E-COMMERCE PRODUCT REVIEW MANAGEMENT SYSTEM BASED ON OPINION MINING

Submitted in partial fulfillment of the requirement

of the degree of

Bachelor in Engineering

by

Sujith Nair	BE-3-26
Apresh Pandit	BE-3-28
Sanal Pillai	BE-3-32
Onkaar Sawant	BE-3-39

Under the guidance of

Prof. Rupali Kale

DEPARTMENT OF COMPUTER ENGINEERING

SHAH AND ANCHOR KUTCHHI ENGINEERING COLLEGE

CHEMBUR, MUMBAI-400088.

2017 - 2018

www.shahandanchor.com Tel: 022 2558 0854

Mahavir Education Trust's

SHAH & ANCHOR KUTCHHI ENGINEERING COLLEGE

Mahavir Education Trust Chowk, W.T. Patil Marg, Chembur, Mumbai 400 088
Affiliated to University of Mumbai, Approved by D.T.E. & A.I.C.T.E.
Awarded provisional accreditation for Computer & Electronics Engineering by NBA
(for 2 years from 06-08-2014)

Certificate

This is to certify that the report of the project entitled

E-COMMERCE PRODUCT REVIEW MANAGEMENT SYSTEM BASED ON OPINION MINING

is a bonafide work of

Sujith Nair	BE-3-26
Apresh Pandit	BE-3-28
Sanal Pillai	BE-3-32
Onkaar Sawant	BE-3-39

submitted to the

UNIVERSITY OF MUMBAI

during semester VII in partial fulfilment of the requirement for the award of the degree of

BACHELOR OF ENGINEERING
in
COMPUTER ENGINEERING.

(Prof. Rupali Kale)
Guide

(Prof. Uday Bhave)
I/c Head of Department

(Dr. Bhavesh Patel)
Principal

Approval for Project Report for B. E. semester VII

This project report entitled E-COMMERCE PRODUCT REVIEW MANAGEMENT SYSTEM BASED ON OPINION MINING by Sujith Nair, Apresh Pandit, Sanal Pillai, Onkaar Sawant is approved for semester VII in partial fulfilment of the requirement for the award of the degree of Bachelor of Engineering.

Examiners

1._____

2._____

Guide

Prof Rupali Kale:_____

Date: 9/11/2017

Place: MUMBAI

Attendance Certificate

Date:9/11/17

To,
The Principal
Shah and Anchor Kutchhi Engineering College,
Chembur, Mumbai-88

Subject: Confirmation of Attendance

Respected Sir,

This is to certify that Final year (BE) students
Sujith Nair
Apresh Pandit
Sanal Pillai
Onkaar Sawant

have duly attended the sessions on the day allotted to them during the period from July to October for performing the Project titled E-COMMERCE PRODUCT REVIEW MANAGEMENT SYSTEM BASED ON OPINION MINING .

They were punctual and regular in their attendance. Following is the detailed record of the student's attendance.

Attendance Record:

Date	Sujith Nair	Apresh Pandit	Sanal Pillai	Onkaar Sawant
16/08/17	Present	Present	Present	Present
13/09/17	Present	Present	Present	Present
27/09/17	Present	Present	Present	Present
04/10/17	Present	Present	Present	Present

Signature and Name of Internal Guide

Prof. Rupali Kale

Declaration

We declare that this written submission represents our ideas in our own words and where others' ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Name of student	Roll No.	Signature
Sujith Nair	BE-3-26	_____
Apresh Pandit	BE-3-28	_____
Sanal Pillai	BE-3-32	_____

Onkaar Sawant BE-3-39 _____

Date:

Abstract

The growth of E-commerce has led to the invention of several websites that market and sells products as well as allows users to post reviews. It is typical for an online buyer to refer to these reviews before making a buying decision. Hence, automatic summarization of users' reviews has a great commercial significance. However, since the product reviews are written by non experts in an unstructured, natural language text, the task of summarizing them is challenging. This Project presents a semi supervised approach for mining online user reviews to generate comparative feature-based statistical summaries that can guide a user in making an online purchase. It includes various phases like preprocessing and feature extraction and pruning followed by feature based opinion summarization and overall opinion sentiment classification. Empirical studies indicate that the approach used in the paper can identify opinionated sentences from blog reviews with a high average precision of 91% and can classify the polarity of the reviews with a good average accuracy of 86%.

Table of contents

List of Diagrams

CHAPTER 1

Introduction

There are many users who purchase products through E-commerce websites. Through online shopping many E-commerce enterprises were unable to know whether the customers are satisfied by the services provided by the firm. This boosts us to develop a system where various customers give reviews about the product and online shopping services, which in turn help the E-commerce enterprises and manufacturers to get customer opinion to improve service and merchandise through mining customer reviews. An algorithm could be used to track and manage customer reviews, through mining topics and sentiment orientation from online customer reviews.

1.1 Objective

The objective of this project is to develop a system which will help customers who are buying products online to understand the popularity and efficiency of the product through the sentiments of customers who have earlier used that product and reviewed them.

1.2 Problem Statement

The main goal of our project is to provide a prototype system that could be used to track and manage customer reviews, the opinion of customers who bought that product, provide popularity of the customer's searched product, and also provide detailed comparison of features through data mining techniques and sentiment analysis from online customer reviews.

1.3 Methodology used

We generated an opinion review database by crawling Some popular websites that categorically post product reviews by actual users. Our product opinion summarizer has three main phases. These phases are

(1) preprocessing phase,

(2) feature extraction phase, and

(3) opinion summarization and classification phase.

These phases are briefly described next.

Preprocessing Phase:

Online blog reviews posted by users frequently contain spelling errors and incorrect punctuation. Our next phase—the feature-extraction phase—requires parts-of- speech tagging which works at the sentence level. Thus, it becomes important to detect end of sentences. So, in this phase we performed basic cleaning tasks like sentence boundary detection and spell-error correction. Sentences normally end with punctuations like period (.), question mark (?), or exclamation mark (!). Sometimes bloggers overuse the "?" and "!" symbols for emphasis. For example, a blogger may post a review that says

"It's surprising that the eBook reader does not have a touch screen !!!!"

In such cases we conflate the repetitive punctuation symbols to a single occurrence (i.e., "!!!!" is replaced by a single "!").

Several other considerations arise during the Preprocessing phase. The period (.) requires to be disambiguated as it may mean a full stop or a decimal point or an abbreviation (e.g., "Dr.," "Ltd."). Sometimes a single sentence straddles multiple lines as the user presses unnecessary return keys. In such cases we apply the sentence merge rules as proposed by Dey and Haque [14]. After sentence boundary detection, we perform spell-error correction using a word processor.

Feature Extraction Phase:

In this phase we extract opinion features from the pre-processed review text obtained From the previous phase. We treat frequently occurring nouns (N) and noun phrases (NP) as possible opinion features and associated adjectives describing them as indicators of their opinion orientation.

We perform parts-of- speech (POS) tagging on the review sentences using the Link Grammar Parser [34]. The Link Grammar Parser is a well-known and efficient syntactic parser for English language (http://www.abisource.com/projects/link- grammar/). First, we extract all nouns (N) and noun phrases (NP) tagged by the Link Grammar Parser and identify the frequently occurring N and NP as possible opinion features. By frequently occurring N and NP we mean those Ns and NP which occur at least five times in the users' reviews. We do not extract frequent itemsets from review sentence database using the Apriori based approach, since this method mines frequent features using a BOW (bag-of- words) approach and does not take into account the order in which the words of a phrase occur. Moreover, mining in this way would require ordering besides compactness and redundancy pruning [4, 8, 19]. We also do not use the seed-set expansion approach as it would require prior domain knowledge to specify a seed set. Instead we generate a frequent feature set using the multiword approach.

Opinion Summarization and Classification Phase:

In the previous phase we extracted opinion features, adjectives describing them, and any modifiers if present. We also generate a statistical feature-wise summary for each product which enables comparison of different brands selling similar products. In order to determine the sentiment polarity of an adjective describing an opinion feature we make use of SentiWordNet which is a lexical resource for opinion mining. SentiWordNet assigns three normalized sentiment scores: positivity, objectivity, and negativity to each synset of WordNet. Let us revisit the review sentence:

"The processor[.n] is[.v] significantly faster[.a], and the text[.n] is[.v] clear[.a]."

In this example, the SentiWordNet scores assigned to the appropriate usage of adjective clear is indicated as (P:0.625; O: 0.375; N: 0). Since the value of the positive polarity is highest, the adjective "clear" can be assigned a positive polarity. In this way, we generate a feature-orientation table (FO table) that records the opinion features and their corresponding descriptors of positive and negative polarities. The Table 1 shows the FO table entries for some of the features of product "Tablet." The FO table, thus generated, enables us to generate feature-wise summary of a product or comparative summaries of different brands of similar products.

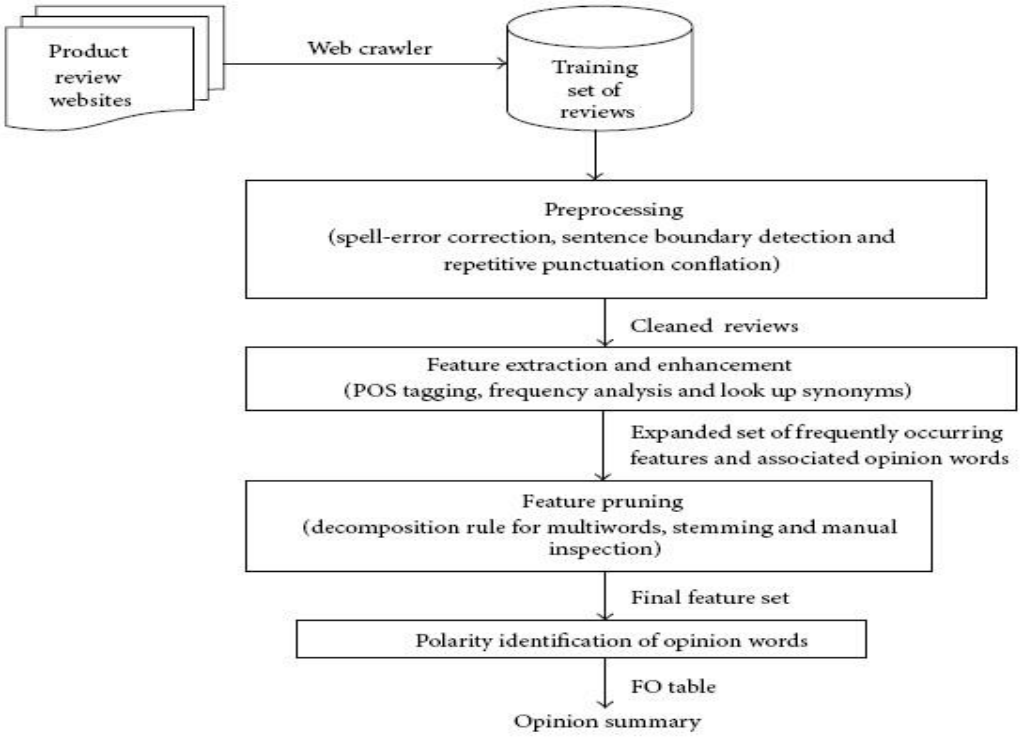


Fig 1.1: System Architecture Diagram [1]

1.4 Organization of the report

The main body of the report is preceded by detailed contents including lists of figures, tables, and annexes followed by units used in the report. This is followed by executive summary giving briefly the scope and objectives of the study, importance of the topic, methodology, limitations, major observations/findings, and recommendations & action plan.

Chapter 1: Gives an introduction to the project.

Chapter 2: Discusses the current implementation and research done to tackle the problems faced.

Chapter 3: Lists the Hardware and Software requirements for the proposed system.

Chapter 4: Explains the design for the system.

The main report is followed by a glossary, giving the acronyms and abbreviations used in the report, a listing of all the keywords corresponding to various chapters. References which have been used for certain inputs are listed after the keywords.

CHAPTER 2

Literature review

Classification and summarization of online blog reviews are very important to the growth of E-commerce and social networking applications. Earlier work on automatic text summarization has mainly focused on extraction of sentences that are more significant in comparison to others in a document corpus. The main approaches used to generate extractive summaries are (1) combinations of heuristics such as cue words, key words, title words or position (2) lexical chains, and (3) rhetorical parsing theory. [1]

However, it is important to note that the task of summarizing online product reviews is very different from traditional text summarization, as it does not involve extracting significant sentences from the source text. Instead, while summarizing user reviews, the aim is to first of all identify semantic features of products and next to generate a comparative summary of products based on feature-wise sentiment classification of the reviews that will guide the user in making a buying decision. In the authors have demonstrated that traditional unsupervised text classification techniques like naive Bayes, maximum entropy, and support vector machine do not perform well on sentiment or opinion classification and pointed out the necessity for feature-oriented classification. Thus, recent research work in opinion mining has focused on feature based extraction and summarization. [3]

Opinion mining from users' reviews involves two main tasks—(1) identification of the opinion feature set and (2) sentiment analysis of users' opinions based on the identified features.

It has been observed that nouns and noun phrases (N and NP) frequently occurring in reviews are useful opinion features, while the adjectives and adverbs describing them are useful in classifying sentiment. [4]

In order to extract nouns, noun phrases, and adjectives from review text, parts-of-speech (POS) tagging is performed. However, all nouns and noun phrases are not useful in mining and cannot directly be included in the feature set. So, the feature set is subsequently extracted using approaches that involve frequency analysis and/or use of domain knowledge as is discussed next. [4]

Various methods exist in the literature to associate features with their corresponding descriptors. Hu and Liu proposed the nearby-adjective heuristic. Although this method is simple and fast, it may result in inaccuracies. So, supervised approaches to determine association have been proposed in recent years such as syntactic dependency parsing and syntactic tree templates. [2]

CHAPTER 3

System Requirements

3.1 Hardware Requirements:

- 1 GB RAM.
- 200 GB HDD.
- Intel 1.66 GHz Processor Pentium 4

3.2 Software Requirements:

- Windows XP, Windows 7,8
- Visual Studio 2010
- MS SQL Server 2008
- Windows Operating System

CHAPTER 4

DESIGN

This section describes the fundamental design aspects of the project. Software design is a process to transform user requirements into suitable form.

4.1. ER Diagram

An entity–relationship model (ER model) describes inter-related things of interest in a specific domain of knowledge. An ER model is composed of entity types (which classify the things of interest) and specifies relationships that can exist between instances of those entity types.

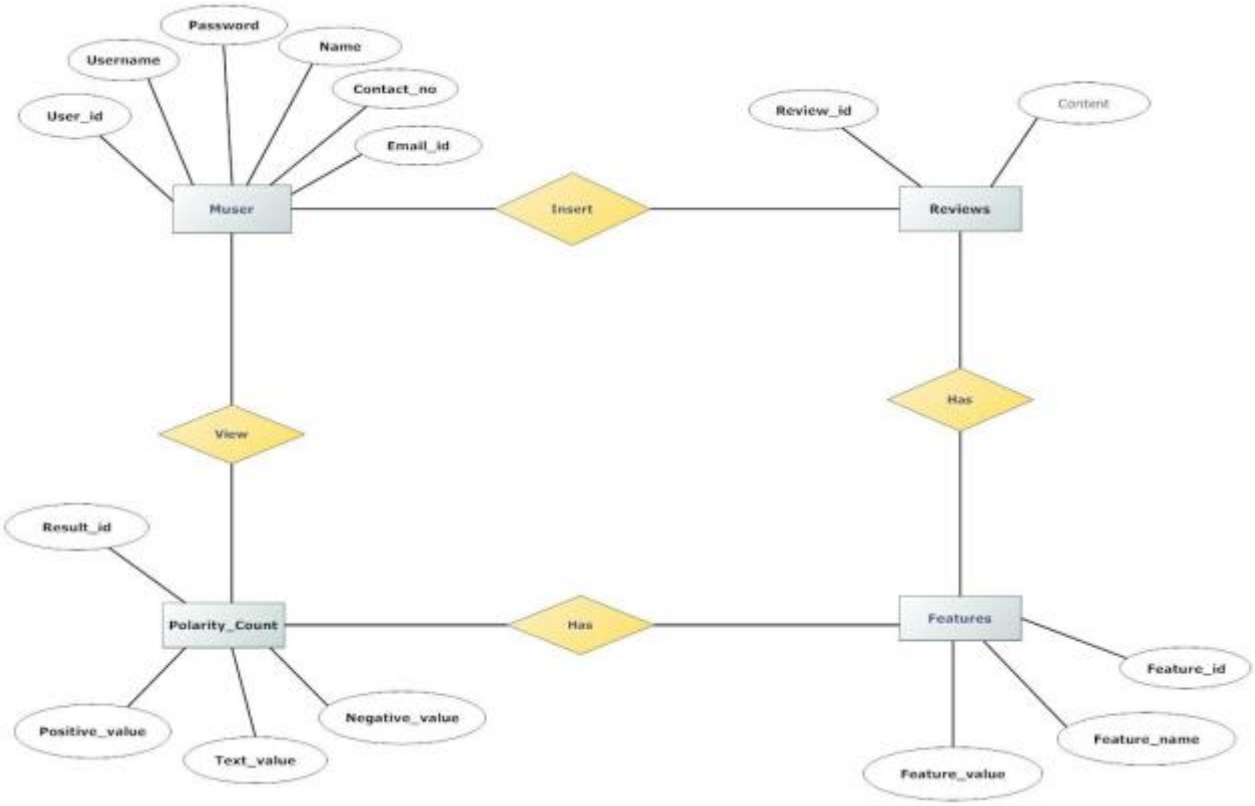


Fig 4.1: ER Diagram

Here, there are 4 entities, namely, user, reviews, polarity_count and features. The User entity has several attributes such as username, password, and contact numbers. Review entity has two attributes, review_id and content. Polarity_count stores the polarity values for a review. It has attributes such as result_id and values. Features store the features extracted from reviews. It stores basically the feature name, feature id and its polarity. A user can “insert” reviews into the system and can also “view” the polarity count. Reviews contains features with a one-to-many relationship.

4.2. Data Flow Diagram

A data flow diagram (DFD) is a graphical representation of the "flow" of data through an information system, modelling its process aspects. A DFD shows what kind of information will be input to and output from the system, how the data will advance through the system, and where the data will be stored. It does not show information about process timing or whether processes will operate in sequence or in parallel.

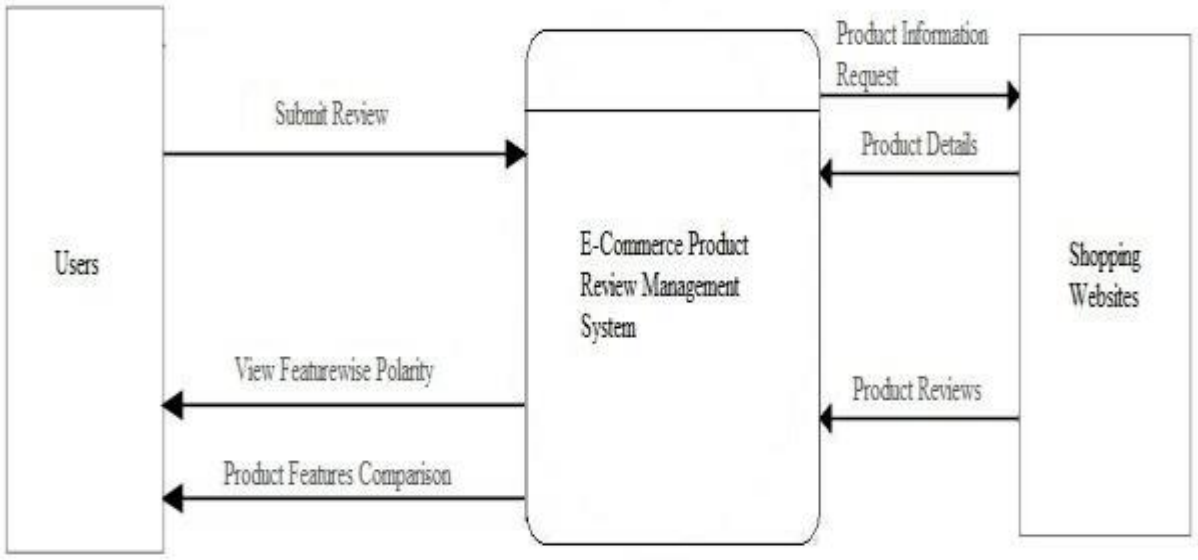


Fig 4.2: Data Flow Diagram

Above is a simple DFD of the system which shows the flow of data between user and system. User can submit login details and reviews to the system. The system returns with feature wise polarity for the product being viewed.

4.3. User Interface Design:

This section shows how the Graphical User Interface is being designed for the system. It shows the flow of pages present in the user interface to manage the system. The system will allow the user to submit reviews only after logging in to the system using login credentials. It will also display the result in a separate portion of the GUI.

4.3.1 System Block Diagram

A block diagram is a diagram of a system in which the principal parts or functions are represented by blocks connected by lines that show the relationships of the blocks. They are heavily used in engineering in hardware design, electronic design, software design, and process flow diagrams.

Block diagrams are typically used for higher level, less detailed descriptions that are intended to clarify overall concepts without concern for the details of implementation. Contrast this with the schematic diagrams and layout diagrams used in electrical engineering, which show the implementation details of electrical components and physical construction.

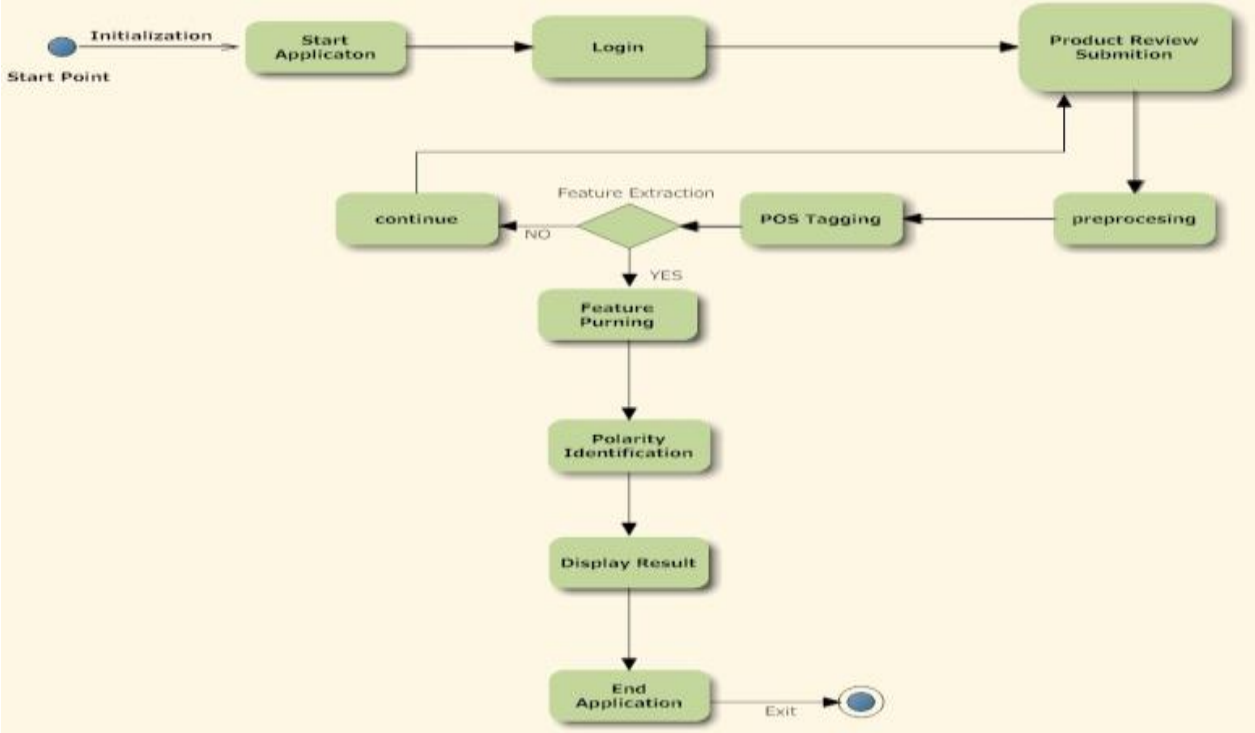


Fig 4.3.1: System Block Diagram

The system basically starts with the user logging in to the system. It then moves to "Product Review Submission" which allows the user to submit his/her own reviews on a product. The review is then pre-processed and POS (part-of-speech) tagging is performed. The system further performs feature identification from the review, followed by polarity identification of the review.

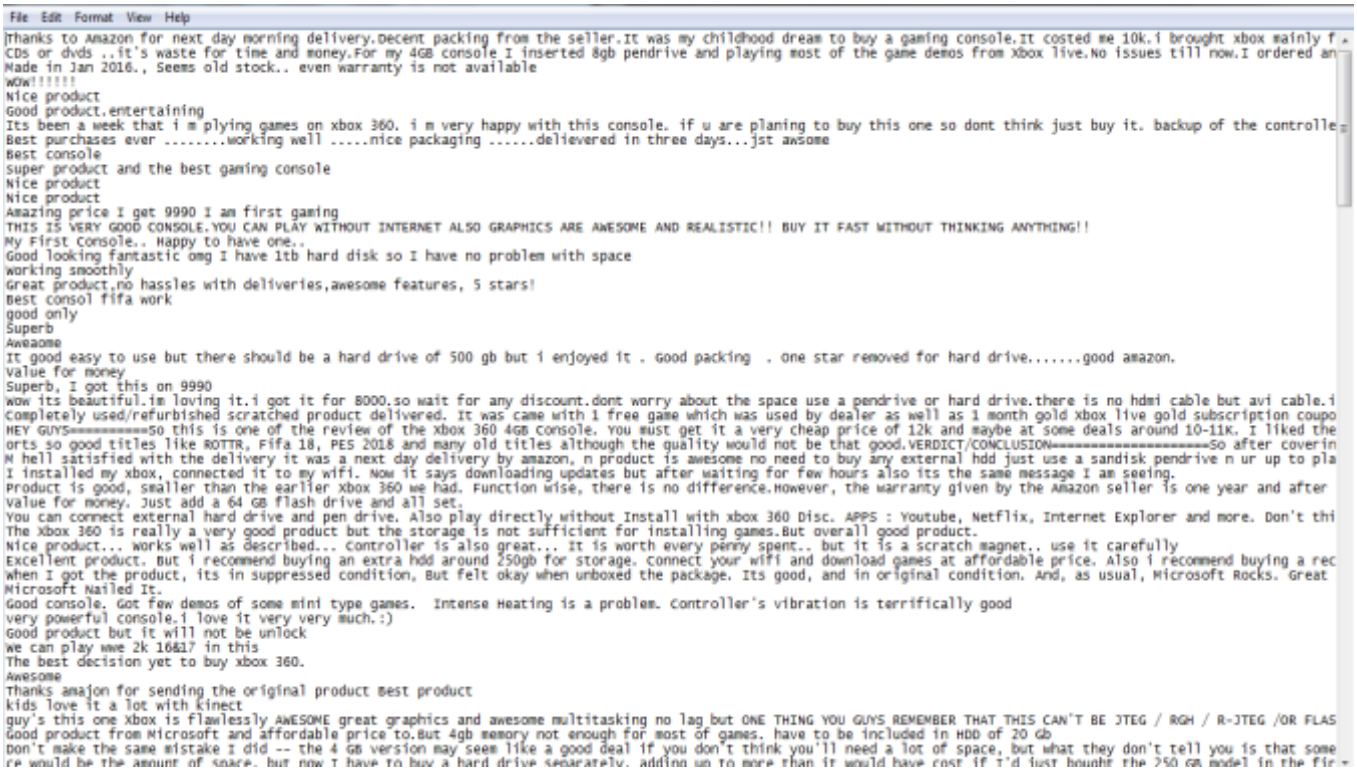
The result of the analysis is displayed as a graph showing polarity analysis or the polarity of the review is displayed. The system stops after user logs out of the system.

4.3.2 Input/Output Form

```
from bs4 import BeautifulSoup
import urllib, ast, sys, time
def crawl_amazon():
    p1 = "https://www.amazon.com/PlayStation-4-500GB-Console-Old-Model/product-reviews/B00BGA9WK2/ref=cm_cr_arp_d_paging_btm?ie=U
    p2 = "&showViewpoints=0&sortBy=byRankDescending"
    posts = []
    file_name = "ps4_us.txt"
    with open(file_name, 'w') as fp:
        for pn in xrange(1, 1170):
            url = p1 + str(pn) + p2
            print url
            amazon_url = p1 + str(pn) + p2
            ur = urllib.urlopen(amazon_url)
            soup = BeautifulSoup(ur.read())
            posts_=soup.find_all("span", class_="a-size-base review-text")
            #posts_ = soup.select("span data-hook="review-body" class="a-size-base review-text"")
            print len(posts_)
            for x in posts_: posts.append(x.text)
            for x in posts_: fp.write(x.text.encode('utf-8') + "\n")
            time.sleep(1)
```

```
In [13]: crawl_amazon()
```

```
https://www.amazon.com/PlayStation-4-500GB-Console-Old-Model/product-reviews/B00BGA9WK2/ref=cm_cr_arp_d_paging_btm?ie=UTF8&re
viewerType=all_reviews&pageNumber=1&showViewpoints=0&sortBy=byRankDescending
10
https://www.amazon.com/PlayStation-4-500GB-Console-Old-Model/product-reviews/B00BGA9WK2/ref=cm_cr_arp_d_paging_btm?ie=UTF8&re
viewerType=all_reviews&pageNumber=2&showViewpoints=0&sortBy=byRankDescending
10
https://www.amazon.com/PlayStation-4-500GB-Console-Old-Model/product-reviews/B00BGA9WK2/ref=cm_cr_arp_d_paging_btm?ie=UTF8&re
viewerType=all_reviews&pageNumber=3&showViewpoints=0&sortBy=byRankDescending
10
https://www.amazon.com/PlayStation-4-500GB-Console-Old-Model/product-reviews/B00BGA9WK2/ref=cm_cr_arp_d_paging_btm?ie=UTF8&re
viewerType=all_reviews&pageNumber=4&showViewpoints=0&sortBy=byRankDescending
10
https://www.amazon.com/PlayStation-4-500GB-Console-Old-Model/product-reviews/B00BGA9WK2/ref=cm_cr_arp_d_paging_btm?ie=UTF8&re
viewerType=all_reviews&pageNumber=5&showViewpoints=0&sortBy=byRankDescending
10
https://www.amazon.com/PlayStation-4-500GB-Console-Old-Model/product-reviews/B00BGA9WK2/ref=cm_cr_arp_d_paging_btm?ie=UTF8&re
viewerType=all_reviews&pageNumber=6&showViewpoints=0&sortBy=byRankDescending
10
https://www.amazon.com/PlayStation-4-500GB-Console-Old-Model/product-reviews/B00BGA9WK2/ref=cm_cr_arp_d_paging_btm?ie=UTF8&re
```


File Edit Format View Help
Thanks to Amazon for next day morning delivery.Decent packing from the seller.It was my childhood dream to buy a gaming console.It costed me 10k.i brought xbox mainly f
CDs or dvds ..it's waste for time and money.For my 4GB console I inserted 8gb pendrive and playing most of the game demos from Xbox live.No issues till now.I ordered an
Made in Jan 2016., Seems old stock.. even warranty is not available
WOW!!!!!!
Nice product
Good product.entertaining
Its been a week that i m plying games on xbox 360. i m very happy with this console. if u are planing to buy this one so dont think just buy it. backup of the controlle
Best purchases ever ....working well .....nice packaging ......delievered in three days...jst awsome
Best console
super product and the best gaming console
Nice product
Nice product
Amazing price I get 9990 I am first gaming
THIS IS VERY GOOD CONSOLE.YOU CAN PLAY WITHOUT INTERNET ALSO GRAPHICS ARE AWESOME AND REALISTIC!! BUY IT FAST WITHOUT THINKING ANYTHING!!
My First Console.. Happy to have one..
Good looking fantastic omg I have 1tb hard disk so I have no problem with space
working smoothly
Great product,no hassles with deliveries,awesome features, 5 stars!
Best consol fifa work
good only
Superb
Awesome
It good easy to use but there should be a hard drive of 500 gb but i enjoyed it . Good packing . one star removed for hard drive.....good amazon.
Value for money
Superb, I got this on 9990
Wow its beautiful.im loving it.i got it for 8000.so wait for any discount.dont worry about the space use a pendrive or hard drive.there is no hdmi cable but avi cable.i
Completely used/refurbished scratched product delivered. It was came with 1 free game which was used by dealer as well as 1 month gold Xbox live gold subscription coupo
HEY GUYS==========So this is one of the review of the Xbox 360 4GB console. You must get it a very cheap price of 12k and maybe at some deals around 10-11K. I liked the
orts so good titles like ROTTR, Fifa 18, PES 2018 and many old titles although the quality would not be that good.VERDICT/CONCLUSION==========So after coverin
N hell satisfied with the delivery it was a next day delivery by amazon, n product is awesome no need to buy any external hdd just use a sandisk pendrive n ur up to pla
I installed my xbox, connected it to my wifi. Now it says downloading updates but after waiting for few hours also its the same message I am seeing.
Product is good, smaller than the earlier Xbox 360 we had. Function wise, there is no difference.However, the warranty given by the Amazon seller is one year and after
Value for money. Just add a 64 GB flash drive and all set.
You can connect external hard drive and pen drive. Also play directly without Install with xbox 360 Disc. APPS : Youtube, Netflix, Internet Explorer and more. Don't thi
The Xbox 360 is really a very good product but the storage is not sufficient for installing games.But overall good product.
Nice product... Works well as described... Controller is also great... It is worth every penny spent.. but it is a scratch magnet.. use it carefully
Excellent product. But i recommend buying an extra hdd around 250gb for storage. Connect your wifi and download games at affordable price. Also i recommend buying a rec
When I got the product, its in suppressed condition, But felt okay when unboxed the package. Its good, and in original condition. And, as usual, Microsoft Rocks. Great
Microsoft Nailed It.
Good console. Got few demos of some mini type games. Intense Heating is a problem. Controller's vibration is terrifically good
very powerful console.i love it very very much.:)
Good product but it will not be unlock
We can play wwe 2k 16&17 in this
The best decision yet to buy xbox 360.
Awesome
Thanks amajon for sending the original product Best product
kids love it a lot with kinect
guy's this one Xbox is flawlessly AWESOME great graphics and awesome multitasking no lag but ONE THING YOU GUYS REMEMBER THAT THIS CAN'T BE JTEG / RGH / R-JTEG /OR FLAS
Good product from Microsoft and affordable price to.But 4gb memory not enough for most of games. have to be included in HDD of 20 Gb
Don't make the same mistake I did -- the 4 GB version may seem like a good deal if you don't think you'll need a lot of space, but what they don't tell you is that some
ce would be the amount of space, but now I have to buy a hard drive separately, adding up to more than it would have cost if I'd just bought the 250 GB model in the fir

CHAPTER 5

References

[1] L. Zhao and C. Li, “Ontology based opinion mining for movie reviews,” in Proceedings of the 3rd International Conference on Knowledge Science, Engineering and Management, pp. 204–214, 2009.

[2] A. Balahur, Z. Kozareva, and A. Montoyo, “Determining the polarity and source of opinions expressed in political debates,” in Proceedings of the 10th International Conference on Intelligent Text Processing and Computational Linguistics, vol. 5449 of Lecture Notes in Computer Science, pp. 468–480, Springer, 2009.

[3] Y. H. Gu and S. J. Yoo, “Mining popular menu items of a restaurant from web reviews,” in Proceedings of the International Conference onWeb Information Systems andMining (WISM’11), vol. 6988 of Lecture Notes in Computer Science, pp. 242–250, Springer, 2011.

[4] M.Hu and B. Liu, “Mining and summarizing customer reviews,” in Proceedings of the 10th ACMSIGKDD International Conference on Knowledge Discovery and Data Mining (KDD ‘04), pp. 168–177, August 2004.

[5] M. A. Jahiruddin, M. N. Doja, and T. Ahmad, “Feature and opinion mining for customer review summarization,” in Proceedings of the 3rd International Conference on Pattern Recognition and Machine Intelligence (PReMI ‘09), vol. 5909 of Lecture Notes in Computer Science, pp. 219–224, 2009.
[6] S. Shi and Y. Wang, “A product features mining method based on association rules and the degree of property co-occurrence,” in Proceedings of the International Conference on Computer Science and Network Technology (ICCSNT ’11), pp. 1190–1194, December 2011.
[7] S. Huang, X. Liu, X. Peng, and Z. Niu, “Fine-grained product features extraction and categorization in reviews opinion mining,” in Proceedings of the 12th IEEE International Conference on Data MiningWorkshops (ICDMW’12), pp. 680–686, 2012.
[8] C.-P.Wei,Y.-M.Chen,C.-S.Yang, and C. C.Yang, “Understanding what concerns consumers: a semantic approach to product feature extraction fromconsumer reviews,” Information Systems and e-Business Management, vol. 8, no. 2, pp. 149–167, 2010.

Chapter 6

Acknowledgement

We are personally indebted to a number of people who gave us their useful insights to aid in our overall progress for this project. A complete acknowledgement would therefore be encyclopedic. First of all, we would like to give our deepest gratitude to **our parents** for permitting us to take up this course.

Our sincere thanks and heartfelt sense of gratitude goes to our respected Principal, **Dr. Bhavesh Patel** for all his efforts and administration in educating us in his premiere institution.

We take this opportunity to convey our sincere thanks to our Head of the Department and our project mentor, **Prof. Rupali Kale** for his constant encouragement and invaluable suggestions throughout the project and for his technical support rendered during the course of our project.