CHAPTER 1

INTRODUCTION

1. INTRODUCTION

1.1 SYSTEM OVERVIEW

The pervasive use of social media has generated unprecedented amount of social data. Social media provides easily and accessible platform for users to share information .Mining social network has its potential to extract actionable pattern that can be beneficial for business, users & consumers. Social network mining using nosql gives a faster way to store, access & modify data. The proposed approach is to achieve an efficient way for mining social networks by relational models i.e. graph based nosql, text mining and crawling for predicting poll results depending on the views of the people predicting the reach among the public for an ad or a product. Avoiding juvenile delinquencies by finding out the kids who have developed hatred towards school or community etc Avoid cyber crime by monitoring abuse words .

1.2 OBJECTIVES

To do Data mining using nosql database helps discovering useful and actionable knowledge or information in large scale data. It gives user an easy-to-use to communicate and network with each other.

1.3 PROJECT OVERVIEW

A technique that discovers the useful knowledge or information from social network database is used to store the large scale of data with preprocessing of text data. To the modified text, mining is performed. The mined data is displayed in the form of word cloud or graph. The output can be used for predictive analysis. The data can also be stored in graph like nosql database like neo4j and make it to represent relationships in graph. The output is visualized either through word cloud or graph . The visualized output of social network mining can be used for predictive analysis

1.4 SYSTEM STUDY

1.4.1 Existing System

In the traditional social network mining has dealt with the data which are not networked. The dimensional(increasing size) problem is of critical scenario for analysis of social networks, dealing in aggregation in tables with terabytes of data prohibiting, if the analysis has to useful for fast decision making.

1.4.2 Literature Survey

Based on the proposed system various literature surveys were done. The papers referred in developing the project and the ideas inferred from them are discussed here. In the paper titled Mining the network value of customers developed a method for mining social network models from collaborating filtering databases and using this model to optimize the marketing decision. Maximizing the spread of influence in a social network. Models for the processes by which ideas and influence propagate through a social network have been studied in a number of do mains, including the diffusion of medical and technological innovations, the sudden and widespread adoption of various strategies in game-theoretic settings, and the effects of “word of mouth” in the promotion of new products .

1.4.3 Proposed System

The proposed system is predicts poll results depending on the views of the people. Predicting the reach among the public for an advertisement or a product. Avoiding juvenile delinquencies by finding out the kids who have developed hatred towards school or community etc. cyber crime by monitoring abuse words. A visualization of output of social network mining represented in either word cloud or graph. This visualization of social network analysis can be used for predictive analysis.

1.5 ORGANISATION OF THE REPORT

The Chapter 1 presents review of literature. The goal tended to be attained in the project is explained in objectives. The problem description tells the need for the system with the advantage of proposed system over existing system. The Chapter 2 explains the system requirement for both, by feature and by functional hierarchy. The Chapter 3 describes the System Design which includes the decomposition description, dependency description and detailed design of modules. The Chapter 4 describes the implementation of the project, which includes the modules and components used in this project. The Chapter 5 deals with the test plan and testing of the project. The Chapter 6 describes the results of the implementations. The Chapter 7 contains the conclusion of the work done and also the extension of the work.

CHAPTER 2

SOFTWARE REQUIREMENTS SPECIFICATION

2. SOFTWARE REQUIREMENT SPECIFICATION

2.1 OVERALL DESCRIPTION

2.1.1 Product Perspective

The pervasive use of social media has generated unprecedented amount of social data. Social media provides easily and accessible platform for users to share information. Mining social network has its potential to extract actionable pattern that can be beneficial for business, users & consumers.

2.1.2 Product Functions

The project allows the users to find important information regarding interest, trends, relations and also gives freedom to users to build a network according to his area of interest. Searching about particular person, community, products in efficient ways ontologically.

2.1.3 User Classes and Characteristics

The user privileges vary according to their designation. Knowledge of how to handle the system is necessary. The user interface will be friendly enough to guide the user.

2.2 EXTERNAL INTERFACE REQUIREMENTS

2.2.1 User Interfaces

The main interface of the user is extremely user friendly. The desktop environment is necessary for the showing results based on mining. Relationship between different objects, items or user will be generated in a graphical model. The keywords are taken as the queries from the user to select the information from the database.

2.2.2 Hardware Interfaces

There is no distinct interface for hardware upon which the project needs to reside. For this project, these requirements are important.

Hardware Specification

- Processor : X84 Architecture
- RAM : 1GB
- Hard Disk Drive : 80GB

2.2.3 Software Interfaces

Software Specification

- Platform : R 2.15
- IDE : RStudio (optional)
- Operating System : Windows or Linux

2.3 SYSTEM FEATURES

2.3.1 Data Collection

Description and Priority

- This module is responsible for collecting data from social networking site.
- The keyword is obtained as user input via GUI , when the user types the search term and gives enter.
- Corresponding search takes place through twitter API to collect required data. Once the package connects with twitter API it returns the collection of required tweets. The tweets collected are stored in data frame which are meant to be preprocessed.

Stimulus/Response Sequences

Once the data is collected from social networking site, it can be used to process the data and mine the data, which are the subsequent modules

Functional Requirements

- Purpose : To collect comments from social networking site.
- Inputs : Search terms.
- Output : List of comments from the social networking site

2.3.2 Data Preprocessing

Description and Priority

- To preprocess the data.
- Removes Stopwords , special characters.
- Converts all words to lower cases.

Stimulus/Response Sequences

Once the data has been preprocessed it can be dumped into corpus and can be proceeded with mining, before this process all the stopwords has to be removed such as hash tags , special characters, non ASCII input and number. The tweets are then converted into lower case with separate words.

Functional Requirements

- Purpose : To pre process the data
- Input : Data
- Processing : The pre processed data is processed and then various algorithms are implemented.
- Output : Preprocessed data

2.3.3 Mining the data

Description and Priority

Mining the data is to find out and discover the useful patterns, trends that is present in the data. It can be classified either by emotion or by polarity.

Stimulus/Response Sequences

Once the data is preprocessed, it is converted into corpus where several mining operations can be performed, such as text mining, sentimental mining and social network analysis etc.

Functional Requirements

- Purpose : To mine the useful patterns in the data.
- Input : Preprocessed data.
- Output : Graphical/Word cloud output to know user's activity.

2.4 OTHER NON FUNCTIONAL REQUIREMENTS

2.4.1 Performance Requirements

The data frame can store large amounts of data with it which allows the admin to update it regularly. It will not crash or freeze at the moment of high activity and implementation. The data frames can be passed through error handling functions which is capable of removing non ASCII inputs by using regular expression, which results in reduced consumption of memory and hence improving the performance of the overall system. It can work both on 32 bit and 64 bit architecture and since it is cross platform, it can work in any environment with basic dependencies.

2.4.2 Software Quality Attributes

As this project is based on the mining public data available over social networking sites on the web any user can use this system without any trouble.

2.4.3 Safety Requirements

The system possibly will not make any loss, damage or harm to user loaded data. The system will not have any unreasonable effect on operating system and the performance of the system on the whole.

2.4.4 Security Requirements

This project can be strictly used for non-commercial purpose. This project is created in the hope that it will be useful, but without any warranty without even the implied warranty of merchantability.

CHAPTER 3

SOFTWARE DESIGN

3. SOFTWARE DESIGN

3.1 SYSTEM ARCHITECTURE

3.1.1 Architectural Design

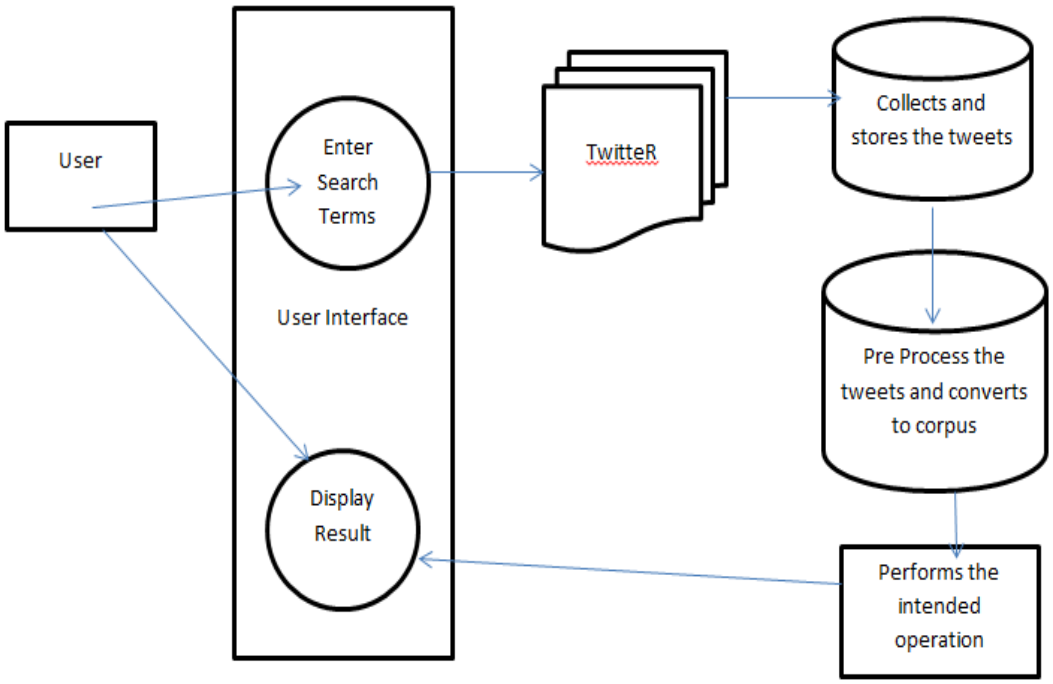

Fig 3.1:System Architecture

In this architectural design, social network mining is done with a simple process with a user interface. User enters the keyword to search. Corresponding method for the operation specified is selected and passed through twitteR package which in turn connects to Twitter API. It results in tweets, which are pre processed and then processed for mining the data. Once the mining of the data has been completed the output of the mining is displayed to the user . The output may be either of the type graph or word cloud. Refer Fig 3.1.

3.1.2 Use Case Diagram

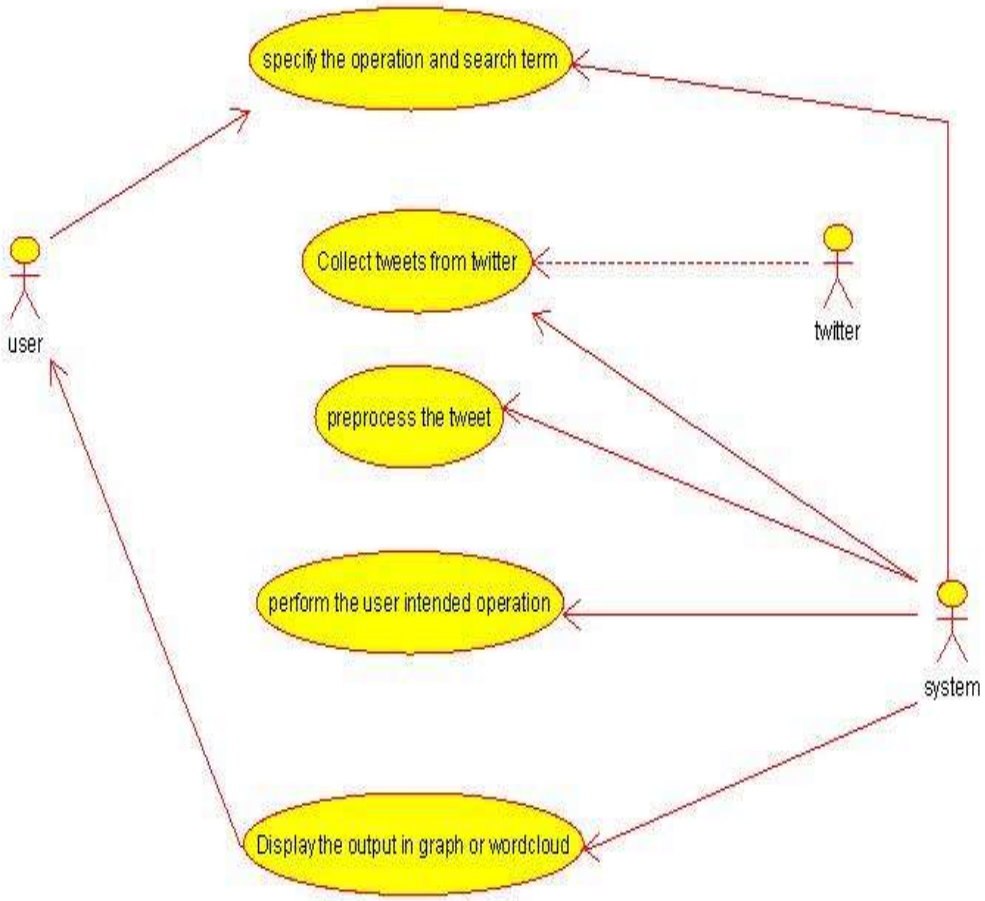

Fig 3.2: Use Case Diagram

An usecase diagram is a type of behavioral diagram defined by and created from a use-case analysis.Its purpose is to present a graphical overview of the functionality provided by a system in terms of actors, their goals and any dependencies between those usecase. The usecases that are involved here are User specifying operation and search term, collecting tweets, preprocessing tweets, performing intended operation and displaying the output The actors are user , system and twitter API . Refer Fig 3.2.

3.1.3 Class Diagram

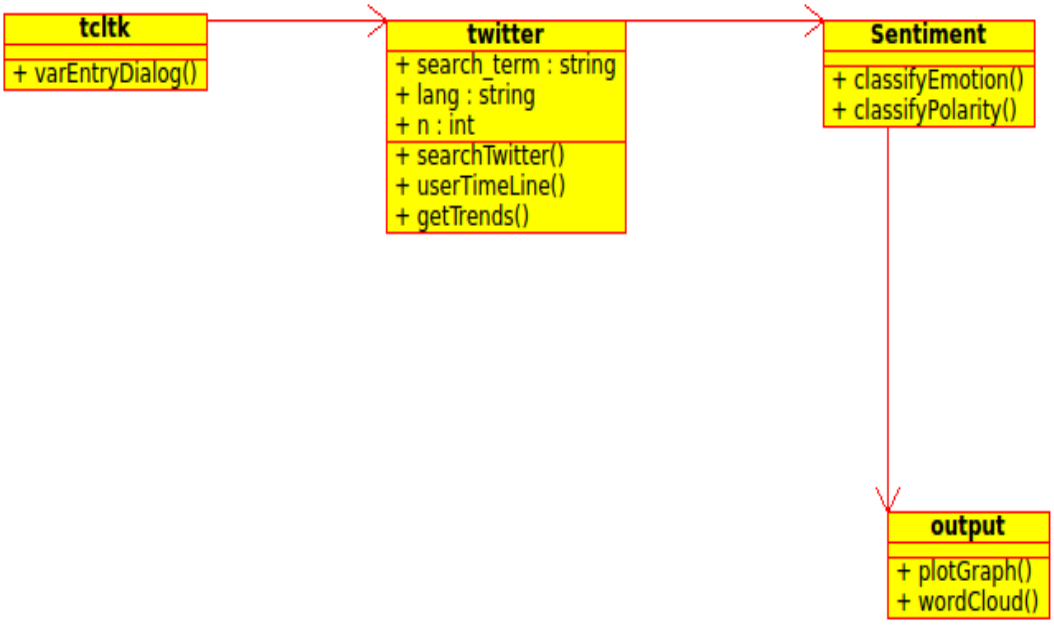

Fig 3.3 : Class Diagram

A class diagram is a type of static structure diagram that describes the structure of a system by showing the system's classes ,their attributes ,operations(or methods),and the relationships among the classes. The class diagram of the project is describing that the modules of each class and functions like search twitter, classify polarity. The class diagram will be used for creation of methods and members of a class during implementation stage. Classes include TwitteR, Sentiment, the various methods that are used here includes varEntryDialog, searchTwitter, userTimeLine, getTrends, plotGraph(), wordCloud() etc. Refer Fig 3.3.

3.1.4 Activity Diagram

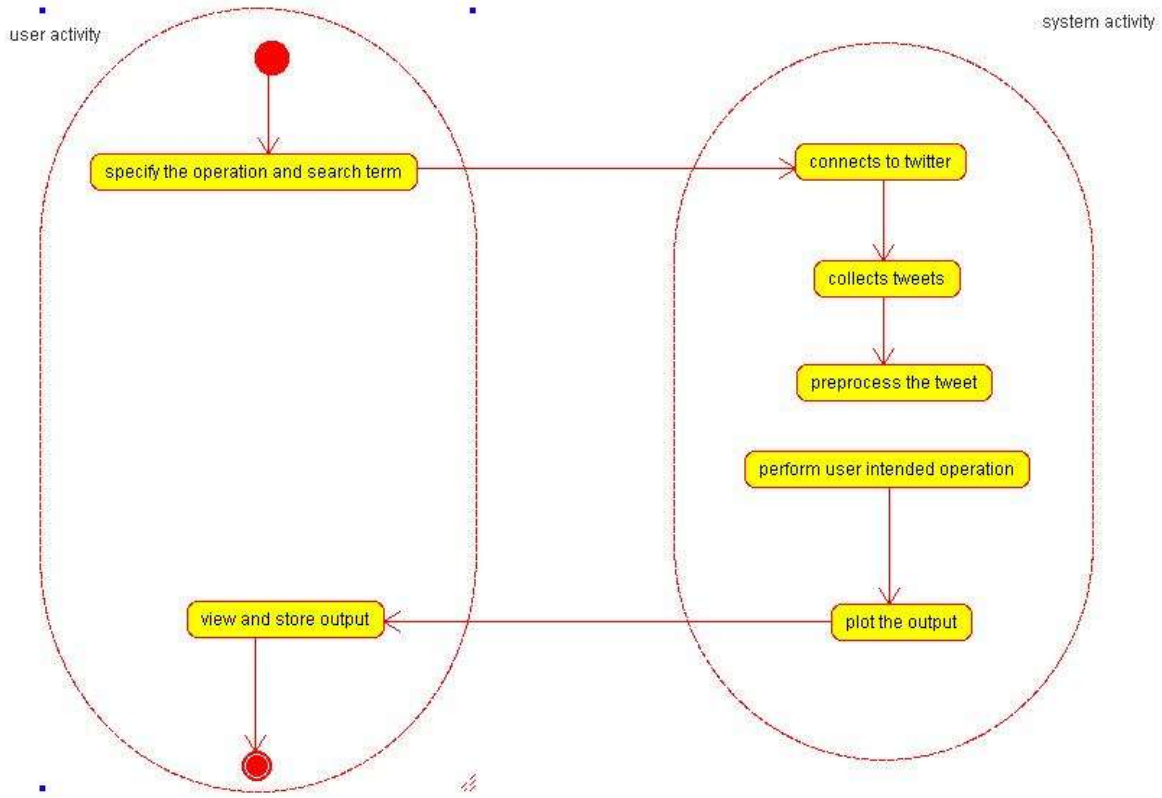

Fig 3.4: Activity diagram

The activity diagram shows the overall activities that occur in the program. User enters the keyword to search. Corresponding method for the operation specified is selected and passed through twitteR package which in turn connects to Twitter API. It results in tweets, which are pre processed and then processed for mining the data. The activity diagram also depicts the overall activities that are involved in the project conveniently which would help in better understanding of the project .Once the mining of the data has been completed the output of the mining is displayed to the user . The output may be either of the type graph or word cloud, as shown if Fig 3.4. Then predictive analysis can be performed based on the output . The output can be saved in either jpg or pdf format.

3.1.5 Sequence Diagram

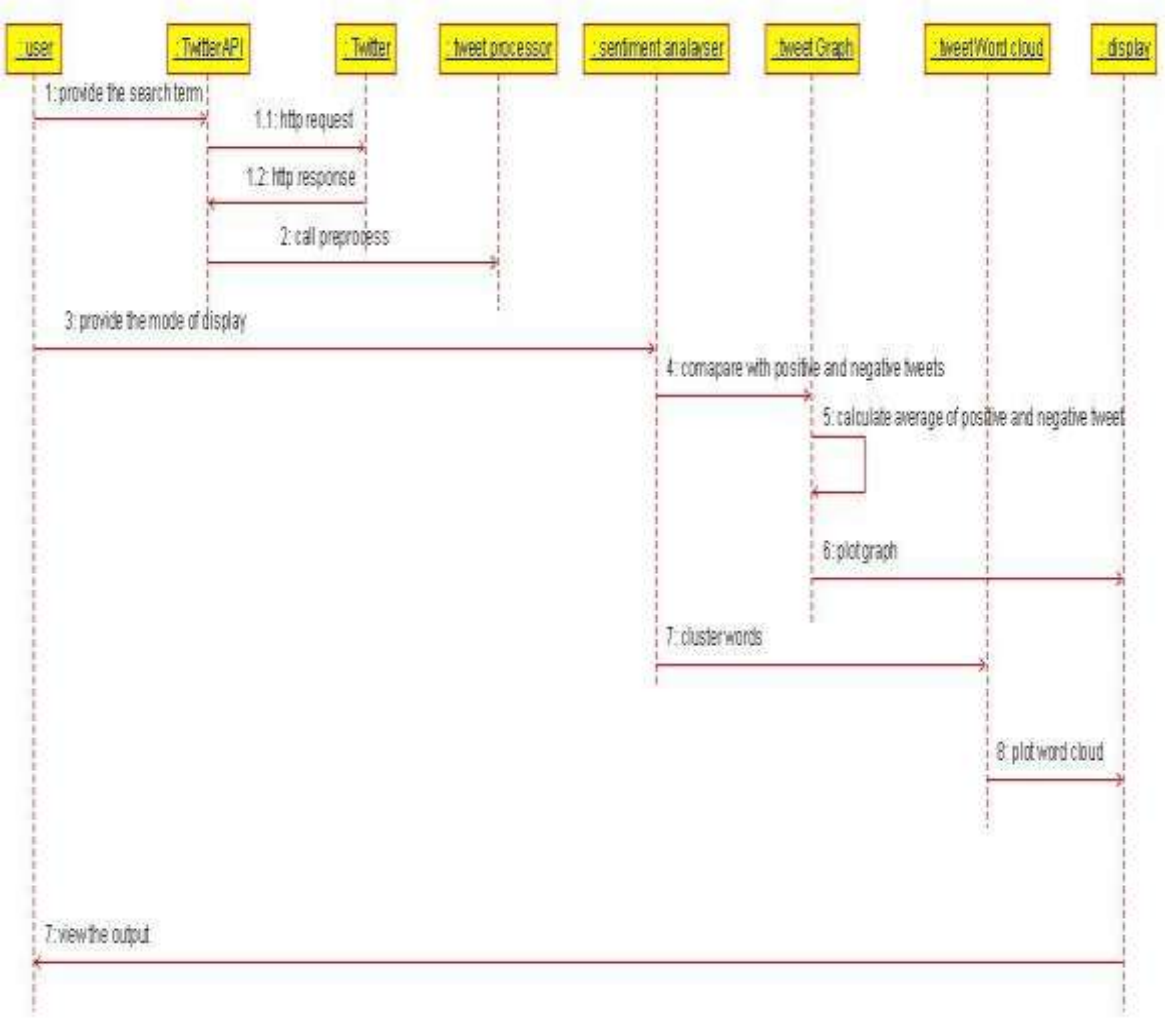

Fig3.5 : Sequence Diagram

The sequence diagram is a kind of interaction diagram that shows how processes operate with one another and in what order. It is a construct of a, message sequence chart. It depicts the objects and classes involved in the scenario and the sequence of messages exchanged between the objects needed to carry out the functionality of the scenario. The responses are provided to reveal interesting rules and patterns through analysis and discovery. Refer Fig 3.5. The sequence ends when the visualization is given as the output to the user .

3.1.6 Data Flow Diagram

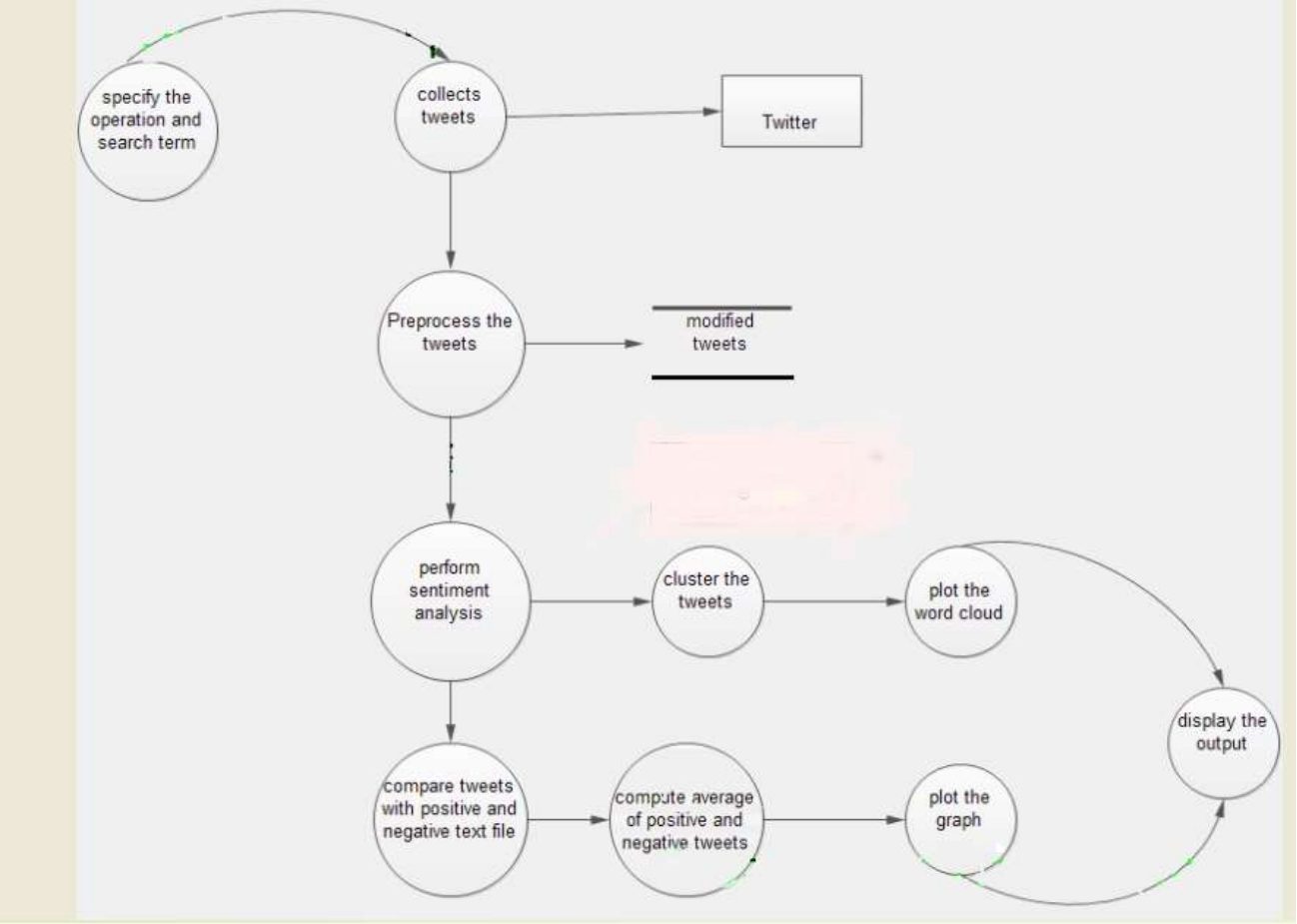

Fig 3.6: Data Flow Diagram

The user interface is used to get the keywords from person or the sentiment analysis the desire keywords like product name, trends etc is taken. The required request is processed using various data mining techniques like preprocessing ,categorization ,clustering. The processed data is displayed as the result of social network analysis in graphical form with removable of all possible stopwords. Then the data flows to the mining .Appropriate mining algorithm for the user intended operation is applied here. The mining results are stored as output. The output is displayed to the user either in the form of word cloud or graph for visualization of the data in word cloud or graph. Refer Fig 3.6. From the visualization of social network analysis which can be used for predictive analysis

3.2 DECOMPOSITION DESCRIPTION

The decomposition of subsystems in the architectural design is explained in this section by the functional diagram like data flow diagram(DFD) and by various UML diagram

3.2.1Module1: Data Collection

The user interface is used to get the tweets for the sentiment analysis the desire keywords like product name ,trends etc, is taken. Twitter API is loaded and the tweets are returned as result. The results that is tweets are in turn given as the input to the module 2 which is data preprocessing.

3.2.2 Module2: Data Preprocessing

The output of the module 1 is given as the input to the module data preprocessing. Here the data is preprocessed by removing non ASCII characters, special characters, re-tweets, numbers and all the words in tweets are converted into lowercase. Once the collection of tweets is preprocessed by removing all noisy characters, the tweets are separated into tokens by splitting into each word based on space. The preprocessed data is then passed on to the next module for performing mining action

3.2.3Module3: Mining the Data

The preprocessed data processed using various data mining techniques like sentimental mining, opinion mining, classification by emotion and classification by polarity. The processed data is stored as the result of social network analysis. Appropriate mining algorithm is selected when the corresponding operation is invoked The result can be manipulated for predictive analysis. Visualization of the results is represented in either word cloud form with removable of all possible stop words or graphical output such as bar chart or histogram etc. The outputs can then be saved as image and pdf formats for later use. The visualized output can be used for predictive analysis and decision making.

3.3 DATA DESIGN

3.3.1 Data Description

The dataset used in this project is default R data frame , which if of vector list . Other data used are search , which is the term to be searched , tweets refers to the micro blogging comments , corpus refers to the collection of required text , emotion and polarity refers to the output of Bayes algorithm of classification .

3.3.2 Data Dictionary

Data	Description
Text	Text of the DM
Recipient	A user object representing the recipient of the message
Recipient ID	ID number of the recipient
Recipient SN	Screen name of the recipient
Sender	A user object representing the sender of the message
Sender SN	Screen name of the sender
Sender ID	ID number of the sender
Created	When the messages was created

Table 3.6: Data Dictionary

The Data Dictionary consists of all the terms associated with their description . The common terms associated here are given in Table 3.6. The terms are stored as data in data dictionary is to serve the purpose of understanding the contextual meaning of their corresponding description which is given . These terms are used through out this project and in the explanation given in the report .

3.4 COMPONENT DESIGN

3.4.1 Data Collection

In this module, the system reads the input and connects to twitter API via twitteR package and collects the tweets .

BEGIN

READ the input

CONNECT to twitter API via twitteR

PASS the input as arguments

COLLECT the required tweets

CONSTRUCT data frame

SAVE the tweets in data frames

END

3.4.2 Data Preprocessing

Once the required tweets are collected, preprocessing of tweets to be carried out. Load the tweets and use regular expression to remove special characters and retweets. Extracts the required content from tweets and save.

BEGIN

LOAD tweets

REMOVE special characters

REMOVE non ASCII characters

REMOVE numbers

REMOVE retweets

CONVERT into lowercase

SPLIT into words based on space

EXTRACT required content from tweets

CREATE a list containing only text data

SAVE preprocessed tweets

END

3.4.3 Mining the Data

The data frames are stored in the lexical corpus. The matrixes term document matrix and document term matrix are created .For mining the text Bayes algorithm is used to classify either based on emotion or polarity.

BEGIN

CREATE lexical corpus

CREATE of term document matrix

CONSTRUCT document term matrix

APPLY appropriate mining algorithm

CLASSIFY either based on emotions or polarity

CONSTRUCT comparison of two or more mined collection

END

3.5 HUMAN INTERFACE DESIGN

3.5.1 Overview of user interface

The user interface I simple and designed in the way to ease the use of the application. The system is made up of two parts of UI for selection of process and to give input parameter.

3.5.2 Screen Images

Selection of process

The first main GUI is the selection process, in which user can select their choice by clicking radio button, the options are search term in tweets, display tweets. Operation specific graphical user interfaces are displayed once the operation is selected from the main window that is the selection of process window. This window is called when the script is invoked hence the first thing which is displayed in user interface is selection of process GUI, which consists of several radio buttons, with each radio button denoting a single process. Each radio button denotes an operation , on selecting it the operation is invoked.

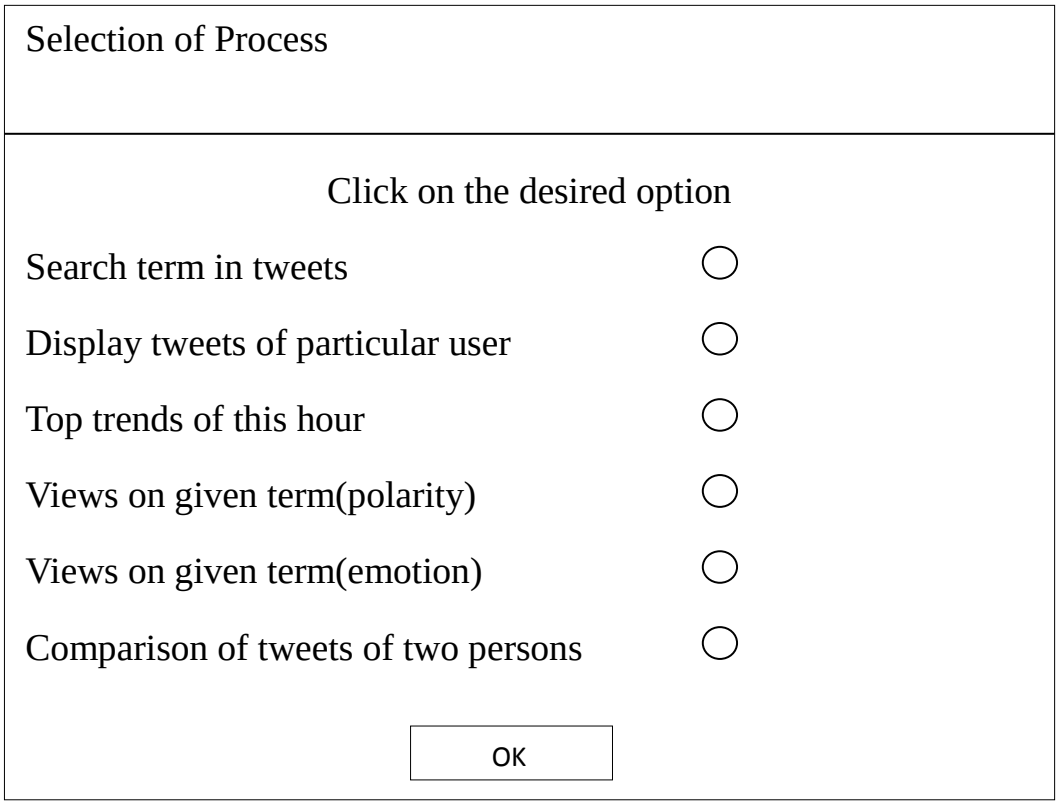

Fig 3.7: Selection of Process

Search parameter

On clicking 'Submit' displays GUI, in which user can give the search term to be searched in twitter, based on the choice the user selection

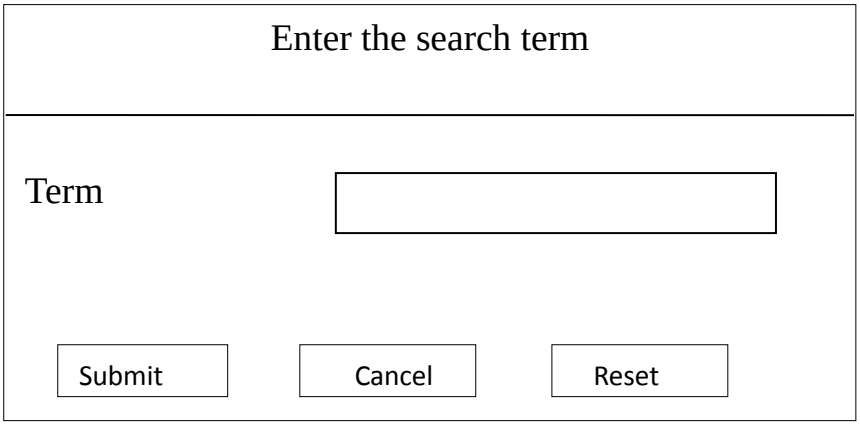

Fig 3.8: Search Parameter

Username Of The Required User :

This GUI is used to get the input for the username, this is used as parameter for userTimeline function.

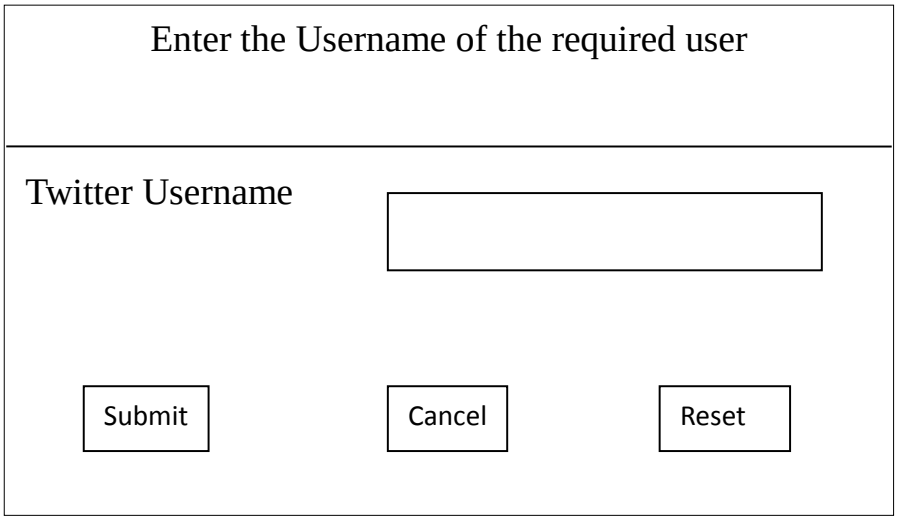

Fig 3.9: Username of the Required User

Search Term to Classify Based on Emotion :

The search term to classify the tweets based on emotion is entered in order to collect tweets and classify accordingly.

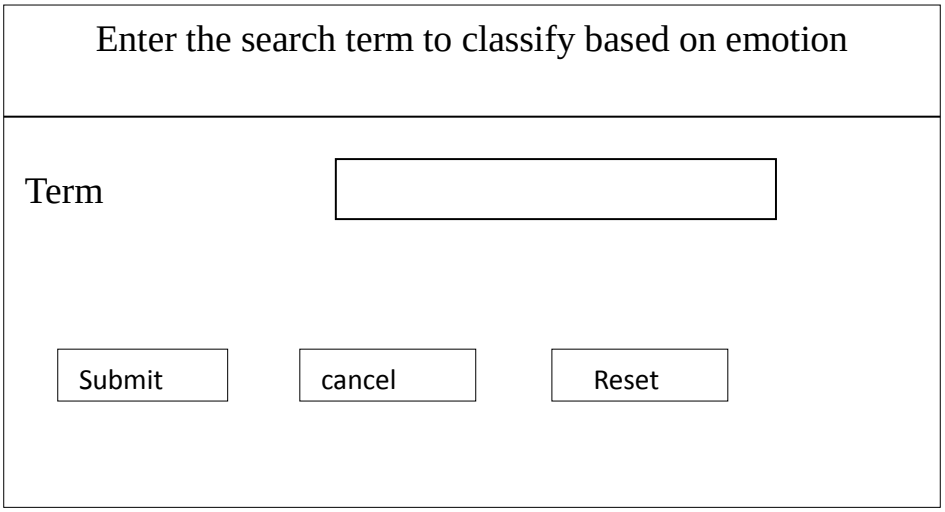

Fig 3.10: Search Term to Classify Based on Emotion

Search Term To Classify Based On Polarity :

The search term to classify the tweets based on polarity is entered in order to collect tweets and classify accordingly

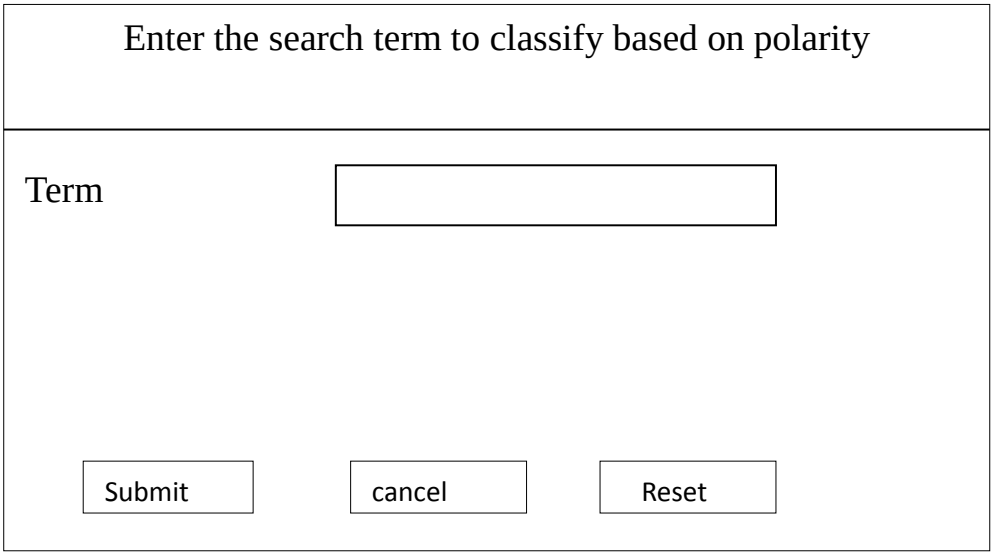

Fig 3.11: Search Term to Classify based on Polarity

Comparison of Two Profiles :

To compare the profiles of two twitter users enter the usernames of both the profiles and click submit.

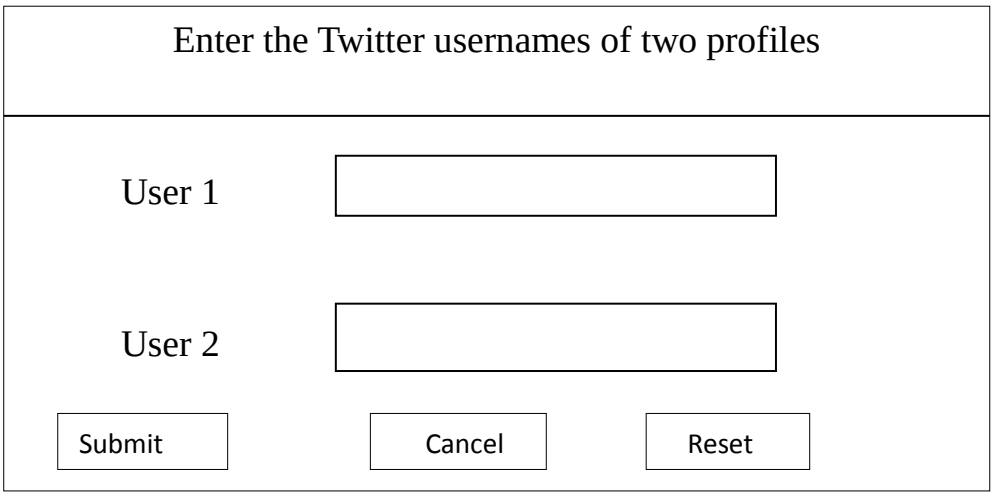

Fig 3.12: Comparison of two profiles

3.5.3 Objects & Actions.

Selection Of Process

The GUI chooser which is shown in fig 3.7 is used to get the selection process of desire option through radio button, from this user can go for any one of the selection at a time, on clicking OK button the corresponding operation is selected

Search Word

If the user selects choice 1 from the GUI chooser, search words has to be entered. In the text box to find that words in posts. This process is shown in fig 3.8.

Username

In fig 3.9, the user enters the username of a particular person, or community to get the related posts. If user wants to reset the text button then “reset“ button is entered.

Search Term to Classify Based on Emotion :

In fig 3.10 user enters some keywords, then the tweets containing that term is classified according to emotion.

Search Term To Classify Based On Polarity :

In fig 3.11 user enters some keywords, then the tweets containing that term is classified according to polarity.

Compare Tweets

In fig 3.12 the user has to enter usersnames of any two persons whose tweets are going to be compared. Clicking ok, posts are searched in Twitter for the comparison.

CHAPTER 4

IMPLEMENTATION

4. IMPLEMENTATION

Implementation is the stage of the project when the theoretical design is turned out into working system. Thus it can be considered to be the most critical stage in achieving a successful new system and in giving the user, confidence that the new system will work and be effective. The implementation stage involves careful planning, investing of the existing system and its constraints on implementation, designing of methods to achieve changeover and evaluation of changeover methods.

Implementation is the process of converting a new system design into operation. To implement the concept of social networking mining, twitter mining is considered in this project. The tweets are collected, preprocessed and classified using Bayes algorithm.

The Risk Assumption here is that, to run the code successfully needs internet connection to collect tweets from twitter.

DATA COLLECTION

Collection of data involves connecting to the twitter via twitter API by loading the package. Tweets and trends are then collected and stored in data frame once this process is over.

```
library(twitteR)

library(RJSONIO)

library(rjson)
```

The above two libraries are required in order to connect to twitter API and to collect the JSON results that is the JavaScript Object Notation results. Once the connection is established to twitter API , the required methods based

on the operation selected by the user is invoked , they are searchTwitter , userTimeline and getTrends

tweets = searchTwitter (searchterm)

User_tweets = userTimeLine (username)

Trends = getTrends(period)

DATA PREPROCESSING

Processing of data involves conversion of tweets into R data frames and storing in nosql . Initially the dumping twitter data into data frame using the method twListTodF has to be done. Getting text from a tweet, getting text from a tweet ,getting screen name of the publisher of the tweet, getting text from a tweet, getting screen name of the publisher of the tweet, getting screen name of the publisher of the tweet.

tweets_df = twListToDF(tweets)

#Transformation of data includes stripping white spaces

mycorpus_without_whitespaces = tm_map(mycorpus , stripWhitespace)

#Conversion of all the text into lower case

mycorpus_in_lowercase = tm_map(mycorpus , tolower)

#Removing of stopwords given in stopwords list

mycorpus_without_stopwords = (mycorpus ,removeWords , stopwords)

#Removing punctuations

mycorpus_without_punctuation = (mycorpus ,removePunctuation)

mycorpus_without_numbers = (mycorpus ,removeNumbers)

MINING THE DATA

Mining the data i.e processed tweets includes construction of lexical corpus ,applying appropriate algorithm and displaying the output of mining .

```
sapply (tweets , function(x) , x$getText())

#Getting screen name

sapply (tweets , function(x) , x$getScreenName())

#Getting user id

sapply (tweets , function(x) , x$getId())

#Construction of Lexical Corpus

myCorpus = Corpus(VectorSource(object))

#Getting characters per tweet

chars_per_tweet = sapply(results , nchar)

Summary(chars_per_tweet)

#Splitting words

words_list = strsplit  (results , " ")

Applying algorithm

# removeretweet entities

some_txt = gsub("(RT|via)((?:\\b\\W*@\\w+)+)", "", some_txt)

# remove at people

some_txt = gsub("@\\w+", "", some_txt)

some_txt = gsub("[[:punct:]]", "", some_txt)
```

```
# remove numbers

some_txt = gsub("[[:digit:]]", "", some_txt)

# remove html links

some_txt = gsub("http\\w+", "", some_txt)

# remove unnecessary spaces

some_txt = gsub("[ \t]{2,}", "", some_txt)

some_txt = gsub("^\\s+|\\s+$", "", some_txt)

# define "tolower error handling" function

tryTolower = function(x)

{

# create missing value

y = NA

# tryCatch error

try_error = tryCatch(tolower(x), error=function(e) e)

# if not an error

if (!inherits(try_error, "error"))

y = tolower(x)

# result

return(y)

}
```

Now in order to handle errors that might occur , exception handling is implemented via the method tryTolower, which would eventually handles all the exception that arises during conversion of words into lower case.

```
some_txt = sapply(some_txt, tryTolower)

# remove NAs in some_txt

some_txt = some_txt[!is.na(some_txt)]

names(some_txt) = NULL

# Perform Sentiment Analysis

# classify emotion

class_emo = classify_emotion(some_txt, algorithm="bayes", prior=1.0)

# get emotion best fit

emotion = class_emo[,7]

# substitute NA's by "unknown"

emotion[is.na(emotion)] = "unknown"

# classify polarity

class_pol = classify_polarity(some_txt, algorithm="bayes")

# get polarity best fit

polarity = class_pol[,4]

# data frame with results

sent_df = data.frame(text=some_txt, emotion=emotion,

polarity=polarity, stringsAsFactors=FALSE)
```

```
# sort data frame

sent_df = within(sent_df,

emotion<- factor(emotion, levels=names(sort(table(emotion),
decreasing=TRUE))))
```

Visualization of the analysis is represented in the form of graph here . The code involved for gpplot mechanism for plotting bar chart is as follows.

```
ggplot(sent_df, aes(x=emotion)) +

geom_bar(aes(y=..count.., fill=emotion)) +

scale_fill_brewer(palette="Dark2") +

labs(x="emotion categories", y="number of tweets") +

opts(title = "Sentiment Analysis of Tweets \n(classification by emotion)",

plot.title = theme_text(size=12))
```

CHAPTER 5

TEST PLAN AND TESTING

5. TEST PLAN AND TESTING

Test Plan

Test plan contains the various testing by analyzing the systems each and every module of the project. Test plan is the process of analyzing the system for any possible errors of all kinds. A good test case is the one that has a high probability of finding an undiscovered error.

Testing Objectives

Testing contains various testing phases and methods by analyzing the systems, each and every module of the project. Testing majorly concentrates on the logical errors and the flow of the modules. This will be done individually to every module and then tested for the whole system. The kinds of tests to be performed are unit testing and integration testing. Each testing will be having separate strategies.

5.1 UNIT TESTING

Unit testing involves independent analysis of the project in parts means as modules. The main advantage of this testing approach is that its high level of accuracy in finding errors for each unit. In unit testing is for each and every module. Each module depends on the other and requires input from other modules. So all are combined and used by the main module.

The main functions in modules are taken and the sample input for them is provided. This will lead to an output which is compared to the expected output. Then the result will show whether the function is working well or not. The status of the functional output and the error message if any tabulated. The results of test reports are taken into consideration for the subsequent modifications.

Data Collection

S. No.	Function	Sample input	Expected output	Actual output	Status	Error message
1	Connect to twitter	Twitter API Auth ID	Connection established	Can’t connect to twitter	Failure	No network
2	Collect twitter	Keywords	Tweets	Tweets	success	

Table 5.1: Unit Testing-Data Collection

Data Preprocessing

S. No	Function name	Sample input	Expected output	Actual output	Status	Error message
1	Remove white space	Tweet	Modified tweet	Modified tweet	success	
2	Remove retweets	Tweet	Modified tweet	Modified tweet	success	
3	Remove stop words	Tweet	Modified tweet	Modified tweet	success	

Table 5.2: Unit Testing- Data Preprocess

5.2 INTEGRATION TESTING

After finishing the unit testing process, next is integration testing process. In this testing process we put we put our focus on identifying the interfaces between components and their functionality as dictated by the DFD diagram. The Bottom up incremental approach was adopted during the testing.

S.No.	Function	Description	Input	Expected Output	Implementation	Status
1	Data loading	Updating in database	Twitter link	updated	YES	success
2	Data preprocessing	Preprocess the tweets	Tweets	Preprocessed tweets	NO	Failure
3	Mining	User view graph	Modified data	Classified data	YES	Success

Table 5.3: Integration Testing

Chapter 6

RESULTS AND OBSERVATION

6. RESULTS AND OBSERVATION

The Result for social network mining and analysis using sentimental mining and classification are displayed in graphical and word cloud representations for easy visualization . Powerful observations can be done through the outputs which is the first step towards predictive analysis.

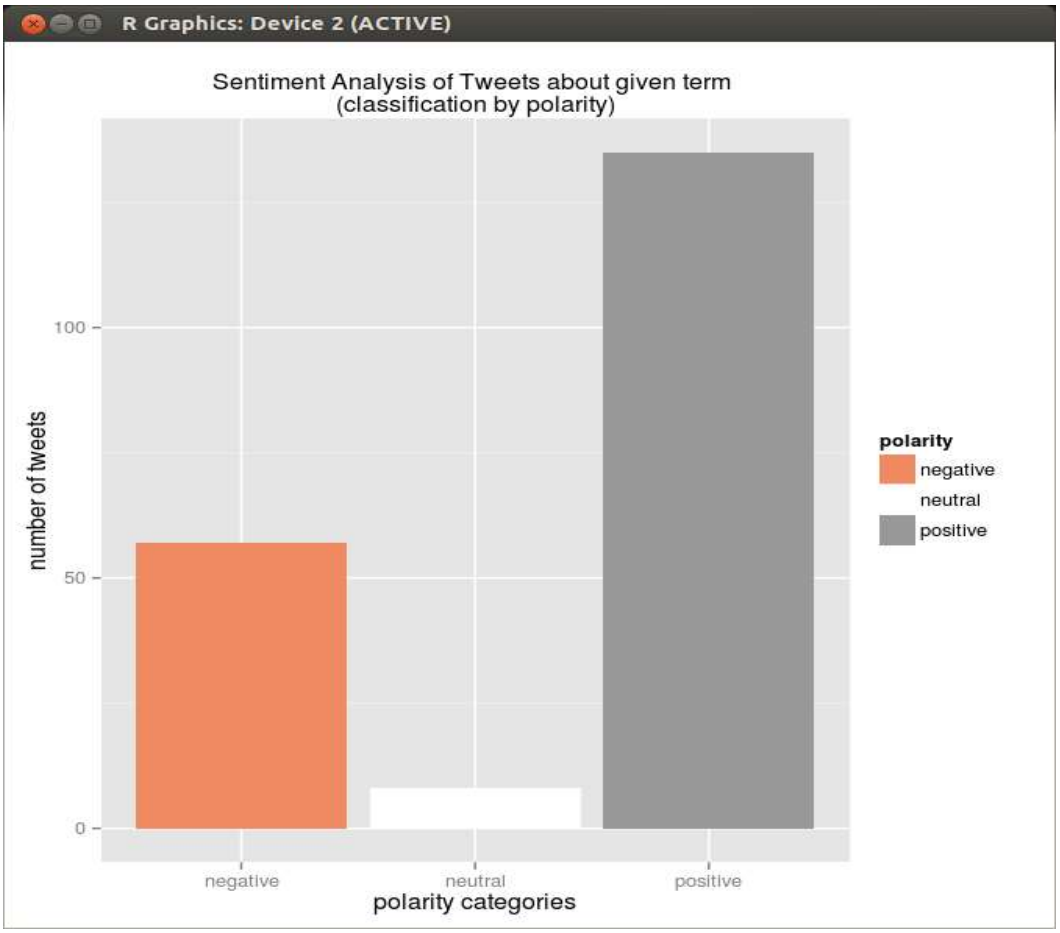

Table 6.1 Output

Chapter 7

CONCLUSION AND FUTURE WORK

7. CONCLUSION AND FUTURE WORK

7.1 CONCLUSION

Visual representation of social networks is important to understand the network data and convey the result of the analysis. Many of the analytic software have modules for network visualization. Exploration of the data is done through displaying nodes and ties in various layouts, and attributing colors, size and other advanced properties to nodes. Visual representations of networks may be a powerful method for conveying complex information. Usage of sentimental analysis and polarity and emotions makes better understanding of user’s perception and views.

7.2 FUTURE WORK

Social network mining has been implemented via sentimental mining and classification only to twitter , the same can be applied to other social networks as well . And integration of all the results of social networks and can be stored in graphical database like Neo4j which can easily depict social relations and can be accessed through cipher queries .

APPENDIX

APPENDIX

A1 SELECTION OF PROCESS

Fig A1: Selection of Process

In home page select the process of sentiment mining in twitter, click your choice in any one of the radio buttons. The options given using radio buttons are search term in tweets, display tweets of particular user , top trends of this hour , views on given term (by emotion), views on given term (by polarity), comparison of tweets of two person . The user has to select any of the choice by clicking the corresponding radio button and then click OK.

A2 ENTER THE SEARCH TERM

Fig A2: Enter the Search Term

Enter the search term to search in Twitter .

A3 ENTER THE SEARCH TERM TO CLASSIFY BASED ON POLARITY

Fig A3: Enter Term for Polarity Classification

Enter the search term to classify based on polarity, which is based on positive and negative word.

A4 ENTER THE USERNAME TO SEARCH

Fig A4: Enter Twitter Username

To search the particular user tweets , type the user name of the twitter user .

A5 OUTPUT OF VIEWS ON GIVEN TERM(EMOTION)

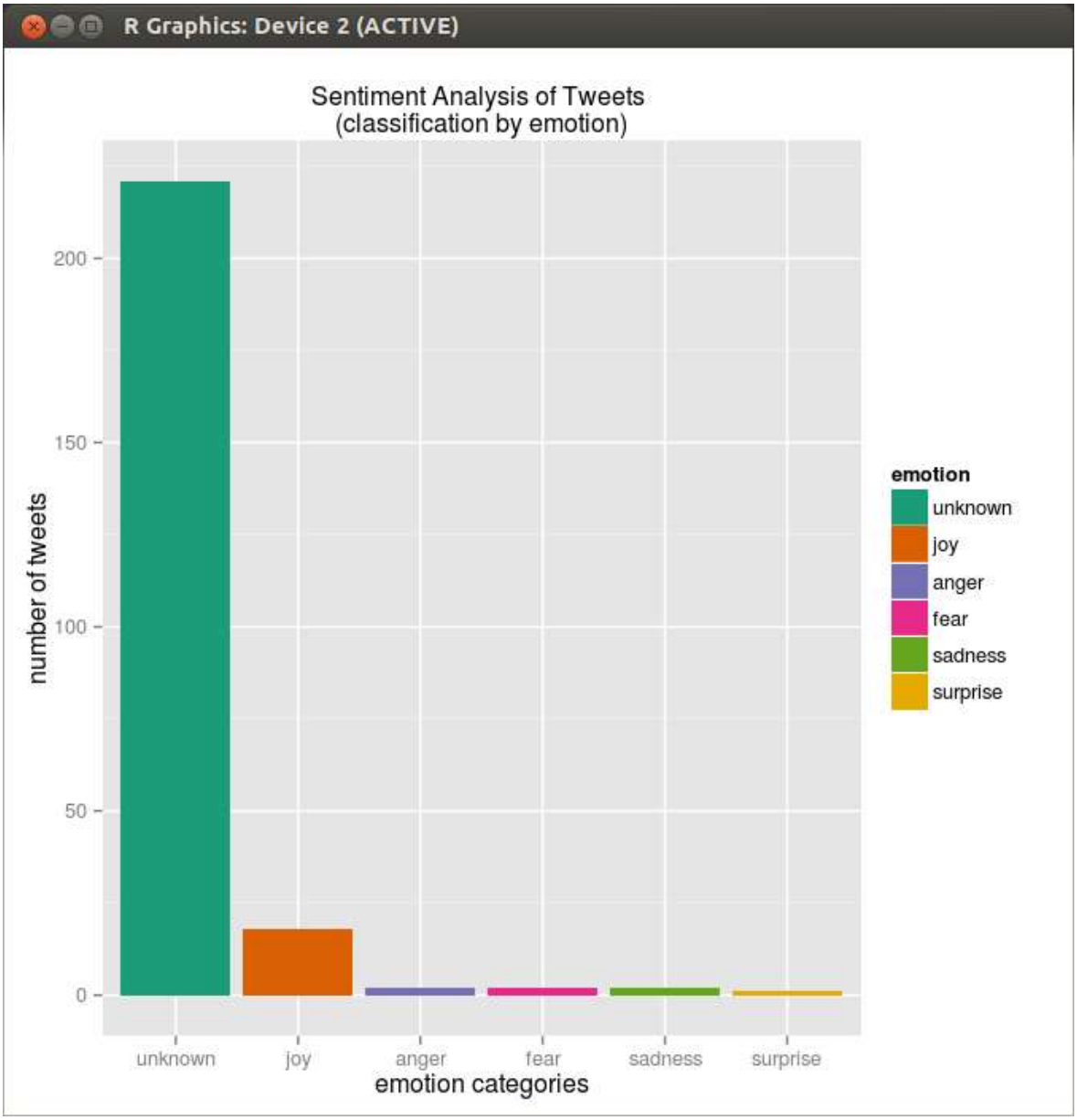

Fig A 5: Output Based on Emotion

The figure A5 shows the output of emotion classification of tweets in sentiment mining. Here the emotions joy, anger, fear, sadness, surprise and unknown emotions are classified and depicted as graph.

A6 OUTPUT OF VIEWS ON GIVEN TERM (POLARITY)

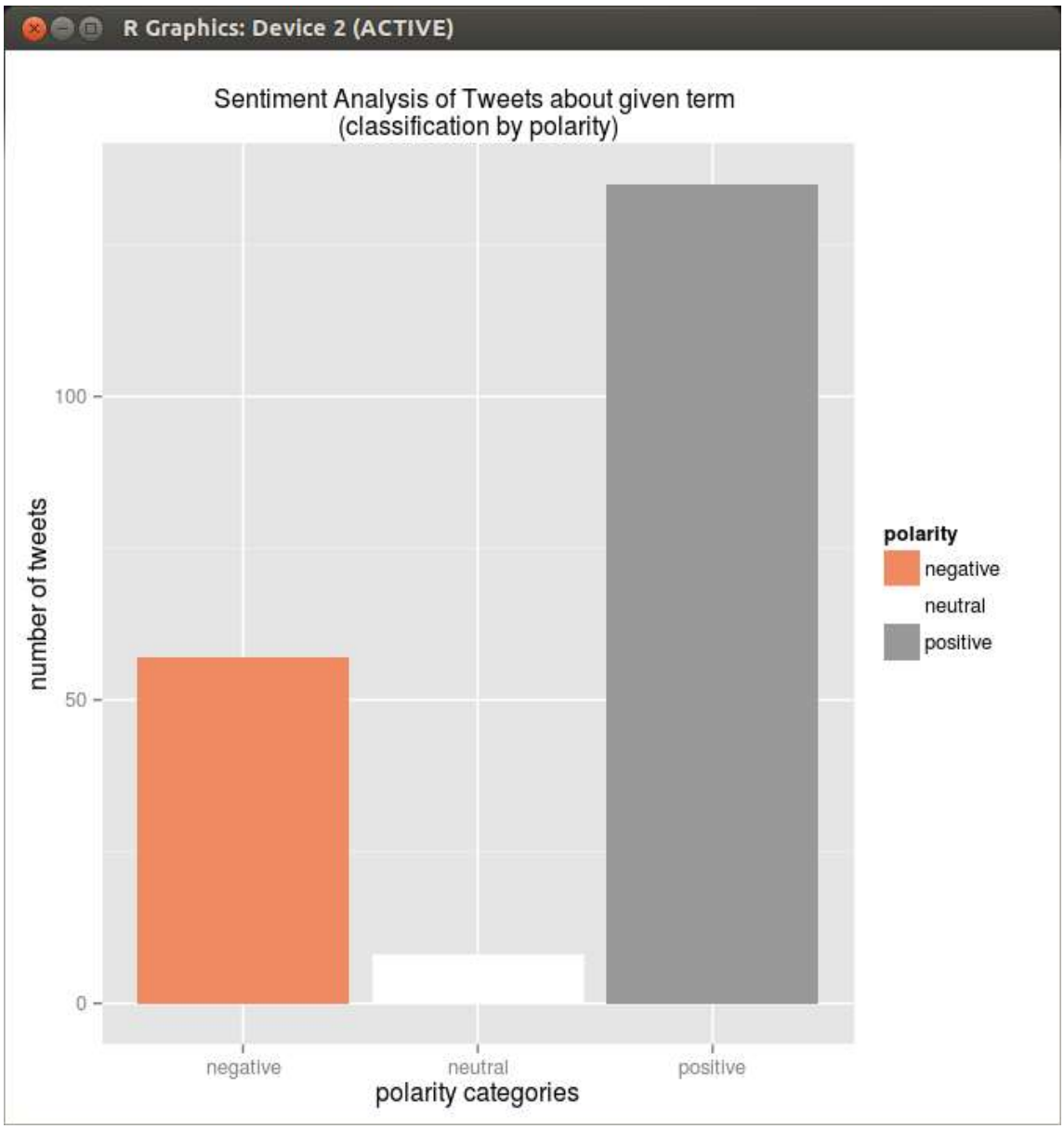

Figure A6: Output Based on Polarity

The figure shows the output of tweets classification based on polarity. For classification, the polarity categories considered are negative, positive and neutral.

A7 OUTPUT OF COMPARISON TWEETS OF TWO USERS

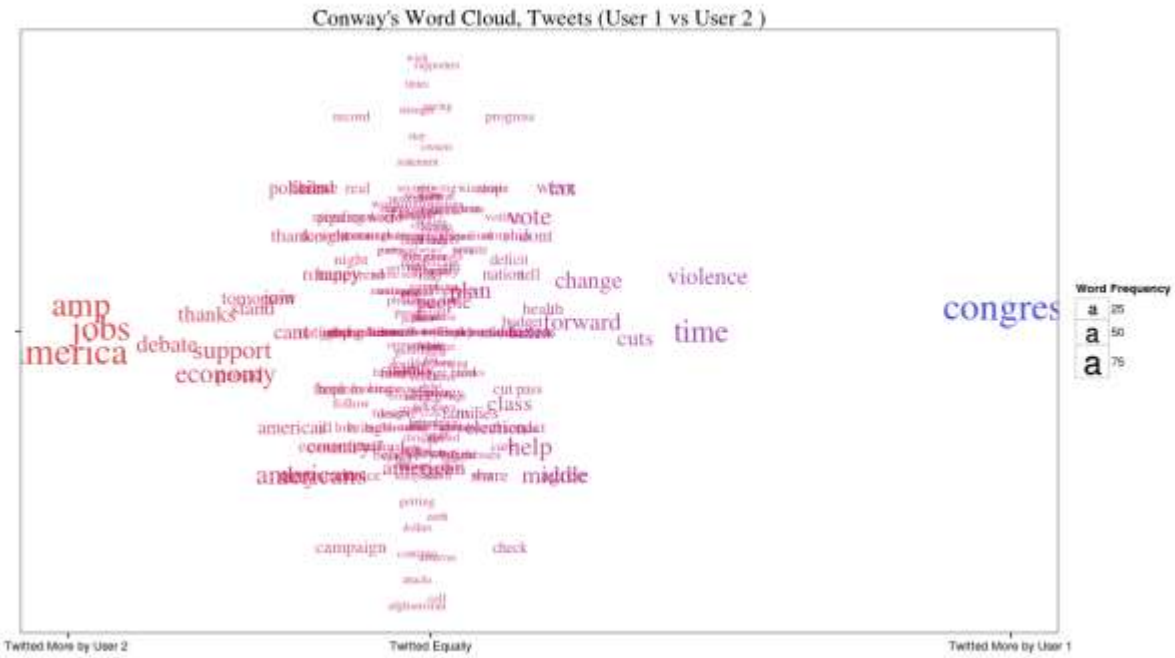

Fig A 7: Output Based on Comparison of two users

The figure shows the output which is a word cloud of two twitter users based on the frequency of words used in tweets. Tweets which are used equally by two users are in middle of word cloud, tweets which are twitted frequently by the user1 is in right, tweets which are twitted frequently by user 2 is in left..Ref fig A7.

A8 OUTPUT OF TWEETS OF PARTICULAR USER

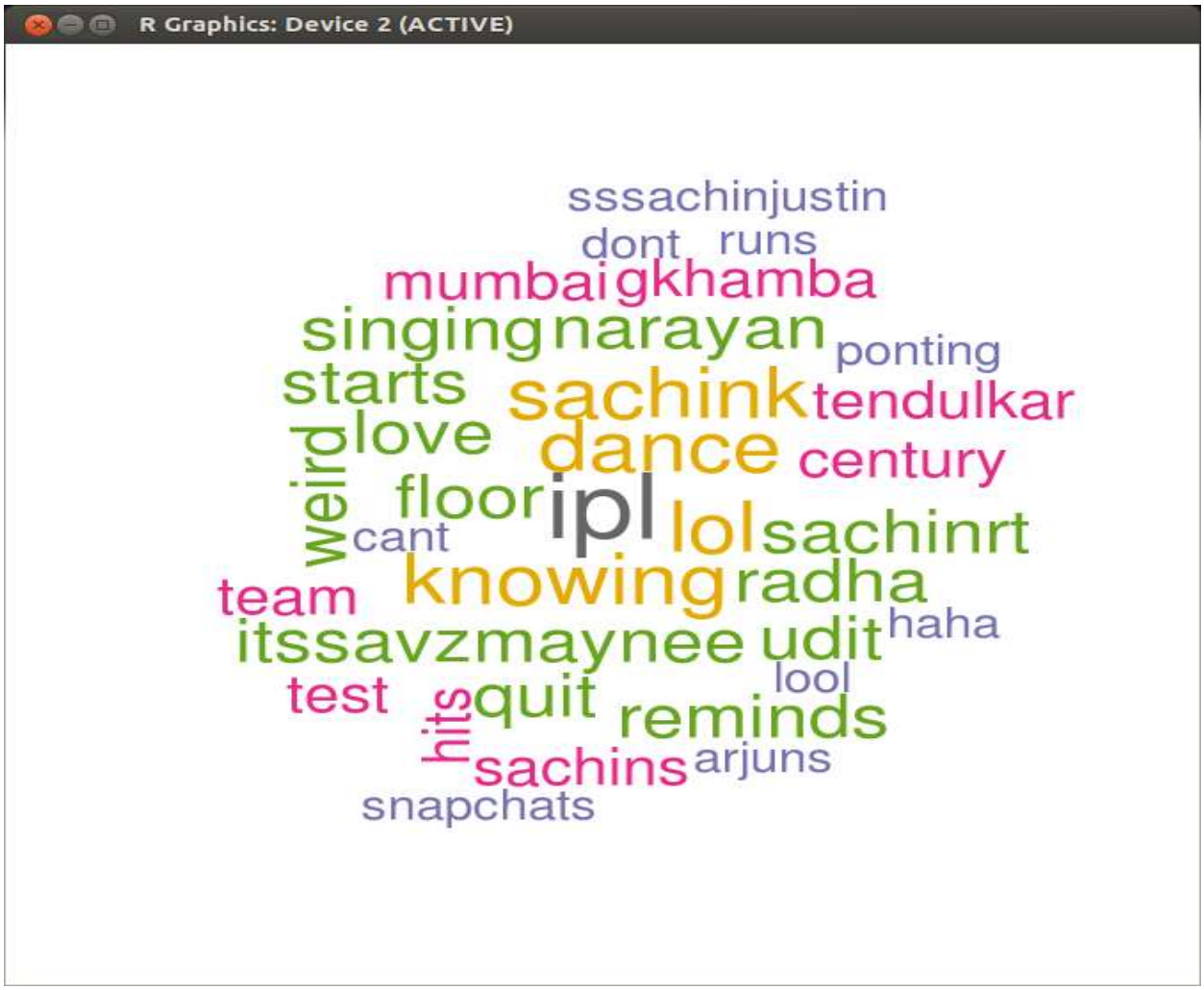

Fig A 8 : Output Based on Particular Use

Tweets of particular user is displayed as output. The output is displayed in the form of word cloud . Word cloud is the clustered way of displaying words. Different color and different font size refers to the difference in occurrence of words. Ref fig A 8.

A9 OUTPUT OF TRENDS OF AN HOUR

Figure A9: Output of Trends of an Hour

Displays the tweets which is frequent within an hour, displays top tweets based on term. Ref fig A 9.

REFERENCES

REFERENCES

1. Alexander Pak and Patrick Parobubek (2010), "***Twitter as a Corpus for Sentiment Analysis and Opinion Mining"***Universit´e de Paris-Sud, Laboratoire LIMSI-CNRS, Bˆatiment 508, F-91405 OrsayCedex, France.

2. EfthymiosKouloumpis , TheresaWilson and Johanna Moore (2011) ***,"Twitter Sentiment Analysis:The Good the Bad and the OMG"***,In Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media.

3. Dmitry Davidov,Oren Tsurand Ari Rappoport(2010), ***"Enhanced Sentiment Learning Using Twitter Hashtags and Smileys"***, CNC Institute of Computer ScienceThe Hebrew University.

4. David Jensen and Jennifer Neville(2010) ***,"Data Mining in Social Networks",*** Knowledge Discovery LaboratoryComputer Science Department, University of Massachusetts, Amherst, MA 01003.

5. Wil M.P. van der Aalst and Minseok Song(2010***),"Mining Social Networks: Uncovering InteractionPatterns in Business Processes",***Department of Technology Management, Eindhoven University of Technology, P.O.Box 513, NL-5600 MB, Eindhoven, The Netherlands.