SLIIT

COMPUTING | BUSINESS | ENGINEERING

Software Requirements Specification

Emotion Extraction for Chat-Reviews

Project ID: 17- 066

IT 14 0320 66 – Cooray B.S.U.M

B.Sc. Special (Honors) Degree in Information Technology
Sri Lanka Institute of Information Technology
Date of Submission: 02-05-2017
Emotion Extraction for Chat-Reviews

Software Requirements Specification Report

Member	Student ID	Signature
Cooray B.S.U.M	IT 14 0320 66	

Supervisor: Ms. Nethmini Weerawarna

.....

Co-Supervisor: Dr.Darshana Kasthurirathna

.....

B.Sc. Special (Honors) Degree in Information Technology
Sri Lanka Institute of Information Technology
May 2017

DECLARATION

We declare that this is our own work and this project proposal does not incorporate without acknowledgement any material previously submitted for a Degree or Diploma in any other University or institute of higher learning and to the best of our knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text

Cooray B.S.U.M

…….

ABSTRACT

The use of Internet chat applications has benefited many different segments of society. It also creates opportunities for criminal enterprise, terrorism, and espionage. We present a study of a real-world application of chat analysis which will analyze chat messages in four ways such as topic detection, Emotion Extraction, Evaluate healthy and Personal information sharing analysis. Also analyzing chat traffic has important applications for both the military and the civilian world. Here on this document, it compares the results of an unsupervised learning approach with those of a supervised classification approach with regards to chat review application. The paper also discusses some of the specific challenges presented by this chat review application.

Unsupervised learning techniques such as clustering are very popular for analyzing text for topic identification as well as emotion extraction. These techniques have several attractive features, the most significant being that they do not require labeled training examples. This however is also a disadvantage under some circumstances. Therefor meantime we do this research we will discover more and more technologies required for analyzing chat massages based on four different categories such as topic detection, Emotion Extraction, evaluate healthy and Personal information sharing analysis.

With use of this chat analysis application user will be able identify the chatting partner in analytical way. And system will keep an analytical review for each chat session user interacted. Also system will be capable of showing its analytical data in a user friendly manner (in a graphical way).

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

1 Introduction

The introduction of the Software Requirements Specification (SRS) provides an overview of the entire SRS with purpose, scope, definitions, acronyms, abbreviations, references and overview of the SRS. The aim of this document is to gather and analyze and give an in-depth insight of the Emotion Extraction from Chat Messages in Chat-Reviews by defining the problem statement in detail.

1.1 Purpose

The purpose of this document is to give a detailed description of the requirements for the "Chat-Reviews" software. It will illustrate the purpose and complete declaration for the development of system. It will also explain system constraints, interface and interactions with other external applications. This document is primarily intended to be proposed to a customer for its approval and a reference for developing the first version of the system for the development team.

In short, the purpose of this SRS document is to provide a detailed overview of our software product, its parameters and goals. This document describes the project's target audience and its user interface, hardware and software requirements.

1.2 Scope

The system which is named as "Chat-Reviews" is proposed to develop as a real-world application of chat analysis which will analyze chat messages in four ways such as topic detection, Emotion Extraction, Evaluate healthy and Personal information sharing analysis. This document covers the requirements for release "Emotion Extraction" of "Chat-Reviews".

The analysis shows that this model, with a higher degree of specialization, can adapt better to identify the emotion of the people about the topic. It can be negative one or positive one according to information which is in chat messages. In the emotion extraction process we analyze the sentence for the existence of any emotion or not. If any we categorize the emotions under the identified emotion set. Then we show that details in

Graphical review. Also we hope to categorize that identified emotion set under the negative and positive. Those details also show in graphical manner.

So user can analyze the chat, this situation chatting partner in this type of emotion and that situation chatting partner in that type of emotion.

The final outcome of the project is an intelligent chat analysis application which user will be able to identify the chatting partner in analytical way. And system will keep an analytical review for each chat session user interacted. Also system will be capable of showing its analytical data in a graphical review.

The Objective is to design a flexible intelligent, efficient and real time chat analysis application and it's a kind of a cost effective, time saving way to a profitable business with new technologies.

1.3 Definitions, Acronyms, and Abbreviations

Term	**Definition**
User	Someone who interacts with the web application
AI	Artificial Intelligence
NLP	Natural Language Processing
ML	Machine learning is a type of artificial intelligence (AI) that provides computers with the ability to learn without being explicitly programmed
SRS	A software requirements specification is a comprehensive description of the intended purpose and environment for software under development
UML	Unified Modeling Language is an industry standard modeling language with a rich graphical notation, and comprehensive set of diagram and elements
Use Case Diagram	A broad level diagram of the project showing a basic overview

API	Application Programmer Interface
IM	Instant Messaging
DESC	Description
SLIIT	Sri Lanka Institute of Information Technology
DEP	Dependency

Table 1.1 - List of definitions, acronyms and abbreviation

1.4 Overview

The remaining sections of this document provide a general description, including characteristics of the users of this project, the product's hardware, and the functional and data requirements of the product. General description of the project is discussed in section 2 of this document. Section 3 gives the functional requirements, data requirements and constraints and assumptions made while designing the Chat-Reviews. It also gives the user viewpoint of product. Section 3 also gives the specific requirements of the product. Section 3 also discusses the external interface requirements and gives detailed description of functional requirements. Section 4 is for supporting information.

2 Overall Descriptions

This section will give an overview of the whole system. The system will be explained in its context to show how the system interacts with other systems and introduce the basic functionality of it. It will also describe what type of stakeholders that will use the system and further lists and briefly describes the major features and a brief description of each of the proposed system. At last, the constraints and assumptions for the system will be presented.

2.1 Product perspective

This system will consist of four parts

i) Topic Detection

ii) Emotion Extraction

iii) Evaluate Healthy

iv) Analysis of personal information sharing

Each of above mentioned parts will be developed as separate modules and it will be able to customize based on the outcomes required by client end. Also it will be feasible to add more modules and remove modules are allowed. This product would be easily integrated with other chatting application for the purpose of analyzing messages going through the system.

	Intelligent Diagnosis system	Honey Chatting	GroupWize	Chat-Reviews
Topic Identification	Yes		Yes	Yes
Emotion Extraction			Yes	Yes
Message Encryption		Yes		Yes
Detect Personal Information				Yes
Evaluate Healthy				Yes

Table 2.1 - Comparing the proposed system with existing systems

2.1.1 System interfaces

Chat messages on a chat will be accessed by our system. Since the system is collecting data from chat session and not doing any changes to the data. There is no need for a particular application interface.

2.1.2 User interfaces

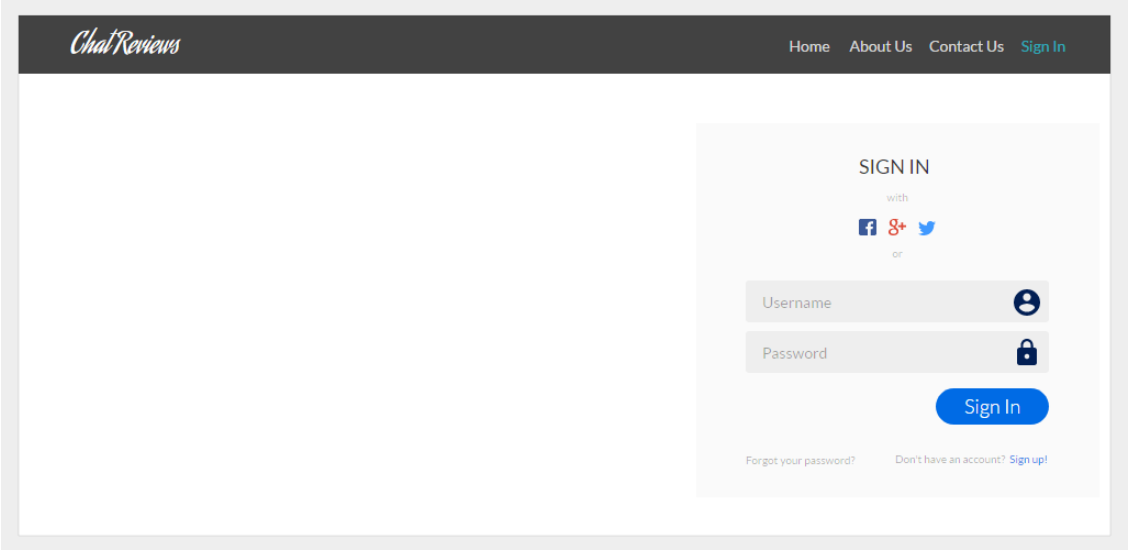

Figure 2 .1- Login page

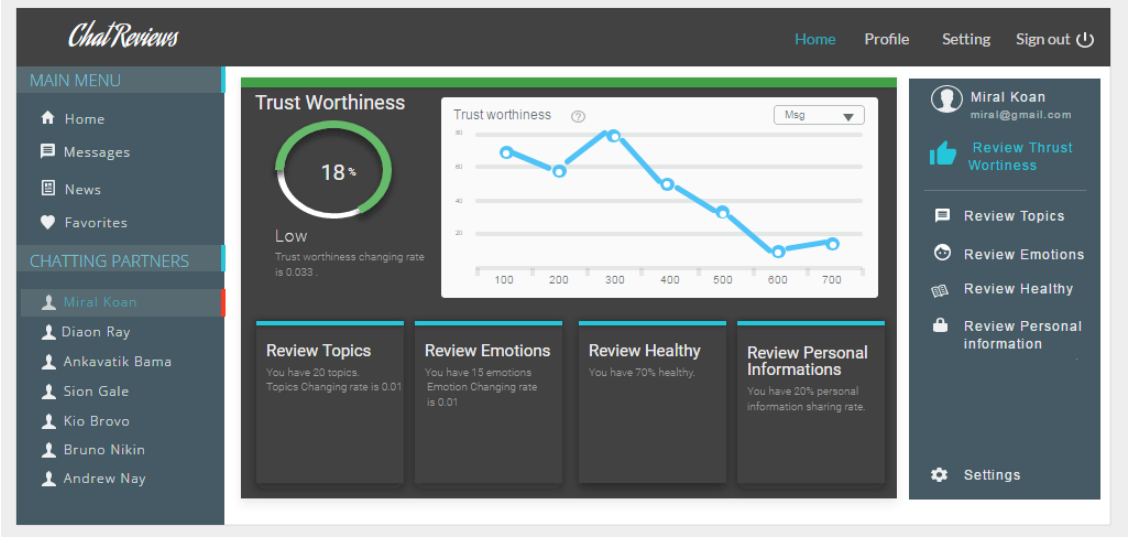

Figure 2.2 - Home page

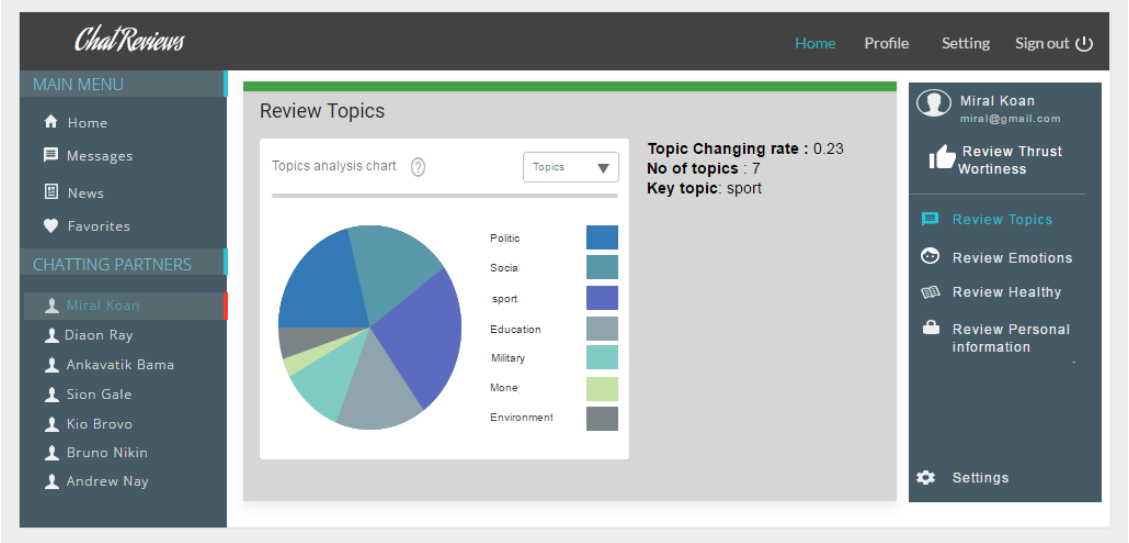

Figure 2.3 - Topic Reviews page

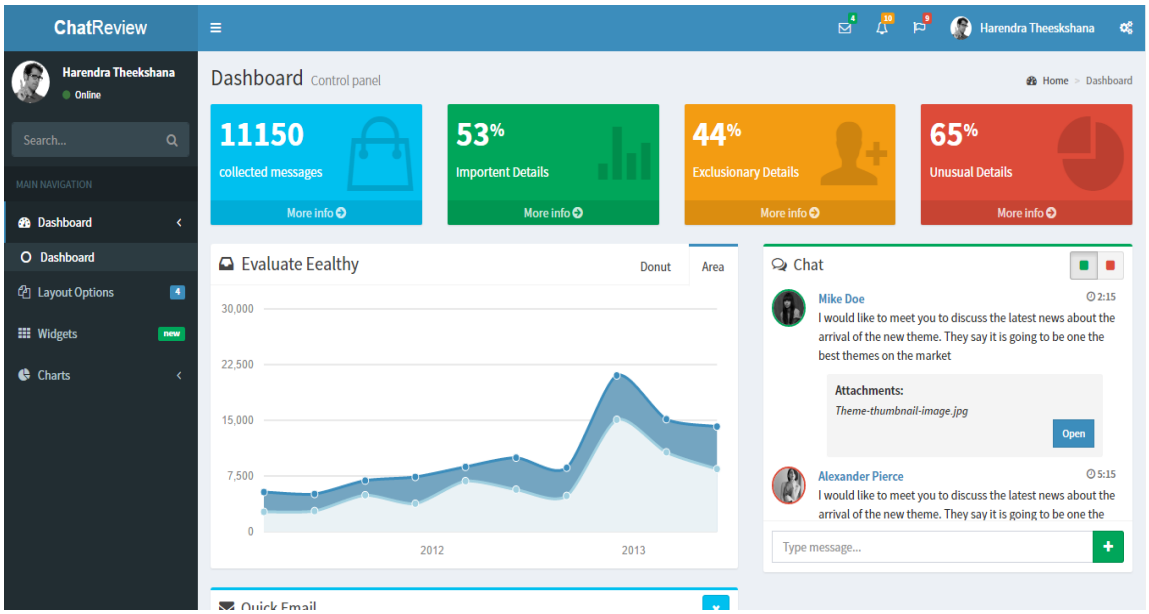

Figure 2.4 - Evaluate Healthy of particular chat

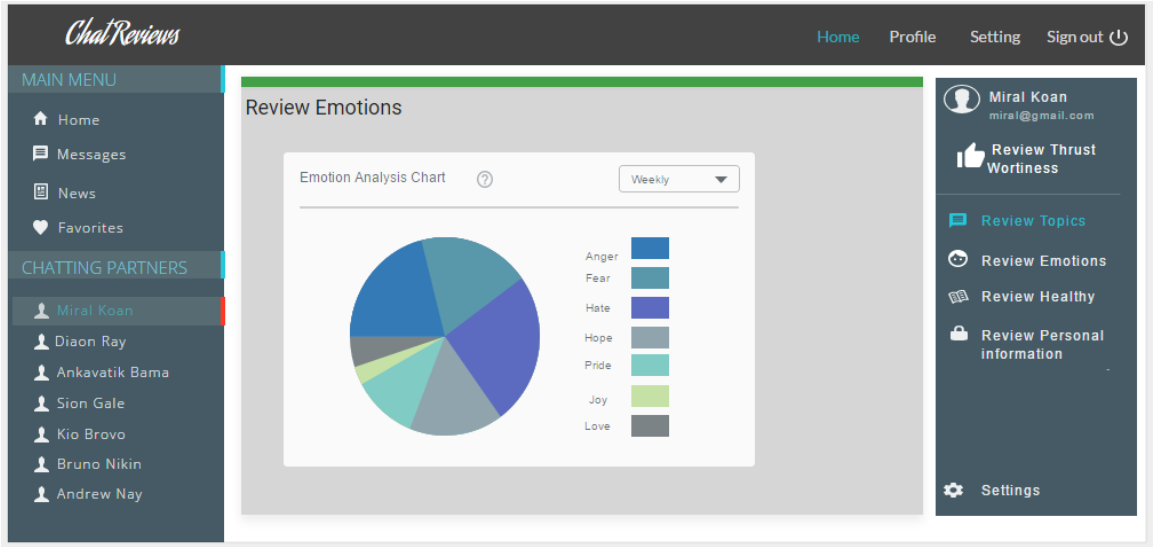

Figure 2.5 - Topic Reviews page

2.1.2 Hardware interfaces

- Web application runs on the Desktop PC or Laptop which should be able to high bandwidth internet access.
- A router or 3G dongle for internet access. It should be firewall enabled and should be able to handle excessive traffic.
- Server application to control the web application.

2.1.3 Software interfaces

The end users of the main web based system needs a web-browser to access the web based application. The system will be compatible with the default browser of Microsoft Windows operating system - Internet Explorer version 5.0 or above, Mozilla Firefox version 3.5 or above, Google Chrome etc.

2.1.5 Communication interfaces

A modem or a router will be required to access the system. Therefore, internet connectivity is essential to make use of the system.

- Dialup modem
- Broadband internet
- Dialup or Broadband connection with an internet provider

2.1.6 Memory constraints

The product would not have any special memory constraints. It is enough to have a web browser, internet connection and a device which can connect to internet.

2.1.7 Operations

- User should make sure that all the hardware equipment is connected properly and can run the web browser.
- Then user should be making sure that internet connection working properly.
- Then user should open the browser and load the Chat-Reviews.

2.1.8 Site adaptation requirements

Implementation of this system is carried out with English language because English is the international language that the majority speaks and comprehends.

2.2 Product functions

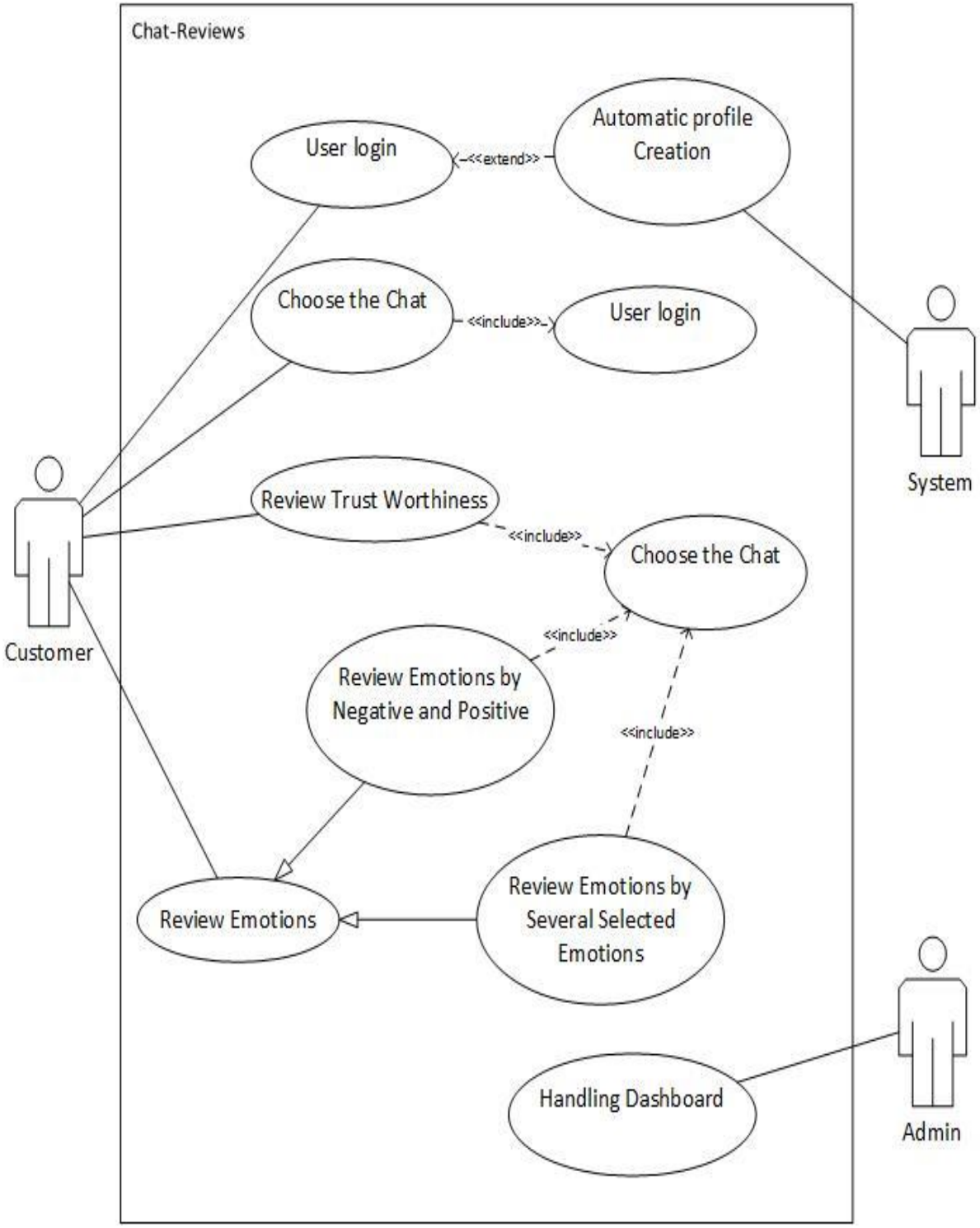

Figure 2.6 - Use Case Diagram

Use case Name: Automatic profile Creation
Actor: System, Admin
Pre-condition: User should have sent a valid request.
Main Flow: 1. Use case begins when Admin open the case. 2. System will create profiles automatically for a valid request. 3. Customer will redirect to the profile when logged to the system 4. Customer can add, update, and can perform basic operations to his / her profile 5. Use case ends when particular user logged off from the system

Table 2.2 - Use case scenario for Automatic profile Creation

Use case Name: User login
Actor: Admin, Customer
Pre-condition: User must have a user account.
Main Flow: 1. Use case starts when home page displays the login region 2. User enters the username and password. 3. System validates the username and password. 4. Use case ends when system redirects the user to the relevant profile page.
Extensions: 4.a System displays an error message by asking user to reenter the username and password.

Table 2.3 - Use case scenario for User login

Use case Name: Choose The Chat.
Actor: Customer.
Pre-condition: Customer should be logged to the web application.
Main Flow: 1. Use case begins when Customer open the case. 2. Customer can add particular chat session. 3. Use case ends when particular chat session added to the system.

Table 2.4 - Use case scenario for Choose the Chat

Use case Name: Review Trust Worthiness.
Actor: Customer.
Pre-condition: Customer should be added particular chat to the web application.
Main Flow: 1. Use case begins when Customer open the case. 2. Customer can review trust worthiness of particular chat session within graphs. 3. Use case ends when particular chat session reviewed by system.

Table 2.5 - Use case scenario for Review Trust Worthiness

Use case Name: Review Emotions by Negative and Positive.
Actor: Customer.
Pre-condition: Customer should be logged to the web application.
Main Flow: 1. Use case begins when Customer open the case. 2. Customer can review variation of positive and negative emotions of particular chat session within graphs. 3. Use case ends when particular chat session reviewed by system.

Table 2.6 - Use case scenario for Review Emotions by Negative and Positive

Use case Name: Review Emotions by Several Selected Emotions.
Actor: Customer.
Pre-condition: Customer should be logged to the web application.
Main Flow: 1. Use case begins when Customer open the case. 2. Customer can review variation of several selected emotions of particular chat session within graphs. 3. Use case ends when particular chat session reviewed by system.

Table 2.7 - Use case scenario for Review Emotions by Several Selected Emotions

2.3 User characteristics

User of this system should have knowledge about the computer and handle web browser and load the application. Understand the English language and website features such as buttons, icons and other similar tools operations.

In this system there are 2 users

- **Customer**

Customer will have the permission to access the system with login approvals. System will allow all operations which customer entitled to a registered customer.

- **Administrator**

Administer can have the permission to manage this application process.

2.4 Constraints

Developers should also be careful about the privacy of users. Since product will be cloud application, all user data will be kept on cloud server and necessary precautions should be taken to protect user data.

Since product will be cloud application and all user programs will be executed on cloud server, developers should limit the privileges of the users so that they cannot harm other users' data and system server.

The developers have to be extra careful about the accuracy and the availability of the data to provide reliability service to users. The Internet is wide area, so the connection speed should be concern in prime.

2.5 Assumptions and dependencies

- We assume in operating systems versions will support for future web server applications.
- Users who don't have basic computer and website knowledge will be unable to use this system.
- Security of the data is also considerable since inside the database there will be personal data and these data must be protected without losing the privacy.
- Web server will run 24 hours a day without any interruptions.
- Analyze is only for English language.
- We assume every customer who interacts with this system will use this system meaningfully.

2.6 Apportioning of requirements

Due to the fact that this is a research type project, some features indicated in the specification may change in the final product.

3 Specific requirements

This section contains all of the functional and quality requirements of the system. It gives a detailed description of the system and all its features.

3.1 External interface requirements

This section provides a detailed description of all inputs into and outputs from the system. It also gives a description of the hardware, software and communication interfaces and provides basic prototypes of the user interface.

3.1.1 User interfaces

Login page

1)

Name of item: User Name

Description of purpose: Enter username

Source of input: keyboard input

Valid range: Should be an existing username

Units of measure: Number of characters (100 Characters)

Timing: Enter once for each login

Relationships to other inputs: Strongly coupled with the password

Data formats: String

2)

Name of item: Password

Description of purpose: Enter password

Source of input: keyboard input

Valid range: Should be valid to the relevant login

Units of measure: Number of characters (100 Characters)

Timing: Enter once for each login after entered the username

Relationships to other inputs: Should be matched with entered username

Data formats: String

Home page

1)

Name of item: Review Emotion

Description of purpose: review emotional informations on chat session graphically.

Source of input: Mouse click

Valid range: Should contain with industry-wise words

Timing: Click once for each reviews.

Relationships to other inputs: No strong relationships with other inputs

Data formats: String

3.1.2 Hardware interfaces

- PC- PC should be able to run the Web Browser
- Router - Router should be firewall enabled. It should be able to handle excessive traffic. Server - Server application controls the entire system of client request response.

3.1.3 Software interfaces

Operating System:

- Microsoft Windows 7

Software for Documentation

- Microsoft Word 2010
- Microsoft Power Point

3.1.4 Communication interfaces

- Modem
- Router
- Internet connection

3.2 Classes/Objects

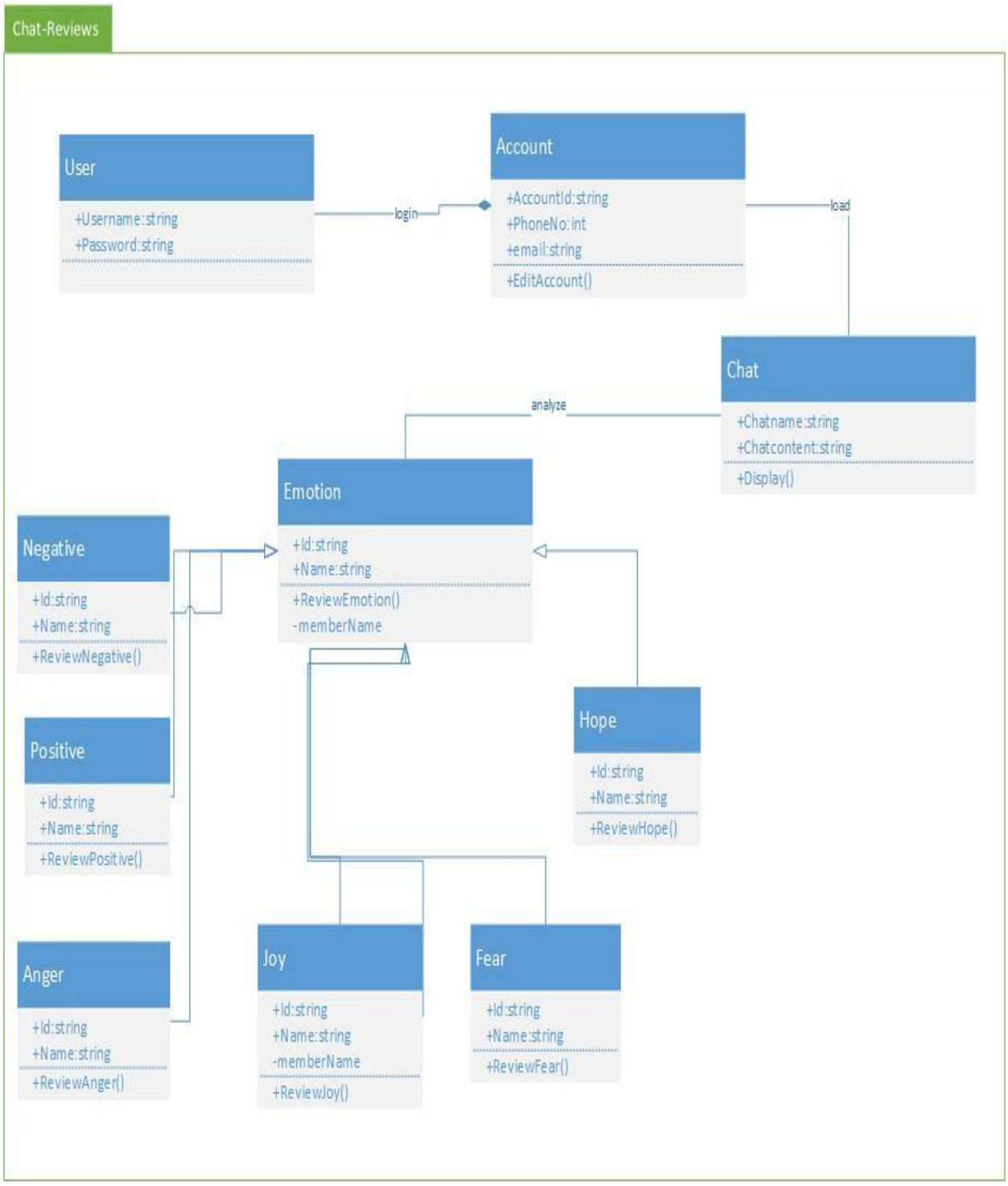

Figure 3.1 - Class Diagram

3.3 Performance requirements

We will do our best to maximize the performance of the system. Performance of the Chat-Reviews varies according to the several factors. According to our minimum hardware requirements, we expect the systems loading time within 30 seconds and every page loading time within maximum 2-3 seconds. But it will be changed depending on the hardware performance also. Moreover quick and accurate response generation should be performed in a minimum period of time.

3.4 Design constraints

During the design stages the major constraint that will be faced is the limitation of available time. The project group is expected to complete this project during the period of 10 months. Another constraint that will be faced is that it is difficult to find past research and model projects that relate to the some parts of this project.

3.5 Software system attributes

3.5.1 Reliability

When consider any system; System reliability is the most important part of it. Because, the entire system process is depend on that. If there is any crash or error occurs those can be recovered without any harm to the system or without any data losses. A system may come up with some hardware or software failures within its processing time. Hardware failures can be occurs more frequently than the software failures. In order to enhance that reliability in this system backup system will be used in order to ensure the system data security. It is very important to the user. Because if any case if the system crash user can rely on the backup. This system should be a highly reliable system, with a Mean Time to Failure greater than 8000 hours. The system should be tested for errors while developing units or modules. Then testing will be done while integrating the modules and then the final testing will also be performed on the final system assuring high reliability. Multiple users can access system simultaneously.

3.5.2 Availability

When it comes to the availability it plays major role in this system. The system must be available for one user at a time for modifications. But many users can analyze chat sessions at once.

3.5.3 Security

Security one another important thing which must highly consider with the implementation of the system, the process of Chat-Reviews depends on the chat sessions data. Therefore we mainly concern about security requirement in order to protect the chat messages details. We facilitate that system requirement by controlling the access to updating system. Only authorized people can access to that system by giving their username and password to do some changes. Unauthorized access to the database data is restricted.

3.5.4 Maintainability

In development site, System will need to maintain a own database Because social chat messengers will not give every details that we asked like old data. So the system should be ready to detect these issues in future. So system will always observe by the system administrator. Every new change will be tested by the developers before deliver to the users. Database of the system will be maintained by the system administrator.

3.6 Other requirements

User friendliness

For web based system user friendliness is important. If it is really easy to navigate through the web application; those web applications can have more user attractiveness

Reusability

Functionalities, database and other new ideas of this system will help to update the system with new features. Reusability will reduce the system cost and complex of the system functionalities.

Correctness

Correctness of the analysis of text will check by tester before move to next implementation. Final result also will check by administrator in every updates.

Testability

System testing, acceptance testing, regression testing, smoke testing and etc. It will be applied when doing testing.

Accuracy

Our system all functions are related with other functions. So A single error can harm the whole project.

Speed

System must have good speed when load in the web browser.

Safety Requirements

System will have backup functions and backup databases for safety. If the system will fail in a development site or after delivered all backups will help to re test the system.

4 Supporting information

4.1 Appendices

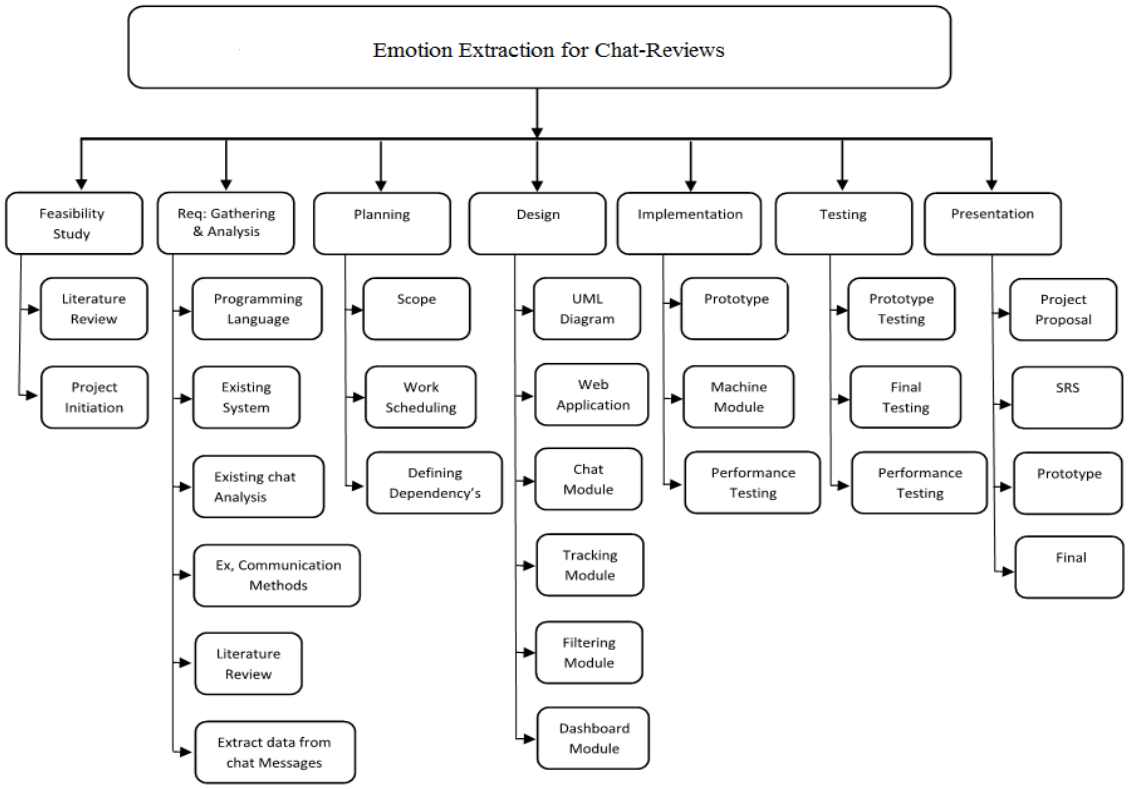

Figure 4.1 - Work Breakdown Structure

4.2 References

[1] D. Jurafsky and J. H. Martin, *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. London: Prentice Hall, Pearson Education International, 2009.

[2] S. C. Herring, "Computer-Mediated Communication Linguistic, Social, and Cross-Cultural Perspectives," 1996.

[3] E. N. Forsyth, "Improving automated lexical and discourse analysis of online chat dialog," M.S. thesis, Naval Postgraduate School, Monterey, CA, 2007.

[4] M. Zitzen and D. Stein, "Chat and conversation: a case of transmedial stability?" *Linguistics,* vol. 42, pp. 983–1021, 2004.

[5] D. M. Blei, A. Y. Ng, M. I. Jordan, and J. Lafferty, "Latent dirichlet allocation," *Journal of Machine Learning Research,* vol. 3, 2003.

[6] M. R. Freiermuth, "Features of electronic synchronous communication: a comparative analysis of online chat, spoken and written texts," 2002.

[7] C. D. Manning and H. Schütze, *Foundations of Statistical Natural Language Processing,* 6th printing ed. Cambridge, Mass.: MIT Press, 2003.

[8] J. G. Shanahan and N. Roma, "Improving SVM text classification performance through threshold adjustment," in *Machine Learning: ECML 2003,* vol. 2837/2003, Berlin / Heidelberg: Springer Berlin / Heidelberg, 2003.

[9] W. Cohen, V. R. Carvalho and T. M. Mitchell, "Learning to Classify Email into Speech Acts," *EMNLP,* 2004.

[10] J. Lin, "Automatic author profiling of online chat logs," M.S. thesis, Naval Postgraduate School, Monterey, CA, 2007.

[11] J. Tam, "Detecting age in online chat," M.S. thesis, Naval Postgraduate School, Monterey, CA, 2009.

[12] D. Avrahami and S. E. Hudson. “Communication characteristics of instant messaging: effects and predictions of interpersonal relationships.” In CSCW’06, pages 505–514, 2006.

[13] X. Shi, L. A. Adamic, and M. J. Strauss. “Networks of strong ties. Physica A Statistical Mechanics and its Applications”, 378:33–47, May 2007.

[14] Instant Messaging Traces , http://tstat.tlc.polito.it/traces-IM.shtml.

[15] Stellar Technologies, “Stellar Technologies, Inc. – web, e-mail & IM management Stellar Internet global employee management”. Available at http://www.stellarim.com/.

[16] Spectorsoft.com, “SpectorSoft.com – Spector Pro for Windows”. Available at http://www.spectorsoft.com/products/SpectorPro_Windows/index.html.

[17] Zemerick Software Inc., “Zemerick software – home of Chat Watch”.

Available at http://www.zemericks.com/.

[18] Exploreanywhere.com, “SpyBuddy spy software”. Available at http://www.exploreanywhere.com/sb-intro.php.

[19] InvisibleKeylogger.com, “Invisible Keylogger – the perfect stealth keylogger”. Available at http://www.invisiblekeylogger.com/.