FRAMEWORK FOR LAPTOP REVIEW ANALYSIS

Mr. Thanapat Chatchaithanawat

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science Program in Computer Science and Information Technology
Department of Mathematics and Computer Science
Faculty of Science
Chulalongkorn University
Academic Year 2015

กรอบงานสำหรับการวิเคราะห์บทวิจารณ์แล็ปท็อป

นายธนพัฒน์ ฉัตรชัยธนวัฒน์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาวิทยาการคอมพิวเตอร์และเทคโนโลยีสารสนเทศ ภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์
คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2558

Thesis Title	FRAMEWORK FOR LAPTOP REVIEW ANALYSIS
By	Mr. Thanapat Chatchaithanawat
Field of Study	Computer Science and Information Technology
Thesis Advisor	Pakawan Pugsee, Ph.D.

Accepted by the Faculty of Science, Chulalongkorn University in Partial Fulfillment of the Requirements for the Master's Degree

....Dean of the Faculty of Science

(Associate Professor Polkit Sangvanich, Ph.D.)

THESIS COMMITTEE

....Chairman

(Associate Professor Nagul Cooharojananone, Ph.D.)

....Thesis Advisor

(Pakawan Pugsee, Ph.D.)

....External Examiner

(Kanokwan Atchariyachanvanich, Ph.D.)

ธนพัฒน์ ฉัตรชัยธนวัฒน์ : กรอบงานสำหรับการวิเคราะห์บทวิจารณ์แล็ปท็อป (FRAMEWORK FOR LAPTOP REVIEW ANALYSIS) อ.ที่ปรึกษาวิทยานิพนธ์หลัก: ดร. ภควรรณ ปักษี, 103 หน้า.

งานวิจัยนี้พัฒนากรอบงานที่ทำให้ผู้ใช้เข้าใจถึงสิ่งที่บทวิจารณ์แล็ปท็อปกล่าวถึง วัตถุประสงค์ของงานวิจัยคือ นำเสนอกรอบงานสำหรับการวิเคราะห์บทวิจารณ์แล็ปท็อป กรอบงานนี้ประกอบด้วยสี่ส่วนหลัก ได้แก่ การเตรียมข้อมูลสำหรับการวิเคราะห์ การตรวจจับย่อหน้าข้อความที่มีประโยคแสดงความคิดเห็น การระบุหมวดหมู่ให้กับข้อความแต่ละย่อหน้า และการจำแนกอารมณ์ความรู้สึกของข้อความแต่ละย่อหน้า โดยย่อหน้าที่มีประโยคแสดงความคิดเห็นใช้การตรวจจับคำอัตวิสัย (subjective words) ของประโยคในย่อหน้านั้น ถัดมาย่อหน้าที่มีประโยคแสดงความคิดเห็นจะถูกแบ่งออกเป็นหมวดหมู่ตามแต่ละลักษณะ (aspect domain) โดยการเปรียบเทียบกับคำศัพท์ในแต่ละหมวดหมู่ ขั้นตอนสุดท้ายข้อความในแต่ละย่อหน้าจะถูกจำแนกออกเป็นย่อหน้าที่มีความคิดเห็นด้านบวกหรือย่อหน้าที่มีความคิดเห็นด้านลบ โดยใช้ตัวจำแนกข้อความนาอีฟเบย์ (Naive Bayes classifier) ผลการทดสอบกรอบงานนี้แสดงให้เห็นว่าการตรวจจับข้อความที่มีคำอัตวิสัยและการระบุหมวดหมู่ของย่อหน้าข้อความมีความถูกต้องมากกว่า 90% และการจำแนกอารมณ์ความรู้สึกมีความถูกต้องมากกว่า 77% สรุปได้ว่ากรอบงานนี้เป็นประโยชน์สำหรับนำไปพัฒนาระบบวิเคราะห์บทวิจารณ์แล็ปท็อปเพื่อช่วยในการตัดสินใจของผู้บริโภค (consumer) ก่อนการเลือกซื้อ

ภาควิชา	คณิตศาสตร์และวิทยาการคอมพิวเตอร์	ลายมือชื่อนิสิต	....
		ลายมือชื่อ อ.ที่ปรึกษาหลัก	....
สาขาวิชา	วิทยาการคอมพิวเตอร์และเทคโนโลยีสารสนเทศ		
ปีการศึกษา	2558		

5772613623 : MAJOR COMPUTER SCIENCE AND INFORMATION TECHNOLOGY
KEYWORDS: SENTIMENT ANALYSIS / REVIEW ANALYSIS / OPINION MINING / MACHINE LEARNING

THANAPAT CHATCHAITHANAWAT: FRAMEWORK FOR LAPTOP REVIEW ANALYSIS. ADVISOR: PAKAWAN PUGSEE, Ph.D., 103 pp.

This research aims to develop a framework which enables users to be aware that what the reviews on the laptop mentioned. The objective of the research is to propose the framework for the analysis of laptop reviews. The framework consists of four main processes: preparing data for analysis, detecting subjective text paragraphs, identifying the aspect of each text paragraph and classifying the sentiment of each text paragraphs. The subjective text paragraphs are found by detecting subjective words in the sentences of each paragraph. Then, only the subjective paragraphs will be categorized into each category by comparing with the vocabulary in each aspect domain. Finally, the sentiment of paragraphs will be classified into positive and negative opinions by the Naïve Bayes classifier. The test results show that the accuracy of the subjective detection and the aspect identification of the text paragraph are more than 90% and the accuracy of sentiment classification is more than 77%. In summary, this framework is helpful in the development of the analysis system of reviews on a laptop to help consumers in making a decision before purchasing.

Department:	Mathematics and Computer Science	Student's Signature	....
		Advisor's Signature	....
Field of Study:	Computer Science and Information Technology		
Academic Year:	2015		

ACKNOWLEDGEMENTS

This Thesis has been completed with supports from many people, I would like to express my profound gratitude to Lect. Dr.Pakawan Pugsee, my advisor, for a suggestion, inspiration and enthusiastic encouragement to guide me through any issues happened during the making of this Thesis.

I would like to thank the Thesis committee, Assoc. Prof. Dr.Nagul Cooharojananone and Lect. Dr.Kanokwan Atchariyachanvanich for an exhortation in this Thesis.

My special thanks are reached out to my friends, junior and senior in my laboratory and department, who had helped in regards to examinations and documentation.

Finally, I wish to thank my parents for their support and consolation all through my study.

CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

1.1 Problem Statement

Currently, the market of laptop has become more competitive due to the rapid growth of mobile devices and tablets. There are many brand manufactures in the laptop industry and they continually create many new laptop series to compete each other. For this reason, consumers have to deal with the problem in making decision for purchasing laptops. Although consumers can find information about laptops in the marketplace from review forums on blogs, websites or online communities, consumers have to spend lots of time to read and search for information they need. The previous reasons are the motivations for developing the system that help consumers choose the right laptop to buy. The system will gather review data from various review forums and analyze the useful information for the consumer by using the sentiment analysis.

In order to read people’s mind through text messages, sentiment analysis is an essential process to identify people's opinions expressed through texts. The process of sentiment analysis includes different fields of knowledge ranging from natural language processing, artificial intelligence and text mining. The main objective of sentiment analysis is to distinguish the opinion of a source text into positive opinion or negative opinion. The opinions expressed in the texts could be judgments, evaluations, affective states, beliefs or wishes [1].

For those purposes, sentiment analysis will be applied to analyze reviews in this paper. Nowadays sentiment analysis can be used to analyze consumer opinions in order to check consumers’ satisfaction. Moreover, this technique can be used for a market survey in order to understand consumer's needs and increase the efficiency of consumer service and the company’s ability in competing. For example, if a laptop company is able to perceive consumer opinions in both pros and cons, then the company can make more improvement in their products and fulfill consumer needs.

Most reviews on community websites about laptops, such as notebookreview.com, laptopmag.com, cnet.com, pcmag.com, and notebookcheck.net, are composed of laptops' information about performance, design and features. Therefore, this research studies on review analysis about laptops in three aspects which are the product performance, the design of a product and the features of a product. A framework of laptop review analysis is proposed to implement the automatic analysis system.

1.2 Objective

Propose a framework for laptop review analysis to classify the sentiments of text paragraphs of laptop reviews, including the aspect identification.

1.3 Scope of Thesis

In this study, the classification system is constrained as follows:

1. The reviews have been gathered from 9 major laptop brands; Apple, Dell, HP, MSI, Samsung, Lenovo, Asus, Toshiba and Acer. Information source is from www.notebookreview.com and www.cnet.com because these websites are more visitors than other websites. Each brand contains 40 review topics which make up 360 review topics in total.
2. The input data is only in text format (*.txt)
3. Price of the laptop is not included in this research.

1.5 Expected Outcomes

This framework can automatically classify sentiments of subjective paragraphs in each aspect domain of laptop reviews and show results into 2 categories; positive and negative paragraphs.

1.4 Structure of the Thesis

The structure of this Thesis will be described in the following; the literature reviews and related works will be presented in Chapter 2. The Chapter 3 will show the methodology. The experimental results will be described in Chapter 4. Finally, conclusion, discussions and limitations will be shown in Chapter 5.

CHAPTER 2
FUNDAMENTAL KNOWLEDGE AND LITERATURE REVIEW

This chapter consists of information about fundamental knowledge and literature reviews which support this thesis. The background of sentiment analysis, Natural Language Processing, Machine Learning and literature review are provided in Section 2.1 – Section 2.4 respectively.

2.1 Sentiment Analysis

The goal of sentiment analysis is to distinguish comments or the attitude on various topics in the natural language, so that this analysis can classify the emotional aspects of communication.

The research in this field is about grouping of words or messages as the positive attitude or the negative attitude [2, 3]. Some sentences or phrases can express opinions or attitudes, positive or negative. These sentences or phrases also help identify the groups of reviews or comments more easily. Therefore, [2, 3] developed two approaches in the sentiment analysis to identify comment messages on a social network into the positive or the negative group.

There are some lexicons containing only sentiment words such as The MPQA (Multi-Perspective Question Answering) Subjectivity Lexicon [4] and Sentiwordnet [5]. The MPQA Subjectivity Lexicon and Sentiwordnet are a publicly available lexical for opinion mining. Sentiwordnet assigns to each synset of WordNet three sentiment scores: positivity, negativity, objectivity [6, 7]. The MPQA Subjectivity Lexicon can be used to score words or phrases of words to determine whether they are positive or negative. For every entry, the lexicon creates a result to indicate if an entry is positive, neutral or negative in its opinion [8].

Our research used Sentiwordnet because it contains more words than the MPQA Subjectivity Lexicon and it can also be used for not only question answering system, but also the general purposed.

2.2 Natural Language Processing (NLP)

Natural Language Processing is the study on computer science, artificial intelligence and linguistics in term of the interaction between humans and computers. Effort has been made in order to make computers understand natural language or human language involving natural language comprehension and making computers understand human or natural language input.

The Natural Language Processing consists of 3 processes [9]: syntactic analysis, semantic analysis and pragmatic analysis.

- **Syntactic Analysis**

  Syntactic Analysis will check the grammatical structures and the position of various groups of words that make up the sentence. If the incoming input sentence is not grammatically correct. The computer should tell that it is wrong.

- **Semantic Analysis**

  Semantic Analysis is the accuracy verification in term of the meaning of the sentence. The grammatical sentences normally have exact meaning. However, some grammatical sentences considered in this field might have ambiguous meaning or no meaning at all.

- **Pragmatic Analysis**

  Sometimes the sentences might not be able to interpret directly. To interpret these sentences, the situation is needed to be considered also. In this case, the sender, the receiver and the content have to be in the same situation in order to have the same comprehension.

The program that helps to prepare data for syntactic analysis in this thesis is Stanford Part-of-speech Tagger (POS Tagger) [10]. Stanford POS Tagger is open source software that reads text and assigns parts of speech to each word and other token such as noun, verb, adjective and adverb. This thesis brings the definition of part of speech from Penn Treebank which symbols and meanings are shown in Table 2-1.

Table 2-1 Symbols and Meanings of Penn Treebank

Penn Treebank Tag set	Meaning
CC	Coordinating Conjunction
CD	Cardinal Number
DT	Determiner
EX	There is or There are
FW	Foreign Word
IN	Preposition or Subordinating Conjunction
JJ	Adjective
JJR	Comparative Adjective
JJS	Superlative Adjective
LS	List Item Marker
MD	Modal Verb
NN	Singular Noun or Mass Noun
NNP	Singular or Proper Noun
NNPS	Plural Proper Noun
NNS	Plural Noun
RB	Adverb
RBS	Superlative Adverb
SYM	Symbol
TO	To
UH	Interjection
VB	Verb in Base Form
VBD	Past Tense Verb
VBG	Gerund or Present Participle Verb
VBN	Past Participle Verb
VBP	Non-3rd Person Singular Present Verb

Table 2-1 Symbols and Meanings of Penn Treebank (Cont.)

Penn Treebank Tag set	Meaning
VBZ	3rd Person Singular Present Verb
WDT	Wh-determiner
WP	Wh-pronoun
WP$	Possessive wh-pronoun
WRB	Wh-adverb

2.3 Machine Learning

Machine learning is a type of artificial intelligence [11] that makes computers have the self-learning ability. Machine Learning can be categorized into 2 main types: Supervised Learning and Unsupervised Learning.

Supervised Learning is a learning of the input data in which the answers are already given such as the stock price at a particular time or e-mail spam detection. Supervised Learning is prepared for the data prediction involving the problems like Regression and Classification.

Unsupervised Learning is a learning of the input data in which the answers are still unknown. The learning makes us getting closer to the answers or understanding more problems by arranging the data structure. The model will be prepared to use in the data structure in order to reduce duplication and categorize data into the same group, for example, the problem about Clustering.

One of the popular machine learning tool is WEKA (Waikato Environment for Knowledge Analysis) [12]. WEKA is a suite of software tools implemented in JAVA by University of Waikato. This software is free to use under General Public License-GPL. It is a collection of machine learning algorithms for data mining tasks. Weka has tools for data pre-processing, classification, regression, clustering, association rules, and visualization. It is also suitable for developing a new machine learning patterns. There are many classifiers in WEKA such as Decision Tree and Naïve Bayes classifier.

- **Decision Tree**

Decision tree classification will separate a dataset into littler and littler subsets while at the same time a related decision tree is incrementally created. The final result is a tree with decision nodes and leaf nodes. A decision node has two or more branches. Leaf node represents a classification or decision.

The algorithm for building decision trees called ID3 by Ross Quinlan [13]. ID3 uses Entropy and Information Gain to construct a decision tree. Entropy H(S) is a measure of the amount of uncertainty in set S which can be decomposed as

$$H(S) = -\sum_{x \in X} p(x) \log_2 p(x)$$

S is the current (data) set for which entropy is being calculated.

X is a set of classes in S.

p(x) is a proportion of the number of elements in class x.

When H(S) = 0, the set S is perfectly classified

In ID3, entropy is calculated for each remaining attribute. The attribute with the smallest entropy is used to split the set S on this iteration. The higher the entropy, the higher the potential to improve the classification here [14].

Information gain IG(A, S) is the measure of the difference in entropy from before to after the set S is split on an attribute A which can be decomposed as

$$IG(A, S) = H(S) - \sum_{t \in T} p(t) H(t) .$$

H(S) is the entropy of set S.

T is subsets created from splitting set S by attribute A.

p(t) is the proportion of the number of elements in t to the number of elements in set S.

H(t) is an entropy of subset t.

In ID3, information gain can be calculated for each remaining attribute. The attribute with the largest information gain is used to split the set S on this iteration [14].

- **Naive Bayes classifier**

The Naive Bayes classifier depends on Bayes theorem with independence presumptions between predictors. A Naive Bayes model is easy to create with no complicated iterative parameter estimation, which makes it useful for huge datasets.

Bayes theorem gives a method for calculating the posterior probability, P(c|x), from P(c), P(x), and P(x|c). Naive Bayes classifier assumes that the impact of the value of a predictor (x) on a given class (c) is independent of the values of other predictors. This assumption is called class conditional independence [15] which can be decomposed as

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)} .$$

P(c|x) is the posterior probability

P(c) is the prior probability of class.

P(x|c) is the likelihood which is the probability of predictor given class.

P(x) is the prior probability of predictor.

The value of P(x) is a constant of every class. The most appropriate equation is

$$P(c|x) = P(x_1|c) \times P(x_2|c) \times \ldots \times P(x_n|c) \times P(c)$$

- **Model Evaluation**

To evaluate the performance of a model, a confusion matrix is used for this thesis. A confusion matrix also known as an error matrix for the system implemented the machine learning technique is shown in Table 2-2.

Table 2-2 The Confusion Matrix

Actual Class	Predicted Class	
	Positive	***Negative***
Positive	True Positive (TP)	False Negative (FN)
Negative	False Positive (FP)	True Negative (TN)

According to all parameters in Table 2-2, if the actual class is positive and the result of predicted class is positive, the value of result data will be called True Positive (TP). Also, if actual class is negative and the result is the correct prediction, the value of result data will be called True Negative (TN). Whereas, if the result is the incorrect prediction and the actual class is positive, the value of result data will be called False Negative (FN). If the actual class is negative and the predicted class is positive, the value of result data will be called False Positive (FP).

Accuracy is a common measure for the performance of classification. While accuracy is the ratio of correct examples to the total examples, error rates is the ratio of incorrect examples to the total examples. All examples might be labelled as dominant class and will effect to other class in classification. The formula of accuracy can be shown in (1).

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \qquad (1)$$

Otherwise, using accuracy with skewed data, the data which has one class significantly than the other, needs to be very careful. Therefore, other measurement values are considered.

Precision and recall [16] are two basic and widely-used metrics in evaluating search strategies like text mining, information retrieval, etc. Precision and recall often used as an extension of accuracy. The combination of them can be used with skewed data problem in classification problem. Precision is the ratio of correct examples to the total of positive-classified examples. It often used to measure the exactness. Recall is the ratio of correct example to the total of truly-positive examples. It often used to measure the completeness. The formulas of precision and recall can be shown in (2) and (3).

$$\text{Precision} = \frac{TP}{TP + FP} \qquad (2)$$

$$\text{Recall} = \frac{TP}{TP + FN} \qquad (3)$$

2.4 Literature Review

There are many researches about the sentiment analysis. The definition of sentiment analysis is processed to identify subjective information in source materials by using Natural Language Processing (NLP) to analyze sentences. Sentiment analysis can be used to determine the sentiment by scoring individual words in those documents; a positive score or a negative score [1]. This section describes about the research involved the sentiment analysis.

P. Pugsee et al. [17] analyzed opinions about airline services on Twitter by collecting the subjective words in messages on Twitter and studying the sentiment messages by the machine learning. The machine learning can classify messages into two groups which are subjective and objective messages. This technique can apply to our proposed framework for classifying the sentiment of paragraphs by the machine learning.

Y. Yamamoto et al. [18] presented the method for calculating sentiment values of messages on Twitter based on emoticons and emoticon roles. In addition, words and emoticons in messages are detected by the sentiment lexicon and the emoticon lexicon to analyze the meaning of emoticons. Our proposed framework will apply this technique to find emoticons in review paragraphs.

C. Bhadane et al. [19] developed the system composed of six processes that are preprocessing, lexical analysis, stemming, part of speech tagging and machine learning to classify text files into different aspects. All these processes will be applied to this thesis in order to classify sentences into different aspects.

CHAPTER 3
METHODOLOGY

In this research, all of the above studies in Chapter 2 will be applied to analyze laptop reviews. This framework consists of four main processes: data preparation in section 3.1, subjective detection in section 3.2, aspect identification in section 3.3 and sentiment classification in section 3.4. The overview of this framework is shown in Fig. 3-1.

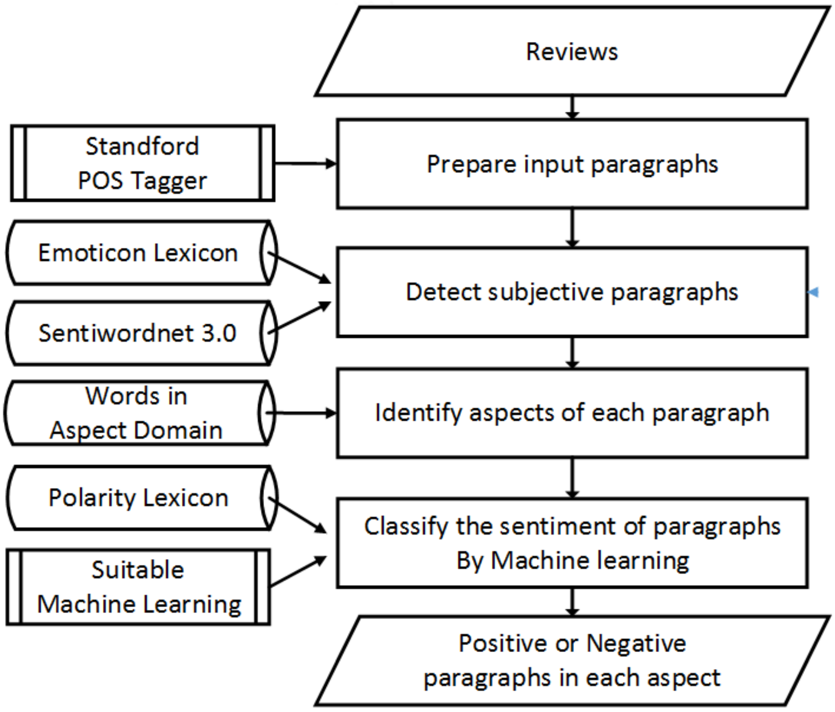


Figure 3-1 The framework overview

According to Fig. 3-1, the input of this framework is contents from community websites and the output is two groups of review paragraphs which are positive or negative paragraphs for each aspect. The emoticon texts from the emoticon lexicon and words from Sentiwordnet [5] are applied for detecting subjective paragraphs. Then, only subjective paragraphs are identified to the aspects of laptop reviews. To classify the sentiment of paragraphs, the library of machine learning from WEKA is included in our technique. An example of input data is shown in Fig. 3-2.

The touchpad is able to recognise even the complex 3-finger gestures with great precision. During about 2 hours of use I've only had 3 times when the mouse didn't do what I was expecting, mostly when trying to select text (which is tricky business on touchpads anyway). I didn't have trouble with palm rejection either, though it might be because my hands don't touch the touchpad while typing :p .

Figure 3-2 An example of input data for the proposed framework

Before starting the following processes (first of all), this framework can separate reviews and filter brands, models and screen size with the term that the input text must be tagged by "#" and a filter word after the symbol, for example, #apple, #dell or #17inch. In addition, the reviews can be selected by the interesting keyword. After that, the review must be ended with "#####end#####". The reviews that are already separated and filtered will be analyzed review by review later. An example of multiple reviews is shown in Fig. 3-3.


```
#apple #macbookpro #15inch
Review 1...
#####end#####

#dell #Inspiron #14.1inch
Review 2...
#####end#####
```

Figure 3-3 An example of multiple reviews

3.1 Prepare input paragraphs

The first process of this framework is to prepare input paragraphs. The steps of this process are shown in Fig. 3-4.

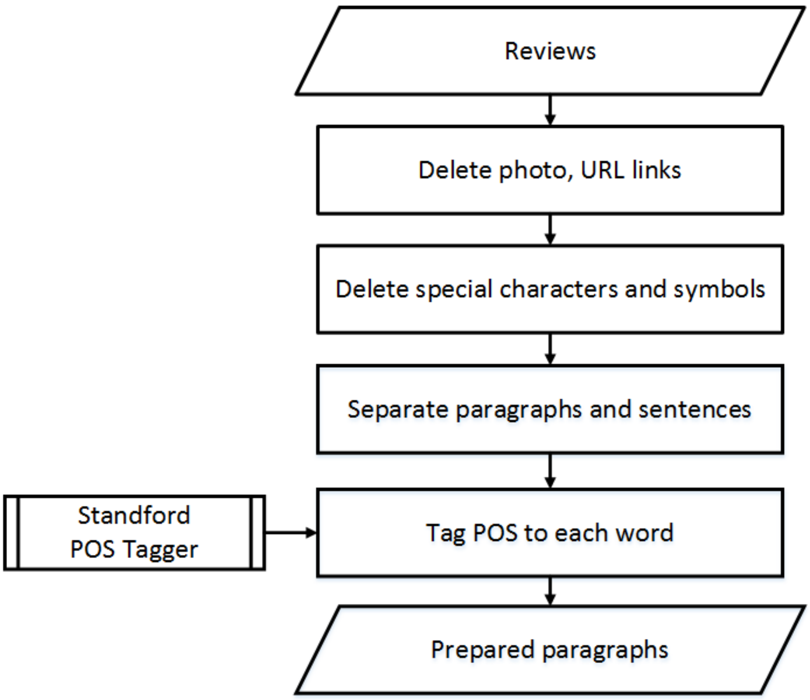


Figure 3-4 The steps of preparing input paragraphs

According to Fig. 3-4, this process is composed of four methods: deleting photo and URL links, deleting special characters and symbols, separating paragraphs and sentences, and tagging parts of speech to each word. The inputs of this process are reviews from community websites and the outputs are prepared paragraphs.

3.1.1 Delete photo and URL

When the photos from reviews in community website are saved into text format it will be saved as [IMG] tag. This process will delete [IMG] tag from the original reviews. In addition, normal URL links will also be deleted from the reviews by detecting "http" and "www". Moreover, picture links like a "photobucket.com/albums/l320/kingblast2/Photo49.jpg" will be deleted from reviews by detecting the ".jpg", ".gif" and ".png". The examples of URL links found in our collected reviews are "http://www.thinkgeek.com", "www.intel.com" and "http://www.3dmark.com/3dm11/6549824".

3.1.2 Delete special character and symbols

Special characters and special symbols will be deleted from input paragraphs. The examples of special characters and symbols are shown in Table 3-1. Likewise different characters separated from English letters, symbols and numbers will be categorized as special characters. List of all special characters and special symbols are shown in APPENDIX A.

Table 3-1 Example of special characters and special symbols

Special Characters	Special Symbols
Å	✓
Ç	☑
™	✗
©	☒

3.1.3 Separate paragraphs and sentences

This method will detect a newline and a full stop character to separate paragraphs and sentences respectively from one another. To separate paragraphs, the system will detect a newline character representing that the reviewer wants to write a new paragraph. After breaking the source content into individual paragraphs, the system will separate sentences in paragraphs by detecting the full stop. The examples in this step are shown in Fig. 3-5 and Fig. 3-6.

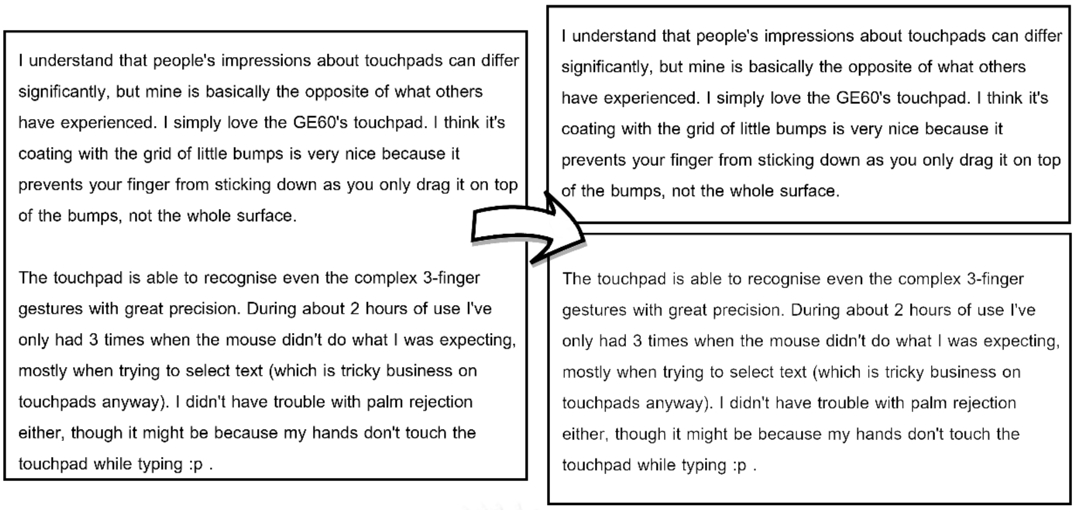


Figure 3-5 A source content will be separated into individual paragraphs.

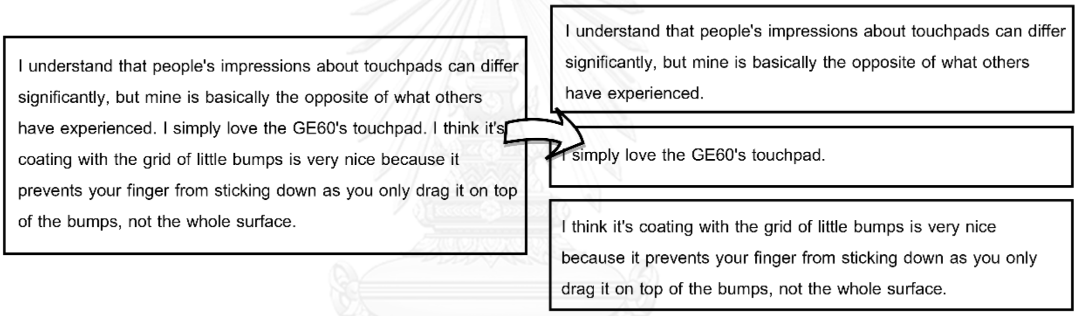


Figure 3-6 An individual paragraph will be separated into individual sentences.

3.1.4 Tag parts of speech to each word

At this stage, the method will tag words with their parts of speech. Each word in the individual sentence will be tagged its parts of speech by Stanford POS Tagger [10]. This tagger identifies parts of speech of words into four main groups that are noun, verb, adjective, and adverb. The example of tagged paragraph is shown in Table 3-2.

Table 3-2 The example of tagged words and sentences

Source data	Tagged data
The touchpad is able to recognise even the complex 3-finger gestures with great precision.	The_DT touchpad_NN is_VBZ able_JJ to_TO recognise_VB even_RB the_DT complex_NN 3-finger_NN gestures_NNS with_IN great_JJ precision_NN ._.
During about 2 hours of use I've only had 3 times when the mouse didn't do what I was expecting, mostly when trying to select text (which is tricky business on touchpads anyway).	During_IN about_RB 2_CD hours_NNS of_IN use_NN I_PRP 've_VBP only_RB had_VBN 3_CD times_NNS when_WRB the_DT mouse_NN did_VBD n't_RB do_VB what_WP I_PRP was_VBD expecting_VBG ,_, mostly_RB when_WRB trying_VBG to_TO select_VB text_NN -LRB-_-LRB- which_WDT is_VBZ tricky_JJ business_NN on_IN touchpads_NNS anyway_RB -RRB-_-RRB- ._.
I didn't have trouble with palm rejection either, though it might be because my hands don't touch the touchpad while typing :p .	I_PRP did_VBD n't_RB have_VB trouble_NN with_IN palm_NN rejection_NN either_CC ,_, though_IN it_PRP might_MD be_VB because_IN my_PRP$ hands_NNS do_VBP n't_RB touch_VB the_DT touchpad_NN while_IN typing_NN :p_NN ._.

3.2 Detect subjective paragraphs

The word information from Sentiwordnet [5] will be applied to identify whether those words are subjective or objective words. Every paragraph, which has at least one subjective word and emoticon text, will be considered as a subjective paragraph. In addition, emoticon texts will also be found by compare emoticon texts in the paragraph with data from an emoticon lexicon [18]. The steps of this process are shown in Fig. 3-7

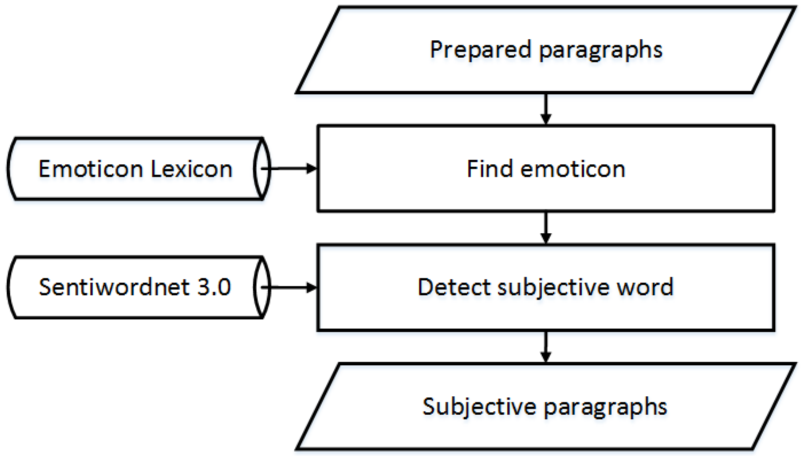


Figure 3-7 The steps of detect subjective paragraphs process

The inputs of this process are prepared paragraphs from the previous process. Then, the emoticons in prepared paragraphs will be detected by comparing emoticon found in prepared paragraph to emoticon lexicon. If the emoticons match with the emoticon lexicon, then those paragraphs will be collected as subjective paragraphs. And if the emoticons cannot be detected in the prepared paragraphs, then this process will detect the subjective words from paragraphs in the previous step by comparing to words in Sentiwordnet. If the subjective words are found, then the paragraphs will be collected as subjective paragraphs.

However, only adjective words and adverb words of review paragraphs will be compared with words from Sentiwordnet to detect subjective paragraphs. Sentiwordnet consists of 17,335 adjective words and 3,092 adverb words. If there is at least one subjective word, this paragraph will be detected as a subjective paragraph. The examples of adjective words and adverb words are shown in Table 3-3 and Table 3-4.

Table 3-3 The examples of adjective words

different	used	important	every
large	available	popular	able
basic	known	various	difficult
several	united	historical	hot
useful	mental	scared	additional
emotional	old	political	similar
healthy	financial	medical	traditional
federal	entire	strong	actual
significant	successful	electrical	expensive
pregnant	intelligent	interesting	poor
happy	responsible	cute	helpful
recent	willing	nice	wonderful
impossible	serious	huge	rare
technical	typical	competitive	critical
electronic	immediate	aware	educational

Table 3-4 The examples of adverb words

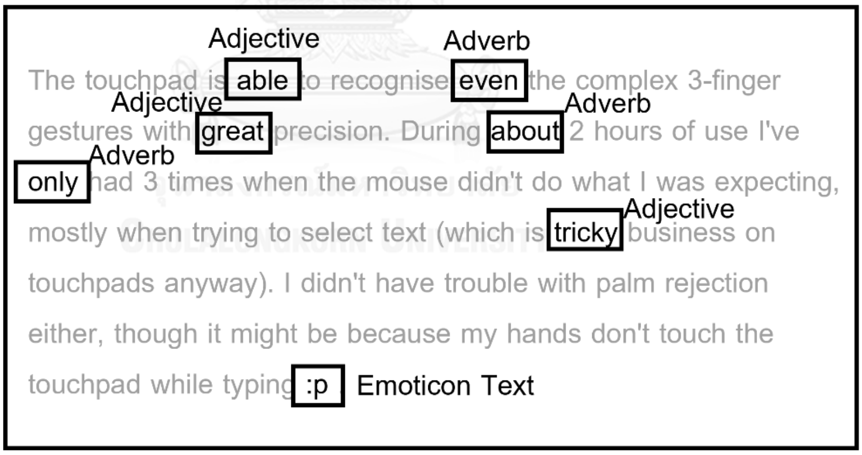
not	also	very	often
however	too	usually	really
early	never	always	sometimes
together	likely	simply	generally
instead	actually	again	rather
almost	especially	ever	quickly
probably	already	below	directly
therefore	else	thus	easily
eventually	exactly	certainly	normally
currently	extremely	finally	constantly
properly	soon	specifically	ahead
daily	highly	immediately	relatively
slowly	fairly	primarily	completely
ultimately	widely	recently	seriously
frequently	fully	mostly	naturally

This research will use only adjectives and adverbs to detect the subjectivity in this process because many subjective words can be found in these part of speech and they have higher performance than using all part of speech (noun, verb, adjective and adverb) to detect subjective paragraphs. In addition, other types of word, i.e. nouns cannot express the subjectivity clearly. Furthermore, although there are verbs which can be considered as subjective words i.e. to love, to like and to hate, these subjective verbs are minor comparing to other verbs so this part of speech will not be used in this research. All of the subjective words which are adjectives and adverbs found in all collected paragraphs are shown in APPENDIX B.

The output of this process focuses on only the subjective paragraphs. This process will detect subjective words and emoticon texts in the paragraph, such as paragraph in Fig. 3-2. Fig. 3-8 (a) shows the detected adjectives and adverbs on tagged data automatically by our framework and Fig. 3-8 (b) shows the detected subjective words and emoticon texts, so this paragraph is defined as the subjective paragraph.

The_DT touchpad_NN is_VBZ **able_JJ** to_TO recognise_VB **even_RB** the_DT complex_NN 3-finger_NN gestures_NNS with_IN **great_JJ** precision_NN ._. During_IN **about_RB** 2_CD hours_NNS of_IN use_NN I_PRP 've_VBP **only_RB** had_VBN 3_CD times_NNS when_WRB the_DT mouse_NN did_VBD n't_RB do_VB what_WP I_PRP was_VBD expecting_VBG ,_, mostly_RB when_WRB trying_VBG to_TO select_VB text_NN -LRB-_-LRB- which_WDT is_VBZ **tricky_JJ** business_NN on_IN touchpads_NNS anyway_RB -RRB-_-RRB- ._. I_PRP did_VBD n't_RB have_VB trouble_NN with_IN palm_NN rejection_NN either_CC ,_, though_IN it_PRP might_MD be_VB because_IN my_PRP$ hands_NNS do_VBP n't_RB touch_VB the_DT touchpad_NN while_IN typing_NN **:p_NN** ._.

(a)



(b)

Figure 3-8 The detected subjective words and emoticon texts in the paragraph

According to Fig. 3-8, although the words ‘n’t_RB’, ‘mostly_RB, and ‘anyway_RB’ (adverb) are found in the paragraph, they are not detected as subjective words because the polarity scores of these words are zero (the sentiment score = neutral) as referred from the Sentiwordnet.

3.3 Identify aspects of each paragraph

The identify aspects of each paragraph process can be illustrated as Fig. 3-9

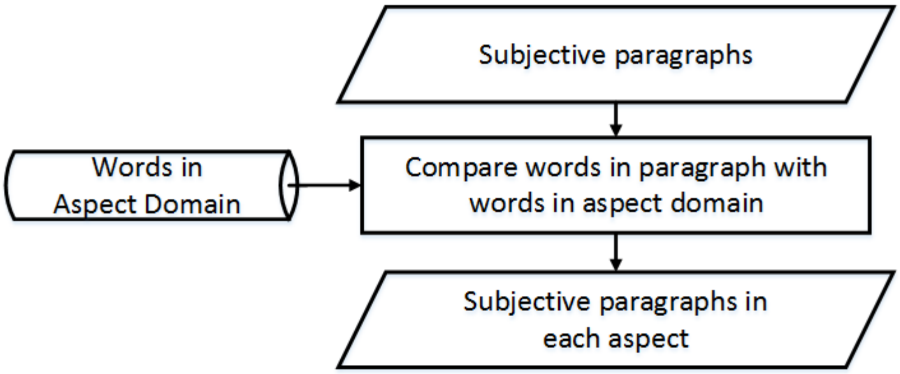
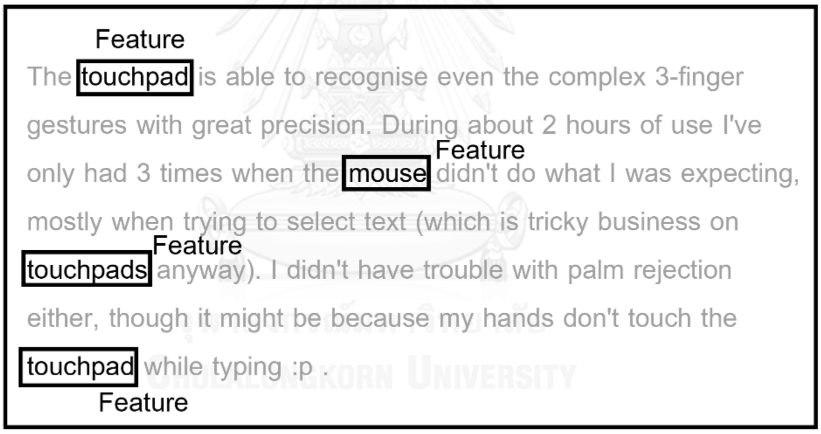


Figure 3-9 The step of identify aspects of each paragraph process

In this process, the subjective paragraphs from the previous process will be categorized into different aspects ("Performance", "Design" and "Feature") by comparing words with words in each aspect domain. Individual subjective paragraph can match more than one aspect. However, some subjective paragraphs, which cannot be identified into these groups, will be classified into "Other" aspect. The words of each aspect domain are listed by analyzing the popular words found in laptop reviews. Mentioning to Fig. 3-2, this process will detect words in aspect domains for identifying types of aspect. The detected words are shown in Fig. 3-10.

The_DT **touchpad**_NN is_VBZ able_JJ to_TO recognise_VB even_RB the_DT complex_NN 3-finger_NN gestures_NNS with_IN great_JJ precision_NN ._. During_IN about_RB 2_CD hours_NNS of_IN use_NN I_PRP 've_VBP only_RB had_VBN 3_CD times_NNS when_WRB the_DT **mouse**_NN did_VBD n't_RB do_VB what_WP I_PRP was_VBD expecting_VBG ,_, mostly_RB when_WRB trying_VBG to_TO select_VB text_NN -LRB-_-LRB- which_WDT is_VBZ tricky_JJ business_NN on_IN **touchpads**_NNS anyway_RB -RRB-_-RRB- ._. I_PRP did_VBD n't_RB have_VB trouble_NN with_IN palm_NN rejection_NN either_CC ,_, though_IN it_PRP might_MD be_VB because_IN my_PRP$ hands_NNS do_VBP n't_RB touch_VB the_DT **touchpad**_NN while_IN typing_NN :p_NN ._.

(a)



(b)

Figure 3-10 The detected words in the feature aspect.

The list of words in each aspect domain are shown in Table 3-5. These words will be collected from all review paragraphs by using “AntConc”[20]. AntConc helps to find frequency of words in each paragraph. The example of the result generated by AntConc is shown in Fig. 3-11. Then, the aspect words with high frequency from all reviews will be categorized into each aspect domain by the researcher’s judgment to classify the aspect of paragraphs.

Table 3-5 List of words in each aspect

Performance	Design	Feature
processor	weight	usb
ghz	width	hdmi
mhz	height	vga
cpu	size	dvi
gpu	display	touchpad
memory	stylish	trackpad
ram	materials	firewire
Framerate	aluminum	webcam
resolution	unibody	bluetooth
battery	inch	mic
brightness	lcd	camera
graphics	led	microphone
fullhd	screen	keyboard
	solid	dvd
	glossy	speaker
	plastic	wireless
	models	mouse
	widescreen	port
	wide	fingerprint
		headphone
		ethernet
		bluray
		blu ray

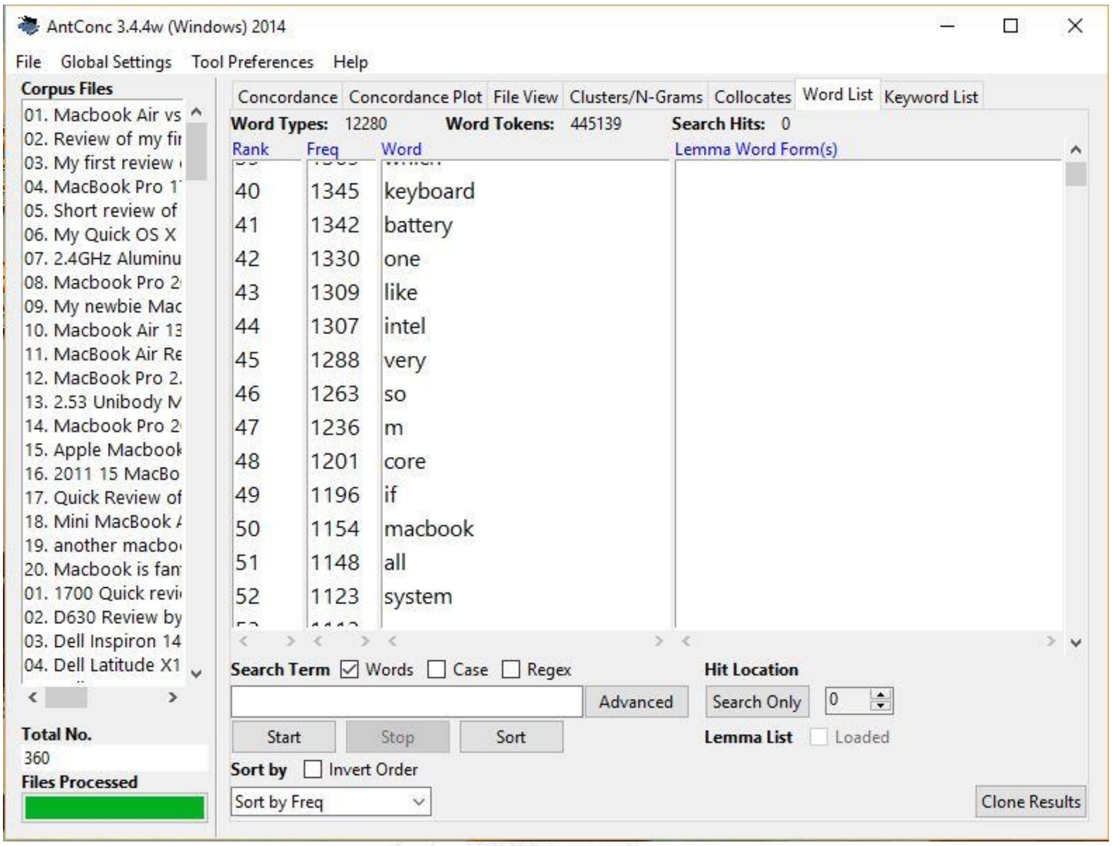


Figure 3-11 AntConc result

According to Fig 3-11, there are 445,139 words in the review paragraphs which are 12,280 different words. The aspect words (words in the aspect domains) will be selected from these top 200 high frequency words which are 56 different words divided into 13 words in a performance group, 19 words in a design group and 24 words in a feature group.

3.4 Classify the sentiment of paragraphs

In this process, subjective paragraphs in each aspect will be classified into the sentiment types of paragraphs using the machine learning. The classification model will be created from the training data and implemented to classify the sentiments by the machine learning. Therefore, this section can be divided into two parts: the model creation and the sentiment classification. The model creation will be discussed in 3.4.1 and the sentiment classification of this framework will be described in 3.4.2.

3.4.1 The model creation

The model creating process is consisted of importing all subjective paragraphs in order to tag POS to each word in the paragraphs, finding polarity of each word in paragraphs in order to be used as feature sets, and create the model of selected machine learning using those feature sets. The overall process of model creation can be illustrated as Fig. 3-12.

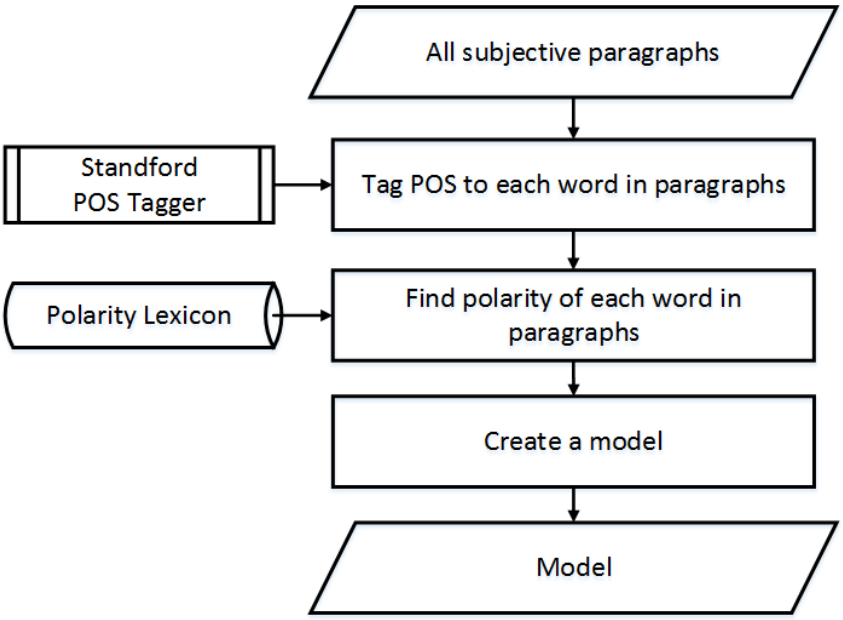


Figure 3-12 The model creation process

According to Fig 3-12, the polarity lexicon is used to find the polarity of the words in paragraphs instead of using the Sentiwordnet. The reason is that the percent of accuracy, precision and recall rate of the sentiment classification are less than 55% when implemented by the model created from the information about word polarities in Sentiwordnet. There are too many defects because there are many words that do not appear in the Sentiwordnet, but appear in the reviews. So, the polarity lexicon is created by using words' information and polarity scores from Sentiwordnet, including added extra words and their polarity scores. After using the polarity lexicon to replace the Sentiwordnet, the percent of accuracy, precision and recall rate of sentiment classification can be up to about 70%.

The examples of added words are shown in Table 3-6. All extra words can be divided into 3 main types:

- adjectives and adverbs in comparative form : 76 words
- adjectives and adverbs in superlative form : 49 words
- words in concatenated form : 1,081 words

Table 3-6 The examples of added words in Sentiwordnet

Comparative	Superlative	Concatenated words
bolder	biggest	battery-drain
brighter	brightest	battery-friendly
broader	coolest	cost-saving
closer	dimmest	finger-control
cooler	easiest	finger-swipe
darker	fanciest	greater-than
happier	fastest	half-inch
harder	largest	high-quality

According to Table 3-6, for the comparative and superlative group, the polarity scores will be referred from the base form of those words in Sentiwordnet. Therefore, these scores will be increased or decreased by 0.25 for positive scores or negative scores. For example, the word ‘bright’ has the polarity score of +0.125 then ‘brighter’ and ‘brightest’ will be +0.375 (+0.125+0.25) and +0.625 (+0.375+0.25), respectively. Moreover, if there is no base form of those words in Sentiwordnet, then the polarity score of those words can be found from using The Natural Language Toolkit (NLTK) [21]. The polarity scores of subjective words from NLTK are similar to polarity word scores from Sentiwordnet. NLTK scores can be divided into +0.1 to +1.0 for positive scores and -0.1 to -1.0 for negative scores. Then, the extra words with its polarity score will be added to the polarity lexicon. Finally, the polarity score will be changed to the polarity level with these following rules:

Score > 0.75 means strong positive

0.5 < Score <= 0.75 means very positive

0 < Score <= 0.5 means positive

Score = 0 means neutral

0 > Score >= -0.5 means negative

-0.5 > Score >= -0.75 means very negative

Score < -0.75 means strong negative.

All of extra words are shown in APPENDIX C.

However, the significant points that must be considered in the model selection are suitable machine learning and feature sets.

To select the suitable machine learning in the classification model, the basic features of the classification model are all adjectives, adverbs and their parts of speech of all subjective words from the previous process. All features will be learned and classified by WEKA Machine Learning [12]. This feature set is tested with four classifications; Naive Bayes, RBF Network, ZeroR and J48. The results of each machine learning classification are shown in Table 3-7.

Table 3-7 The evaluation results of all classifiers

Class	**Percent of**		
	Accuracy	*Precision*	*Recall*
Naive Bayes	68.61	68.40	68.60
RBF Network	65.93	65.50	65.90
ZeroR	59.35	35.20	59.40
J48	73.37	73.33	73.40

According to Table 3-7, it can be concluded that the Naive Bayes and J48 are two highest performance classification models for all three evaluation values (accuracy, precision, and recall). So, we choose these techniques to classify the sentiment of paragraphs in the learning process.

To select the feature set, there are all 3 experiments to determine the appropriate feature sets for each machine learning technique.

In the first experiment, all words and their parts of speech in adjectives and adverbs will be used as a feature set. The example of this feature set is shown in Fig. 3-13.

Relation: All

No.	Row 1 Nominal	Row 2 Nominal	Row 3 Nominal	Row 4 Nominal	Row 5 Nominal	Row 6 Nominal	Row 7 Nominal	Row 8 Nominal	Row 9 Nominal	Row 10 Nominal	Row 11 Nominal	Row 12 Nominal	Row 13 Nominal	Row 14 Nominal	Row 15 Nominal	Row Nomi
31	not_RB	minor_JJ	stand...	not_RB	nope_RB	minor_JJ	irritati...									
32	bad_JJ	least_...	so_RB	far_RB	now_RB	good_JJ										
33	really_...	video_JJ	quickly...	easily...	not_RB	proper...	fantas...	under...	quite_RB	often_...	able_JJ	full_JJ	VERY_RB	useful...		
34	full_JJ	just_RB	comfo...													
35	on/off...	not_RB	perso...	able_JJ	quickly...	easily...	wirele...	else_RB								
36	just_RB	wirele...														
37	Overal...	really_...	nice_JJ	sleek_JJ	solid_JJ	well_RB	very_RB	happy...	new_JJ							
38	very_RB	fast_JJ	hot_JJ	up_RB	higher...											
39	absolu...	great_JJ	very_RB	bright_JJ	clear_JJ	dead_JJ										
40	very_RB	nice_JJ	just_RB	not_RB	loud_JJ											
41	vey_RB	good_JJ	especi...													
42	not_RB	great_JJ	little_JJ	anywa...												
43	Hard_JJ	quiet_JJ														
44	pc270...	not_RB	due_JJ													
45	great_JJ	not_RB	yet_RB	yet_RB												
46	other_JJ	only_JJ	full_JJ	alread...	also_RB	not_RB										
47	availa...															
48	big_JJ	well_RB														
49	very_RB	good_JJ														
50	64mb_JJ	right_RB	away_...	Also_RB	still_RB	variou...	very_RB	slowly...	not_RB	sure_JJ	well_RB					
51	yet_RB	great_JJ														
52	great_JJ	small_JJ														
53	USB_JJ	great_JJ														
54	Overal...	first_JJ	very_RB	impres...	fine_JJ	overall...	great_JJ	only_JJ	8mb_JJ							
55	more_...	compl...														
56	Fast_JJ	respo...	Even_RB	respo...												
57	Great_JJ	overall...	nice_JJ													
58	crisp_JJ															

Figure 3-13 The example of the first experiment feature set

The percent of accuracy, precision, and recall rate of the sentiment classification will be calculated in WEKA Explorer by the confusion matrix. The confusion matrix of the result in the first model and the percent of three evaluation values are shown in Table 3-8 and Table 3-9.

Table 3-8 The confusion matrix of the result in the first experiment

Naïve Bayes			
Actual Class		**Predicted Class**	
		Positive	*Negative*
Positive	2,534	1,499	1,035
Negative	3,700	922	2,778
Total	6,234	2,421	3,813

J48			
Actual Class		**Predicted Class**	
		Positive	*Negative*
Positive	2,534	1,695	839
Negative	3,700	821	2,879
Total	6,234	2,516	3,718

Table 3-9 The percent of accuracy, precision, and recall rate of the first experiment

Class	**Percent of**		
	Accuracy	*Precision*	*Recall*
Naive Bayes	68.61	68.40	68.60
J48	73.37	73.33	73.40

According to Table 3-9, the percent of accuracy, precision and recall rate of J48 classification are higher than those of Naïve Bayes classifier, but all evaluation values are less than 75%. Therefore, another feature set will be selected to be training data for the classification model in the second experiment.

In the second experiment, the polarity scores from the polarity lexicon are applied to define the polarity level of each word. The word polarity level (“strong positive”, “very positive”, “positive”, “neutral”, “negative” “very negative” and “strong negative”) is included into feature set with words and their parts of speech. The example of this feature set is shown in Fig. 3-14. The confusion matrix of the result in this model and the percent of accuracy, precision, and recall rate are shown in Table 3-10, Table 3-11, respectively.

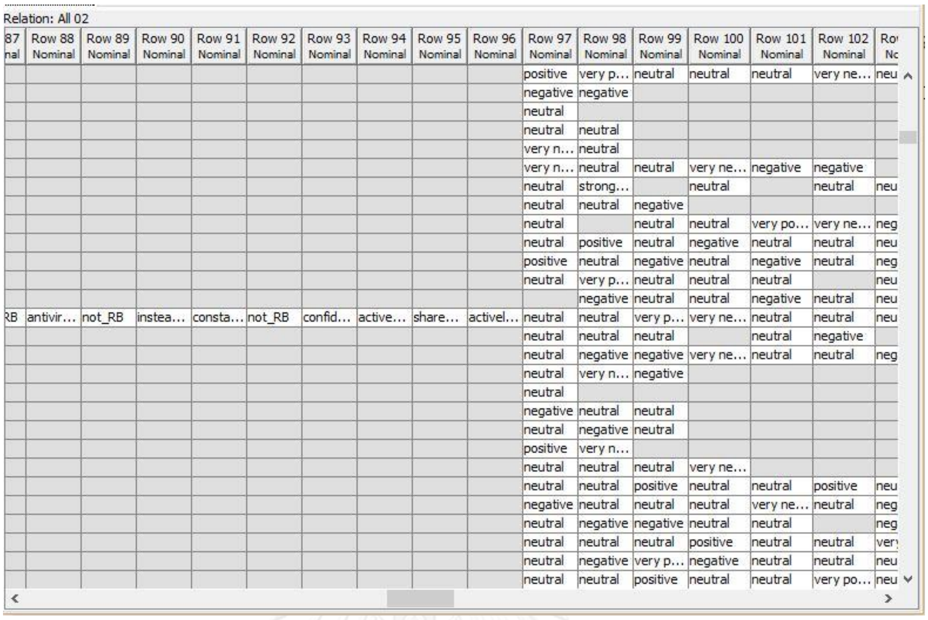


Figure 3-14 The example of the second experiment feature set

Table 3-10 The confusion matrix of the result in the second experiment

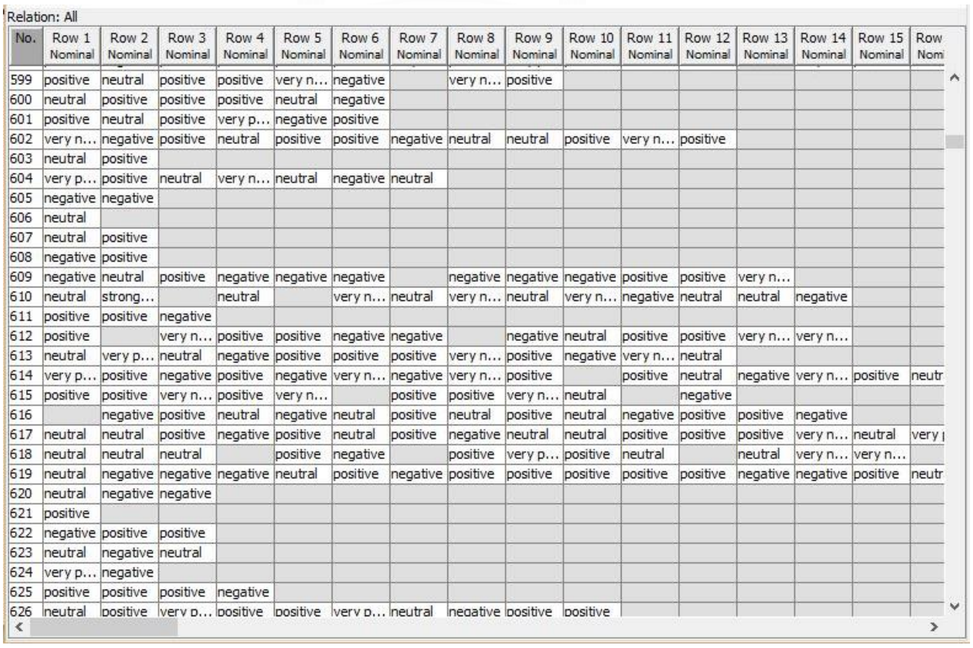
Naïve Bayes			
Actual Class		Predicted Class	
		Positive	*Negative*
Positive	2,534	1,676	858
Negative	3,700	860	2,840
Total	6,234	2,536	3,698
J48			
Actual Class		Predicted Class	
		Positive	*Negative*
Positive	2,534	958	1,576
Negative	3,700	368	3,332
Total	6,234	1,326	4,908

Table 3-11 The percent of accuracy, precision, and recall rate of the second experiment

Class	Percent of		
	Accuracy	*Precision*	*Recall*
Naive Bayes	72.49	72.40	72.40
J48	68.82	69.70	66.10

According to Table 3-11, the percent of accuracy, precision and recall rate of Naïve Bayes classifier are higher than those of J48 classification, however their evaluation values are still less than 75%. As a result, the new feature set are tried to discover for classification in the third experiment.

In the third experiment, only the polarity level of words in adjective and adverbs from polarity scores, which are strong positive, very positive, positive, neutral, negative, very negative and strong negative, will be used in feature set. Fig. 3-15 shows the example of the third experiment feature set. The confusion matrix of the result are displayed in Table 3-12 and the percent of accuracy, precision, and recall rate are shown in Table 3-13.



Relation: All

No.	Row 1 Nominal	Row 2 Nominal	Row 3 Nominal	Row 4 Nominal	Row 5 Nominal	Row 6 Nominal	Row 7 Nominal	Row 8 Nominal	Row 9 Nominal	Row 10 Nominal	Row 11 Nominal	Row 12 Nominal	Row 13 Nominal	Row 14 Nominal	Row 15 Nominal	Row Nomi
599	positive	neutral	positive	positive	very n...	negative		very n...	positive							
600	neutral	positive	positive	positive	neutral	negative										
601	positive	neutral	positive	very p...	negative	positive										
602	very n...	negative	positive	neutral	positive	positive	negative	neutral	neutral	positive	very n...	positive				
603	neutral	positive														
604	very p...	positive	neutral	very n...	neutral	negative	neutral									
605	negative	negative														
606	neutral															
607	neutral	positive														
608	negative	positive														
609	negative	neutral	positive	negative	negative	negative		negative	negative	negative	positive	positive	very n...			
610	neutral	strong...		neutral		very n...	neutral	very n...	neutral	very n...	negative	neutral	neutral	negative		
611	positive	positive	negative													
612	positive		very n...	positive	positive	negative	negative		negative	neutral	positive	positive	very n...	very n...		
613	neutral	very p...	neutral	negative	positive	positive	positive	very n...	positive	negative	very n...	neutral				
614	very p...	positive	negative	positive	negative	very n...	negative	very n...	positive		positive	neutral	negative	very n...	positive	neutr
615	positive	positive	very n...	positive	very n...		positive	positive	very n...	neutral		negative				
616		negative	positive	neutral	negative	neutral	positive	neutral	positive	neutral	negative	positive	positive	negative		
617	neutral	neutral	positive	negative	positive	neutral	positive	negative	neutral	neutral	positive	positive	positive	very n...	neutral	very
618	neutral	neutral	neutral		positive	negative		positive	very p...	positive	neutral		neutral	very n...	very n...	
619	neutral	negative	negative	negative	neutral	positive	negative	positive	positive	positive	positive	positive	negative	negative	positive	neutr
620	neutral	negative	negative													
621	positive															
622	negative	positive	positive													
623	neutral	negative	neutral													
624	very p...	negative														
625	positive	positive	positive	negative												
626	neutral	positive	very p...	positive	positive	very p...	neutral	negative	positive	positive						

Figure 3-15 The example of the third experiment feature set

Table 3-12 The confusion matrix of the result in the third experiment

Naïve Bayes			
Actual Class		**Predicted Class**	
		Positive	*Negative*
Positive	2,534	1,907	627
Negative	3,700	781	2,919
Total	6,234	2,688	3,546
J48			
Actual Class		**Predicted Class**	
		Positive	*Negative*
Positive	2,534	1,811	723
Negative	3,700	830	2,870
Total	6,234	2,641	3,593

Table 3-13 The percent of accuracy, precision, and recall rate of the third experiment

Class	**Percent of**		
	Accuracy	*Precision*	*Recall*
Naive Bayes	77.41	77.70	77.40
J48	75.09	75.30	75.10

According to Table 3-13, the percent of accuracy, precision and recall rate of Naïve Bayes classifier are higher than those of J48 classification, and all values of both classification models are also higher than 75%.

In conclusion, the results of all three experiments indicated that the performance of Naïve Bayes classifier (accuracy, precision and recall) is higher than those of J48 classification for the second and the third feature set. Moreover, the performance of Naïve Bayes classifier with the feature set in the third experiment is the highest performance. Therefore, this research selects the feature set in the third experiment and Naïve Bayes classifier to be the model for classifying the sentiments of paragraphs.

3.4.2 Sentiment classification

The sentiment of paragraphs will be classified by the suitable model from section 3.4.1. The use of polarity level as feature sets and the use of Naïve Bayes classifier as the machine learning are the highest performance of the sentiment classification model. The results of this process are two groups of text paragraphs (positive and negative paragraphs). The flow of classifying the sentiment of paragraphs is shown in Fig. 3-16.

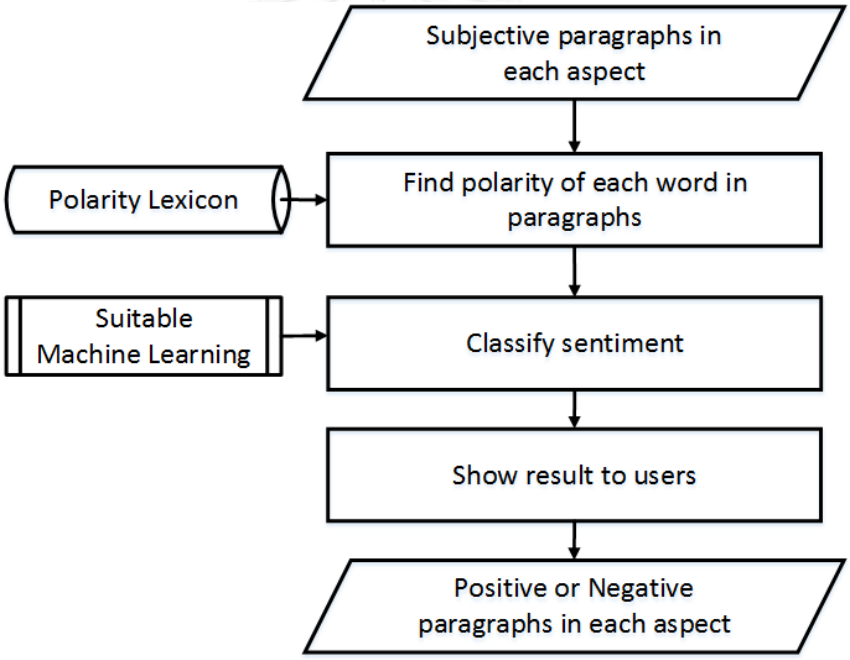
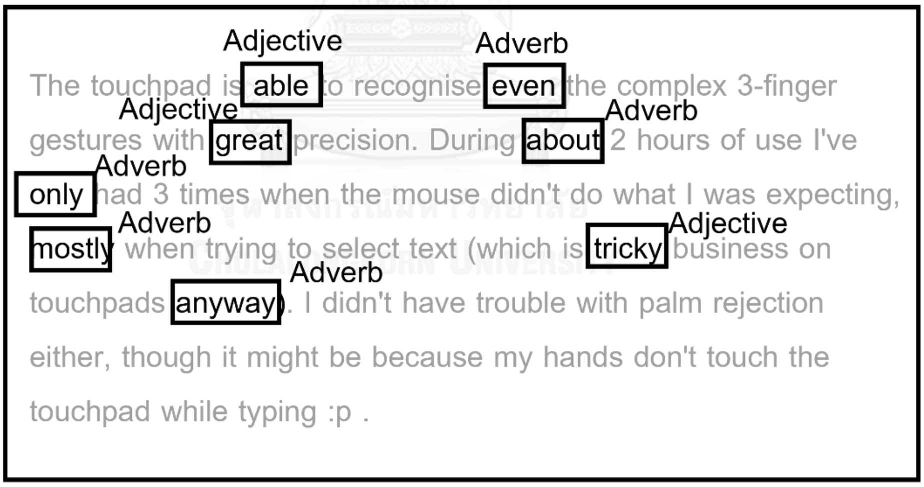


Figure 3-16 The steps of classifying the sentiment of paragraphs

First, all adjective and adverb words will be found by detecting their part of speech from tagged word list of the input paragraph in this process (subjective paragraphs in each aspect). The examples of detected adjectives and adverbs in this process are shown in Fig 3-17. Although, the word 'not' following the verbs, 'to be' and 'to do', is tagged as adverb words, it cannot express the sentiments. So, this word and its abbreviation form (n't) are removed from the detected adverb list. These words' polarity levels are not also included in the feature set for the sentiment classification.

The_DT touchpad_NN is_VBZ **able_JJ** to_TO recognise_VB **even_RB** the_DT complex_NN 3-finger_NN gestures_NNS with_IN **great_JJ** precision_NN ._. During_IN **about_RB** 2_CD hours_NNS of_IN use_NN I_PRP 've_VBP **only_RB** had_VBN 3_CD times_NNS when_WRB the_DT mouse_NN did_VBD n't_RB do_VB what_WP I_PRP was_VBD expecting_VBG ,_, **mostly_RB** when_WRB trying_VBG to_TO select_VB text_NN -LRB-_-LRB- which_WDT is_VBZ **tricky_JJ** business_NN on_IN touchpads_NNS **anyway_RB** -RRB-_-RRB- ._. I_PRP did_VBD n't_RB have_VB trouble_NN with_IN palm_NN rejection_NN either_CC ,_, though_IN it_PRP might_MD be_VB because_IN my_PRP$ hands_NNS do_VBP n't_RB touch_VB the_DT touchpad_NN while_IN typing_NN :p_NN ._.

(a)



(b)

Figure 3-17 The examples of detected adjectives and adverbs in the process

Next, all detected adjective and adverb words will be changed into the polarity scores from the polarity lexicon and these scores will be transformed into the polarity level by rules in section 3.4.1. All of the above-mentioned replacement is shown in Table 3-14 and Fig. 3-18.

Table 3-14 Polarity score and polarity level of detected words

Words	Polarity score	Polarity level
able_JJ	+0.125	Positive
even_RB	-0.125	Negative
great_JJ	+0.250	Positive
about_RB	+0.095	Positive
only_RB	-0.032	Negative
mostly_RB	0.0	Neutral
tricky_JJ	-0.320	Negative
anyway_RB	0.0	Neutral

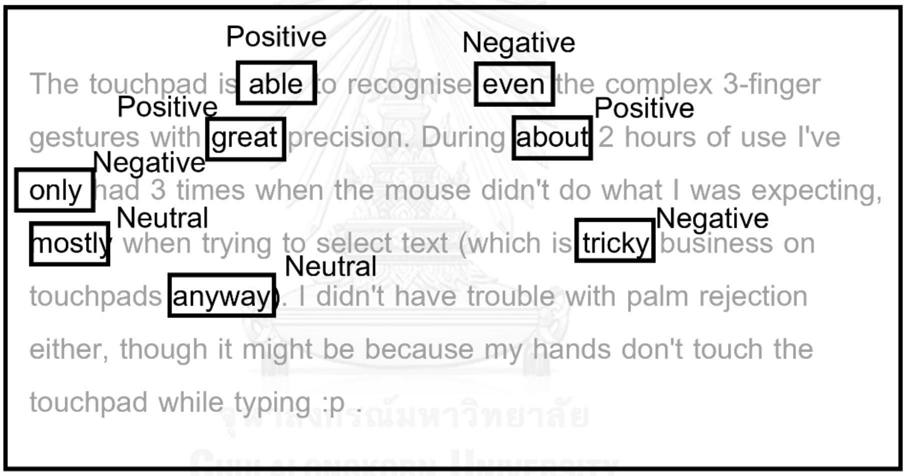


Figure 3-18 The polarity level of detected words by the replacement process

After that, the polarity level shown in Table 3-14 will be input to the generated classification model to classify the sentiment of this paragraph. The feature set of this paragraph is shown in Fig.3-19.

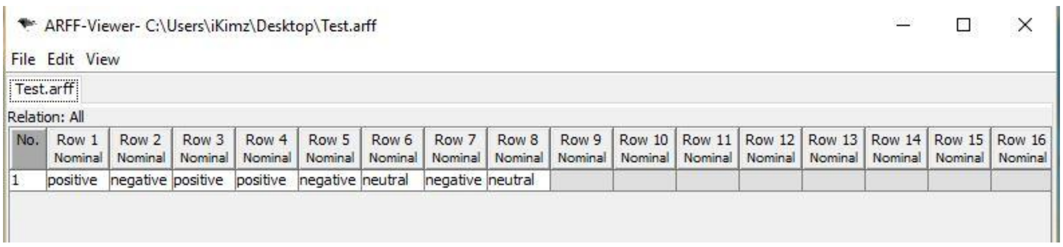


Figure 3-19 The feature set of input paragraph

Finally, the sentiment of this paragraph will be classified and shown on the user interface of the software implementing the framework. The result of this example paragraph shown in Fig. 3-20 is the positive in the feature aspect. The next section will present the software implementing this framework.

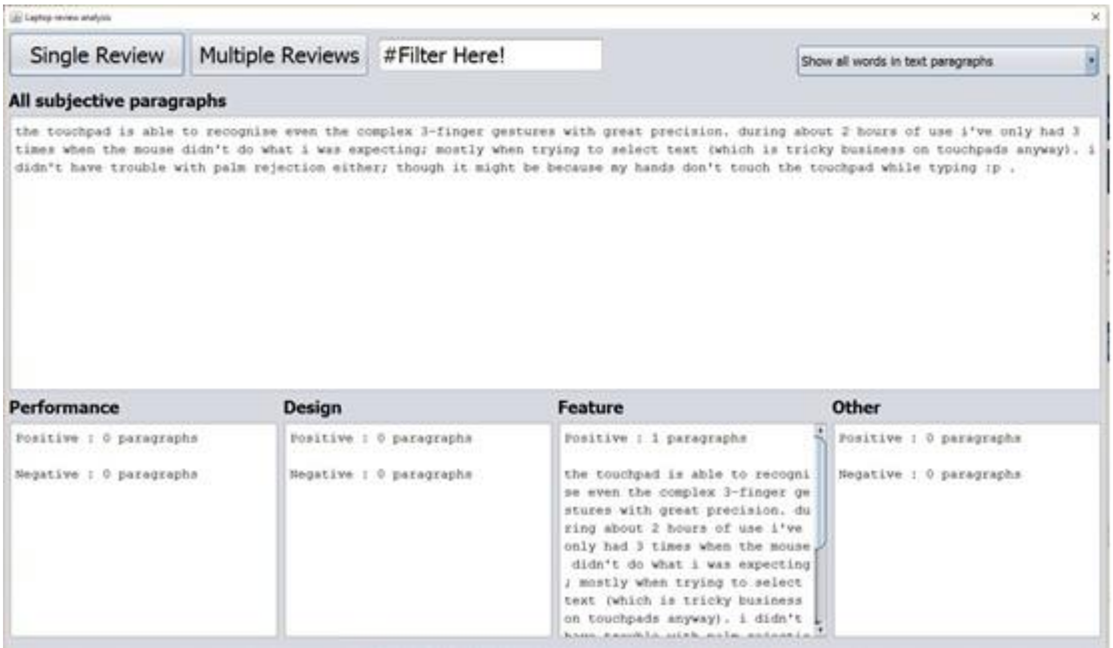


Figure 3-20 The result of the sentiment classification process

3.5 The software implementing the proposed framework

This software is an easy way to apply this framework for analyzing laptop reviews. In this research, the researcher has designed the layout of the user interface for this software as one page to make the software easier to use. The main screen consists of three areas: the menu bar, a middle text area and four bottom text areas. The menu bar includes “Single Review” button for analyzing a review, “Multiple Review” button for analyzing reviews, text box for inputting a filter word and drop down list for selecting output types. The middle text area shows only subjective paragraphs in the review and four bottom text areas show the subjective paragraphs in each aspect domain. The output on the screen of this software implementing the framework can be divided into 3 main types: show all words in text paragraphs show the keywords of text paragraphs and show the keywords with their polarity level of text paragraphs. The main user interface is shown in Fig. 3-21.

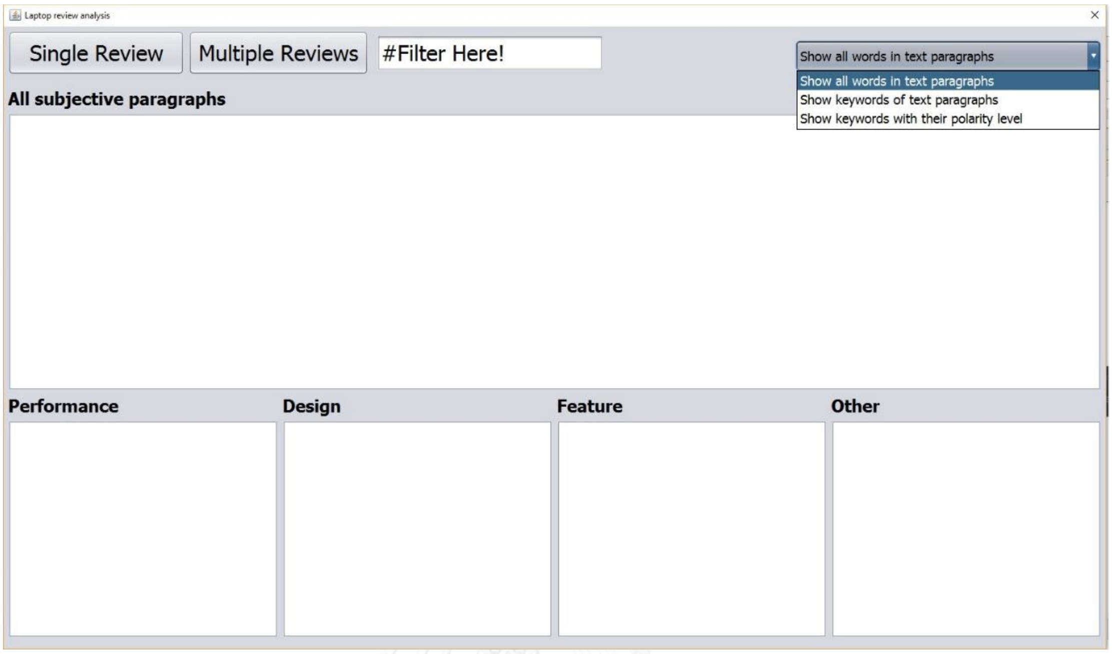


Figure 3-21 The main user interface

According to Fig. 3-21, the main user interface is the first page that the users will see when they start the software implementing the proposed framework. This page can be divided into 2 main function parts as:

1. Single review: The single review button is used to analyze only one text review in one text file. The example output of the single review analysis is shown in Fig. 3-22.
2. Multiple reviews: The multiple reviews button is used to analyze more than one text reviews in one text file. The format of multiple text reviews is already shown in Fig. 3-3. The example output of the multiple review analysis is shown in Fig. 3-23.

Figure 3-22 The single review analysis result

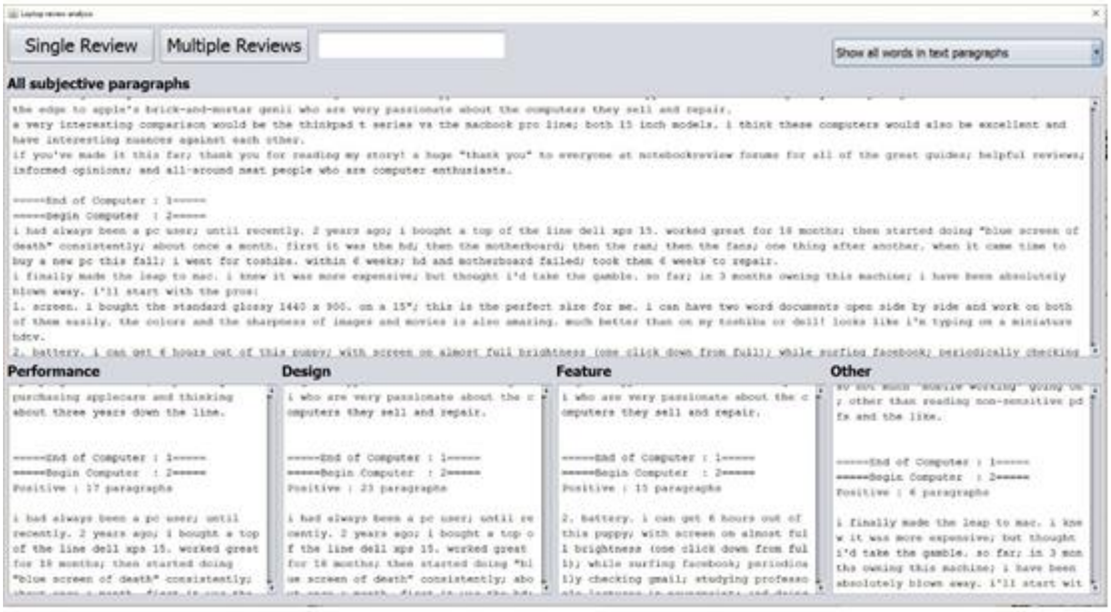


Figure 3-23 The multiple reviews analysis result

The default output type of this software shows all words in paragraphs. The bottom text area shows all positive paragraphs and all negative paragraphs in the review separated by the aspect domain. For each aspect text area , there are the total number of positive paragraphs, all positive text paragraphs, the total number of negative paragraphs and all negative text paragraphs, respectively. In the case of multiple review analysis, a product will be discussed in on review normally. So, this software shows the result of a product review by review. That means the displayed list of positive and negative text paragraphs are divided by each review as shown in Fig 3-23.

The next output type shows the keywords of the text paragraphs. All text areas show only adjective words, adverb words and some words displaying the aspect of paragraphs instead of all words in paragraphs. These words are the keywords for detecting aspect and the keywords to generate feature set for classifying sentiments. Therefore, aspect words found on paragraphs are shown and all adjective and adverb words, which their polarity scores are discovered to identify the polarity levels of them for the sentiment classification, are revealed. The examples are shown in Fig. 3-24.

Figure 3-24 The result showing only keywords

The last output type shows the keywords with their polarity level. All adjective and adverb words with their polarity levels generated from their polarity scores are displayed, including words in each aspect domain. The keywords' polarity levels are the feature set of the sentiment classification by the machine learning. The examples are shown in Fig. 3-25.

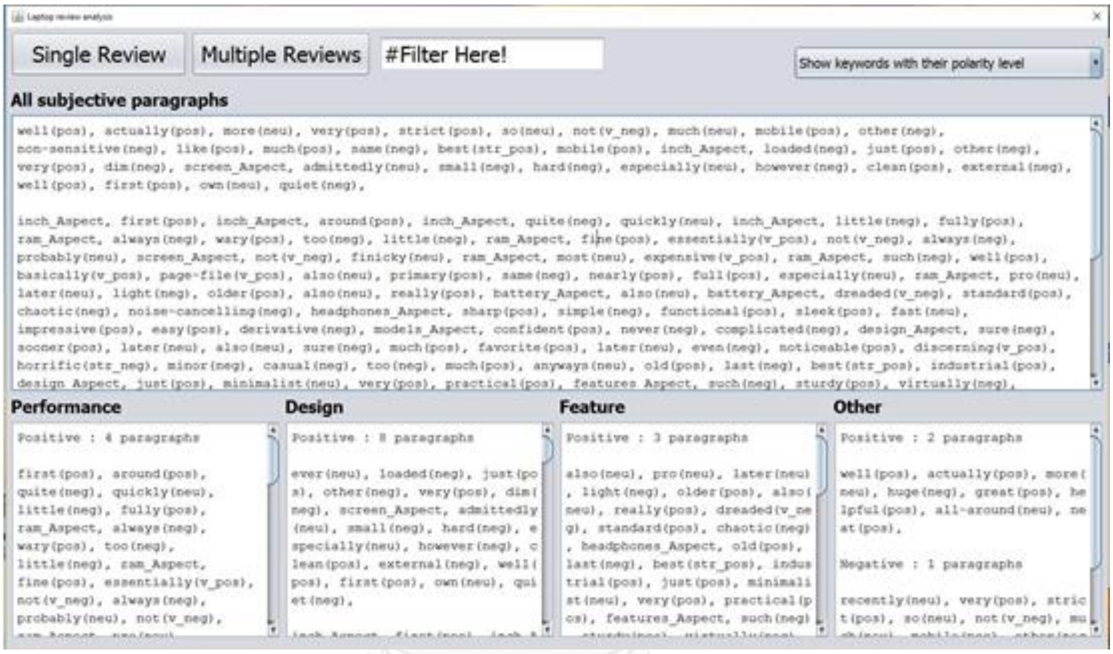


Figure 3-25 The result showing keywords and their polarity level

The output examples of review analysis by the software implementing the proposed framework are shown in Table 3-15 – Table 3-17.

Table 3-15 The example output of show all words in the paragraphs

All subjective paragraphs
sorry it took so long to review this; have a lot going on. anyways here goes. i received the vostro a couple days ago and the first thing i can say is that i love the build. the case is durable and looks good in all black. i also love the keyboard; i am not a good typer but i can type fast without looking at the keyboard on this which is something i could never do before on a laptop and still have trouble on other keyboards. the screen looks incredible. i upgraded to the badass screen with truelife and i do not see any graininess. the screen is viewable from all angles; looks great up close and at a distance. the dvd burner burned a 4 gig dvd as fast as i would expect it to at max speed. vista is running great; i have had very few issues with programs that were not compatible and could find updates for all but one relatively fast. vista business has some interesting things; but most are more novelties that are not too important. i do think that the gadget bar is pretty useful. at least for a person working for a small business. keeps you updated with news; currency; stock and other things really easily. the webcam is pretty good. as good as i have seen any web cam but i don't know any specific tests for it. it blurs if i move fast but otherwise looks good. also has settings to view what is happening in the webcam from other locations very easily. i don't know if i would ever use it but it sounds interesting.

Table 3-15 The example output of show all words in the paragraphs (Cont.)

Performance
Positive : 1 paragraphs vista is running great; i have had very few issues with programs that were not compatible and could find updates for all but one relatively fast. vista business has some interesting things; but most are more novelties that are not too important. i do think that the gadget bar is pretty useful. at least for a person working for a small business. keeps you updated with news; currency; stock and other things really easily. Negative : 0 paragraphs
Design
Positive : 1 paragraphs the screen looks incredible. i upgraded to the badass screen with truelife and i do not see any graininess. the screen is viewable from all angles; looks great up close and at a distance. Negative : 0 paragraphs

Table 3-15 The example output of show all words in the paragraphs (Cont.)

Feature
Positive : 4 paragraphs i received the vostro a couple days ago and the first thing i can say is that i love the build. the case is durable and looks good in all black. i also love the keyboard; i am not a good typer but i can type fast without looking at the keyboard on this which is something i could never do before on a laptop and still have trouble on other keyboards. the dvd burner burned a 4 gig dvd as fast as i would expect it to at max speed. vista is running great; i have had very few issues with programs that were not compatible and could find updates for all but one relatively fast. vista business has some interesting things; but most are more novelties that are not too important. i do think that the gadget bar is pretty useful. at least for a person working for a small business. keeps you updated with news; currency; stock and other things really easily. the webcam is pretty good. as good as i have seen any web cam but i don't know any specific tests for it. it blurs if i move fast but otherwise looks good. also has settings to view what is happening in the webcam from other locations very easily. i don't know if i would ever use it but it sounds interesting. Negative : 0 paragraphs
Other
Positive : 0 paragraphs Negative : 1 paragraphs sorry it took so long to review this; have a lot going on. anyways here goes.

Table 3-16 The example output of show the keywords of the text paragraphs

All subjective paragraphs
sorry_JJ so_RB long_JJ anyways_RB here_RB ago_RB first_JJ durable_JJ good_JJ black_JJ also_RB keyboard_NN not_RB good_JJ fast_RB keyboard_NN never_RB before_RB still_RB other_JJ keyboards_NNS screen_NN incredible_JJ screen_NN not_RB screen_NN viewable_JJ great_JJ close_RB dvd_NN dvd_NN fast_JJ great_JJ very_RB few_JJ programs_NNS not_RB compatible_JJ relatively_RB fast_RB interesting_JJ most_JJS more_RBR not_RB too_RB important_JJ pretty_RB useful_JJ least_JJS small_JJ other_JJ really_RB easily_RB webcam_NN pretty_RB good_JJ as_RB good_JJ n't_RB specific_JJ fast_RB otherwise_RB good_JJ also_RB webcam_NN other_JJ very_RB easily_RB n't_RB ever_RB interesting_JJ
Performance
Positive : 1 paragraphs great_JJ very_RB few_JJ programs_NNS not_RB compatible_JJ relatively_RB fast_RB interesting_JJ most_JJS more_RBR not_RB too_RB important_JJ pretty_RB useful_JJ least_JJS small_JJ other_JJ really_RB easily_RB Negative : 0 paragraphs

Table 3-16 The example output of show the keywords of the text paragraphs (Cont.)

Design
Positive : 1 paragraphs screen_NN incredible_JJ screen_NN not_RB screen_NN viewable_JJ great_JJ close_RB Negative : 0 paragraphs
Feature
Positive : 4 paragraphs ago_RB first_JJ durable_JJ good_JJ black_JJ also_RB keyboard_NN not_RB good_JJ fast_RB keyboard_NN never_RB before_RB still_RB other_JJ keyboards_NNS dvd_NN dvd_NN fast_JJ great_JJ very_RB few_JJ not_RB compatible_JJ relatively_RB fast_RB interesting_JJ most_JJS more_RBR not_RB too_RB important_JJ pretty_RB useful_JJ least_JJS small_JJ other_JJ really_RB easily_RB webcam_NN pretty_RB good_JJ as_RB good_JJ n't_RB specific_JJ fast_RB otherwise_RB good_JJ also_RB webcam_NN other_JJ very_RB easily_RB n't_RB ever_RB interesting_JJ Negative : 0 paragraphs
Other
Positive : 0 paragraphs Negative : 1 paragraphs sorry_JJ so_RB long_JJ anyways_RB here_RB

Table 3-17 The example output of show the keywords with their sentiment

All subjective paragraphs
sorry(v_neg), so(neu), long(neg), anyways(neu), here(neu) ago(neu), first(pos), durable(neu), good(v_pos), black(neg), also(neu), keyboard_Aspect, not(v_neg), good(v_pos), fast(neu), keyboard_Aspect, never(neg), before(neu), still(neg), other(neg), keyboards_Aspect, screen_Aspect, incredible(neu), screen_Aspect, not(v_neg), screen_Aspect, viewable(v_pos), great(pos), close(pos), dvd_Aspect, dvd_Aspect, fast(pos), great(pos), very(pos), few(neu), programs_Aspect, not(v_neg), compatible(pos), relatively(neu), fast(neu), interesting(pos), most(neu), more(neu), not(v_neg), too(neg), important(pos), pretty(neg), useful(neu), least(neu), small(neg), other(neg), really(pos), easily(pos), webcam_Aspect, pretty(neg), good(v_pos), as(neg), good(v_pos), specific(neu), fast(neu), otherwise(pos), good(v_pos), also(neu), webcam_Aspect, other(neg), very(pos), easily(pos), ever(neu), interesting(pos),
Performance
Positive : 1 paragraphs great(s_pos), very(pos), few(neu), programs_Aspect, not(v_neg), compatible(pos), relatively(neu), fast(neu), interesting(pos), most(neu), more(neu), not(v_neg), too(neg), important(pos), pretty(neg), useful(neu), least(neu), small(neg), other(neg), really(pos), easily(pos), Negative : 0 paragraphs

Table 3-17 The example output of show the keywords with their sentiment (Cont.)

<table>
<tr><th>Design</th></tr>
<tr><td>Positive : 1 paragraphs

screen_Aspect, incredible(neu), screen_Aspect, not(v_neg), screen_Aspect, viewable(v_pos), great(pos), close(pos),

Negative : 0 paragraphs</td></tr>
<tr><th>Feature</th></tr>
<tr><td>Positive : 4 paragraphs

ago(neu), first(pos), durable(neu), good(v_pos), black(neg), also(neu), keyboard_Aspect, not(v_neg), good(v_pos), fast(neu), keyboard_Aspect, never(neg), before(neu), still(neg), other(neg), keyboards_Aspect,

dvd_Aspect, dvd_Aspect, fast(pos),

great(pos), very(pos), few(neu), programs_Aspect, not(v_neg), compatible(pos), relatively(neu), fast(neu), interesting(pos), most(neu), more(neu), not(v_neg), too(neg), important(pos), pretty(neg), useful(neu), least(neu), small(neg), other(neg), really(pos), easily(pos),

webcam_Aspect, pretty(neg), good(v_pos), as(neg), good(v_pos), specific(neu), fast(neu), otherwise(pos), good(v_pos), also(neu), webcam_Aspect, other(neg), very(pos), easily(pos), ever(neu), interesting(pos),

Negative : 0 paragraphs</td></tr>
</table>

Table 3-17 The example output of show the keywords with their sentiment (Cont.)

Other
Positive : 0 paragraphs Negative : 1 paragraphs sorry(v_neg), so(neu), long(neg), anyways(neu), here(neu)

CHAPTER 4
EXPERIMENTAL RESULTS AND EVALUATIONS

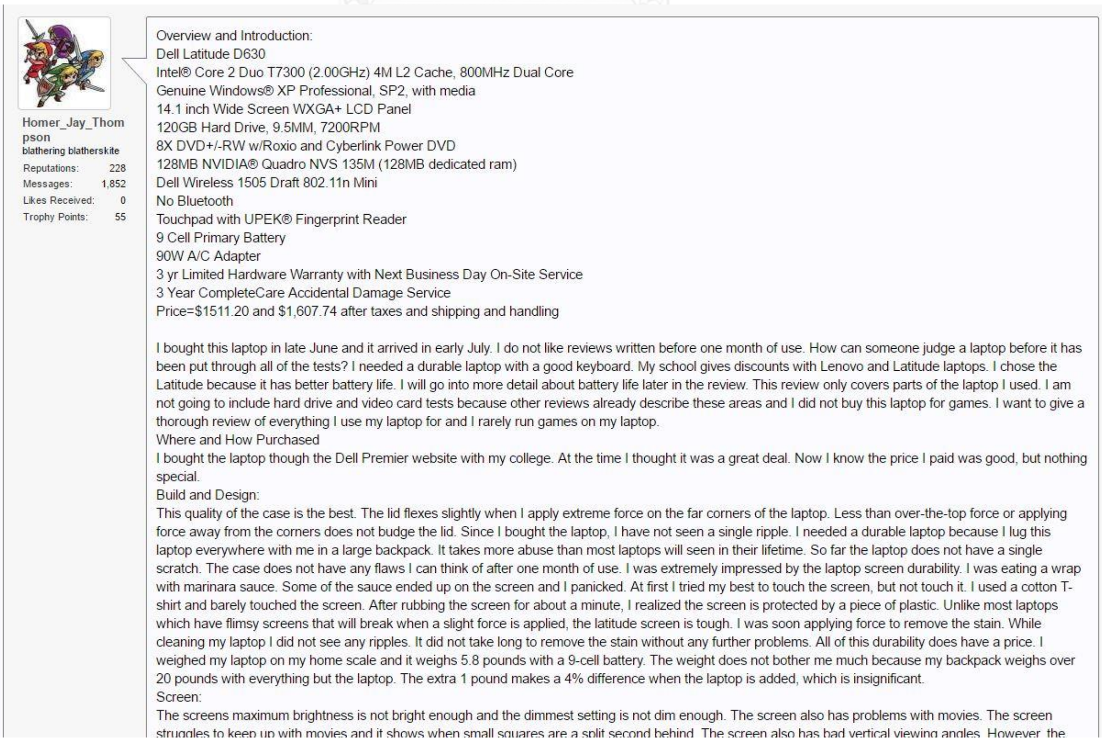
This chapter contains the evaluations by using the methodology from the previous chapter. This chapter is divided into 2 sections; Data Gathering is in Section 4.1 and Evaluations in Section 4.2.

4.1 Data Gathering

The input data gathering is performed by collecting the data from www.notebookreview.com and www.cnet.com between June and August 2015. Both websites are more popular than other websites because the number of visitors is higher than other websites. Details of the sampling scope are expanded below.

- The collected data take from 9 major brands in the community forum.
- Data contain 40 review topics per brands
- The sample data is in text format (*.txt)

The example of data from the community website is shown in Fig. 4-1



Homer_Jay_Thompson
blathering blatherskite
Reputations: 228
Messages: 1,852
Likes Received: 0
Trophy Points: 55

Overview and Introduction:
Dell Latitude D630
Intel® Core 2 Duo T7300 (2.00GHz) 4M L2 Cache, 800MHz Dual Core
Genuine Windows® XP Professional, SP2, with media
14.1 inch Wide Screen WXGA+ LCD Panel
120GB Hard Drive, 9.5MM, 7200RPM
8X DVD+/-RW w/Roxio and Cyberlink Power DVD
128MB NVIDIA® Quadro NVS 135M (128MB dedicated ram)
Dell Wireless 1505 Draft 802.11n Mini
No Bluetooth
Touchpad with UPEK® Fingerprint Reader
9 Cell Primary Battery
90W A/C Adapter
3 yr Limited Hardware Warranty with Next Business Day On-Site Service
3 Year CompleteCare Accidental Damage Service
Price=$1511.20 and $1,607.74 after taxes and shipping and handling

I bought this laptop in late June and it arrived in early July. I do not like reviews written before one month of use. How can someone judge a laptop before it has been put through all of the tests? I needed a durable laptop with a good keyboard. My school gives discounts with Lenovo and Latitude laptops. I chose the Latitude because it has better battery life. I will go into more detail about battery life later in the review. This review only covers parts of the laptop I used. I am not going to include hard drive and video card tests because other reviews already describe these areas and I did not buy this laptop for games. I want to give a thorough review of everything I use my laptop for and I rarely run games on my laptop.
Where and How Purchased
I bought the laptop though the Dell Premier website with my college. At the time I thought it was a great deal. Now I know the price I paid was good, but nothing special.
Build and Design:
This quality of the case is the best. The lid flexes slightly when I apply extreme force on the far corners of the laptop. Less than over-the-top force or applying force away from the corners does not budge the lid. Since I bought the laptop, I have not seen a single ripple. I needed a durable laptop because I lug this laptop everywhere with me in a large backpack. It takes more abuse than most laptops will seen in their lifetime. So far the laptop does not have a single scratch. The case does not have any flaws I can think of after one month of use. I was extremely impressed by the laptop screen durability. I was eating a wrap with marinara sauce. Some of the sauce ended up on the screen and I panicked. At first I tried my best to touch the screen, but not touch it. I used a cotton T-shirt and barely touched the screen. After rubbing the screen for about a minute, I realized the screen is protected by a piece of plastic. Unlike most laptops which have flimsy screens that will break when a slight force is applied, the latitude screen is tough. I was soon applying force to remove the stain. While cleaning my laptop I did not see any ripples. It did not take long to remove the stain without any further problems. All of this durability does have a price. I weighed my laptop on my home scale and it weighs 5.8 pounds with a 9-cell battery. The weight does not bother me much because my backpack weighs over 20 pounds with everything but the laptop. The extra 1 pound makes a 4% difference when the laptop is added, which is insignificant.
Screen:
The screens maximum brightness is not bright enough and the dimmest setting is not dim enough. The screen also has problems with movies. The screen struggles to keep up with movies and it shows when small squares are a split second behind. The screen also has bad vertical viewing angles. However, the

Figure 4-1 Example of data from the community website

According to Fig.4-1, only texts from the community site written by reviewers will be manually saved into a text format (the automatic extraction process will be implemented in the future work). The review data will be saved one review per one text file. There are 360 reviews collected from two community websites and divided into 15,384 paragraphs. The answers of these review paragraphs (subjective or objective paragraphs, in which aspect domain, and positive or negative paragraphs) are decided by the researcher based on the following rules:

1. Subjective or objective paragraphs
   - If the paragraphs have at least one emoticon text or subjective words with four parts of speech (nouns, verb, adjectives, adverbs), then those paragraphs will be considered as subjective paragraphs.
2. The aspect domain
   - Some review writing styles already separated the content into each aspect (performance, design, feature and others). So text paragraphs with these writing styles will be identified the aspect group clearly.
   - If the paragraphs contain one or more words in the aspect word list of each domain, then these paragraphs will be identified to that aspect.
3. Positive or negative paragraphs

   3.1 If the paragraphs have only positive subjective words, then those paragraphs will be considered as positive paragraphs. On the other hand, if the paragraphs have only negative subjective words, then those paragraphs will be considered as negative paragraphs.

   3.2 If there are both positive and negative subjective words in the same paragraph, then the number of positive words is compared to the number of negative words in the paragraph.

   - If the number of positive words is exactly more than the number of negative words, this paragraph will be the positive.
   - If the number of negative words is exactly more than the number of positive words, this paragraph will be the negative.

- If the number of positive words and negative words are hardly different, the total of word polarity scores from the Sentiwordnet is calculated to classify the sentiment.
  - If the total score is more than 0, this paragraph will be the positive.
  - If the total score is less than 0, this paragraph will be the negative.
  - If the total score equal to 0, the NLTK will decide whether it is a positive or negative paragraph.
  - If the total score conflicts with researcher's feeling, it will depend on the researcher's judgment whether it is a positive or negative paragraph. For example, one subjective word 'small' in the design aspect is a positive word, while 'small' in Sentiwordnet has a negative score.

4.2 Experimental Result

There are 3 parts of the experiment in this research which are detecting subjective paragraph, identifying aspects and classifying sentiment. All parts will be evaluated by the confusion matrix [22].

4.2.1 Subjective Detection

Firstly, all selected review topics will be separated into 15,384 paragraphs by detecting a new line. These paragraphs can be divided into 6,399 subjective paragraphs and 8,985 objective paragraphs. To detect subjectivity, words in the experimental data will be compared with words in Sentiwordnet and our emoticon lexicon. The result of the subjective detection is classified into 6,992 subjective paragraphs and 8,392 objective paragraphs. The confusion matrix of the result in the subjective detection is shown in Table 4-1.

Table 4-1 The confusion matrix of the result in the subjective detection

Actual Class		**Predicted Class**	
		Subjective	*Objective*
Subjective	6,399	6,367	32
Objective	8,985	625	8,360
Total	15,384	6,992	8,392

According to Table 4-1, the correct subjective prediction of this process is 6,367 of 6,399. The percent of accuracy, precision and recall rate of the subjective detection are shown in Table 4-2.

Table 4-2 The percent of accuracy, precision and recall rate of the subjective detection

Class	**Percent of**		
	Accuracy	*Precision*	*Recall*
Subjective paragraphs	95.73	91.06	99.50
Objective paragraphs	95.73	99.62	93.04

Referring to Table 4-2, the accuracy, precision and recall rates are more than 90% for subjective paragraph detection. This means that the proposed framework can detect subjective review paragraphs effectively.

Nevertheless, one error type of detection occurred by word style writing, such as “Hard disk” and “Harddisk”. This style writing can make the error of part of speech tagging because POS Tagger can correctly identify “Harddisk” as noun, but “Hard disk” will be incorrectly identified into two words “Hard” as adjective and “disk” as noun. Then, “Hard” will be detected to be a subjective word. As a result, some objective paragraphs containing “Hard disk” in the paragraphs will be classified into subjective paragraphs.

4.2.2 Aspect Identification

Secondly, subjective paragraphs will be identified into each aspect group (Performance, Design, and Feature). However, some subjective paragraphs, which cannot be identified into these groups, will be classified into "Other" aspect. The subjective paragraphs can be categorized into 1,347 performance-paragraphs, 1,796 design-paragraphs and 1,658 feature-paragraphs by researcher reading manually. The confusion matrix of the result in the aspect identification is shown in Table 4-3. The percent of accuracy, precision and recall rate of the aspect identification are also shown in Table 4-4.

Table 4-3 The confusion matrix of the result in the aspect identification

Class	Actual	Predicted	TP	FP	TN	FN
Performance	1,347	1,489	1,334	155	5,490	13
Design	1,796	2,150	1,737	413	4,783	59
Feature	1,658	1,811	1,614	197	5,137	44

Table 4-4 The percent of accuracy, precision and recall rate of the aspect identification

Class	Percent of		
	Accuracy	*Precision*	*Recall*
Performance	97.60	89.59	99.03
Design	93.25	80.79	96.71
Feature	96.55	89.12	97.35

Referring to Table 4-3 and Table 4-4, the percent of accuracy and recall rate of all classes are more than 90%, although the precision rate is about 80%. As a result, the aspect identification has high accuracy to identify the aspect of collected reviews in this research.

However, the results of the aspect identification show that all aspect groups of paragraphs are exaggeratingly identified because words in text paragraphs are compared to defined words in each aspect domain. One reason of this defect is that a few words in the domain such as “solid” and “size” including in the design aspect, but the paragraphs consisting of those words are not in the design aspect. The first example is “solid state drive faster than HDD 7200rpm” and the second example is “I also like the keyboard that is of a decent size”. In the first example, “solid state drive” is a storage device so this paragraph should be identified into the performance aspect. Unfortunately “solid” are detected and this paragraph is identified as the design aspect. In addition, “size" is a common word from the second example paragraph and the review writer discussed keyboard, so it should be identified as the feature aspect. However, this framework found “size” and detected the second example into the design aspect.

4.2.3 Sentiment Classification

Finally, all paragraphs in each aspect will be classified the sentiment by using Naive Bayes in WEKA. Only polarity level of adjective and adverb words are learning data in the sentiment classification. The confusion matrix and the percent of accuracy, precision and recall rate of the sentiment classification are shown in Table 4-5 and Table 4-6.

Table 4-5 The confusion matrix of the result in sentiment classification

Actual Class		**Predicted Class**	
		Positive	*Negative*
Positive	2,534	1,907	627
Negative	3,700	781	2,919
Total	6,234	2,688	3,546

Table 4-6 The percent of accuracy, precision and recall rate of sentiment classification

Class	**Percent of**		
	Accuracy	*Precision*	*Recall*
Positive	77.41	82.30	78.90
Negative		70.90	75.30
Average		77.70	77.40

Referring to Table 4-6, the average accuracy, precision and recall rates are more than 75% for sentiment classification. This means that the proposed framework can classify the sentiment of subjective review paragraphs acceptably on collected reviews.

One reason for incorrect sentiment classification is that the Naïve Bayes classifier uses the probabilities of features to make a classification. In this research, the features are the polarity level of words appearing in text paragraphs. Therefore, the position and the frequency of occurring polarities levels of words in paragraphs will affect the classification. For example, two paragraphs consists three positive words, three negative words and two neutrals in the feature set. The sentiments of these paragraphs may be a positive or negative paragraph depending on the sequence of words' polarities levels in paragraphs.

CHAPTER 5
CONCLUSIONS AND DISCUSSIONS

In this chapter, the discussion will be described in Section 5.1. Section 5.2 will be started the limitation of the experiment and finally conclusion will be discussed in Section 5.3.

5.1 Discussion

There are three main processes of the proposed framework: the subjective detection, the aspect identification and the sentiment classification.

This research detects the subjective paragraphs by comparing subjective words in the adjective and adverb groups only as most of the words that indicate the subjectivity is in the adjective and adverb groups. This is enough for detecting the subjective words from the collected reviews even though there are words indicating feelings in other part of speech also such as verb. However, the words indicating feeling in the verb group are rare comparing to other words in the same group. Even though the percent of recall rate of the subjective detection is nearly 100% and the percent of precision rate is among 90%, the error is still occurred. The error can be happened is that this process will consider some objective paragraphs as subjective paragraphs. For example, the technical word such as “Harddisk” which was written as “Hard disk” will be considered as a subjective paragraph. As a result, some objective paragraphs containing “Hard disk” in the paragraphs will be classified into subjective paragraphs.

To identify aspects, the difference between this research and other articles is the aspect classification divided into 3 aspects: performance, design and feature. These 3 aspects are normally the top significant matters domain for the laptop reviews. This process will distinguish the aspect by comparing words to words in the aspect domain which generated from the selected high frequency words appearing in the text review paragraphs. Nevertheless, the selected words will be categorized into each aspect by the researcher’s consideration as explained in section 3.3. However, the results show that all aspect groups of paragraphs are exaggeratingly identified. One reason of this error is that the subjective words in one aspect domain are the proper nouns or compound nouns in another aspect domain. For example, “solid” is the subjective words in the design aspect, while “solid state drive” is a storage device mostly appearing on review paragraphs in the performance aspect.

For the sentiment classification, the difference between this research and others is that this research uses the polarity level of words (defining by the sentiment word scores from the polarity lexicon) as the feature for creating a model to classify sentiment. This feature set is more efficient than the other feature sets as explained in section 3.4.1. In addition, the Naïve Bayes classifier is used to be the machine learning in this framework because the Naïve Bayes classifier is more efficient than other classification techniques (J48, RBF Network and ZeroR) Nevertheless, the limitation of this sentiment classification model is that the probabilities of occurred words’ polarity levels are calculated to classify the sentiment. If the word information is not enough for defining the correct polarity level of words found on reviews, the performance of this classification model will be reduced.

5.2 Limitation

A major limitation of this framework is unseen words which are not included in the polarity lexicon, but may be found on laptop reviews. The reason is that this review analysis framework focuses on words and the polarity of words to identify the aspect and to classify the sentiment of review paragraphs. If there are some missing input words from our lexicons, such as words with wrong spelling or technical words, the polarity level finding in the sentiment classification process cannot give the correct polarity level of words. Therefore, the performance of the sentiment classification process will be reduced by this error.

5.3 Conclusion

Nowadays many laptops are manufactured with various features. When consumers decide to purchase a laptop, they normally search for laptop reviews in order to get the information first. Moreover, many reviews are created to let the consumers know more about each laptop. For those reasons, this research then developed a framework which helps users to know what is mentioned in the laptop reviews. The framework consists of four main processes: preparing data for analysis, detecting subjective text paragraphs, identifying the aspect of each text paragraph and classifying the sentiments of each text paragraph. Firstly, photos, URL links, special characters and symbols are deleted from input reviews. Then, the reviews are separated into individual paragraphs and words in the paragraph are tagged their parts of speech. Secondly, the subjective paragraphs are detected by discovering at least one emoticon text or one subjective word in prepared paragraphs. Thirdly, only the subjective paragraphs are categorized into four aspects (performance, design, feature, others) by comparing words with words in each aspect domain because these aspect groups are normally the top significant matters domain for the laptop reviews. Finally, the sentiment of paragraphs will be classified by the polarity levels of words as a feature and Naïve Bayes classifier as a machine learning. The results of performance evaluation show that the subjective detection and the aspect identification has high accuracy and precision, including acceptably accurate and precise sentiment classification. In conclusion, this framework is useful for developing the review analysis system of laptops in order to help consumers gain information before purchasing a laptop.

APPENDIX A

LIST OF ALL SPECIAL CHARACTERS AND SPECIAL SYMBOLS

Table A-1 List of all special characters

À	×	Á	á	à	Â	â
Ä	ä	Ǎ	ǎ	Ă	ă	Ā
ā	Ã	ã	Å	å	Ą	ą
Ǡ	ǡ	Ǻ	ǻ	Ǟ	ǟ	Æ
æ	Ǽ	ǽ	Ǣ	ǣ	ƀ	Ƀ
Ƃ	ƃ	Ƅ	ƅ	Ć	ć	Ċ
ċ	Ĉ	ĉ	Č	č	Ç	ç
Ƈ	ƈ	Ɔ	Ď	ď	Đ	đ
Ƌ	ƌ	Ɖ	Ð	ð	ƍ	Ǳ
ǲ	ǳ	Ǆ	ǅ	ǆ	Đ	É
é	È	è	Ė	ė	Ê	ê
Ë	ë	Ě	ě	Ĕ	ĕ	Ē
ē	Ę	ę	Ǝ	ə	Ə	Ɛ
Ƒ	ƒ	Ǵ	ǵ	Ġ	ġ	Ĝ
ĝ	Ğ	ğ	Ğ	ğ	Ģ	ģ
Ǥ	ǥ	Ɠ	Ɣ	Ĥ	ĥ	Ħ
ħ	ƕ	Ƕ	ı	Í	í	Ì
ì	İ	Î	î	Ï	ï	Ǐ
ǐ	Ĭ	ĭ	Ī	ī	Ĩ	ĩ
Į	į	Ɨ	Ɩ	Ĵ	ĵ	ǰ
ĸ	Ǩ	ǩ	Ķ	ķ	Ƙ	ƙ
Ĺ	ĺ	Ŀ	ŀ	Ľ	ľ	Ļ
ļ	ł	Ł	ł	ƛ	Ɯ	Ń
ń	Ǹ	ǹ	Ň	ň	Ñ	ñ
Ņ	ņ	Ɲ	ŉ	ƞ	Ŋ	ŋ

Table A-1 List of all special characters (Cont.)

Ó	ó	Ò	ò	Ô	ô	Ö
ö	Ǒ	ǒ	Ŏ	ŏ	Ō	ō
Õ	õ	Ǫ	ǫ	Ő	ő	Ɵ
Ø	Ǭ	ǭ	Ơ	ơ	Ƣ	ƣ
Œ	œ	Ƥ	ƥ	Ʀ	Ŕ	ŕ
Ř	ř	Ŗ	ŗ	Ś	ś	Ŝ
ŝ	Š	š	Ş	ş	Ʃ	Ƨ
ƨ	ƪ	ß	ſ	Ť	ť	Ţ
ţ	Ƭ	ƭ	ƫ	Ʈ	Þ	þ
Ŧ	ŧ	Ú	ú	Ù	ù	Û
û	Ü	ü	Ǔ	ǔ	Ŭ	ŭ
Ū	ū	Ũ	ũ	Ů	ů	Ų
ų	Ű	ű	Ǘ	ǘ	Ǜ	ǜ
Ǚ	ǚ	Ǖ	ǖ	Ư	ư	Ʊ
Ʋ	Ŵ	ŵ	ƿ	Ƿ	Ý	ý
Ŷ	ŷ	ÿ	Ÿ	Ƴ	ƴ	Ź
ź	Ż	ż	Ž	ž	Ƶ	

Table A-2 List of all special symbols

™	®	©	☑	√	✓	☒
╳	✗	⊗	×	✕	ヅ	ツ
ッ	シ	♂	♀	☿	♮	♯
℃	℉	♪	♫	♩	♬	♭

APPENDIX B

ALL SUBJECTIVE WORD USED IN THIS THESIS

Table B-1 All subjective words of adjective in this thesis

able	above	actual	additional
advanced	affected	afraid	all-around
amazing	animated	annoying	anti-static
apparent	appropriate	audible	audio
automatic	available	average	aware
baby-filled	backlit	bad	balanced
beautiful	best	better	big
black	bottom	brick-and-mortar	bright
brown	bummed	canadian	capable
capacitive	careful	carnvial	casual
centered	certain	chaotic	cheap
cheapest	clean	clear	clocked
close	cold	comfortable	common
comparable	compatible	complicated	confident
confused	conservative	cool	cross-pacific
current	dark	darn	dead
decent	decorative	dedicated	demographic
derivative	developed	different	difficult
dim	direct	discerning	distinct
disturbing	divx/xvid	dreaded	dual
due	durable	dynamic	early
easier	easy	efficient	empty
enough	entire	equal	ergonomic
everquest	evil	excellent	exceptional
expensive	express	expresscard	extended

Table B-1 All subjective words of adjective in this thesis (Cont.)

External	extra	extreme	fantastic
fast	favorite	few	final
fine	finicky	first	flat
flexable	flush	former	fortunate
forward	four-speaker	free	fresh
friendly	front	full	functional
general	gimmicky	glad	glossy
glowing	good	graphic	great
green	grey	happy	hard
heavier	heavy	hefty	helpful
high	higher	horizontal	horrific
hot	huge	hyperlinked	ideal
idle	immediate	important	impressed
impressive	in-built	incredible	individual
indoor	industrial	informative	initial
integrated	intelligent	intense	intensive
interested	interesting	ips-wide	irritating
key	laptop	large	larger
largest	last	later	latest
least	left	less	light
lightest	like	limited	little
loaded	local	located	long
longer	loose	loud	low
lower	lowest	main	manageable
many	maximum	med	medium
mental	microsoft	mid	middle
minimalist	minor	mobile	modern
more	most	much	multiple
Native	neat	necessary	negative

Table B-1 All subjective words of adjective in this thesis (Cont.)

New	news	next	nice
noise-cancelling	noisier	non-sensitive	normal
notebookreview	noteworthy	noticeable	observable
obvious	odd	ok	okay
old	older	omega	on/off
one-fan	only	only-adequate	open
optical	optional	original	other
outdoor	outer	outside	outstanding
outward	overall	own	page-file
particular	passionate	perfect	personal
plastic	play/pause	playable	pleased
portable	possible	powerful	power-on
practical	present	previous	primary
prime	prior	pro	professional
protective	quick	quiet	random
ready-boost	real	rear	recent
regular	remote	resident	responsive
rewind	right	same	sanlitun
sata	satelite	satisfactory	satisfied
saturated	secure	semi-knowledgeable	sensitive
serial	serious	seriuos	several
shader	sharp	short	side
silent	similar	simple	simplistic
single-minded	sleek	slight	slower
small	smooth	snappy	snuggly
soft	solid	sony	sophisticated
sorry	sound	special	specific
Splendid	standard	stiff	straight

Table B-1 All subjective words of adjective in this thesis (Cont.)

Strict	stupid	sturdy	subcategory
subtle	such	sufficient	super
superfluous	sure	surprising	textured
thermal	thick	thicker	thin
thinkpad	thinnest	toasty	top
total	touching	traditional	transparent
tricky	trivial	true	type
unassuming	unchanged	uncomfortable	underwater
unibody	unmute	unreal	unscathed
unusual	upgrade	upgradeable	useful
user-replaceable	valuable	variable	various
video	viewable	virtual	visible
warm	warmer	warmest	wary
wash-free	weak	weakest	white
whole	wireless	worldwide	worried
worst	worth	worthwhile	wrong

Table B-2 All subjective words of adverb in this thesis

abnormally	about	absolutely	accidentally
accidently	actually	additionally	admittedly
again	ago	ahead	almost
already	also	altogether	always
amazingly	anytime	anyway	anyways
apart	around	as	automatically
away	back	badly	barely
basically	before	better	better
carefully	certainly	clearly	close
comfortably	completely	constantly	definately
definitely	directly	down	easily
else	enough	especially	essentially
even	ever	everywhere	excessively
faily	fairly	far	fast
faster	finally	firmly	first
fluently	forth	fortunately	forward
frequently	fully	further	furthermore
hard	hardly	hence	here
hopefully	hotter	however	immediately
indeed	indoors	inside	instead
internally	just	later	later
likely	little	long	louder
mainly	maybe	merely	more
most	mostly	much	nearly
never	nicely	nope	normally
not	notably	noticeably	now
nowadays	n't	obviously	occasionally
off	often	once	only
originally	otherwise	outstandingly	over

Table B-2 All subjective words of adverb in this thesis (Cont.)

overall	overly	passively	personally
pretty	previously	probably	properly
quickly	quite	rarely	rather
really	recently	relatively	right
roughly	seemingly	separately	significantly
simply	slightly	slowly	so
softly	somehow	sometimes	somewhat
somewhere	soon	sooner	specifically
still	surprisingly	technically	thankfully
then	there	therefore	though
thus	tightly	together	too
typically	unfortunately	up	usually
vastly	vertically	very	virtually
well	wirelessly	yet	

APPENDIX C

ALL EXTRA WORDS IN POLARITY LEXICOM

Table C-1 The extra words in comparative form

blacker	bolder	brighter	broader
bulkier	cheaper	chunkier	cleaner
clearer	closer	cooler	crisper
darker	deeper	denser	dimmer
dual-finger	easier	edgier	fancier
faster	feather	flashier	friendlier
fuller	fuzzier	glossier	happier
harder	harsher	heavier	heftier
hotter	lighter	longer	looser
louder	lower	narrower	newer
nicer	noisier	odder	poorer
quicker	richer	riskier	rougher
safer	sexier	sharper	shorter
simpler	skinnier	sleeker	slighter
slimmer	slower	smarter	smoother
snappier	softer	squatter	starter
stiffer	stronger	sturdier	taller
thicker	thinner	tighter	tougher
warmer	weaker	wider	yellower

Table C-2 The extra words in superlative form

barest	beefiest	biggest	boldest
brightest	broadest	bulkiest	cheapest
clearest	closest	coolest	dimmest
easiest	fanciest	fastest	finest
fullest	hardest	heaviest	highest
highest	hottest	largest	lightest
longest	loudest	lousiest	nearest
newest	nicest	palm-rest	priciest
quietest	sharpest	shortest	simplest
sincerest	sketchiest	slightest	slimmest
slowest	smallest	strongest	sturdiest
thickest	thinnest	toughest	warmest
weakest			

Table C-3 All concatenated words the extra words

above-average	accident-proof	add-on	aggressive-looking
air-alike	air-light	air-like	all-aluminum
all-angles	all-black	all-day	all-in-all
all-in-one	all-in-ones	all-matte-black	all-over
all-over-the-place	all-plastic	all-ssd	all-too-common
all-white	alphabet-spanning	already-stellar	also-excellent
also-flat	also-recently-reviewed	aluminum-and-plastic	aluminum-clad
always-on	amazing-looking	angst-ridden	anti-aliasing
anti-alising	anti-blue	anti-ghosting	anti-glare
anti-reflective	anytime-vista	arcade-style	arm-based
artificial-feeling	atom-based	atom-powered	attention-grabbing
audio-shaping	automatic-switching	auto-snapping	average-size
baby-filled	back-and-forth	back-breaking	back-light
back-lighting	back-lit	back-to-school	bang-for-the-buck
bare-bones	bargain-basement	bass-heavy	bass-worthy
battery-drain	battery-friendly	battery-life	bd-enabled
beats-branded	best-built	best-designed	best-feeling
best-forgotten	best-in-class	best-in-show	best-kept
best-looking	best-performing	best-suited	better-built
better-detailed	better-made	better-quality	better-sounding
better-than-1080p	better-than-average	better-than-cd	better-than-hd
bezel-less	big-brand	big-name	big-screen
big-screened	blade-thin	blocked-up	borderline-alienating
bottom-mounted	bottom-rung	boutique-level	brand-spanking-new
brick-and-mortar	brick-like	brittle-feeling	bro-tastic
browser-based	brushed-aluminum	brushed-black	brushed-metal

Table C-3 All concatenated words the extra words (Cont.)

brushed-metal-like	budget-breaking	budget-driven	budget-feeling
budget-focused	budget-friendly	budget-looking	budget-minded
budget-priced	budget-range	budget-targeted	build-it-yourself
bumped-up	bushed-metal	business-friendly	business-minded
business-oriented	business-rugged	business-targeted	butter-smooth
button-bar	button-free	buzz-heavy	carbon-fiber
carry-all-day-every-day	cd-maker	center-right	ceo-level
cheap-feeling	cheap-looking	chiclet-style	chintzy-feeling
cinemascope-wide	clack-free	class-leading	classy-looking
cleanest-feeling	clean-looking	clevo-based	clevo-chassis
click-free	clickpad-style	click-pad-style	closest-performing
cloud-based	cloud-heavy	cloud-storage	coach-class
coffee-shop	coke-bottle-glasses	college-bound	color-accented
comfortable-feeling	coming-soon	conservative-looking	console-style
consumer-friendly	consumer-level	consumer-oriented	consumer-targeted
context-sensitive	cooler-running	core-m-powered	corporate-friendly
cost-saving	counter-strike	cpu-extensive	cpu-intensive
cpu-upgraded	cramped-but-cozy	cross-comparable	crossover-friendly
cross-pacific	cross-platform	crystal-like	current-gen
current-generation	custom-programmed	cut-off	cutting-edge-technology
cyber-effect	data-save	dated-looking	deal-breakers
deal-breaking	deal-killing	decent-but-pedestrian	decent-enough
decent-looking	decent-size	decent-sized	decked-out
deep-sounding	department-issued	design-heavy	desktop-dominating

Table C-3 All concatenated words the extra words (Cont.)

desktop-replacement	desktop-replacement-level	desktop-sized	detachable-screen
dialed-down	diamond-textured	dis-asembling	discrete-class
display-style	display-type	distortion-free	double-width
dragon-and-floral	dust-attracting	dust-shielding	dvd-drive
dvd-ripping	ear-buds	early-adopter	easier-on-the-eye
easy-access	easy-fold	easy-to-lose	easy-to-miss
easy-to-see	easy-to-use	eco-button	edge-lit
edge-of-pad	edge-to-edge	education-targeted	energy-efficient
energy-saving	enthusiast-grade	enthusiast-oriented	entry-level
entry-point	error-free	esata-only	even-higher-than-hd
even-more-expensive	ever-lower	ever-popular	ever-so-slightly
ever-so-slightly-off	every-day	everyday-use	executive-level
extended-cell	extended-life	extra-deep	extra-high
extra-large	extra-long	extra-portable	extra-slim
extra-wide	extra-wide-screen	eye-candy	eye-scorching
fabric-like	family-friendly	fast-booting	faster-than-air
faster-than-usb	faster-throughput	faux-wood	feature-free
feature-rich	feature-wise	feedback-free	fiber-like
finger-control	finger-friendly	finger-print	fingerprint-reading
fingerprint-resistant	fingerprint-revealing	finger-swipe	finger-swiping
finger-tapping	fixed-configuration	flash-capable	flat-matte
flat-out	flex-free	flip-and-fold	flip-and-rip
flip-down	flippy-convertible	flip-screen	foam-lined
fold-back	forward-thinking	four-cell	four-color
four-column	four-finger	four-speaker	four-star
fourth-gen	fourth-generation	four-way	fragile-feeling

Table C-3 All concatenated words the extra words (Cont.)

frill-free	front-facing	frustration-free	full-colored
fuller-featured	full-feature	full-featured	full-function
full-hd	full-on	full-power	full-powered
full-price	full-screen	full-sized	full-uhd
full-voltage	full-width	function-key	function-reversed
future-proofed	game-centric	game-playing	gamer-centric
gamer-friendly	gamer-level	gamer-like	gamer-oriented
gamer-targeted	gaming-capable	gaming-centric	gaming-class
gaming-focused	gaming-oriented	garish-looking	geared-up
generation-skipping	generic-looking	genre-leading	glare-free
glass-and-plastic	glass-covered	glasses-free	glass-fronted
glass-topped	glossy-only	go-anywhere	gold-plated
gpu-extensive	gpu-intensive	graphic-intensive	graphics-based
greater-than-1080p	great-looking	ground-breaking	ground-shaking
grown-up	half-height	half-inch	half-the-price
hand-assembled	hand-in-hand	handy-dandy	hard-drive
hard-to-beat	hard-to-overlook	hard-to-use	hard-wired
hd-friendly	hd-quality	hd-video	hi-def
high-capacity	high-concept	high-contrast	high-cost
high-definition	high-design	high-detail	high-end
higher-capacity	higher-def	higher-end	higher-priced
higher-quality	higher-res	higher-resolution	higher-than-1080p
higher-than-hd	higher-than-normal	higher-up	highest-end
highest-resolution	high-gloss	high-maintenance	high-price
high-quality	high-quality/high-resolution	high-res	high-water
hinged-forward	hinge-forward	hi-res	hit-and-miss
hot-swappable	image-processing	image-sensing	improved-light-sensitivity

Table C-3 All concatenated words the extra words (Cont.)

in-browser	in-built	indigo-blue	industrial-design
industrial-style	industry-first	industry-leading	industry-standard
in-game	in-house	innocuous-looking	inoffensive-looking
in-plane	in-screen	instant-on	intel-based
intel-coined	intel-powered	in-thing	in-transit
ios-like	ipad-strength	ips-style	island-style
issue...heat-related	it-department-friendly	it-focused	it-friendly
it-oriented	jack-of-all-trades	jaw-dropping	jaw-droppingly
jazzed-up	jbl-branded	jump-out	just-announced
just-released	just-under	keyboard/trackpad/palm-rest	key-squishing
kid-friendly	killer-branded	kind-of	kiosk-style
lag-free	laid-out	laptop-like	large-media
larger-bodied	larger-capacity	larger-scale	larger-screen
larger-than-average	larger-than-normal	larger-than-usual	large-screened
laser-focused	last-all-day	last-gen	latency-sensitive
least-expensive	leather-like	led-backlighting	led-lit
left-click	left-side	lenovo-branded	lesser-featured
less-expensive	less-pricey	less-than-impressive-feeling	less-than-optimal
less-than-premium-feeling	letterbox-bar-free	lightning-bolt	lightning-quick
light-up	light-years	line-in	line-up
live-streaming	loc-tite	log-in	long-awaited
longer-battery-life	longer-form	longer-term	longest-lasting
longest-life	longest-lived	long-form	long-gone
long-lost	long-overdue	long-running	long-serving

Table C-3 All concatenated words the extra words (Cont.)

long-standing	long-time	loudness-adjustment	low-end
lower-cost	lower-end	lower-power-oriented	lower-priced
lower-res	lower-resolution	lower-right	lower-speed
lower-than-expected	lower-voltage	lowest-cost	lowest-density
lowest-end	lowest-power	lowest-price	lowest-priced
low-light	low-medium	low-power	low-price
low-quality	low-res	magnesium-alloy	mailer-envelope-style
mail-in	mainstream-level	mainstream-looking	mainstream-quality
make-or-break	matte-black	matte-finish	matte-finished
matte-metallic	medium-end	medium-high	medium-quality
medium-to-high	mega-hd	memory-card	memory-hogging
metal-and-plastic	metallic-grey	metallic-red	metal-like
me-too	micro-fiber	micro-hdmi	micro-sim
microsoft-approved	micro-to-full-size	micro-usb	middle-brightness
middle-ground	middle-of-the-pack	mid-game	mid-high
mid-level	mid-price	mid-priced	mid-range
midrange-to-high-end	mid-size	mid-to-high-end	mid-to-older
military-spec	mini-desktop	mini-displayport	mini-hdmi-out
mini-pcie	mirror-finish	mirror-like	mission-critical
mis-sized	modern-feeling	modern-looking	months-old
more-advanced	more-affordable	more-challenging	more-demanding
more-detailed	more-distinct	more-expensive	more-intensive
more-powerful	more-recent	more-robust	more-rounded

Table C-3 All concatenated words the extra words (Cont.)

more-standard	more-than-standard	more-typical	most-premium
movie-targeted	movie-viewing	movie-watching	much-improved
much-maligned	mulit-touch	multi-card	multi-channel
multi-gesture	multi-key	multi-monitor	multi-page
multiple-macbook	multiple-program	multi-speaker	multi-tasking
multi-touch	music-encoding	must-buy	must-have
nano-powered	natural-feeling	near-constant	near-future
near-perfect	near-useless	netbook-esque	netbook-grade
netbook-like	netbook-style	network-attached	next-day
next-gen	next-generation	nice-feeling	nice-looking
nicer-looking	nicest-looking	nigh-identical	nine-cell
nit-picking	nit-picky	no-buy	no-click
no-compromise	noise-cancelling	non-3d	non-4k
non-aircon	non-all	non-apple	non-arm
non-backlit	non-bias	non-business	non-clamshell
non-component-related	non-duo	non-existent	non-functioning
non-gamer	non-gaming	non-glare	non-gloss
non-glossy	non-haswell	non-hybrid	non-ips
non-metal	non-modular	non-netbook	non-obtrusive
non-reactive	non-reflective	non-retina	non-retina-display
non-rubberized	non-sensitive	non-ssd	non-stop
non-synaptics	non-tapered	non-textured	non-touch
non-touch/non-yoga	non-touchscreen	non-trubright	non-ulv
non-upgraded	non-widescreen	non-windows	no-power
not-at-all-shabby	not-metro	not-quite-discrete	not-so-secret
not-so-useful	not-too-distant	now-defunct	now-familiar

Table C-3 All concatenated words the extra words (Cont.)

now-outdated	off-angle	off-axis	off-edge
office-based	offline-enabled	off-size	oft-cited
often-confusing	often-overlooked	oft-repeated	oft-requested
oh-so-much	old-school	olufson-designed	on-and-off
on-board	one-generation-behind	on-goings	only-adequate
on-the-move	opened-up	optical-drive-free	optimus-enabled
os-level	otherwise-excellent	out-of-sync	out-of-the-box
outward-facing	over-exposed	over-saturated	over-sized
page-down	page-file	painted-on	palm-check
palm-rest	paper-thin	part-aluminum	parts-and-labor
patent-pending	pcie-based	pc-only	pebble-style
perfect-for-almost-anyone	performance-wise	phablet-like	photo-editing
photoshop-style	plain-looking	plastic-feeling	plugged-in
plug-in	pocket-friendly	pop-up	portability-minded
port-studded	post-ipad	post-ultrabook	powder-coated
power/quick-launch	powered-off	power-efficient	power-hungry
power-on	power-packed	power-related	power-testing
pre-assembled	pre-atom	pre-calibrated	pre-installed
pre-loaded	premium-feeling	premium-level	premium-priced
pre-netbook	pre-order	previous-gen	previous-generation
pre-windows	price-conscious	price-sensitive	price-to-value
price-wise	processor-wise	professional-grade	professional-level
professional-looking	pro-level	proof-of-concept	prosumer-level
pull-apart	pull-down	pull-tab	push-in
put-down	quad-core	quick-access	quick-booting

Table C-3 All concatenated words the extra words (Cont.)

quick-launch	quick-start	race-car-like	ram-only
random-model-number-to-english	rant-like	razor-thin	read-only
ready-boost	real-life	real-world	rear-facing
rear-vented	reasonable-sounding	re-building	red-accented
red-backlit	red-to-silvery-gray	re-imagined	replacement-type
re-sizeable	retail-only	retail-specific	retina-level
retro-modern	rocker-bar	rocker-style	rock-solid
room-filling	rotating-screen	rounded-edge	rubber-insulated
scratch-resistant	self-adjusting	semi-durable	semi-flush
semi-hot	semi-hybrid	semi-integrated	semi-knowledgeable
semi-mobile	semi-offline	semi-opaque	semi-raised
semi-serious	semi-similar-looking	semi-thin	semi-transparent
sepia-toned	shape-shifting	sharper-looking	sharp-looking
sharp-screened	shorter-than-most	side-by-side	side-edge
side-firing	side-release	silky-smooth	silver-and-black
similar-in-concept	similar-looking	similarly-priced	similar-sounding
single-app	single-card	single-core	single-finger
single-gpu	single-input	single-layer	single-letter
single-package	single-press	single-task	single-window
sky-high	slate-style	sleep-and-charge	sleep-and-play
slickest-looking	slick-looking	slimmed-down	slot-loading
slow-spinning	smaller-bodied	smaller-screen	smaller-than-expected
smaller-than-normal	small-form-factor	small-screen	small-surface-area
smart-looking	smile-shape	smudge-prone	snap-on

Table C-3 All concatenated words the extra words (Cont.)

softly-lit	soft-modding	soft-touch	software-based
software-optimized	software-update	soldered-in	sold-separately
solid-color	solid-feeling	sound-shaping	space-saving
spec-bumped	spill-resistant	spinning-platter	split-key
split-screen	sports-car	spun-metal	squared-off
square-screen	squint-free	standard-issue	standard-looking
standard-size	standard-voltage	stand-out	starting-point
start-up	steelseries-branded	step-down	step-up
still-evolving	still-good	still-impressive	still-new
still-slim	still-small	still-sparse	still-underused
still-welcome	straight-on	straight-up	stretched-out
stutter-free	subwoofer-driven	suede-like	super-high-res
super-high-resolution	super-light	superman-tight	super-powered
super-rugged	super-size	super-slim	super-stretched
super-thin	surround-sound	swivel-screen	swivel-top
system-selling	talked-about	tank-like	tapered-key
teardrop-curved	technicolor-certified	tech-savvy	tent-like
then-empty	then-new	then-quite-impressive	thicker-than-most
thin-and-light	thin-laptop	thinned-down	thin-sounding
third-gen	third-generation	through-the-roof	throw-in-your-luggage
thunderbolt-compatible	tied-together	tight-feeling	tile-based
tinny-sounding	too-early	too-expensive	too-short
top-end	top-firing	top-left	top-level
top-of-the-line	top-rated	top-shelf	top-tier

Table C-3 All concatenated words the extra words (Cont.)

top-to-bottom	touch-centric	touch-controlled	touch-enabled
touch-free	touch-friendly	touch-sensitive	touch-senstive
touch-type	travel-friendly	travel-oriented	travel-ready
tray-loading	tray-table	tri-metal	tri-toned
twisted-nematic	ultra-expensive	ultra-high	ultrahigh-res
ultra-high-resolution	ultralow-voltage	ultra-low-voltage	ultra-mobile
ultra-quality	ultra-thin	ultra-wide	ultra-wide-screen
under-responsive	under-the-hood	un-scrunched	un-used
up-the-nose	us-based	user-adjustable	user-definable
user-replaceable	user-selected	very-important	wait-and-see
wake-up	walk-away	wallet-friendly	wash-free
watchband-like	watt-hour	wear-off	web-based
web-browsing	wee-bit	well-build	well-built
well-constructed	well-designed	well-equipped	well-established
well-featured	well-liked	well-lit	well-machined
well-matched	well-packaged	well-placed	well-put
well-reasoned	well-regarded	well-reviewed	well-sized
well-spaced	well-stocked	well-suited	well-thought-out
well-tuned	white-glove	wide-and-short	wide-gamut
wider-than-usual	wide-tilt	window-powered	windows-based
windows-compatible	windows-wide	wobble-free	woodgrain-like
workhorse-level			

APPENDIX D

EXAMPLE WORD WITH POLARITY SCORE AND POLARITY LEVEL

Table D-1 Example word with polarity score and polarity level

Word	Polarity score	Polarity level	Score from
dysfunctional	0	neutral	Sentiwordnet
especially	0	neutral	Sentiwordnet
bright	+0.125	positive	Sentiwordnet
brighter	+0.375	positive	Polarity lexicon
brightest	+0.625	very positive	Polarity lexicon
close	+0.375	positive	Sentiwordnet
closer	+0.625	very positive	Polarity lexicon
closest	+0.875	strong positive	Polarity lexicon
cool	+0.250	positive	Sentiwordnet
cooler	+0.500	very positive	Polarity lexicon
coolest	+0.750	strong positive	Polarity lexicon
fast	+0.375	positive	Sentiwordnet
faster	+0.625	very positive	Polarity lexicon
fastest	+0.875	strong positive	Polarity lexicon
great	+0.250	positive	Sentiwordnet
greater	+0.500	very positive	Sentiwordnet
greatest	+0.875	strong positive	Sentiwordnet
light	-0.250	negative	Sentiwordnet
lighter	-0.500	very negative	Polarity lexicon
lightest	-0.750	strong negative	Polarity lexicon
long	+0.375	positive	Sentiwordnet
longer	+0.625	very positive	Polarity lexicon
longest	+0.875	strong positive	Polarity lexicon
nice	+0.875	strong positive	Sentiwordnet
nicer	+1.000	strong positive	Polarity lexicon

Table D-1 Example word with polarity score and polarity level (Cont.)

Word	Polarity score	Polarity level	Score from
Nicest	+1.000	strong positive	Polarity lexicon
warn	-0.250	negative	Sentiwordnet
warmer	-0.500	very negative	Polarity lexicon
warmest	-0.750	strong negative	Polarity lexicon
above-average	+0.700	very positive	NLTK
all-too-common	-0.700	very negative	NLTK
also-excellent	+0.700	very positive	NLTK
anti-reflective	+0.600	very positive	NLTK
bass-worthy	+0.600	very positive	NLTK
best-feeling	+0.600	very positive	NLTK
better-detailed	+0.600	very positive	NLTK
buzz-heavy	-0.600	very negative	NLTK
cleanest-feeling	-0.500	very negative	NLTK
cloud-heavy	-0.600	very negative	NLTK
consumer-oriented	+0.500	very positive	NLTK
context-sensitive	+0.500	very positive	NLTK
decent-enough	-0.500	very negative	NLTK
dogtag-like	-0.500	very negative	NLTK
easy-access	+0.500	very positive	NLTK
easy-to-see	+0.600	very positive	NLTK
easy-to-use	+0.600	very positive	NLTK
energy-efficient	+0.500	very positive	NLTK
entry-level	-0.500	very negative	NLTK
extra-wide	+0.500	very positive	NLTK
fiber-like	-0.500	very negative	NLTK
greater-than-1080p	+0.600	very positive	NLTK
hard-to-beat	-0.600	very negative	NLTK
higher-quality	+0.500	very positive	NLTK

Table D-1 Example word with polarity score and polarity level (Cont.)

Word	Polarity score	Polarity level	Score from
jaw-dropping	-0.500	very negative	NLTK
larger-scale	+0.600	very positive	NLTK
low-quality	-0.600	very negative	NLTK
natural-feeling	+0.500	very positive	NLTK
nice-feeling	+0.500	very positive	NLTK
performance-wise	+0.600	very positive	NLTK
phablet-like	-0.500	very negative	NLTK
plastic-feeling	-0.500	very negative	NLTK
sharp-screened	+0.500	very positive	NLTK
smaller-screen	-0.500	very negative	NLTK
sometimes-sluggish	-0.500	very negative	NLTK
still-good	+0.600	very positive	NLTK
too-expensive	-0.700	very negative	NLTK
well-constructed	+0.600	very positive	NLTK
well-sized	+0.600	very positive	NLTK
wider-than-usual	+0.500	very positive	NLTK
woodgrain-like	-0.500	very negative	NLTK

Score from Sentiwordnet means that word is found in Sentiwordnet. Therefore, the score of that word will be collected from Sentiwordnet.

Score from Polarity lexicon means that word is a comparative or superlative which cannot be found in Sentiwordnet. The polarity scores will be referred from the score of that world's base form in Sentiwordnet by increasing for positive words and decreasing for negative words.

Score from NLTK means that word is not found in Sentiwordnet. Then, the score of that word will be collected from the NLTK.

APPENDIX E
EXAMPLE CODE IN THIS THESIS

The example code for buildML() *Method*

This method is functioned to build Naïve Bayes Classifier in order to classify subjective paragraph by reading message from the .arff file. After that, the Naïve Bayes Classifier is made to classify the opinion into 2 results i.e. “0” which indicates the negative opinion and “1” which indicates the positive. The buildML() method is shown in Fig. B-1.

```
public static String buildML() throws Exception
{
    Instances train = buildTrainSet();
    Instances unlabeled = buildTestSet();
    Instances unlabel = new Instances( unlabeled );

    NaiveBayes nB = new NaiveBayes();
    nB.buildClassifier( train );
        for ( int i = 0; i < unlabeled.numInstances(); i++ )
    {
        double clsLabel = nB.classifyInstance( unlabeled.instance( i ) );
        unlabel.instance( i ).setClassValue( clsLabel );

        if ( clsLabel == 0.0 ) {return "NEG";}
                else {return "POS";}
    }
        return "";
}
```

Figure B-1 The example code for buildML() method

The example code for buildTrainSet() *method*

This method uses to prepare the training set by receiving data from files Model.arff. The code is shown in Fig. B-2.

```
private static Instances buildTrainSet() throws FileNotFoundException, IOException
{
    BufferedReader breaderTrain = null;
    breaderTrain = new BufferedReader( new FileReader( "ARFF\\Model.arff" ) );
    Instances train = new Instances( breaderTrain );
    breaderTrain.close();
    train.setClassIndex( train.numAttributes() - 1 );
    return train;
}
```

Figure B-2 The example code for buildTrainSet() method

The example code for buildTestSet() *method*

This method uses to prepare test set by receiving data from files Test.arff. The code is shown in Fig. B-3

```
private static Instances buildTestSet() throws FileNotFoundException, IOException
{
    BufferedReader breaderUnlabeled = null;
    breaderUnlabeled = new BufferedReader( new FileReader( "ARFF\\Test.arff" ) );
    Instances unlabeled = new Instances( breaderUnlabeled );
    breaderUnlabeled.close();
    unlabeled.setClassIndex( unlabeled.numAttributes() - 1 );
    return unlabeled;
}
```

Figure B-3 The example code for buildTestSet() method

APPENDIX F

PUBLICATION

T. Chatchaithanawat and P. Pugsee, “A framework for laptop review analysis,” Proceedings of International Conference on Advanced Informatics: Concepts, Theory and Applications (ICAICTA), pp.1-5, August 19-22, 2015.

REFERENCES

1 Karamibekr, M., and Ghorbani, A.A.: 'Sentiment Analysis of Social Issues', in Editor (Ed.)^(Eds.): 'Book Sentiment Analysis of Social Issues' (2012, edn.), pp. 215-221

2 Pang, B., and Lee, L.: 'Thumbs up? Sentiment Classfication using Machine Learning Techniques'. Proc. Proceedings of EMNLP2002 pp. Pages

3 Turney, P.D.: 'Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews'. Proc. Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)2002 pp. Pages

4 http://mpqa.cs.pitt.edu/lexicons/subj_lexicon/

5 http://sentiwordnet.isti.cnr.it

6 Esuli, A., and Sebastian, F.: 'SENTIWORDNET: A Publicly Available Lexical Resource for Opinion Mining', 2015

7 Baccianella, S., Esuli, A., and Sebastian, F.: 'SENTIWORDNET 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining', 2015

8 Fanzilli, C.L.: 'Stock Price Prediction Using Sentiment Detection of Twitter', Union College, 2015

9 http://exitexam.tsu.ac.th/cst/course/computer_it/AI/nature.html

10 K. Toutanova, D.K., C. Manning and Y. Singer: 'Stanford Part-Of-Speech (POS) Tagger', 2015

11 Hosch, W.L.: 'Machine learning', 2016

12 Fletcher, D., and Reutemann, P.: 'Weka 3: Data Mining Software in Java', 2015

13 Quinlan, J.R.: 'Induction of Decision Trees ', Machine Learning, 1986, 1, pp. 81-106

14 M., M.T.: 'Machine Learning' (McGraw-Hill, 1997. 1997)

15 http://www.saedsayad.com/naive_bayesian.htm

16 Jizba, R.: 'Measuring search effectiveness', 2015

17 P. Pugsee, T.C.a.K.N.N.: 'Subjectivity Analysis for airline services from twitter'. Proc. Proceedings of ITC-CSCC, July 1-4 2014 pp. Pages

18 Yamamoto, Y., Kumamoto, T., and Nadamoto, A.: 'Role of Emoticons for Multidimensional Sentiment Analysis of Twitter'. Proc. Proceedings of the 16th International Conference on Information Integration and Web-based Applications & Services, Hanoi, Viet Nam2014 pp. Pages

19 Bhadane, C., Dalal, H., and Doshi, H.: 'Sentiment Analysis: Measuring Opinions', Procedia Computer Science, 2015, 45, pp. 808-814

20 http://www.laurenceanthony.net/software/antconc/

21 http://www.nltk.org/

22 Fawcett, T.: 'Evaluating performance', ROC Graphs: Notes and Practical Considerations for Researchers, 2015

VITA

Mr. Thanapat Chatchaithanawat

Date of Birth: 27th April 1992

Place of Birth: Bangkok, Thailand

Email address: TChatchaithanawat@gmail.com

Computer Science and Information Technology

Faculty of Science

Chulalongkorn University