive and negative words.

After get calculate positive and negative score, then use positive score to minus negative score to get final result which will not more then +1 or -1. After minus, the final result is more than 0 then will classify as positive sentence. If the final result is less than 0 then will classify as negative sentence.

Final result = positive score – negative score

CHAPTER 2 LITERATURE REVIEWRAL

After the formula implement into sentiment Classification show as below:

$$P(\text{Positive}|\text{Sentiment}) = \frac{P(\text{Sentiment}|\text{Positive}) * P(\text{Positive})}{P(\text{Sentiment})}$$

$$P(\text{Negative}|\text{Sentiment}) = \frac{P(\text{Sentiment}|\text{Negative}) * P(\text{Negative})}{P(\text{Sentiment})}$$

For example:

Sentiment sentence: “This cat is beautiful and cute but she is naughty. “

$$\text{Positive probability} = \frac{1 * P(\text{Positive})}{P(\text{Sentiment})} = \frac{\frac{2}{10}}{\frac{3}{10}} = 0.66$$

$$\text{Negative probability} = \frac{1 * P(\text{Negative})}{P(\text{Sentiment})} = \frac{\frac{1}{10}}{\frac{3}{10}} = 0.33$$

Thus, the probability of being positive is 66% probability. The probability of being negative is 33%. So, the sentence is positive.

2.1.6 Pointwise Mutual Information

Pointwise Mutual Information (PMI) is ways to measure the frequency of positive and negative words occur in a sentence or paragraph. PMI do not restrict to specific part of speech not even for any single words. PMI can be use with multiple word phrases. PMI is calculating sentiment value by detecting positive words and negative words such as “good” and “poor”. If the sentiment value is positive then it can classify as positive sentence but if sentiment value is less than 0 which is negative then it can classify as negative sentence. PMI can be used in many applications and it can be improve certain part if necessary but PMI only considered the co-occurrence intensity. Getting more positive and negative words and fewer noisy words could be improve PMI classify performance.

$$PMI(x, y) = \log_2 \frac{p(x'y')}{p(x')p(y')} = \log_2 \frac{p(x'|y')}{p(x')} = \log_2 \frac{p(y'|x')}{p(y')}$$

CHAPTER 2 LITERATURE REVIEWRAL

PMI is similar with Naïve Bayes Theorem, but the result may more than 1.

X' – Is represent number of positive or negative word detected.

Y' – Is represent total number of word in the sentences.

After get calculate positive and negative score, and then use positive divide negative score. After that, use Log2 to get the final answer. If final answer is more than 0 then will classify as positive sentence. If final answer is less than 0 then will classify as negative sentence.

Moreover, another PMI method is called PMI2, PMI2 is same as PMI, only different is multiple by 0.75 which use to scale down the value which between -5 to +5. The value 0.75 can be change if the rating rate is change such as -10 to +10.

$$PMI2 = \log_2 \frac{p(y' | x')}{p(y')} * 0.75$$

2.1.7 National Institute of Informatics Test Collection Information Retrieval (NTCIR-6)

NTCIR is research on information access technologies and evaluation in area of text retrieval, various documents and cross lingual between Asian language and English. NTCIR can extract opinion from sentences with four factors to be considered which are opinion operator, sentiment words, negation operator and opinion holder to determine their tendency. NTCIR need to collect these four factor words lists to classify if the sentences is opinioned and its polarity. Strength of this method to learn sentiment word is depending on how well sentiment words dictionary is developed.

CHAPTER 2 LITERATURE REVIEWRAL

$$P_{c_i} = \frac{fp_{c_i} / \sum_{j=1}^n fp_{c_j}}{fp_{c_i} / \sum_{j=1}^n fp_{c_j} + fn_{c_i} / \sum_{j=1}^m fn_{c_j}}$$

$$N_{c_i} = \frac{fn_{c_i} / \sum_{j=1}^m fn_{c_j}}{fp_{c_i} / \sum_{j=1}^n fp_{c_j} + fn_{c_i} / \sum_{j=1}^m fn_{c_j}}$$

$$S_{c_i} = (P_{c_i} - N_{c_i})$$

Pci – Is represent to positive score which number of positive word divide by total number of positive and negative words

Nci - Is represent to negative score which number of negative word divide by total number of positive and negative words.

Final answer will be positive score minus negative score and it not more then +1 or -1. After minus, the final result is more than 0 then will classify as positive sentence. If the final result is less than 0 then will classify as negative sentence.

2.1.8 Enhance Sentiment (ES)

Enhance sentiment is kind of formula that combine with NTCIR and PMI method and has been improve to make the system result closely to human result. It calculate the result based on the sentiment word value such as “Clean” sentiment value is 0.33, “Cleaner” is 0.667 and “Cleanest” is 1.00. Based on the sentiment value apply to the Enhance Sentiment formula to get the result. It will not count number of positive and negative words occur in sentence. With the sentiment score will improve the feeling of sentence.

$$\text{Positive} = \frac{\text{Total Positive score}}{\text{Total sentiment words}}$$

$$\text{Negative} = \frac{\text{Total Negative Score}}{\text{Total Sentiment Words}}$$

CHAPTER 2 LITERATURE REVIEWRAL

Total score = Positive – Negative

If total score is less than 0,

$$\text{Total score} = \text{Total Score} * -1$$

$$\text{Total Score} = \text{Log}(\text{Total Score} * 100)$$

$$\text{Final score} = \text{Total Score} + 0.5$$

If total score is more than 0,

$$\text{Total score} = \text{Log}_2(\text{Total Score} * 100)$$

$$\text{Final Score} = \text{Total Score} + 0.7$$

The final score will be the final result. If final result is more than 0 then will classify as positive sentence. If the final result is less than 0 then will classify as negative sentence.

2.2 Fact Finding

2.2.1 Research

The author will do research about sentiment analysis algorithm from online and study to understanding formulas works. The algorithm researches are Semantics Analysis, Nasent Algorithm, Simple Keyword-Base, Text-Parser, PMI, NTCIR and Naive Bayes Theorem. Based on research, author able to figure out how sentiment works and calculate the sentiment scoring for the sentence. Also, consult supervisor to understand and get clear path on the project.

2.2.2 Data Collection

To collect the data, tripadvisor.com website's customer comment is main data resources for my system. I will access tripadvisor.com to collect customer's comment as sample data. The sample data are customer's comment about hotel service and the sentence of comment can be long and short. These data will be use to test system result. Estimate around 20 customer's comment will be use to test in the system on first time.

CHAPTER 2 LITERATURE REVIEWRAL

Besides that, I will access some website to get vocabulary which provides positive and negative words that use to create word dictionary such as www.enchantedlearning.com and positivewordsresearch.com.

2.3 Evaluation planning

Firstly, testing on sentiment result of sentence which done by human. Around 7 people who good in English language such as teacher, they will be invite and request them to analysis some sentiment sentence where get from website such as Tripadvisor.com. They will rate for the sentiment sentence whether the sentence is extreme positive, highly positive, moderate positive, low positive, slightly positive, neutral, slightly negative, low negative, moderate negative, highly negative, extreme negative of the comment.

Secondly, those customer comments will also test on already written few different formulas function in the system to evaluate which formula is close to human result.

Lastly, these sentences will pick for system testing to check whether the system result average same with human and website analysis.

CHAPTER 3 PROPOSED METHOD/ APPROACH

Chapter 3 Proposed Method/ Approach

3.1 Methodology

To develop a system, an appropriate software process model is needed because it can affect the successful of developing system and may slower the developing process. Therefore, 3 types of software process model in this report will bring to discuss, which are Waterfall Model, Incremental Model and V –Model and select the most suitable model for this project

3.1.1 WATERFALL MODEL

Waterfall Model is a simple sequential development approach. It is traditional model and it is first process model that was introduced in early 1950. Waterfall Model has several stages such as requirement analysis, software design, implementation, testing integration, deployment and maintenance. It only allows performing activities from one stage to another stage. Even though, this model also allow to loop if there any previous not done well but it better done well on previous stage before proceed to next stage. This is because it may cause over time or over budget of developing system if rollback to previous stage. So, the development team has to ensure each stage is performing well before go to new stage.

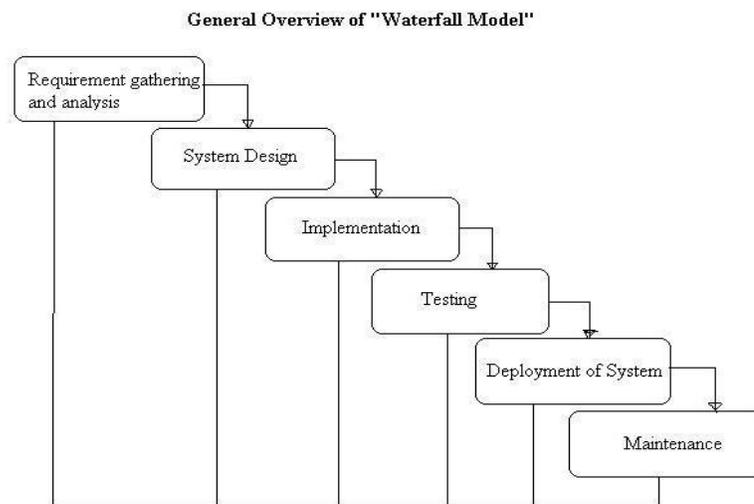


Figure 3.1.1 Waterfall Model Diagram (ISTQB EXAM CERTIFCATION, n.d)

CHAPTER 3 PROPOSED METHOD/ APPROACH

Advantages:

- Simple and easy to use
- It is easy to manage due to rigidity of the model
- Each stage does no overlap because each stage is processed and completed at a time.

Disadvantages:

- If in testing stage, it is difficult to rollback to previous stage to make changes.
- High risk and uncertainty
- Not suitable for complex project

3.1.2 V – MODEL

V-Model is a kind method of software development life cycle. V-Model is an extension of waterfall model which mean that every single phrase in software development life cycle is directly correlative to testing phase. V-Model is creating to overcome disadvantages of waterfall. In this model, programmer only allows to start next phrase after previous stage is completed. This model focuses on quality assurance during early in life of a project. V-Model suggests programmer create system test plan before proceed to development stage. Moreover, V-Model allows performing testing and fixing at any stages in the life cycle.

Firstly, programmer starts working on planning and design stage which are requirement analysis, system requirement, global design and detailed design. After complete design stage, programmer only starts implementation and testing stage which are unit testing, integration testing, system testing and user acceptance testing. During implementation and testing stage, these stages also working parallel with planning and design stage.

CHAPTER 3 PROPOSED METHOD/ APPROACH

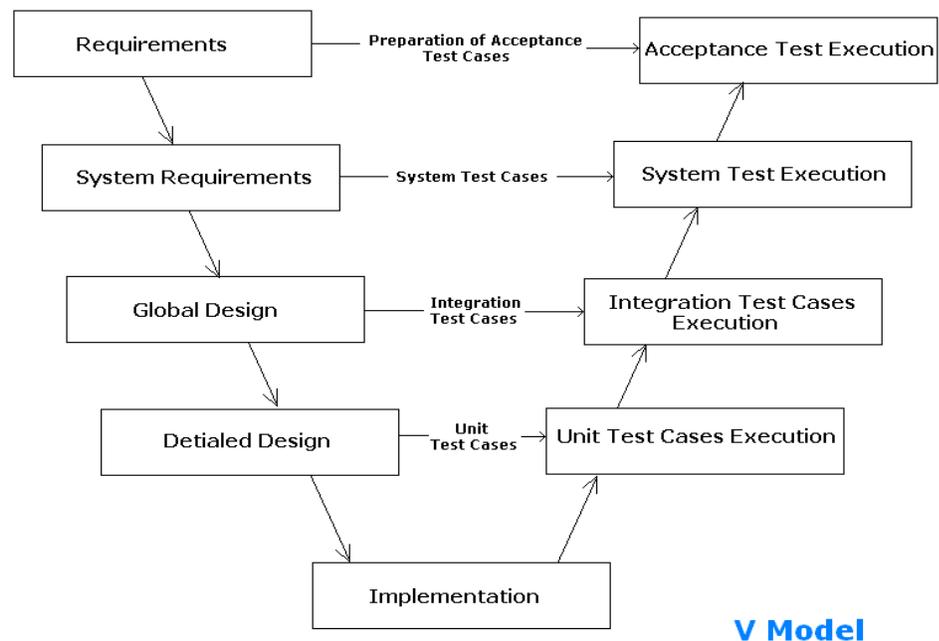


Figure 3.1.2 V-Model Diagram (Software Testing Class, 2012)

Advantages:

- Time saving and quick implement.
- Perform well testing on planning and design activities before coding.
- Able to find defect at early stage.

Disadvantages:

- Risky and least flexible.
- If any changes occur in midway, test document along with requirement document need to be update.
- Doesn't have clear path for problem found during testing phrase.

CHAPTER 3 PROPOSED METHOD/ APPROACH

3.1.3 INCREMENTAL MODEL

Incremental Model is a model that divides a big project into various parts to build. This model consists of multiple development cycles which called multi- waterfall cycle. The model divided into smaller module and each module will go through Requirement, Design, Implementation and Testing stage until the product is finished which is similar to Waterfall Model. The modules are divided separately and deliver to customer separately when it is done. After all modules are done, they will integrate as a complete system.

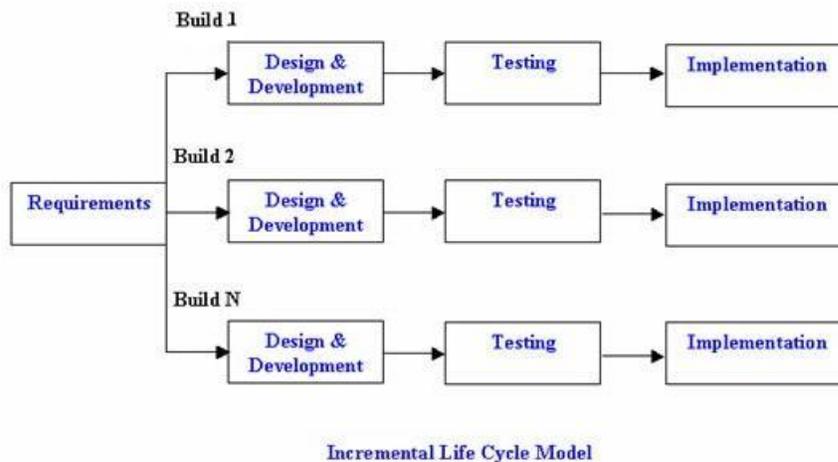


Figure 3.1.3 Incremental Model Diagram (ISTQB EXAM CERTIFICATION, n.d)

Advantages:

- Able to generate working software quickly and early in life cycle
- Less cost required to change scope and requirement
- Low initial delivery cost

Disadvantages:

- Require good planning and design skill
- Need clear understanding on system before divide into module and built incrementally
- Total cost is higher compared to waterfall model.

CHAPTER 3 PROPOSED METHOD/ APPROACH

3.2 Selecting Software Process Model

In this project, V-MODEL will be selected to be used because each stage in software development life cycle is directly correlative to testing phase. So, planning and design stage and implementation and testing stage can perform at same time. Also, the quality of system will able to assured and defect found at early stage. Each stage will be completed first before move from one to another stage, so that during testing stage, each stage able to work parallel and check back with previous stage to make sure the system requirements are meet. Thus, the planning and design activities still can change during implement and testing stage and it is quick implement and time saving.

3.3 Sentiment Analysis Methodology

In this project, there are 4 methods will be used to generate new method of sentiment analysis in this project which includes semantics analysis, Naive Bayes Theorem, Test Parser and Simple keyword-based methodologies but there will only pick few useful parts from these 4 methods.

The new method of sentiment analysis is using semantics analysis method to break paragraph into sentences by indentify the punctuation mark such as semicolon, comma, full stop and exclamation mark.

After that, use text parser method will identify and remove carry with no meaning words in the sentence such as “as”, “the” and “a”, and spot keyword in each sentence such as “good”, “well” and “bad” based on created word dictionary.

At last, it will evaluate the positive or negative word by using Simple keyword-based methodologies and calculate score rate by using Naive Bayes Theorem and show result to user in five star format, or spider web diagram.

CHAPTER 3 PROPOSED METHOD/ APPROACH

The following flow chart is showing how new methodologies analyze sentence:

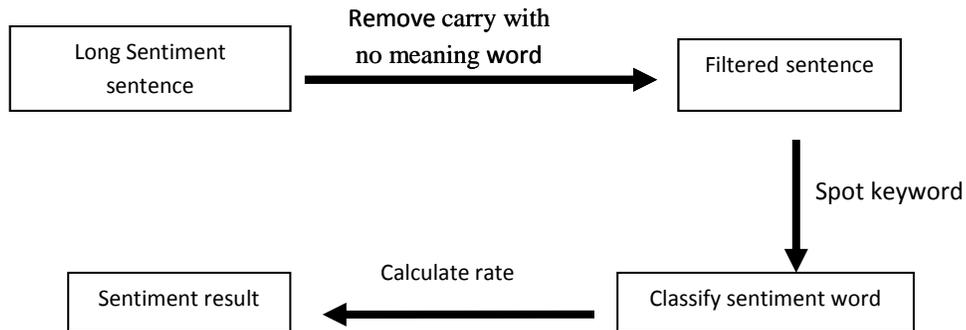


Figure 3.1 Flow Diagram of sentiment analysis methodologies for sentiment analysis.

3.4 Tools to use

3.4.1 C#

C# is an object-oriented language. This language allows developers to create and build powerful applications. C# runs on .NET framework. Many different applications can be created and developed by using C#. For examples, client and server applications, distributed components, XML web services and so on. Besides that, Visual C# does provide enhanced code editor, good user interface designer, integrated debugger and etc. By having these features, it is easier for developer to create advanced and powerful applications.

Moreover, C# language is easy to learn and implement. This is because this language has reduced much complex syntax as compared to C++. Many advanced functionalities are provided by C# such as enumerations, lambda expressions, direct memory access null able value types and etc. Since C# is an object-orientated language, many object-oriented features are included. The features are inheritance, encapsulation and polymorphism as well. Thus, in order to complete the project, C# language will be chosen to develop the system.

CHAPTER 3 PROPOSED METHOD/ APPROACH

The following are the advantages of C# programming language.

- Garbage collection is done automatically.
- Usage of pointer is reduced
- Usage of header files is reduced
- Global variables and functions is no longer exist but it is belongs to a class
- An app is allowed to be executed in a sandbox which is restricted.
- Higher efficiency as compared to java
- Primitive data types are increased
- Accessing objects by using indexes when array is exist
- Multi-threading process is much more simplified
- Structures, Overloading of Operator and Preprocessors directives are supported

3.4.2 Text File

Text file is a kind of computer digital file that can store plain text only. Text file are used for storage of information. Text files easily available and share in different types operating systems such as Linux, Microsoft and MAC OS. So, it prevents some problems occur by using file formats such as different in number of bytes, padding bytes and endianness. Moreover, it is easily to recover and continue processing the rest of content when data corruption occurs in the text file.

In this project, text file is use to create positive and negative word dictionary to perform word classification. So that, word dictionary can easily works in different operating system and it only require less disk space to store data compare to other word file.

Text file will use in data storage which call “PositiveWord.txt”, “NegativeWord.txt”, “GatherComment.txt”, “Short.txt”, “Medium.txt”, “Long.txt” and “HotelName.txt” as basic need for data storage.

CHAPTER 3 PROPOSED METHOD/ APPROACH

3.5 Requirement Specifications

This project is to develop a system that allow user to analyze and evaluate sentiment sentences and hotel comment. This section will list system requirement and system performance definition.

3.5.1 System Requirements

This system doesn't require any hardware equipment. It only needs a computer with Microsoft operating system which installed the system and fast network speed to operate the system. The system require network to built word dictionary and access hotel comments. The network can be wireless or LAN.

3.5.2 System Performance Definition

There is only thing which will define system performance which is network speed. With the fast network speed, the system will not have a problem to access hotel comment. But, if the network speed is slow, it will affect system perform, so every time system access hotel name and comments, so the system have to delay few second for each time after the system click the link and wait to get data. So, at the end it needs take few min to load and save data into file.

CHAPTER 3 PROPOSED METHOD/ APPROACH

3.6 Timeline

3.6.1 Project I schedule

Year: 2015 (Y3Sem1)															
No.	Schedule Task	Jan			Feb				March					April	
		W 1	W 2	W 3	W 4	W 5	W 6	W 7	W 8	W 9	W 10	W 11	W 12	W 13	W 14
1.	Analysis	■													
2.	Design				■										
3.	Prototype								■						
4.	Presentation preparation													■	

Table 3.1 Gantt Chart - Final Year Project 1

3.6.1 Project II schedule

Year: 2015 (Y3Sem2)															
No.	Schedule Task	Jan			Feb				March					April	
		W 1	W 2	W 3	W 4	W 5	W 6	W 7	W 8	W 9	W 10	W 11	W 12	W 13	W 14
1.	Development, Research	■													
2.	Implementation, Research, Modification						■								
3.	Evaluation, Testing										■				
4.	Presentation Preparation, Documentation													■	
5.	Presentation													■	

Table 3.2 Gantt Chart - Final Year Project 2

Chapter 4 System Design

4.1 Block Diagram

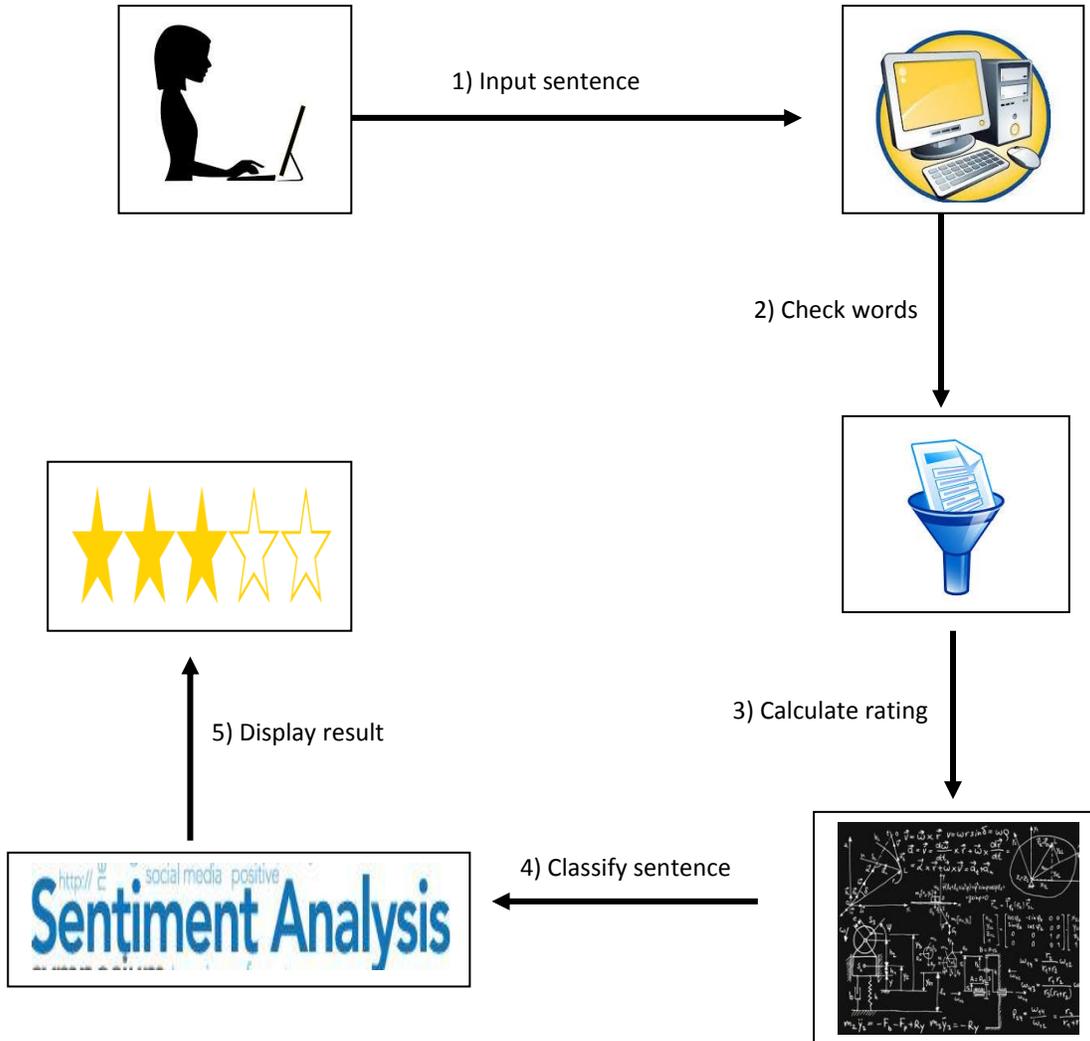


Figure 4.1 Block Diagram of Sentiment Analysis

4.2 Use Case Diagram

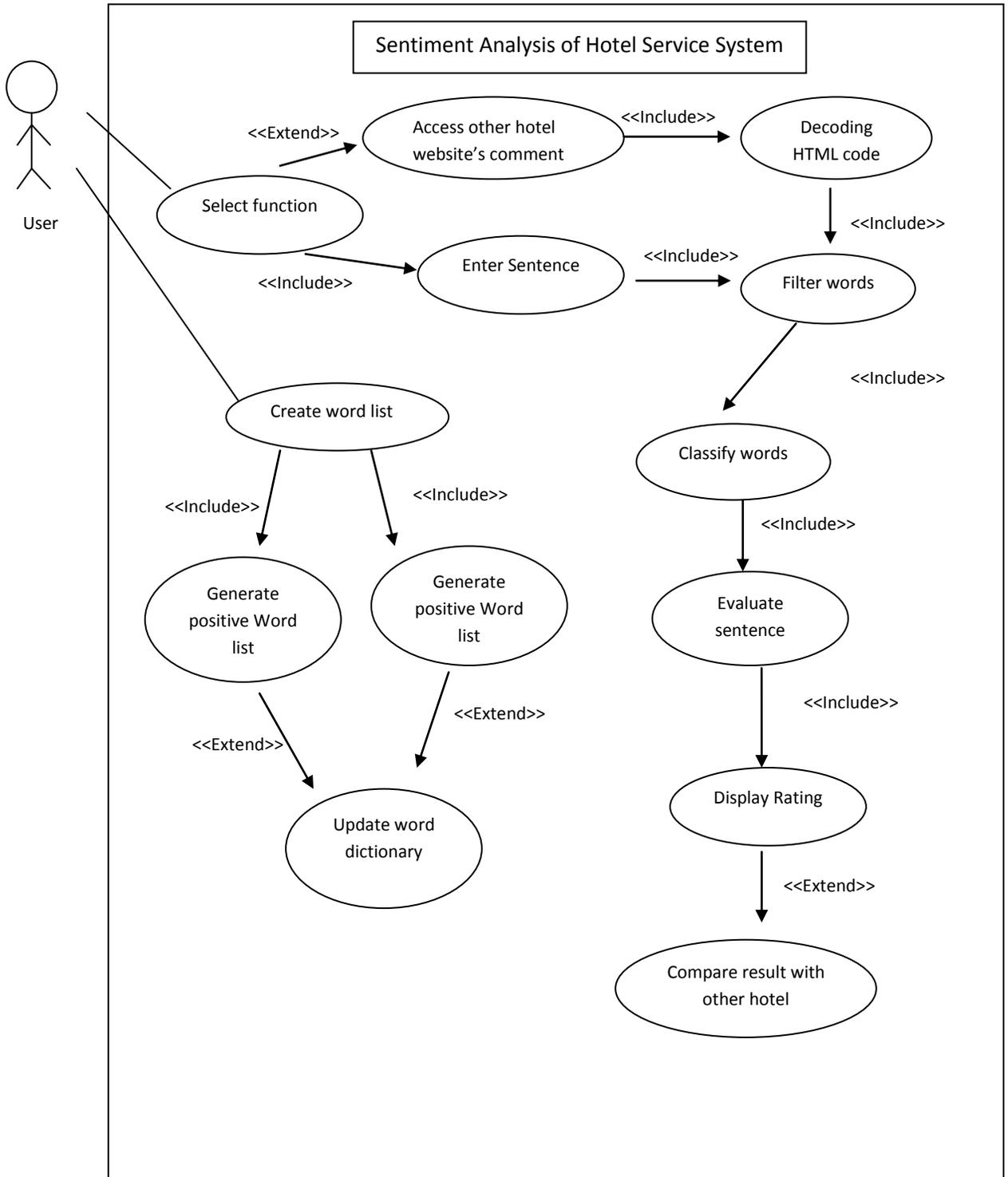


Figure 4.2 Use case diagram of sentiment analysis

4.3 Activity Diagram

4.3.1 Analysis Sentence

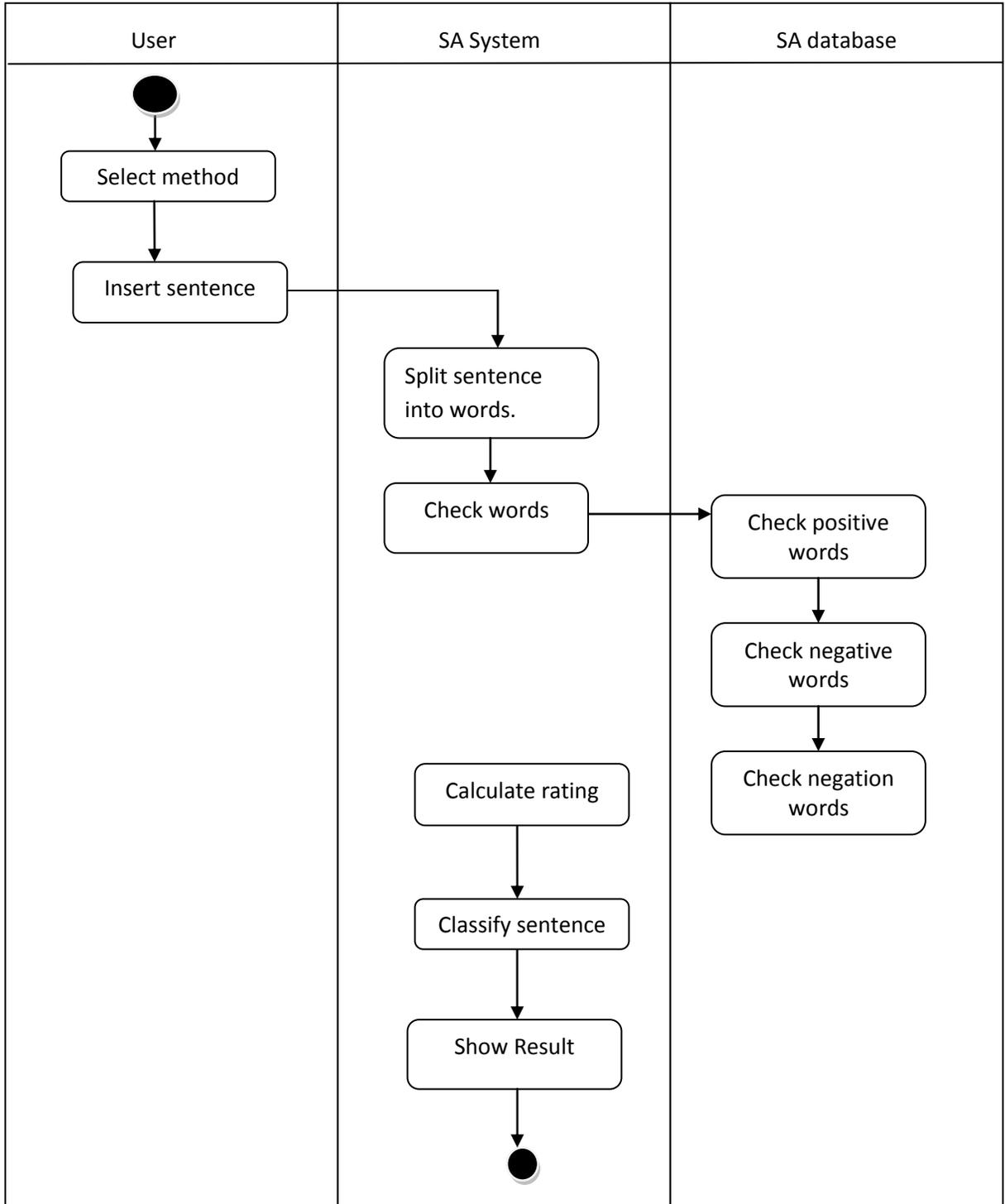


Figure 4.3.1 Activity diagram of analyze sentiment sentence

4.3.2 Create word dictionary

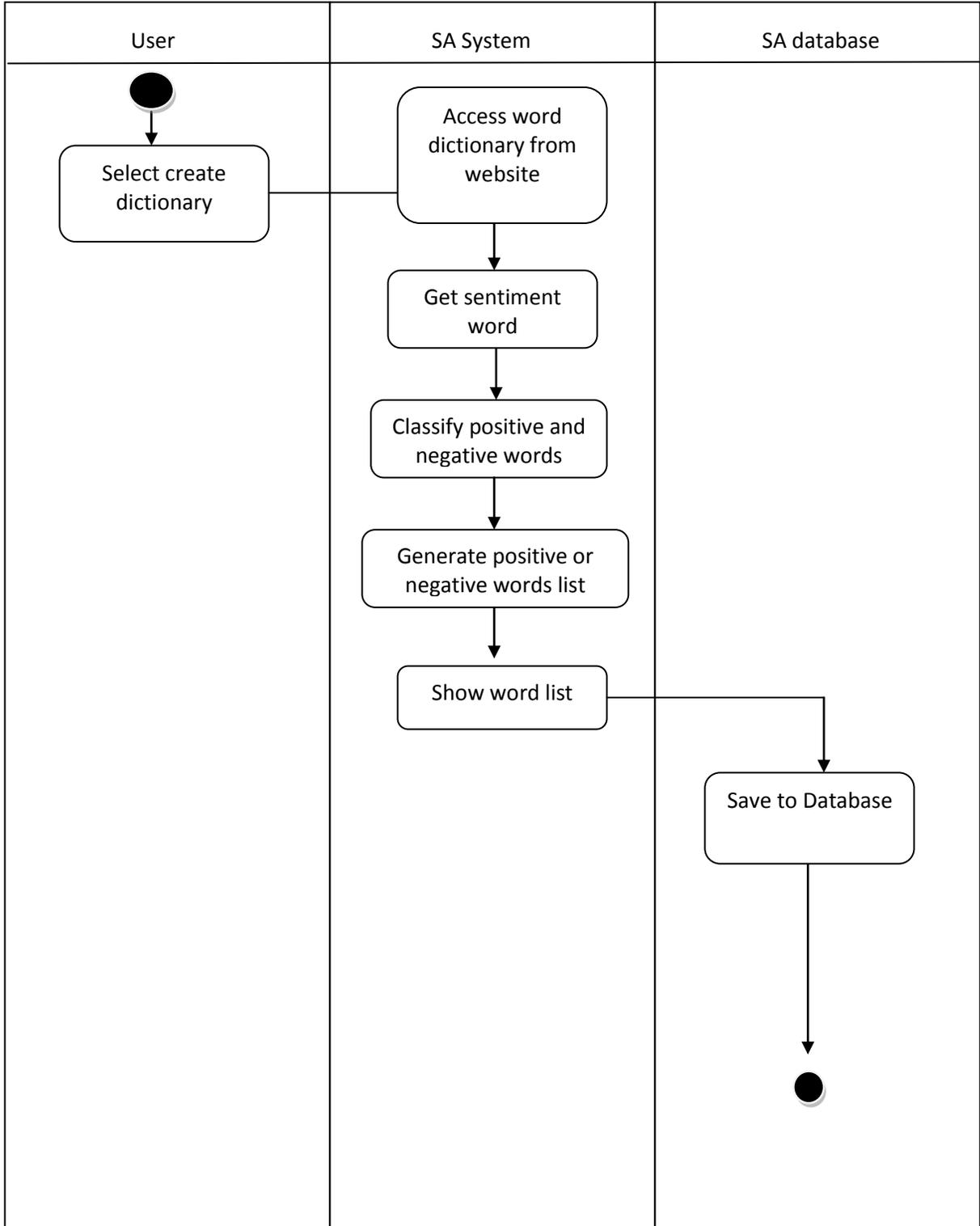


Figure 4.3.2 Activity diagram of create word sentiment dictionary

4.3.3 Analyze Hotel Comment

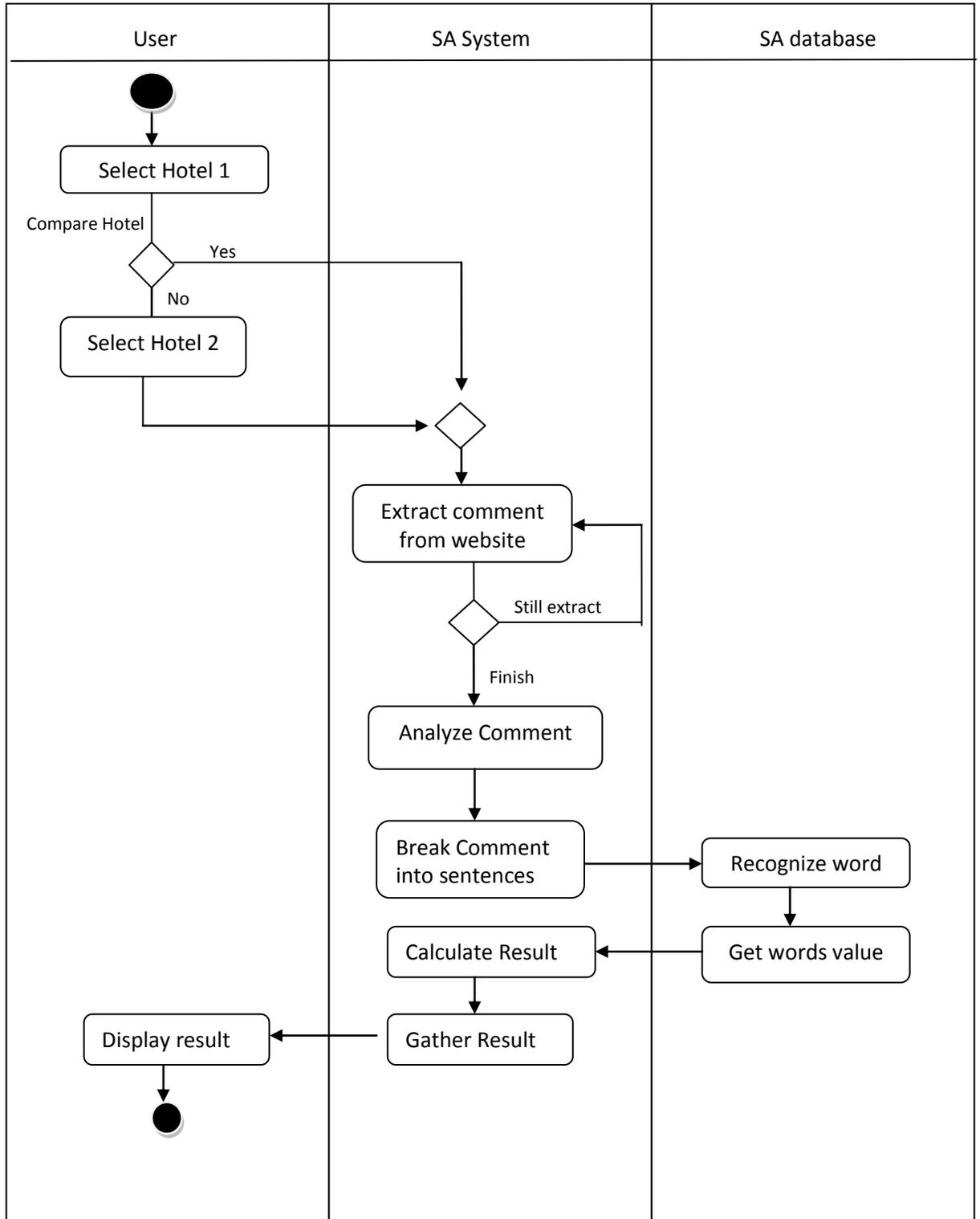


Figure 4.3.3 Activity diagram of analyze hotel comment

CHAPTER 4 SYSTEM DESIGN

4.4 Class Diagram

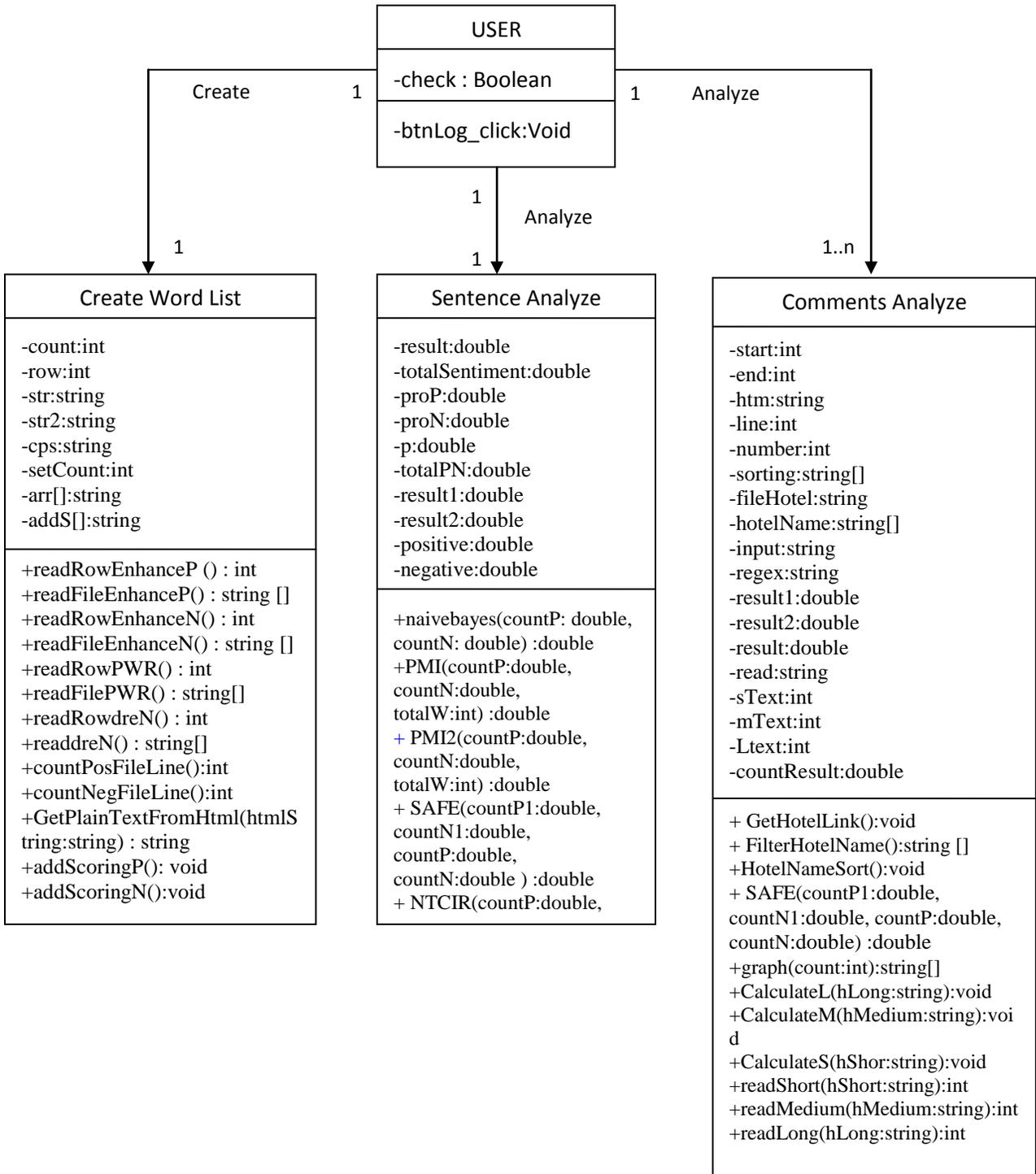


Figure 4.4 Class Diagram of System

CHAPTER 4 SYSTEM DESIGN

4.5 Database Design

4.5.1 Sentiment Word Database

Before system perform calculation, system need recognize sentiment word at first, so system need to built word dictionary for each set of positive and negative opinion database before analyze the sentence. Those words automatically will get from website which provided positive and negative opinion words such as www.enchantedlearning.com and www.winspiration.co.uk. These words will save into txt file instead of Microsoft Access because it is easy to access and there is not requiring any primary key or share key.

The following are example of PositiveWord.txt and NegativeWord.txt file:

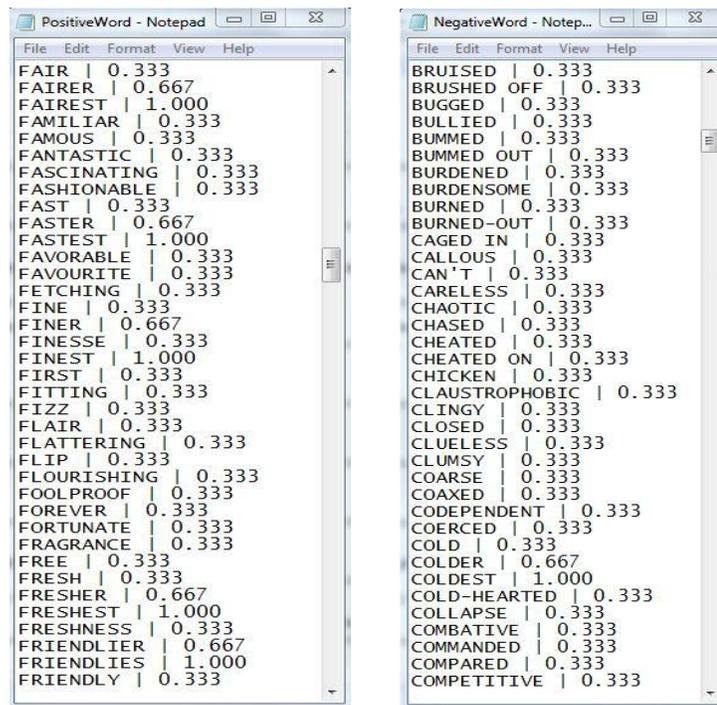


Figure 4.5.1 shows PositiveWord.txt and Negative.txt data file

CHAPTER 4 SYSTEM DESIGN

4.5.2 Hotel Comments Database

To enable system analyze hotel comments, the system will go to the website to get those comments based on user selected hotel. After get the comments, all comments will save into txt file called commentGather.txt , after that those sentences will break into 3 categories which call short sentence, medium sentence and long sentence and these will store in txt file also which called Short.txt, Medium.txt and Long.txt. This purpose is to increase systems result accuracy even the sentences too long. Normally, from the average length of sentence is 30-34 , then medium length of sentence average is 16-29 words and average length of shortest sentence is 2-15 words. These 3 categories will make system easy to calculate.

The following are example of Short.txt, Medium.txt and Long.txt data:

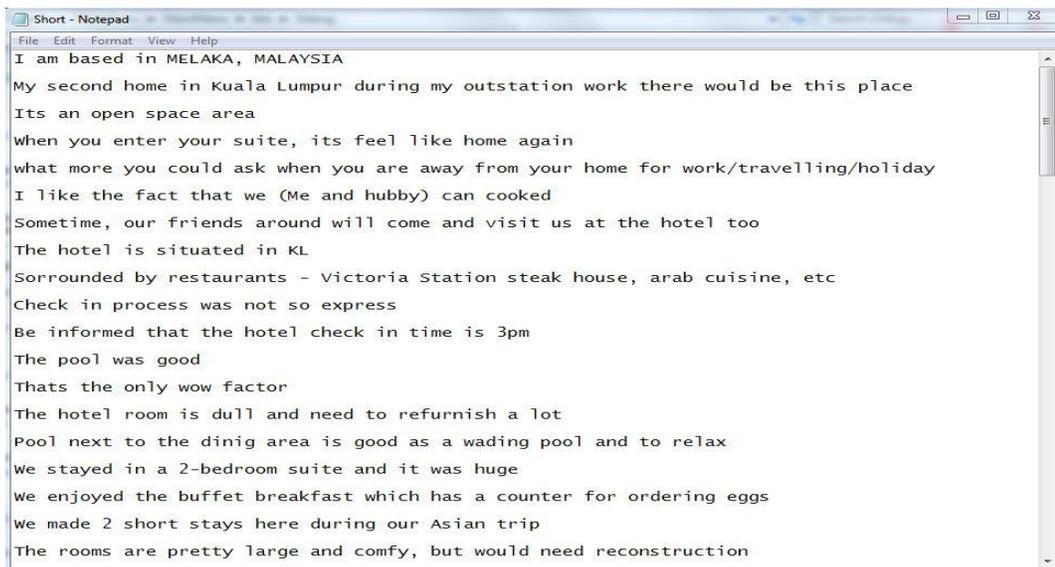


Figure 4.5.2 shows short length of sentence in Short.txt

CHAPTER 4 SYSTEM DESIGN

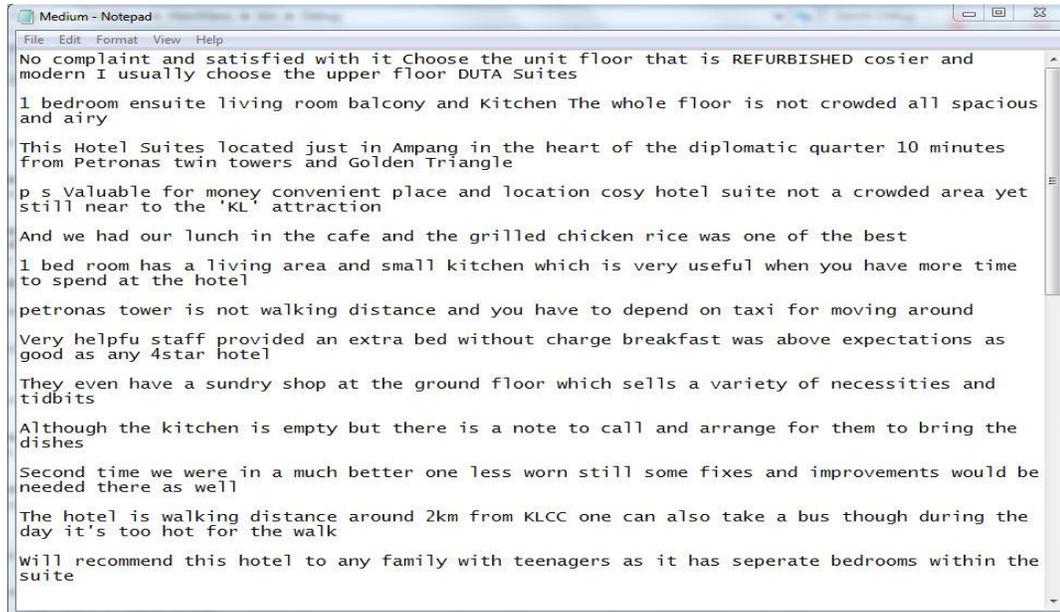


Figure 4.5.3 shows medium length of sentence in Medium.txt

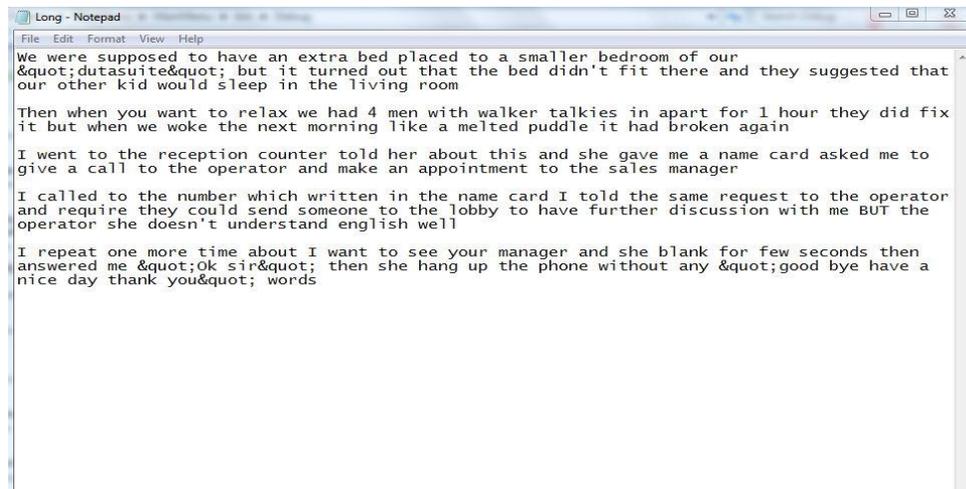


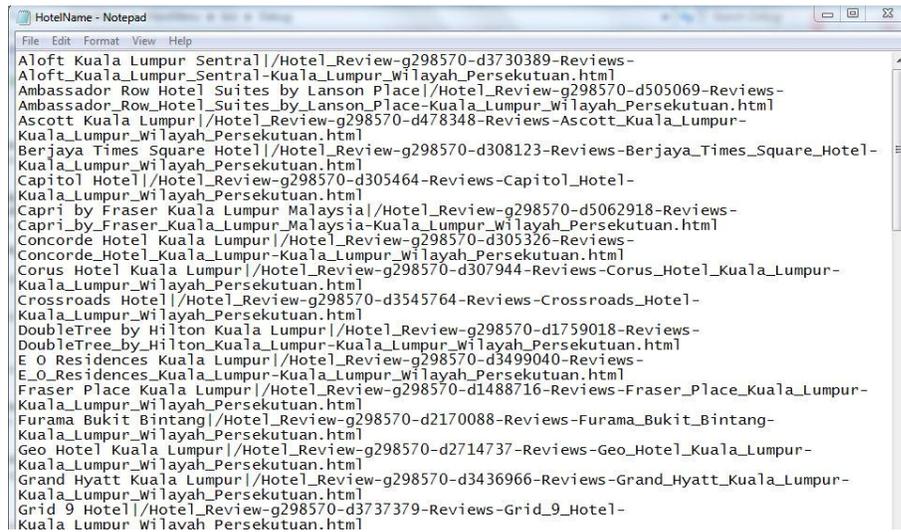
Figure 4.5.4 shows long length of sentence in Long.txt

4.5.3 Hotel Name & Link Database

Besides get hotel comment and sentiment word dictionary, system will also try to get all hotel's name from TripAdvisor. This will reduce user input error and save time to key in hotel name. Moreover, all hotel name show on the list is available to access the comment and it would save time to check whether the hotel is existing.

CHAPTER 4 SYSTEM DESIGN

The following are show hotel name and address link save into HotelName.txt:



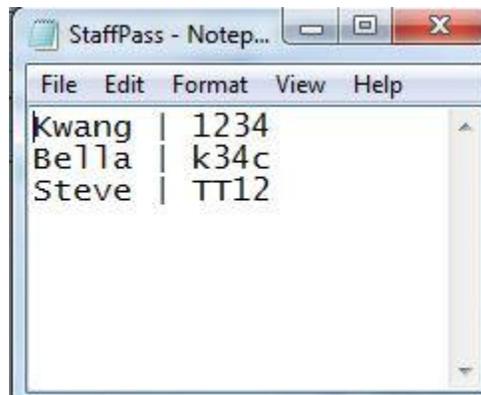
```
HotelName - Notepad
File Edit Format View Help
Aloft Kuala Lumpur Sentral|/Hotel_Review-g298570-d3730389-Reviews-
Aloft_Kuala_Lumpur_Sentral-Kuala_Lumpur_Wilayah_Persekutuan.html
Ambassador Row Hotel Suites by Lanson Place|/Hotel_Review-g298570-d505069-Reviews-
Ambassador_Row_Hotel_Suites_by_Lanson_Place-Kuala_Lumpur_Wilayah_Persekutuan.html
Ascott Kuala Lumpur|/Hotel_Review-g298570-d478348-Reviews-Ascott_Kuala_Lumpur-
Kuala_Lumpur_Wilayah_Persekutuan.html
Berjaya Times Square Hotel|/Hotel_Review-g298570-d308123-Reviews-Berjaya_Times_Square_Hotel-
Kuala_Lumpur_Wilayah_Persekutuan.html
Capitol Hotel|/Hotel_Review-g298570-d305464-Reviews-Capitol_Hotel-
Kuala_Lumpur_Wilayah_Persekutuan.html
Capri by Fraser Kuala Lumpur Malaysia|/Hotel_Review-g298570-d5062918-Reviews-
Capri_by_Fraser_Kuala_Lumpur_Malaysia-Kuala_Lumpur_Wilayah_Persekutuan.html
Concorde Hotel Kuala Lumpur|/Hotel_Review-g298570-d305326-Reviews-
Concorde_Hotel_Kuala_Lumpur-Kuala_Lumpur_Wilayah_Persekutuan.html
Corus Hotel Kuala Lumpur|/Hotel_Review-g298570-d307944-Reviews-Corus_Hotel_Kuala_Lumpur-
Kuala_Lumpur_Wilayah_Persekutuan.html
Crossroads Hotel|/Hotel_Review-g298570-d3545764-Reviews-Crossroads_Hotel-
Kuala_Lumpur_Wilayah_Persekutuan.html
DoubleTree by Hilton Kuala Lumpur|/Hotel_Review-g298570-d1759018-Reviews-
DoubleTree_by_Hilton_Kuala_Lumpur-Kuala_Lumpur_Wilayah_Persekutuan.html
E O Residences Kuala Lumpur|/Hotel_Review-g298570-d3499040-Reviews-
E_O_Residences_Kuala_Lumpur-Kuala_Lumpur_Wilayah_Persekutuan.html
Fraser Place Kuala Lumpur|/Hotel_Review-g298570-d1488716-Reviews-Fraser_Place_Kuala_Lumpur-
Kuala_Lumpur_Wilayah_Persekutuan.html
Furama Bukit Bintang|/Hotel_Review-g298570-d2170088-Reviews-Furama_Bukit_Bintang-
Kuala_Lumpur_Wilayah_Persekutuan.html
Geo Hotel Kuala Lumpur|/Hotel_Review-g298570-d2714737-Reviews-Geo_Hotel_Kuala_Lumpur-
Kuala_Lumpur_Wilayah_Persekutuan.html
Grand Hyatt Kuala Lumpur|/Hotel_Review-g298570-d3436966-Reviews-Grand_Hyatt_Kuala_Lumpur-
Kuala_Lumpur_Wilayah_Persekutuan.html
Grid 9 Hotel|/Hotel_Review-g298570-d3737379-Reviews-Grid_9_Hotel-
Kuala_Lumpur_Wilayah_Persekutuan.html
```

Figure 4.5.5 shows hotel name & address link in HotelName.txt

4.5.4 Staff ID & Password

The system provides security login for user which user needs to login every time start the system. This database only store staff id and staff password.

The following are example of staff id & pass store in database:



```
StaffPass - Notepad
File Edit Format View Help
Kwang | 1234
Bella | k34c
Steve | TT12
```

Figure 4.5.6 shows staff id & password store in StaffPass.txt

CHAPTER 4 SYSTEM DESIGN

4.6 System Graphic User Interface

4.6.1 Login Interface

Login interface which allow authorize user to use the system. This is to make sure other will not access it without permission. To enter the system, user need to enter their staff ID and their own password correctly.



Figure 4.6.1 Login Interface

4.6.2 Main Menu Interface

Main menu interface have three main function which are “WordList”, “Analyze a Sentence” and “Analyze Hotel Comment” , one for exit system and one for edit user ID and password. “WordList” function is allow to create positive and negative wordlist. “Analyze a Sentence” is only analyze single sentnce or paragrp and “Analyze Hotel Comment” is access to website to get comments of the hotel;

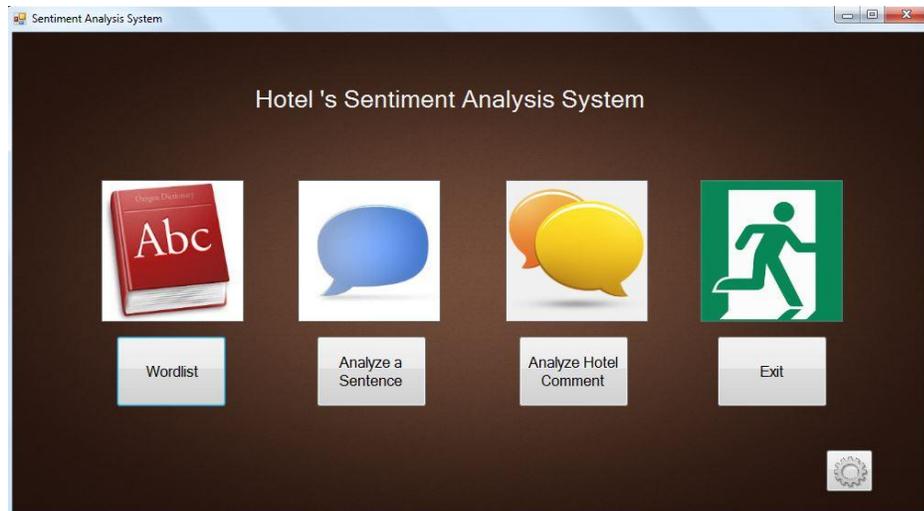


Figure 4.6.2 Main Menu Interface

4.6.3 WordList Interface

WordList interface is to gather positive and negative word and save into text file. The system gathers those words from website which provided positive and negative words. Besides that, at bottom of the window form is allows user to enter their user name and password when accessing proxy network

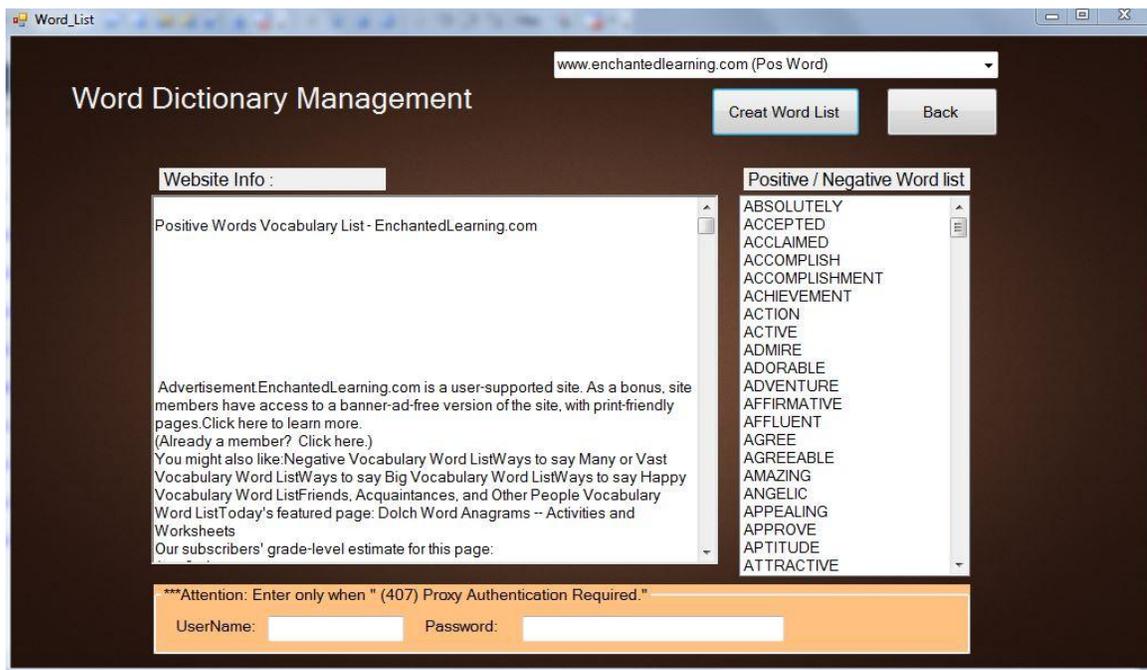


Figure 4.6.3 WordList Interface

CHAPTER 4 SYSTEM DESIGN

4.6.4 Analyze a Sentence Interface

Analyze a Sentence Interface is allow user only analyze a single sentence or paragraph. This form allow user to analyze the sentence with 5 types of solution which are naïve bayes, PMI, PMI2, NTCIR and ES at the drop down list. After analyze, the system shows how many positive and negative words detected and scoring. At last, the system show the 5 rating star, if the result more than zero the star will green color, if result less zero the star will be red.

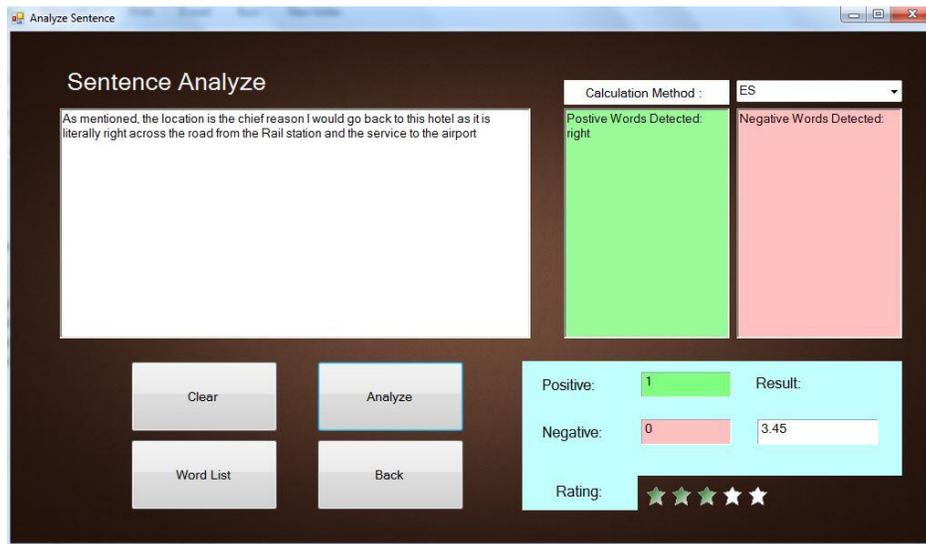


Figure 4.6.4 Analyze a Sentence Interface

4.6.5 Analyze Hotel Comment Interface

This interface is access to website to gather hotel comments which selected by user. The system can compare with 2 hotel comments and show the result to user with graph and rating. Besides that, the system will show overall comment of the hotel which different by extreme positive, highly positive, moderate positive, low positive, slightly positive, neutral, slightly negative, low negative, moderate negative, highly negative, extreme negative.

CHAPTER 4 SYSTEM DESIGN

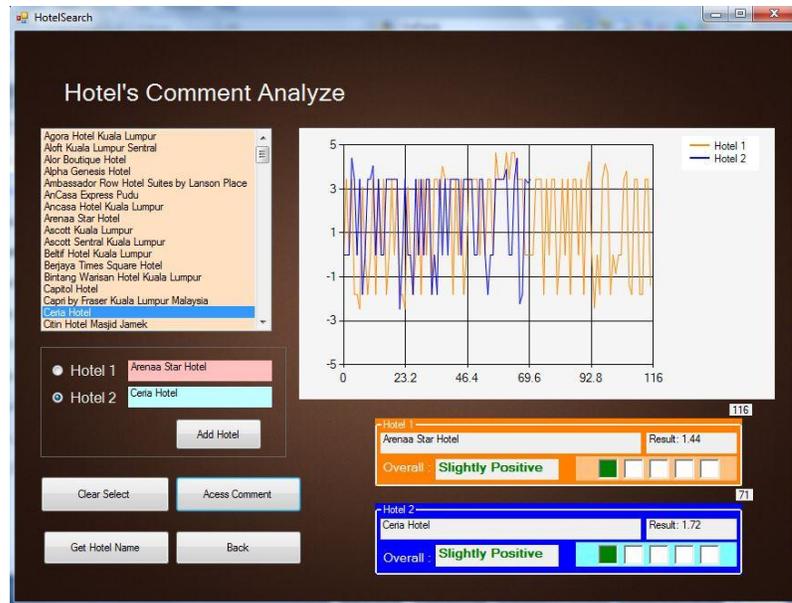


Figure 4.6.5 Analyze Hotel Comment Interface

4.6.6 Edit User Interface

Edit User Interface is a simple form which allow user to add, edit and delete user.



Figure 4.6.6 Edit user Interface

CHAPTER 5 Implementing and Testing

Chapter 5 Implementation and Testing

5.1 Sentiment Value Calculation

To calculate score with sentiment value especially for ES formula, it will need to use sentiment score which categories to 0.333, 0.667 and 1. This is because it would increase system result based on the value.

The following are showing how to get the sentiment value:

Neutral	0	0
Good	1.666667	0.333333
Better	3.333333	0.666667
Best	5	1
1.666667		

Figure 5.1 shows how to calculate sentiment value for the word.

The figure above show there are 3 types of level sentiment which are neutral good, better and best. To identify what value should assign to the words, first need to determine what is the range for rating. In this case, there is 0 until 5 for positive rating range. So, we need to use maximum to minus minimum such as $(5-0)/3$. The reason is because there are only 3 levels of superlative words which not included neutral. The result would be 1.666667 for each level.

After that, change the value to 0 until 1 which represent 0 to 100, to get each value for the level, it would be 1.666667 divide by 5, 3.333333 divide by 5 and 5 divide by 5, so the value within 0 to 1 for each level are 0.333, 0.667 and 1. For example, for “Big” word value is 0.33, “Bigger” word value is 0.667 and “Biggest” word value is 1.

For Naïve Bayes, PMI, PMI2 and NTCIR is only use for number of word detected, so the sentiment value is not required and the value can be ignore which save in data when using these formulas.

CHAPTER 5 Implementing and Testing

These words will save in to word dictionaries database which will use to recognize positive word and negative in the sentences.

The following are example of positive and negative words with value 0.333, 0.667 and 1.00 which save in text file:

Angry 0.333	- angrier 0.667	- angriest 1.000
bad 0.333	- worse 0.667	- worst 1.000
Big 0.333	- bigger 0.667	- biggest 1.000
Brave 0.333	- braver 0.667	- bravest 1.000
Bright 0.333	- brighter 0.667	- brightest 1.000
Broad 0.333	- broader 0.667	- broadest 1.000
Busy 0.333	- busier 0.667	- busiest 1.000
Calm 0.333	- calmer 0.667	- calmest 1.000
clever 0.333	- cleverer 0.667	- cleverest 1.000
clean 0.333	- cleaner 0.667	- cleanest 1.000
clear 0.333	- clearer 0.667	- clearest 1.000
cheap 0.333	- cheaper 0.667	- cheapest 1.000
Cold 0.333	- colder 0.667	- coldest 1.000
common 0.333	- commoner 0.667	- commonest 1.000
Cool 0.333	- cooler 0.667	- coolest 1.000
Curly 0.333	- curlier 0.667	- curliest 1.000
Dark 0.333	- darker 0.667	- darkest 1.000
Dirty 0.333	- dirtier 0.667	- dirtiest 1.000
Dull 0.333	- duller 0.667	- dullest 1.000
Dry 0.333	- drier 0.667	- driest 1.000
Early 0.333	- earlier 0.667	- earliest 1.000
Easy 0.333	- easier 0.667	- easiest 1.000
Evil 0.333	- more evil 0.667	

Figure 5.2 shows assign value for each sentiment word.

At the end, these words will separate and save in to PositiveWord.txt file and PositiveWord.txt file.

5.2 Evaluation Result

There are 5 types of calculation that would use to compare with human result which is Naïve Bayes Theorem, Pointwise Mutual information(PMI), Pointwise Mutual information version 2(PMI2), NTCIR and Enhance Sentiment(ES). Before evaluating the formula, 32 selected questions were sent to 7 people for rating positive and negative based on extremely -5 to extremely +5. The selected 7 people are English based teacher which are primary school teacher, secondary school teacher and university lecture. The human rating result will use as standard to determine which formula is close to human result.

CHAPTER 5 Implementing and Testing

The following are humans rating for 32 questions based on extremely negative/positive -5/+5, highly negative/positive -4/+4, moderate negative/positive -3/+3, low negative/positive -2/+2, slightly negative/positive -1/+1 and neutral 0. Average result is total 7 humans rating score plus together and divided by 7 to the result. For Average_Adj is result that only calculate 6 human rating score which not include human 2 because correlation with system result show human 2 result is below 50%, so it may not able use to compare with system.

Human1	Human2	Human3	Human4	Human5	Human6	Human7	Average	Average_Adj
3	4	3	3	4	4	3	3.43	3.33
4	3	1	4	4	3	5	3.43	3.50
0	-3	-3	-4	-4	2	-2	(2.00)	(1.83)
0	3	-3	-3	-4	-2	-1	(1.43)	(2.17)
3	5	4	3	3	4	4	3.71	3.50
4	0	1	2	5	3	4	2.71	3.17
4	1	4	5	5	4	5	4.00	4.50
3	0	2	4	5	3	5	3.14	3.67
-4	-5	-5	-5	-4	-3	0	(3.71)	(3.50)
-2	-2	-5	-4	-4	-2	-3	(3.14)	(3.33)
-3	-4	-4	-4	-3	-4	-4	(3.71)	(3.67)
-2	-5	-5	-5	-5	-2	-5	(4.14)	(4.00)
-2	4	-5	-3	-5	4	-1	(1.14)	(2.00)
0	0	4	3	4	0	3	2.00	2.33
0	-3	0	-1	-2	-3	-3	(1.71)	(1.50)
-2	-2	-3	-3	-4	-4	-4	(3.14)	(3.33)
-2	-4	0	2	-2	3	0	(0.43)	0.17
0	1	4	2	0	2	2	1.57	1.67
3	3	4	2	2	3	3	2.86	2.83
0	0	-5	-5	-5	-3	-5	(3.29)	(3.83)
3	2	5	4	5	4	5	4.00	4.33
4	2	4	5	5	4	4	4.00	4.33
0	2	-5	-4	-5	-4	-4	(2.86)	(3.67)
-2	-2	-5	-4	-5	-3	-3	(3.43)	(3.67)
0	2	0	4	4	3	5	2.57	2.67
1	2	3	3	4	3	3	2.71	2.83
1	0	-4	1	-5	4	4	0.14	0.17
3	-3	-1	1	-3	4	4	0.71	1.33
0	-3	-1	-5	-5	-3	-1	(2.57)	(2.50)
3	2	2	4	5	4	5	3.57	3.83
4	1	3	5	5	3	5	3.71	4.17
-1	-2	-1	-4	-4	-2	-5	(2.71)	(2.83)

Figure 5.3 shows human rate for 32 questions in survey form.

CHAPTER 5 Implementing and Testing

The following are evaluate for each sentences compare with human result:

	Naïve	PMI	PMI2	NTCIR	ES	HumanAvg
1	1	4.32	3.24	1	4.32	3.43
2	1	5.23	3.92	1	4.32	3.43
3	-1	-2.47	-1.86	-1	-2.25	(2.00)
4	-0.5	-1.58	-1.19	-0.5	-2.04	(1.43)
5	1	2.47	1.86	1	4.32	3.71
6	1	5.06	3.79	1	4.32	2.71
7	1	5.23	3.92	1	4.32	4.00
8	0.33	1	0.75	0.33	3.13	3.14
10	-1	-4.06	-3.04	-1	-2.25	(3.71)
11	-1	-1.6	-1.2	-1	-2.25	(3.14)
12	0	0	0	0	0	(3.71)
13	1	2.4	1.8	1	4.32	(4.14)
14	0	0	0	0	0	(1.14)
15	1	2.56	1.92	1	4.32	2.00
16	-1	-3.98	-2.99	-1	-2.25	(1.71)
17	0	0	0	0	0	(3.14)
19	-0.33	-1	-0.75	-0.33	-1.43	(0.43)
20	-1	-2.8	-2.1	-1	-2.25	1.57
21	1	3.98	2.99	1	4.32	2.86
22	-1	-4.32	-3.24	-1	-2.25	(3.29)
23	1	3.64	2.73	1	4.32	4.00
24	1	5.06	3.79	1	4.32	4.00
25	-1	-2.69	-2.02	-1	-2.77	(2.86)
26	-0.5	-1.58	-1.19	-0.5	-1.73	(3.43)
27	0.33	1	0.75	0.33	3.13	2.57
28	1	2.64	1.98	1	4.32	2.71
29	-1	-2.4	-1.8	-1	-2.25	0.14
30	0	0	0	0	0	0.71
31	-1	-3.06	-2.29	-1	-2.25	(2.57)
32	1	2.47	1.86	1	4.32	3.57
33	1	4.14	3.11	1	4.32	3.71
34	-1	-3.84	-2.88	-1	-2.25	(2.71)

Figure 5.4 shows all formula and human average result.

The human result is average of total rating number. The rest of 5 types formula result is use its own formula method to do the calculation. Naïve Bayes and NTCIR calculation result is only between -1 to +1, but PMI2 and ES result is between -5 to +5 and PMI might more than 5. After the calculation result for each formula, these results will use to compare by using correlation method to check which formula is highest.

There are few questions which all calculation has same result 0 which are question 12, 17 and 30. This situation happened because there are same number of positive and negative occur such as 1 positive and 1 negative word occur in one sentences. But human average result is not 0.

CHAPTER 5 Implementing and Testing

The following are correlation for 5 type's calculation with human result.

	Human1	Human2	Human3	Human4	Human5	Human6	Human7	HumanAvg	HumanAvgAdj
Naïve	0.63	0.48	0.66	0.71	0.80	0.59	0.64	0.74	0.74
PMI	0.70	0.51	0.66	0.74	0.81	0.64	0.68	0.78	0.78
PMI2	0.70	0.51	0.66	0.74	0.81	0.64	0.68	0.78	0.78
NTCIR	0.63	0.48	0.66	0.71	0.80	0.59	0.64	0.74	0.74
ES	0.67	0.49	0.70	0.75	0.86	0.61	0.68	0.78	0.79
	0.67	0.49	0.67	0.73	0.82	0.61	0.66		

Figure 5.5 shows correlation result between formula and human.

Based on the above result, human 2 result is eliminate and will not use to compare with the formula because the average score is below 50%, it define as no accurate. Then other 6 set of human result will use to compare with system which is human1, human3, human4, human5, human6 and human7 and the correlation result is under “HumanAvgAdj” column. On the other hand, “HumanAvg” column is calculating correlation result with 7 human results. Based on the table ES and human3, 4, 5 and 7 correlation result is highest compare to other. Moreover, ES correlation with human1 and 6 also is not lowest. HumanAvgAdj correlation result of ES is highest which is 0.79 compare to PMI and PMI2 which is 0.78, but Naïve and NTCIR are only 0.74. However, correlation result of HumanAvg for PMI, PMI2 and ES are same which 0.78 and only Naïve and NTCIR are remain same value which 0.74. But only HumanAvgAdj correlation result will be use as final result.

However, after add the intensifier grammars recognize function such as very, so and quite, the systems results have been increase.

CHAPTER 5 Implementing and Testing

The following are show increased systems results:

	Naïve	PMI	PMI2	NTCIR	ES	HumanAvg
1	1	4.32	3.24	1	3.45	3.43
2	1	5.23	3.92	1	3.45	3.43
3	-1	-2.47	-1.86	-1	-1.8	(2.00)
4	-0.5	-1.58	-1.19	-0.5	-2.05	(1.43)
5	1	2.47	1.86	1	3.45	3.71
6	1	5.06	3.79	1	3.45	2.71
7	1	5.23	3.92	1	3.45	4.00
8	0.33	1	0.75	0.33	2.5	3.14
10	-1	-4.06	-3.04	-1	-1.8	(3.71)
11	-1	-1.6	-1.2	-1	-1.8	(3.14)
12	0	0	0	0	-2.05	(3.71)
13	1	2.4	1.8	1	3.45	(4.14)
14	0	0	0	0	0	(1.14)
15	1	2.56	1.92	1	3.45	2.00
16	-0.33	-1	-0.75	-0.33	-1.14	(1.71)
17	0	0	0	0	0	(3.14)
19	-0.33	-1	-0.75	-0.33	-1.14	(0.43)
20	-0.33	-1	-0.75	-0.33	-1.14	1.57
21	1	3.98	2.99	1	3.45	2.86
22	-1	-4.32	-3.24	-1	-1.8	(3.29)
23	1	3.64	2.73	1	3.45	4.00
24	1	5.06	3.79	1	4.66	4.00
25	-1	-2.69	-2.02	-1	-2.22	(2.86)
26	-0.5	-1.58	-1.19	-0.5	-1.39	(3.43)
27	0.33	1	0.75	0.33	2.5	2.57
28	1	2.64	1.98	1	3.45	2.71
29	-1	-2.4	-1.8	-1	-1.8	0.14
30	0	0	0	0	0	0.71
31	-1	-3.06	-2.29	-1	-1.8	(2.57)
32	1	2.47	1.86	1	3.45	3.57
33	1	4.14	3.11	1	3.45	3.71
34	-1	-3.84	-2.88	-1	-1.8	(2.71)

Figure 5.6 shows all formula and human average result after recognize intensifier grammars.

Based on each formula result, after added intensifier grammars recognize function into system, ES results have complete change except for the 0 value. For the PMI, PMI2, Naïve Bayes and NTCIR, the results are only partially changed.

The following are improved correlation for 5 type's calculation with human result.

	Human1	Human2	Human3	Human4	Human5	Human6	Human7	HumanAvg	HumanAvgAdj
Naïve	0.64	0.48	0.72	0.75	0.83	0.60	0.65	0.77	0.77
PMI	0.72	0.50	0.72	0.78	0.84	0.63	0.68	0.80	0.80
PMI2	0.72	0.50	0.72	0.78	0.84	0.63	0.68	0.80	0.80
NTCIR	0.64	0.48	0.72	0.75	0.83	0.60	0.65	0.77	0.77
ES	0.71	0.51	0.74	0.78	0.87	0.65	0.70	0.82	0.82
	0.69	0.50	0.72	0.77	0.84	0.62	0.67		

Figure 5.7 shows correlation result between formula and human after recognize intensifier grammars.

CHAPTER 5 Implementing and Testing

Base on correlation results, added intensifier grammars recognize function into system, the human 2 result is at 50% compare to the system without recognize intensifier. Also, the “HumanAvg” and “HumanAvgAdj” result for 5 types of formula also increase. For Naïve Bayes and NTCIR result has increased from 0.74 to 0.77. PMI and PMI2 result are increased from 0.78 to 0.80. At the same time, ES result 0.78 for “HumanAvg” and 0.79 for “HumanAvgAdj” has increased to 0.82 which still remain the highest value in correlation results.

5.3 Challenges

The challenges of developing this project are:

1. The system need to analysis and calculate high accuracy of the result which is 80% compare with human result.
2. Need to find appropriate formulae to calculate sentiment scoring or may need built or edit new formula to improve accuracy.
3. The system needs to access some website that provide positive and negative words by decoding the HTML code and save all the positive and negative words.
4. Need to find English teacher to do the survey form for evaluation part.
5. The system only can access hotel comments from tripadvisor website.
6. The system do not have very complete word dictionary to analyze sentence.

CHAPTER 6 CONCLUSION

Chapter 6 Conclusion

6.1 Conclusion

In conclusion, Sentiment analysis known as opinion mining which is a kind of process indentify large number of posts, comments and online news whether the sentiment is positive, negative or neural through natural language processing. The problem of sentiment analysis is most of the system need to create word dictionary manually and system does not access multiple hotel forums to extract the hotel service comment. Also, the result accuracy may reduce when the sentences are too long and complex. Scope of this project are automatically create positive and negative word dictionary and study sentiment scoring formula to evaluate which is better. The existing of sentiment analysis algorithm are Semantics Analysis, Nasent Algorithm, Simple keyword-based methodology, text parser and Naive Bayes Theorem, Pointwise Mutual information and NTCIR-6. Methodology of developing this project is using V-Model, this is because time saving and quick implement. It also performs well testing on planning and design activities before coding and able to find defect at early stage. Besides that, C# programming language is use to develop the system and text file will be use to create word dictionary. To test system result, the data of comment will get from website such as Tripadvisor.com. To create word dictionary, the system will access to website to get positive and negative words. The implementation issue and challenges of this project are system may need access to other forum website to get customer comment. Moreover, system needs to get positive and negative words from other website by decoding the html codes and save into word dictionary. For evaluation, the question will send to English based teacher for rating, then the result will use to compare with the system result. After comparison, the correlation result shows that the highest value is ES which 0.79 more close to human rating compare to rest of formula. In additional, the result of 5 types of formula has improve after recognize intensifier words and ES result still remain highest value which is 0.82 and Naïve Bayes and NTCIR still lowest value at 0.77 for both “HumanAvg” and “HumanAvgAdj”.

CHAPTER 6 CONCLUSION

6.2 Problem Encountered and Solution

6.2.1 System can't access network with proxy setting

When the computer connects to network that has proxy setting such as UTAR computer lab, the system will not be able to access some website to gather data. For example, the create word list function will not be able to access to website that provide positive and negative words such as “enhancedlearning.com”. To solve the problem, the function needs to use “proxy.Credentials” which to allow user enter authorizes user name and password to access the internet connection.

6.2.2 Error may occur when network speed slow

During accessing hotel name or comments, the system need to wait the “webBrowser” in visual studio finish load, then it only can access the HTML source. This is because when the network connection is slow and the browser not yet finish loaded, the system will run the following code to extract the HTML code, then the error will occur. To solve the problem is to wait “webBrowser” finish load, there will need to be use pop out and close message box automatically to wait the browser finish load. After pop out message box, the time waiting “webBrowser” finish load is around 13second, after that the message box will be close and continue to access other comment and repeat same step.

6.3.3 Internet Explorer browser may pop out

When opening the access hotel comment form and close it immediately, Internet Explorer browser will pop out. Sometimes it will pop out once but sometimes it will continue pop out continuously. The pop out Internet Explorer browser link is same with the system linking website. To avoid this problem is when run the form better is waiting form few second or do not close immediately.

CHAPTER 6 CONCLUSION

6.3 Discussion

Even though current system result is effectively but the system still have limit ability to recognize positive and negative in sentence due to limited bag of words. The ability of system is recognizing normal positive and negative of noun, verb and adjectives words. It also recognizes some of negation words such not and no and also for superlative words such as good, better and best. It stills a challenge to develop complete rule for negation calculation. The system added intensifier grammar would improve the result better. After access the hotel comments the result can be show in rate, numeric result and graph.

6.4 Future Work

To improving system result, word dictionaries should be improve as much as possible and also improve negation word. However, recognize ambiguity words function should added into system instead of just focusing normal words. Also, design the system able to recognize n-grams words such as low price, never lose and great service which could help to differentiate the topic of sentence. For example, “the hotel service is nice and good location”. Mean “nice” is referring to hotel service and “good” refer to location.

References

1. Datapig, 2014. Quantifying Subjective Text with an Excel-Based Sentiment Analysis Tool. Available from: <<http://datapigtechnologies.com/blog/index.php/quantifying-subjective-text-with-sentiment-analysis/>>. [12 November 2014].
2. Enhanced learning, 2010. Available from: <<http://www.enchantedlearning.com/wordlist/positivewords.shtml>>[14 November 2014].
3. On Social Sentiment and Sentiment Analysis, 2013. Available from: <<http://brnr.me/social-sentiment-sentiment-analysis/>>. [11 November 2014].
4. Rober B Heckedorn, n.d., Grammar Tutorial. Available from: <<http://marvin.cs.uidaho.edu/Handouts/grammar.html>>.[18 November 2014].
5. Sentiment Analysis and application: A survey, 2014. Available from: <<http://www.sciencedirect.com/science/article/pii/S2090447914000550>>. [12 November 2014].
6. Slide Hunter, 2014. Available from: <<http://slidehunter.com/powerpoint-templates/spider-diagram-powerpoint-template/>>[14 November 2014].
7. Statistical Parsing of English Sentences, 2006. Available from:<<http://www.codeproject.com/Articles/12109/Statistical-parsing-of-English-sentences>>. [17 November 2014].
8. Tom Abate, 2013, Stanford Algorithm Analyzes Sentence Sentiment, Advances Machine Learning. Available from: <<http://engineering.stanford.edu/news/stanford-algorithm-analyzes-sentence-sentiment-advances-machine-learning>>. [12 November 2014].
9. Tripadvisor,2014.Available from:<http://www.tripadvisor.com.my/Restaurant_Review-g35805-d5600612-Reviews-Eataly-Chicago_Illinois.html>[14 November 2014].
10. Software Testing Class, 2012. Available from:<<http://www.softwaretestingclass.com/v-model/>>. [26 February 2015].

11. Vasilis Vryniotis, 2013, Machine Learning Tutorial: The Naïve Bayes Text Classifier. Available From: <<http://blog.datumbox.com/machine-learning-tutorial-the-naive-bayes-text-classifier/>>. [27 February 2015]
12. ISTQB EXAM CERTIFICATION, n.d. Available from: <<http://istqbexamcertification.com/what-is-waterfall-model-advantages-disadvantages-and-when-to-use-it/>>. [27 February 2015]
13. ISTQB EXAM CERTIFICATION, .n.d. Available from: <<http://istqbexamcertification.com/what-is-incremental-model-advantages-disadvantages-and-when-to-use-it/>>. [27 February 2015]

Appendix A: Survey Form
A-1 Sample of Survey Form

Sentiment Sentence Survey (FYP2)

Name :

Short Description:

Dear Mr / Ms,

I am IB student from FICT faculty. Currently, I develop a system to analysis the sentiment of sentences from hotel website. So, I need some data to compare with my system result. The purpose of this survey form is to gather all human rating rates from the below sentences and it would be used as system result for benchmark and evaluation in FYP2. I appreciate your contribution. Thank you.

1. Slightly <u>Positive</u>	2. Low <u>Positive</u>	3. Moderate <u>Positive</u>	4. Highly <u>Positive</u>	5. Extreme <u>Positive</u>
		0. Neutral		
-1. Slightly <u>Negative</u>	-2. Low <u>Negative</u>	-3. Moderate <u>Negative</u>	-4. Highly <u>Negative</u>	-5. Extreme <u>Negative</u>

QUESTION: Please put a “√” as selection

1. I like the buffet breakfast.

-5	-4	-3	-2	-1	0	1	2	3	4	5

2. Wifi is good, clean room & friendly staff.

-5	-4	-3	-2	-1	0	1	2	3	4	5

3. I was disappointed when I checked in I was informed that the room I choose was without window.

-5	-4	-3	-2	-1	0	1	2	3	4	5

4. When I talk to the reception she said it's the only room is available now after talking for a while she give me another room but wasn't much better than the first one.

-5	-4	-3	-2	-1	0	1	2	3	4	5

5. The hotel is at the center of shopping malls and restaurants, and can easily be accessed thru train.

-5	-4	-3	-2	-1	0	1	2	3	4	5

6. Simply love it.

-5	-4	-3	-2	-1	0	1	2	3	4	5

7. The hotel is lovely, room perfect, breakfast excellent.

-5	-4	-3	-2	-1	0	1	2	3	4	5

8. It was spacious and immaculate.

-5	-4	-3	-2	-1	0	1	2	3	4	5

9. I was disappointed when I checked in I was informed that the room I chose was without window.

-5	-4	-3	-2	-1	0	1	2	3	4	5

10. Toilet was at a bad angle.

-5	-4	-3	-2	-1	0	1	2	3	4	5

11. After checking with the cook, the waiter informed us the cook just started cooking for another table and after that he will start our orders, therefore we were forced to cancel and left.

-5	-4	-3	-2	-1	0	1	2	3	4	5

12. My room was big enough but very, very noisy as it was facing a busy road.

-5	-4	-3	-2	-1	0	1	2	3	4	5

13. The car park stink with urea smell as I believe people are easing themselves there rather than the toilet.

-5	-4	-3	-2	-1	0	1	2	3	4	5

14. I hope that they change some things there or I think they will start to lose more customers.

-5	-4	-3	-2	-1	0	1	2	3	4	5

15. Had the pleasure of staying at the hotel for a couple of days when I visited KL.

-5	-4	-3	-2	-1	0	1	2	3	4	5

16. We noticed the TV channels are limited but as we are not TV fanatic, it did not bother us.

-5	-4	-3	-2	-1	0	1	2	3	4	5

17. Noisy tourists leaving their doors open n shouting on the corridors.

-5	-4	-3	-2	-1	0	1	2	3	4	5

18. Toilet was at a bad angle.

-5	-4	-3	-2	-1	0	1	2	3	4	5

19. There is no in-room safe, however you can utilize the safe deposit boxes at the reception desk.

-5	-4	-3	-2	-1	0	1	2	3	4	5

20. When I complained they did not have any other room of the same type available and I had lost hopes when just in a few hours they managed to relocate me to a room in the upper floors that had a awesome view.

-5	-4	-3	-2	-1	0	1	2	3	4	5

21. I book this hotel online and select my choice with a picture of a nice room with a window.

-5	-4	-3	-2	-1	0	1	2	3	4	5

22. I will never stay there.

-5	-4	-3	-2	-1	0	1	2	3	4	5

23. The location of this hotel is perfect, right next to Pasar Seni LRT station and Kuala Lumpur station, as well as being a one minute walk to the central market and Chinatown.

-5	-4	-3	-2	-1	0	1	2	3	4	5

24. Staff was very accommodating and helpful.

-5	-4	-3	-2	-1	0	1	2	3	4	5

25. Despite the fact we got upgraded when we arrived to the old part of the hotel) this is the worst hotel i have experienced over the last 10 years or so.

-5	-4	-3	-2	-1	0	1	2	3	4	5

26. The lift is cranky and I'm in constant fear it will break down.

-5	-4	-3	-2	-1	0	1	2	3	4	5

27. Good reasonable hotel good area close to all the shopping, clubs, etc remember do not go.

-5	-4	-3	-2	-1	0	1	2	3	4	5

28. He would give me weather updates so I didn't get caught in the evening storms, and also recommended the local Chinese restaurant (next door) to enjoy a tasty, cheap and local dinner.

-5	-4	-3	-2	-1	0	1	2	3	4	5

29. For the price that was charged, could be improved but I doubt I will be back at this hotel.

-5	-4	-3	-2	-1	0	1	2	3	4	5

30. If you don't mind walking from hotel to the local rail station to travel and you want to do KL on a budget then yes.

-5	-4	-3	-2	-1	0	1	2	3	4	5

31. No Newspapers in the lounge area even though the racks are there.

-5	-4	-3	-2	-1	0	1	2	3	4	5

32. We were impressed by the cleanliness of the property and the time staff took to clean the apartment.

-5	-4	-3	-2	-1	0	1	2	3	4	5

33. It was my first visit there and I loved it, the rooms were clean, big and comfortable.

-5	-4	-3	-2	-1	0	1	2	3	4	5

34. Cons- location- not easy to walk from the hotel as lots dual carriage ways.

-5	-4	-3	-2	-1	0	1	2	3	4	5

35. The lift is cranky and I'm in constant fear it will break down.

-5	-4	-3	-2	-1	0	1	2	3	4	5

----- END QUESTION -----

Appendix B: Weekly Report

B-1 Final Year Project 2 Weekly Report

Trimester, Year: Year 3 Sem 2	Study Week no: Week 1
Student Name & ID: Kwang Chyun Yaw 13ACB00854	
Supervisor: MR. Albert Yong Tien Fui	
Project Title: Sentiment Analysis of Hotel Service System	

<p>1. WORK DONE</p> <p>-</p>
<p>2. WORK TO BE DONE</p> <p>- Search sentiment analysis algorithm for deep understanding and discussion with supervisor.</p>
<p>3. PROBLEMS ENCOUNTERED</p> <p>-</p>
<p>4. SELF EVALUATION OF THE PROGRESS</p> <p>- Most of articles are repeat same things or method on same algorithm</p>

Supervisor's signature

Student's signature

Final Year Project 2 Weekly Report

Trimester, Year: Year 3 Sem 2	Study Week no: Week 2
Student Name & ID: Kwang Chyun Yaw 13ACB00854	
Supervisor: MR. Albert Yong Tien Fui	
Project Title: Sentiment Analysis of Hotel Service System	

1. WORK DONE

- Found few sentiment analysis algorithm or formula and understand it

2. WORK TO BE DONE

- Try to apply formula into FYPII and present to supervisor

3. PROBLEMS ENCOUNTERED

- Problem to understand the formula

4. SELF EVALUATION OF THE PROGRESS

- Still have problem on ways of formula works, need to consult with supervisor

Supervisor's signature

Student's signature

Final Year Project 2 Weekly Report

Trimester, Year: Year 3 Sem 2	Study Week no: Week 3
Student Name & ID: Kwang Chyun Yaw 13ACB00854	
Supervisor: MR. Albert Yong Tien Fui	
Project Title: Sentiment Analysis of Hotel Service System	

<p>1. WORK DONE</p> <ul style="list-style-type: none">- Still working with formula
<p>2. WORK TO BE DONE</p> <ul style="list-style-type: none">- Search more formula to work on it and consult supervisor.
<p>3. PROBLEMS ENCOUNTERED</p> <ul style="list-style-type: none">- No sure how to apply formula into FYP2- Much explanation how formulae works but can't find any sample calculation
<p>4. SELF EVALUATION OF THE PROGRESS</p> <ul style="list-style-type: none">- Can't confirm the formula apply is correct.

Supervisor's signature

Student's signature

Final Year Project 2 Weekly Report

Trimester, Year: Year 3 Sem 2	Study Week no: Week 4
Student Name & ID: Kwang Chyun Yaw 13ACB00854	
Supervisor: MR. Albert Yong Tien Fui	
Project Title: Sentiment Analysis of Hotel Service System	

<p>1. WORK DONE</p> <ul style="list-style-type: none">- Found out how formula work
<p>2. WORK TO BE DONE</p> <ul style="list-style-type: none">- Write a system function run the formula
<p>3. PROBLEMS ENCOUNTERED</p> <ul style="list-style-type: none">- Can't confirm the way of formulas works
<p>4. SELF EVALUATION OF THE PROGRESS</p> <ul style="list-style-type: none">- After consult supervisor, figure out how the formula works and make sure it have work correctly

Supervisor's signature

Student's signature

Final Year Project 2 Weekly Report

Trimester, Year: Year 3 Sem 2	Study Week no: Week 5
Student Name & ID: Kwang Chyun Yaw 13ACB00854	
Supervisor: MR. Albert Yong Tien Fui	
Project Title: Sentiment Analysis of Hotel Service System	

1. WORK DONE

- Write formula into system

2. WORK TO BE DONE

- Search sample data and random select sentences to prepare survey form
- Give Supervisor check survey form

3. PROBLEMS ENCOUNTERED

- -----

4. SELF EVALUATION OF THE PROGRESS

- At this moment every things still work fine.

Supervisor's signature

Student's signature

Final Year Project 2 Weekly Report

Trimester, Year: Year 3 Sem 1	Study Week no: Week 6
Student Name & ID: Kwang Chyun Yaw 13ACB00854	
Supervisor: MR. Albert Yong Tien Fui	
Project Title: Sentiment Analysis of Hotel Service System	

1. WORK DONE

- Survey form prepared and formula function in system are ready

2. WORK TO BE DONE

- Invite English lecture or teacher to do survey form and correlation human result with system result
- Use system to evaluate survey questions and show result to supervisor.

3. PROBLEMS ENCOUNTERED

- Mostly lecture no respond to invitation

4. SELF EVALUATION OF THE PROGRESS

- Taking too many days to survey.
- Need to find more lecture or teacher from outside UNI to do the survey form.

Supervisor's signature

Student's signature

Final Year Project 2 Weekly Report

Trimester, Year: Year 3 Sem 1	Study Week no: Week 7
Student Name & ID: Kwang Chyun Yaw 13ACB00854	
Supervisor: MR. Albert Yong Tien Fui	
Project Title: Sentiment Analysis of Hotel Service System	

1. WORK DONE

- Receive done survey form lecture or teacher
- System have evaluate survey question.

2. WORKTO BE DONE

- Enter those survey result into excel to compare with system result.
- Show comparison result to supervisor.

3. PROBLEMS ENCOUNTERED

- The ideal to implement calculation into sentiment analysis

4. SELF EVALUATION OF THE PROGRESS

- Need more research on how to implement and need solve the problem before submission date

Supervisor's signature

Student's signature

Final Year Project 2 Weekly Report

Trimester, Year: Year 3 Sem 2	Study Week no: Week 8
Student Name & ID: Kwang Chyun Yaw 13ACB00854	
Supervisor: MR. Albert Yong Tien Fui	
Project Title: Sentiment Analysis of Hotel Service System	

1. WORK DONE

- Enter human result into excel and compare with system result

2. WORK TO BE DONE

- Figure out which formula result is most close to human result
- Evaluate result to supervisor

3. PROBLEMS ENCOUNTERED

- Even the highest result also not really close to human result.
- Lack of bag of word effect the system result
- Develop own new formula to enhance the result.

4. SELF EVALUATION OF THE PROGRESS

- Some of human evaluate are big different with others.

Supervisor's signature

Student's signature

Final Year Project 2 Weekly Report

Trimester, Year: Year 3 Sem 2	Study Week no: Week 9
Student Name & ID: Kwang Chyun Yaw 13ACB00854	
Supervisor: MR. Albert Yong Tien Fui	
Project Title: Sentiment Analysis of Hotel Service System	

5. WORK DONE

- Built new formula and improve bag of words

6. WORK TO BE DONE

- Make new formula to be most closely to human result compare to others by do some adjustment.
- Show result again to supervisor

7. PROBLEMS ENCOUNTERED

- The new formula result cant close to human result

8. SELF EVALUATION OF THE PROGRESS

- New formula result is almost similar with highest result of existing formula

Supervisor's signature

Student's signature

Final Year Project 2 Weekly Report

Trimester, Year: Year 3 Sem 2	Study Week no: Week 10
Student Name & ID: Kwang Chyun Yaw 13ACB00854	
Supervisor: MR. Albert Yong Tien Fui	
Project Title: Sentiment Analysis of Hotel Service System	

<p>9. WORK DONE</p> <ul style="list-style-type: none">- New formula result is the highest value which more close to human result
<p>10. WORKTO BE DONE</p> <ul style="list-style-type: none">- Come out correlation result for evaluation- Give supervisor check correlation result
<p>11. PROBLEMS ENCOUNTERED</p> <ul style="list-style-type: none">- No problem encountered
<p>12. SELF EVALUATION OF THE PROGRESS</p> <ul style="list-style-type: none">- Progress run well

Supervisor's signature

Student's signature

Final Year Project 2 Weekly Report

Trimester, Year: Year 3 Sem 2	Study Week no: Week 11
Student Name & ID: Kwang Chyun Yaw 13ACB00854	
Supervisor: MR. Albert Yong Tien Fui	
Project Title: Sentiment Analysis of Hotel Service System	

13. WORK DONE

- Evaluate correlation result for each formula in report

14. WORK TO BE DONE

- Finalize all things and complete report.

15. PROBLEMS ENCOUNTERED

- No problem encountered

16. SELF EVALUATION OF THE PROGRESS

- Progress run well

Supervisor's signature

Student's signature

Appendix C: Survey Form Result