

Project Report On

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-4-38
-------------	---------

Apresh Pandit	BE-4-42
---------------	---------

Sanal Pillai	BE-4-46
--------------	---------

Onkaar Sawant	BE-4-57
---------------	---------

Under the guidance of

Prof. Rupali Kale



DEPARTMENT OF COMPUTER ENGINEERING
SHAH AND ANCHOR KUTCHHI ENGINEERING COLLEGE
CHEMBUR, MUMBAI-400088.

2017 - 2018

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-4-38
Apresh Pandit	BE-4-42
Sanal Pillai	BE-4-46
Onkaar Sawant	BE-4-57

submitted to the

UNIVERSITY OF MUMBAI

*during semester VIII 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 Bhawe)
I/c Head of Department

(Dr. Bhavesh Patel)
Principal

Approval for Project Report for B. E. semester VIII

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 VIII in partial fulfilment of the requirement for the award of the degree of Bachelor of Engineering.

Examiners

1. _____

2. _____

Guide

Prof Rupali Kale: _____

Date:

Place: MUMBAI

Attendance Certificate

Date:

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 January to March 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
17/1/18	Present	Present	Present	Present
1/2/18	Present	Present	Present	Present
28/2/18	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-4-38	_____
Apresh Pandit	BE-4-42	_____
Sanal Pillai	BE-4-46	_____
Onkaar Sawant	BE-4-57	_____

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

I.	Title	...i.
II.	Certificates	...ii
III.	Approval	...iii
IV.	Abstract	...iv
V.	Table of Contents	...v
VI.	List of Diagrams	...vi
1.	Introduction	...1
1.1.	Objective	...1
1.2.	Methodology used	...2
1.3.	Organization of the report	...5
2.	Literature Survey	...6
3.	Problem Statement	...8
4.	System Requirements	...9
4.1.	Hardware Requirements	...9
4.2.	Software Requirements	...9
5.	Design	...10
5.1.	ER Diagram	...10
5.2	Data Flow Diagram	...11
5.3.	Component Diagram	...11
5.4.	User Interface Design	...12
5.4.1.	System Block Diagram	...12
6.	Implementation Details	...14
7.	Testing	...17
8.	Result and Analysis	...18
8.1	Results	...18
8.2	Analysis	...20
9.	Conclusion and Future Scope	...21
10.	References	...22
11.	Acknowledgement	...23

List of Diagrams

1. System Architecture Diagram	...4
2. ER Diagram	...10
3. Data Flow Diagram	...11
4. Component Diagram	...12
5. System Block Diagram	...13

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 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 [1]. 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 [3]. 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]. 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.

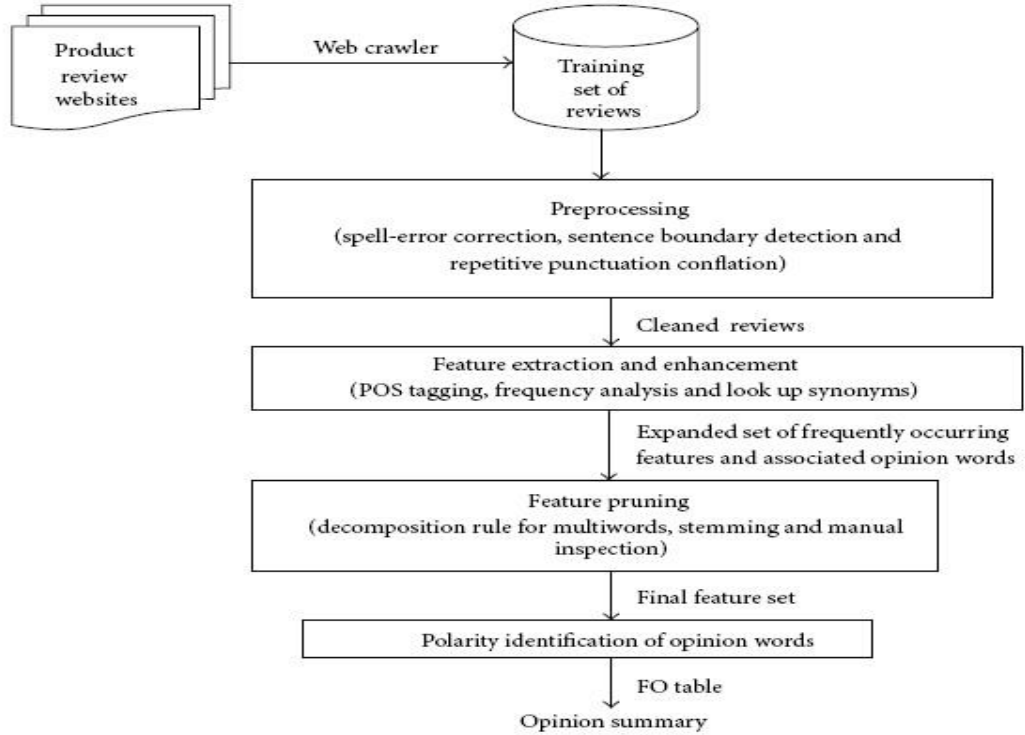


Fig 1.1: System Architecture Diagram [1]

1.3 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: Defines the problem statement

Chapter 4: Lists the Hardware and Software requirements for the proposed system.

Chapter 5: Explains the design for the system.

Chapter 6: Shows the project implementation details.

Chapter 7: Shows the test cases.

Chapter 8: Lists the results and analysis of the results.

Chapter 9: Conclusions and defining the future scope.

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

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, and also provide detailed comparison of features through data mining techniques and sentiment analysis from online customer reviews. The project will help determine the overall sentiment of a user's review and provide a graphical representation of the same.

CHAPTER 4

System Requirements

4.1 Hardware Requirements:

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

4.2 Software Requirements:

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

CHAPTER 5

DESIGN

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

5.1. 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.

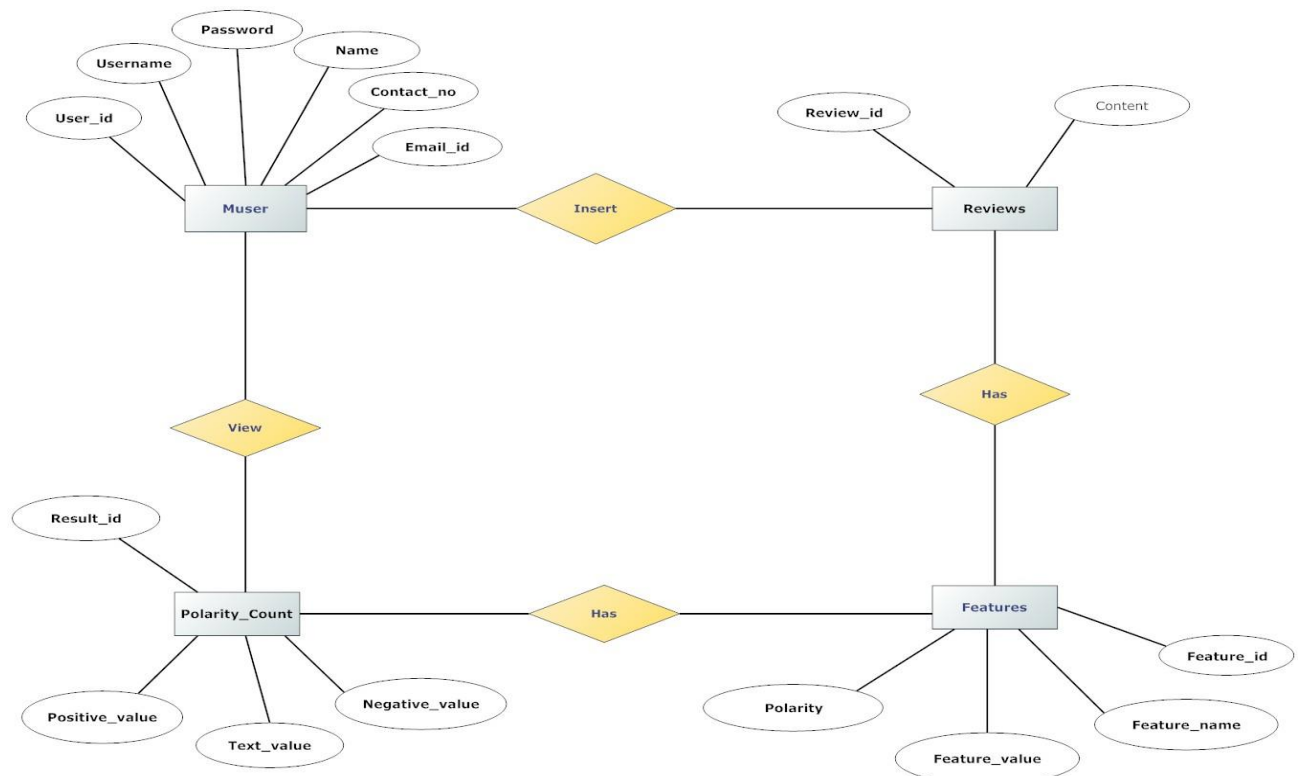


Fig 5.1: ER Diagram

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.

5.2. 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.

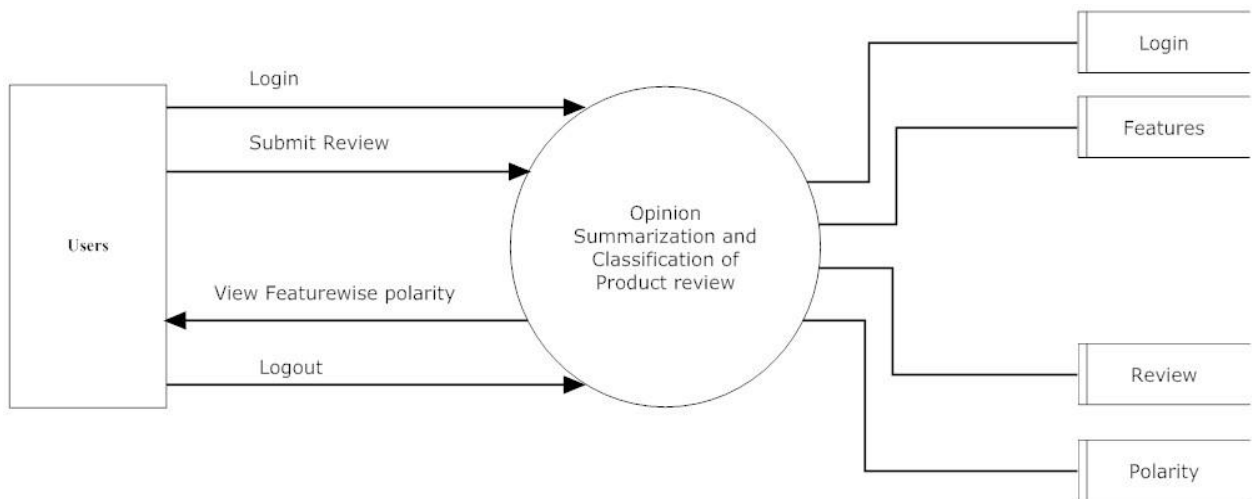


Fig 5.2: Data Flow Diagram

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.

5.3. Component Diagram

A component is a container of logical elements and represents things that participate in the execution of a system. Components also use the services of other components through one of its interfaces. Components are typically used to visualize logical packages of source code (work product components), binary code (deployment components), or executable files (execution components).

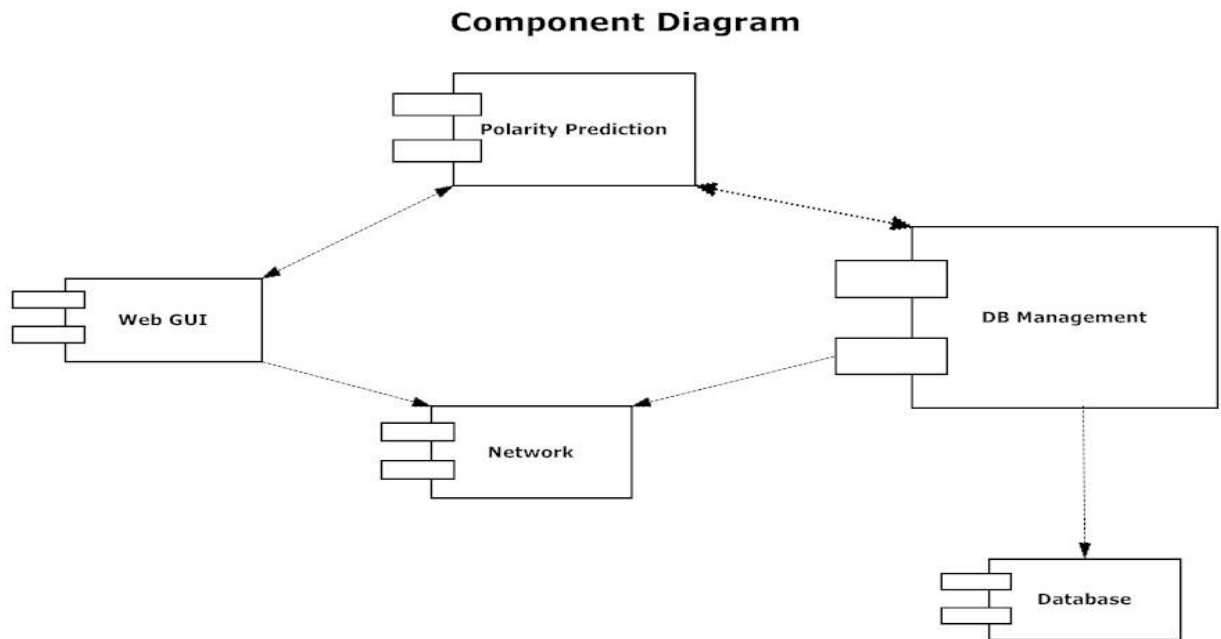


Fig 5.3: Component Diagram

The basic components of the system are the Web GUI, DBMS and Polarity Prediction module.

The DBMS manages a database for storing and retrieving reviews.

5.4. 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.

5.4.1 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.

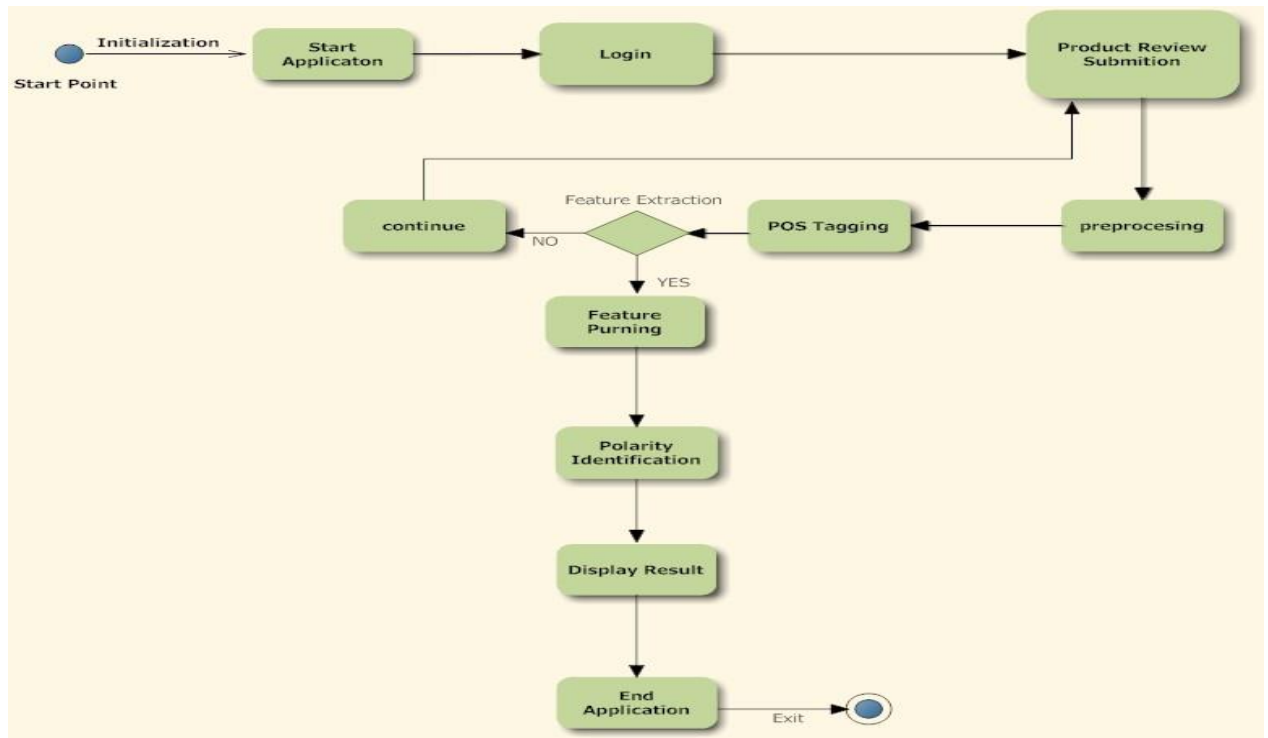


Fig 5.4.1: System Block Diagram

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.

CHAPTER 6

Implementation Details

Web Scrapping Module:

This module is responsible for scraping reviews from a particular website. Module takes product details as input, gains access to the reviews page of that product and then scrapes the review to text form for local usage. This module is implemented in C#. This basically involves extracting only the relevant review part from the HTML file being fetched.

Code:

```
protected void btnSubmit_Click(object sender, EventArgs e)
{
    ProductInfo.ProductName = DropDownList2.SelectedItem.Text;
    //Old Code
    #region
    string htmlCode;
    List<string> lst = new List<string>();

    string uri = String.Format("http://reviewing.net/s/prd/{0}", DropDownList2.SelectedItem);
    var getHtmlWeb = new HtmlWeb();
    var document = getHtmlWeb.Load(uri);
    Literal Literal1 = new Literal();
    Literal1.ID = "Literal1";
    foreach (HtmlNode para in document.DocumentNode.SelectNodes("//div[@class='review-summary']"))
    {
        Literal1.Text += String.Format("<p>{0}</p>", para.InnerText);
        D.Add((string)(para.InnerText));
    }
    // Panel1.Controls.Add(Literal1);
    Literal1.Text = Literal1.Text.Replace("<p>", " ");
    txtPara.Text = Literal1.Text.Replace("</p>", " ");
    allReview = "";
    allReview = txtPara.Text;
    #endregion

    //New Code
    #region
}
```

Parts-of-Speech Tagging Module:

POS tagging is done with the help of Maxent tagger. It is a tool created by Stanford University to tag various parts of a sentence as noun, verb etc.

```
private void TagReader(Reader reader)
{
    var tagger = new MaxentTagger(Model);
    //List obj = (List)MaxentTagger.tokenizeText(reader);
    foreach (ArrayList sentence in MaxentTagger.tokenizeText(reader).ToArray())
    {
        var tSentence = tagger.tagSentence(sentence);
        System.Console.WriteLine(Sentence.listToString(tSentence, false));
        posstring = (Sentence.listToString(tSentence, false));
        newString = newString + posstring;
        System.Console.WriteLine();
    }
}
```

Feature Extraction Module:

In order to determine the sentiment polarity of an adjective describing an opinion feature, SentiWordNet is used, which is a lexical resource for opinion mining. SentiWordNet assigns three normalized sentiment scores: positivity, objectivity, and negativity to each synset of WordNet.

The SentiWordNet scores are assigned to the appropriate usage of adjective indicated as (P:0.625; O: 0.375; N: 0). Since the value of the positive polarity is highest, the adjective can be assigned a positive polarity.

The SentiWordNet score for over 35000 words are stored in a database.

	pos	Id	PosScore	NegScore	Synset Terms
1	a	00001740	0.12500	0.00000	able
2	a	00002098	0.00000	0.75000	unable
3	a	00002312	0.00000	0.00000	dorsal
4	a	00002527	0.00000	0.00000	ventral
5	a	00002730	0.00000	0.00000	acroscopic
6	a	00002843	0.00000	0.00000	basiscopic
7	a	00002956	0.00000	0.00000	abducting
8	a	00003131	0.00000	0.00000	adductive

Query executed successfully. HOME-PC (11.0 RTM) Home-PC\Home (51) master 00:00:00 35552 rows

A sample FO table may look like this:

	id	Feature	PositivePolarity	NegativePolarity	Value	Weight
1	1	music	NULL	1	other	2
2	2	battery	1	NULL	Great	0.666666666666667
3	3	Video	1	NULL	helpful	0.666666666666667
4	4	camera	1	NULL	TIRED	0.666666666666667

A feature-orientation table (FO table) is generated that records the opinion features and their corresponding descriptors of positive and negative polarities. The FO table, thus generated, enables us to generate feature-wise summary of a product or comparative summaries of different brands of similar products.

CHAPTER 7

Development of Test Cases

Test cases:

Test case id	Objective	Steps/Description	Input	Expected output	Actual output	Result	Remark
001	To test login credentials	Give null and wrong credentials to check if appropriate response is given or not.	Null values or incorrect credentials	Error message.	Error message and re-enter option.	Pass	Test passed.
002	To test Retrieval of product information	Select a product from available categories to check its reviews are retrieved from website correctly	A product as a selection parameter from categories	Reviews of the particular product are retrieved from website	Reviews of the particular product are retrieved from website	Pass	Test passed.
003	To test POS tagger Component	Preprocessed reviews are tagged using POS tagger.	Preprocessed review text	The text in the review is tagged appropriately	The text in the review is tagged appropriately	Pass	Test passed.
004	To test feature score generation	Tagged reviews are used to generate feature score and feature opinion table	Tagged reviews and feature words	Feature scored is calculated and feature opinion table is generated	Feature scored is calculated and feature opinion table is generated	Pass	Test passed.

CHAPTER 8

Results and Analysis

8.1 Results

Home About Review Profile Result Logout

Category Name Brand Name Product Name

Select Product: --Select Category--

Submit

POS Taging

Feature Score

Extracted Reviews:

Home About Review Profile Result Logout

Category Name Brand Name Product Name

Select Product: Cell Phones

HTC

Htc Desire 601

Submit

Mar 14, 2014 -
Sure, it doesn't have the latest Android software on board, but if you're after a good-looking HTC phone with powerful speakers and 4G LTE for speedy data downloads, the Desire 601 is an affordable option to consider.

Sep 3, 2013 -
HTC is an odd company, in that it excels in making phones that can so easily sink below the radar of many consumers. For a company that's pledged to make fewer phones, it's come out with a large amount this year thanks to the One, One Mini, Desire 500, Desire 601 and Desire 300 – plus the HTC One Max on its way too in all likelihood. It's hard to see which consumers will plump for the Desire 601 unless it gets a huge marketing boost – there's still

POS Taging

Feature Score

POS Tagging:

Home

About

Review

Profile

Result

Logout

Category Name

Brand Name

Product Name

Select Product:

Cell Phones

HTC

Htc Desire 601

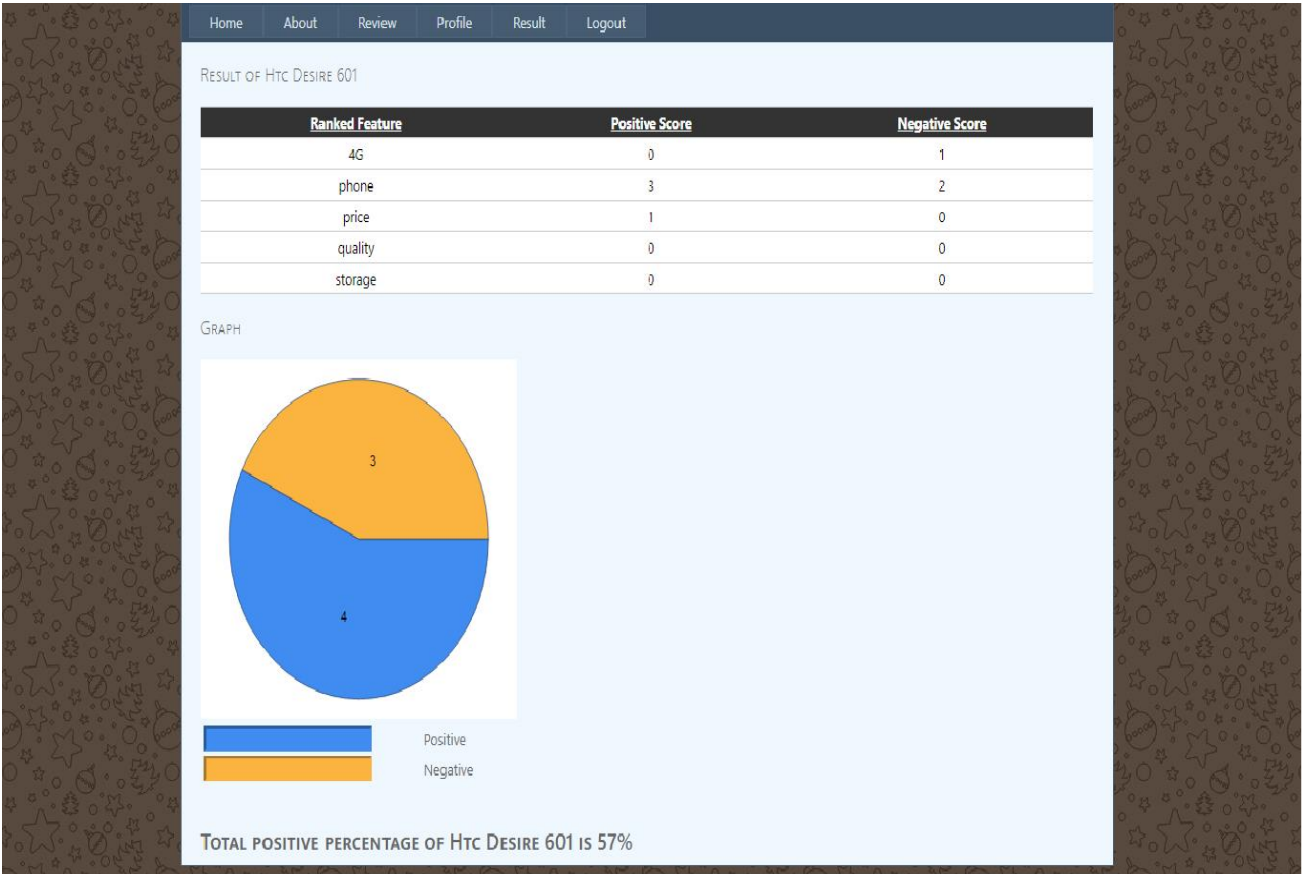
Submit

Mar/NNP 14/CD 2014/CD Sure/JJ it/PRP doesn't/NN have/VBP the/DT latest/JJS Android/NN software/NN on/IN board/NN but/CC if/IN you/PRP re/NNP after/IN a/DT good/JJ looking/VBG HTC/NNP phone/NN with/IN powerful/JJ speakers/NNS and/CC 4G/NN LTE/NN for/IN speedy/JJ data/NNS downloads/NNS the/DT Desire/NN 601/CD is/VBZ an/DT affordable/JJ option/NN to/TO consider/VB ./Sep/NNP 3/CD 2013/CD HTC/NNP is/VBZ an/DT odd/JJ company/NN in/IN that/IN it/PRP excels/VBZ in/IN making/VBG phones/NNS that/WDT can/MD so/RB easily/RB sink/VB below/IN the/DT radar/NN of/IN many/JJ consumers/NNS ./For/IN a/DT company/NN that/WDT s/VBZ pledged/VBN to/TO make/VB fewer/JJR phones/NNS it/PRP s/VBZ come/VBN out/RP with/IN a/DT large/JJ amount/NN this/DT year/NN thanks/NNS to/TO the/DT One/CD One/CD Mini/NN Desire/NN 500/CD Desire/NN 601/CD and/CC Desire/NN 300/CD ndash/NN plus/CC the/DT HTC/NNP One/CD Max/NN on/IN its/PRP\$ way/NN too/RB in/IN all/DT likelihood/NN ./It/PRP s/VBZ hard/JJ to/TO see/VB which/WDT consumers/NNS will/MD plump/JJ for/IN the/DT Desire/NN 601/CD unless/IN it/PRP gets/VBZ a/DT huge/JJ marketing/NN boost/NN ndash/NN there/RB s/VBZ still/RB not/RB enough/JJ appetite/NN for/IN 4G/NN speeds/NNS on/IN a/DT mobile/JJ to/TO make/VB it/PRP the/DT reason/NN to/TO buy/VB a/DT phone/NN and/CC that/IN seems/VBZ to/TO be/VB HTC/NNP s/NNP play/NN ./It/PRP s/VBZ certainly/RB one/CD that/WDT resonates/VBZ with/IN networks/NNS and/CC may/MD serve/VB to/TO bring/VB enough/JJ carrier/NN spending/NN to/TO push/VB the/DT 601/CD but/CC with/IN so/RB many/JJ similar/JJ phones/NNS on/IN the/DT market/NN this/DT could/MD be/VB a/DT tough/JJ sell/NN ./That/DT said/VBD with/IN elements/NNS like/IN a/DT microSD/JJ card/NN slot/NN to/TO back/VB up/RP the/DT 8GB/NN of/IN

POS Taging

Feature Score

Final Output:



8.2 Analysis

In this project, the first step for the user is to create/login to his/her account. After logging in to the application, the user can select the product of his/her choice by browsing through the “Category”, “Brand”, “Product” drop-down menu. After selecting the product of choice, the application then fetches the reviews from a critic rating site known as reviewing.net.

The application after fetching the reviews, displays it to the user in a text field view. The user then should select the “POS Tagging” option so that POS tagging can be performed on the reviews. POS tagging is the process of marking up a word in a text (corpus) as corresponding to a particular part of speech, based on both its definition and its context—i.e., its relationship with adjacent and related words in a phrase, sentence, or paragraph. POS tagging identifies the words as nouns, verbs, adjectives, adverbs, etc.

After this, the user should select “Feature Score” option which gives a sentiment score for the nouns and adjectives tagged by the POS tagger. 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. Hence, according to these scores, a feature opinion table as well as a graph representing positive to negative score are created. Also, the total percentage of positive reviews for the selected product is also displayed. These information help the user to determine whether the product is worth to buy or not.

CHAPTER 9

Conclusions and Future Scope

Conclusion:

Classifying and summarizing opinions of bloggers has several interesting and commercially significant applications. However, this task is much more difficult than classifying regular text and requires intensive Preprocessing. The success of the opinion mining task is mainly dependent on the efficiency and sophistication of the Preprocessing and feature extraction steps. We empirically proved that the proposed approach for product feature set extraction, that is, using frequent multi words with decomposition strategy outperforms other contemporary approaches like the Apriori-based approach and the seed-set expansion approach.

Empirical results indicate that the multistep feature based semi supervised opinion mining approach used in this project can successfully identify opinionated sentences from unstructured user reviews and classify their orientation with acceptable accuracy. This enables reliable review opinion summarization which has several commercially important applications.

Future Scope:

In the future, we want to perform opinion mining on larger and more varied blog data sets. We would also like to extend our work to fuzzy opinion classification with support for fuzzy user querying. We intend to do this by learning the strength of various adjective descriptors along with corresponding linguistic hedges and include them in the feature-orientation table generated during the mining process.

The classification technique proposed in the paper can be naturally extended to support fuzzy classification.

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 on Web Information Systems and Mining (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 ACM SIGKDD 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 Mining Workshops (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 from consumer reviews,” Information Systems and e-Business Management, vol. 8, no. 2, pp. 149–167, 2010.

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.