

## **ABSTRACT**

In social media, information is present in large amount. Extracting information from social media gives us several usage in various fields. In the field of biomedical and healthcare, extracting information from social media is providing number of benefits such as knowledge about the latest technology, updates of current situation in medical field etc.

Twitter is one of the social media which allows the user post tweets of limited number of characters and share the tweet to their followers. Twitter allows application developer to access the tweets for their purpose.

In the implemented system, tweets are collected and sentiment analysis is performed on them. Based on the results of sentimental analysis few suggestions can be provided to the user. The implemented system can perform sentiment analysis on data available for patient and consumer opinions on various treatments and drugs. These results can provide up-to-date information for the hospitals, pharmaceutical industry, and medical staff, on the effectiveness or ineffectiveness of the treatment. In this way, the implemented system can help in improving the healthcare of diabetic patients.

# Table of Contents

ABSTRACT .....	ii
LIST OF FIGURES .....	iv
1. BACKGROUND AND RATIONALE.....	1
1.1 Existing Systems .....	3
1.2 Related Work.....	4
1.3 Introduction to Proposed System.....	5
2. NARRATIVE .....	7
2.1 Problem Statement.....	7
2.2 Motivation .....	7
2.3 Project Goal .....	8
2.4 Project Functionalities .....	8
3 PROPOSED SYSTEM DESIGN.....	9
3.1 System Design .....	9
3.2 Creating Twitter Application.....	12
3.3 Establishing Connections .....	14
3.4 Collecting Tweets .....	14
3.5 Parsing Tweets.....	14
3.6 Pre-processing: .....	15
3.7 Natural Language Processing .....	15
4 REQUIREMENTS.....	16
4.1 Environment .....	16
4.2 System Requirements .....	16
4.2.1 Software Requirements .....	16
4.2.2 Hardware Requirements .....	17
4.3 XAMPP .....	17
4.4 Netbeans IDE.....	17
4.5 Java Swing.....	18
4.6 Glassfish server.....	18
5 IMPLEMENTATION.....	19

5.1 Inductive Content Analysis .....	19
6 TESTING AND EVALUATION .....	22
6.1 Add Category.....	22
6.2 Add Keyword .....	23
6.3 Collect Tweets .....	24
6.4 Pre-process Tweets .....	25
6.5 Positive Tweets.....	26
6.6 Negative Tweets .....	27
6.7 Designing the Forum and Suggesting the Tweets .....	28
6.8 Home and About.....	29
6.9 Tweets and Suggestions.....	29
6.10 Subscription for Newsletter .....	30
6.11 Result without Suggestions .....	32
7 LIMITATIONS.....	33
8 CONCLUSION.....	34
9 FUTURE WORK.....	35
References.....	36

## LIST OF FIGURES

Figure 1.1 Overall Population Effected with Diabetes.....	1
Figure 1.2 Impact of Social Media .....	3
Figure 3.1 Project Architecture.....	10
Figure 3.2 Data Flow Diagram.....	11
Figure 3.3 Obtaining a Request Toke.....	12
Figure 3.4 Obtaining a Secret Token.....	13
Figure 3.5 Consumer Key and Consumer Secret Generated by Twitter.....	13
Figure 5.1 Creating Twitter Application.....	19
Figure 5.2 XAMPP Control Panel.....	20
Figure 5.3 Healthcare Database.....	21
Figure 5.4 Functionalities of the Application.....	21
Figure 6.1 Add Category.....	23
Figure 6.2 Add Keyword.....	23
Figure 6.3 Collecting Tweets from Twitter.....	24
Figure 6.4 Pre-Processed Tweets.....	25
Figure 6.5 List of Positive Tweets.....	26
Figure 6.6 List of Negative Tweets.....	27
Figure 6.8 Home Page of Web Application.....	28
Figure 6.9 Retrieving Results and Sugsesstions.....	29
Figure 6.10 Subscribing for Newslettter.....	30
Figure 6.11 Partial Output.....	31

# 1. BACKGROUND AND RATIONALE

Currently diabetes is the fourth biggest cause of death in the world [1]. According to the International Diabetes Federation (IDF), 30 million people had diabetes in 1985, by 2000, the numbers revised to 150 million. Currently over 300 million people in the world have diabetes and if the current rate of growth continues unchecked then the number will exceed 435 million by 2030 [2]. As shown in Figure 1.1, 0.6 million, 0.8 million and 1.2 million people died because of diabetes in 1985, 2000 and 2015 respectively.

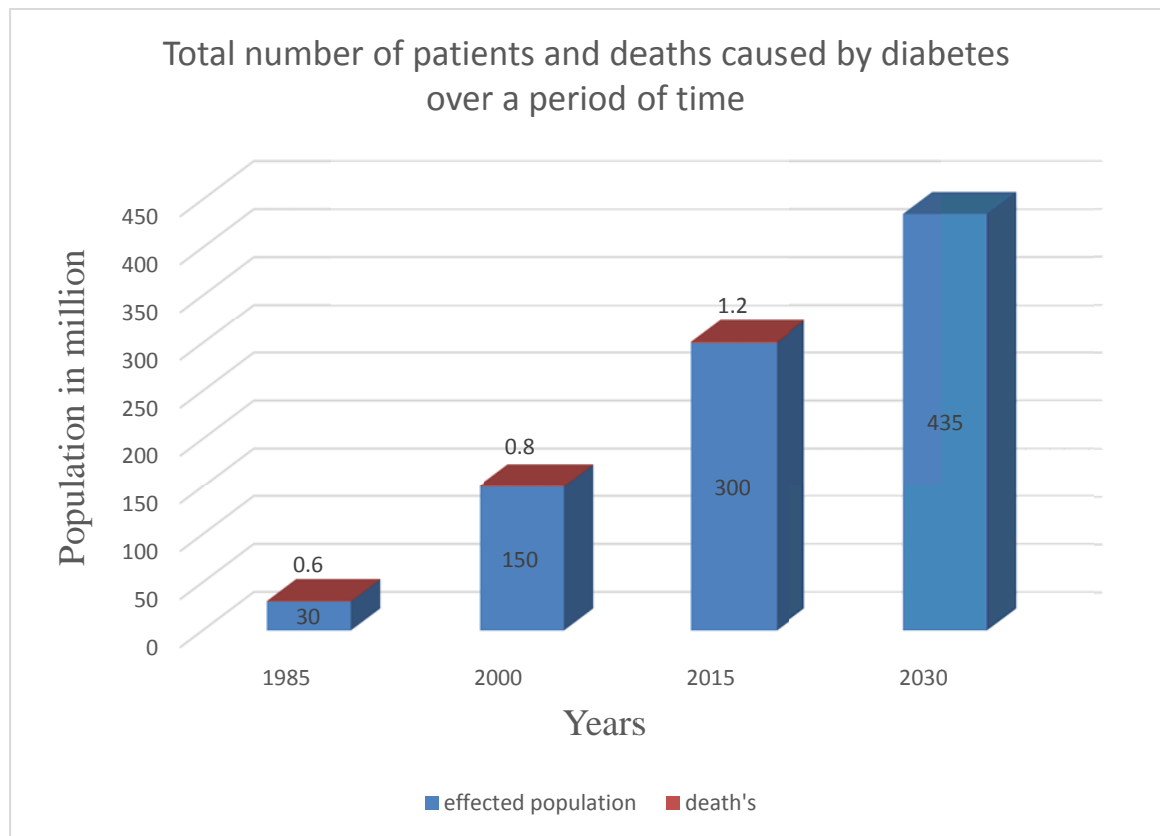


Figure 2.1 Overall Population Effected with Diabetes [2]

Top five countries with the highest number of diabetes sufferers are: India with 50.8 million, china with 43.2 million, USA with 26.8 million, Russia with 9.6 million and Brazil with 7.6 million [2]. Awareness about diabetes, its symptoms, precaution and treatment are very necessary. Healthcare providers could use patient opinion to improve their services. Physicians could collect feedback from other doctors and patients to improve their treatment recommendations and results. Patients could use other consumers' knowledge in making better-informed healthcare decisions. But these methods are time consuming and are not much effective. How to spread this knowledge in an effective and efficient way? Social media is the solution to this problem.

Social media breaks the boundaries between the real world and the virtual world. Social media enables a virtual networking environment [3]. Social media is becoming one of the most comfortable medium for people to share their thoughts and feelings indiscriminately. According to a survey social sites like facebook and twitter are highly used for sharing of data [4]. Figure 1.2 shows the share of different social sites on the web. “Sentiment analysis of data acquired from social media” is the process of collecting, analyzing, and extracting actionable patterns from social media data” [5]. Modeling social media using available network modeling and computational tools is one way of extracting knowledge from the ‘cloud:’ a social network is a structure made of nodes and edges that connect nodes in various relationships [6] [7].

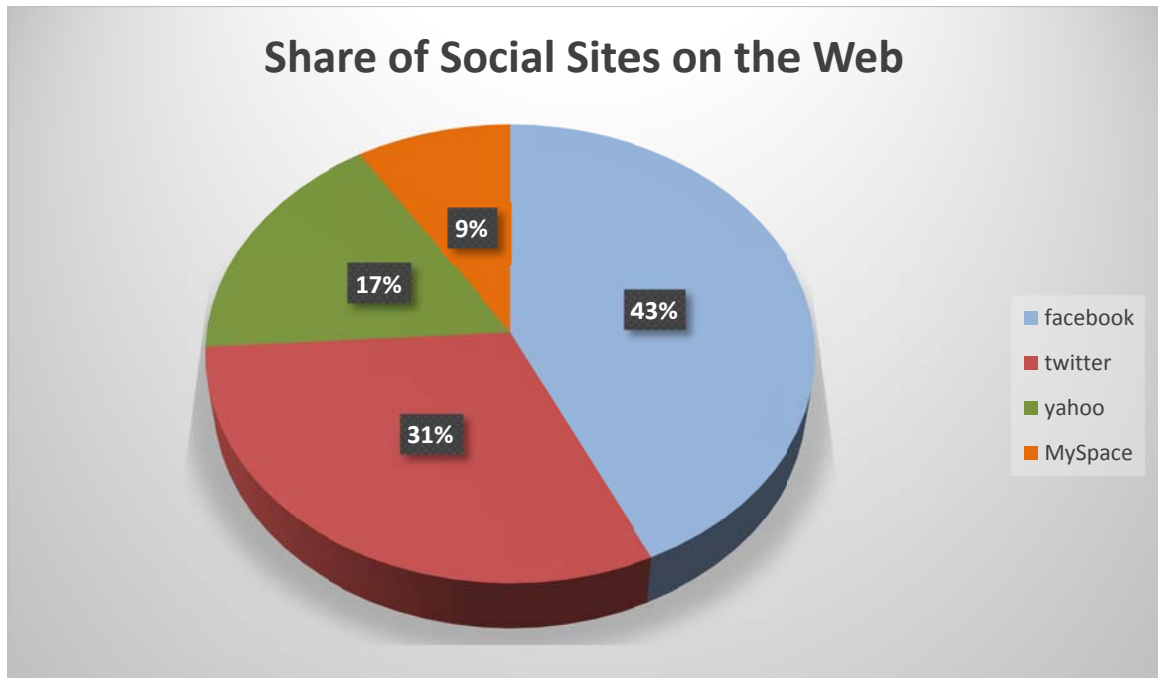


Figure 1.2 Impact of Social Media [4]

## 1.1 Existing Systems

There are many traditional methods which provide the benefit of having knowledge and update on latest technology in medical field. Some of the methods are discussed below:

- **Surveys and Questionnaires**

A set of questions with a choice of answers, devised for healthcare surveys helps a lot to gain knowledge on effectiveness of the treatment. There are various surveys such as patient experience survey which captures every patient's voice using email, phone and mail to provide a deeper data and a clearer picture of patient perception for the entire care. Employee and physician surveys are of equal importance to improve the quality of care.

- **Interviews**

Watching interviews of doctors, patients and drug developers helps in being update with latest technology and effectiveness of the treatments available.

- **Feedback**

Now-a-days many pharmaceutical companies are asking for consumer feedback. These pharmaceutical companies make record of both negative and positive feedback and use these records to come up with a better optimized drugs.

However, the existing traditional methods are time consuming and as they are performed manually, their efficiency decreases with an increase in the frequency. Another important problem is that a patient may not convey information properly in a formal setting when compared to an informal medium like social media.

## **1.2 Related Work**

Previous studies came up with technical solutions to extract user sentiment on government health monitoring system [8], influenza [9]. Despite of the extensive literature, none of the solutions have how forum relationships affect network dynamics. In the proposed system, sentiment analysis is done using Natural Language Processing, which defines a relation between user posted tweets and opinion on the drug, and in addition, suggestions of much better medicines can be provided to the users. Thereby the drawback of identifying a relation between user post and the opinion associated with the post is overcome.



### 1.3 Introduction to Proposed System

In the proposed system, searching the information based on category and keywords from the twitter database is performed. Searching keywords in twitter is one of the hardest tasks because of the diversity of the language and the slangs used on the internet. In the proposed system, the first step involves collection of tweets from twitter and making it as a data set, the second step is preprocessing of the related tweets. In the third step, sentiment analysis is performed using the Natural Language Processing (NLP) algorithm, which is based on numerical statistics [10]. Assigned sentiment value using NLP, is used as a weighting factor in sentiment analysis [11]. In the fourth step, similar data is identified and analyzed, then by using a web application, the final results, which are suggestions of other treatments available for the health issue, is provided. The logics that has been used in the proposed system has the following major steps:

- Collecting tweets
- Pre-processing tweets
- Sentiment analysis
- Suggestion of other treatment

First, a connection is established between a twitter account and the twitter database, and the tweets are collected based on the combination of Keyword and Category provided by the user. In the next step, all the tweets are pre-processed for unwanted words, symbols and characters. Pre-processing consists of three steps which are as follows:

- Removing common stop words and misspelled words.
- Removing numbers, symbols and special characters.
- Converting upper case letters to lower case letters

In the sentiment analysis, the NLP analyzes the sentiment of the collected tweets by performing by the following steps:

- It first performs tokenization.
- Then it performs sentence splitting known as ssplit.
- Next step is to parse the sentence for syntactic analysis.
- Finally it decideds the sentiment value of the tweet based on the results of the above steps [12].

The final step is to design a web forum for providing final results to the users and suggest few other treatment or medicine available for the disease. Users can even subscribe to this forum to get newsletters for latest medicines and its effectiveness. Steps involved in this process are:

- Get the positive tweets from the sentiment analysis result.
- Develop a value comparator logic and apply it to the collected positive tweets, which provides the list of medicines recommended by maximum number of users.

## **2. NARRATIVE**

### **2.1 Problem Statement**

Methods used for improving healthcare include questionnaires, surveys, interviews, and feedbacks about the effectiveness of the treatment and latest technologies. But these methods are formal and time consuming. One of the major disadvantage is - as the frequency increases its effectiveness decreases. Existing sentiment analysis [13] [14] systems are unable to identify the influential user, and how the relationship affect network dynamics [15].

### **2.2 Motivation**

There are 196 countries on our planet, and each country has its own system for improving healthcare of its own citizens. But it is not necessary to study about 196 different systems to get a clear picture on how all countries manage healthcare systems, effectiveness of the current treatments, and updates on the latest technologies. Social media provides a better environment for sharing user experience in an interactive and informal way, so most of the patients, doctors and pharmaceutical companies show interest to post content what they really feel. Twitter is one of the highly used social media. Users in twitter post their thoughts and share their favorite tweets that are posted in the twitter. Twitter allows the developer to access the tweets of the public users. And also twitter allows the developer to access the tweets based upon the queries to collect the list of tweet based upon the query search. In our case, tweets are collected based upon the terms and most used words of medical industry to retrieve the tweets from the twitter.

## **2.3 Project Goal**

The main goal of the proposed system is to perform sentiment analysis on the tweets collected from twitter in order to improve the healthcare of the diabetic affected people. The tweets are filtered on the basis of category and keyword to obtain related tweets for the implemented project. By working with the implemented system, a user can get the resource of social media for the corresponding field of healthcare.

## **2.4 Project Functionalities**

The functionalities of the proposed system are very user friendly and attractive. Some of the functionalities are as follows:

- Add category
- Add keyword
- Select File
- Collect tweets
- Pre-process tweets
- Analysis
- Web Application
- Providing the resultant tweets
- Providing suggestions

### **3 PROPOSED SYSTEM DESIGN**

As twitter provides free APIs, it will be easy to collect and analyze the data using twitter. The implemented system is developed by using Java Swing as front end and MySQL as backend. The system is developed using Intel i3 processor, 2 GB RAM and 256GB Hard disk drive.

#### **3.1 System Design**

Figure 3.1 shows the basic architectural diagram of the implemented system. Basically it consists three modules, they are:

- Twitter application and twitter database
- Sentiment analysis process
- Web application

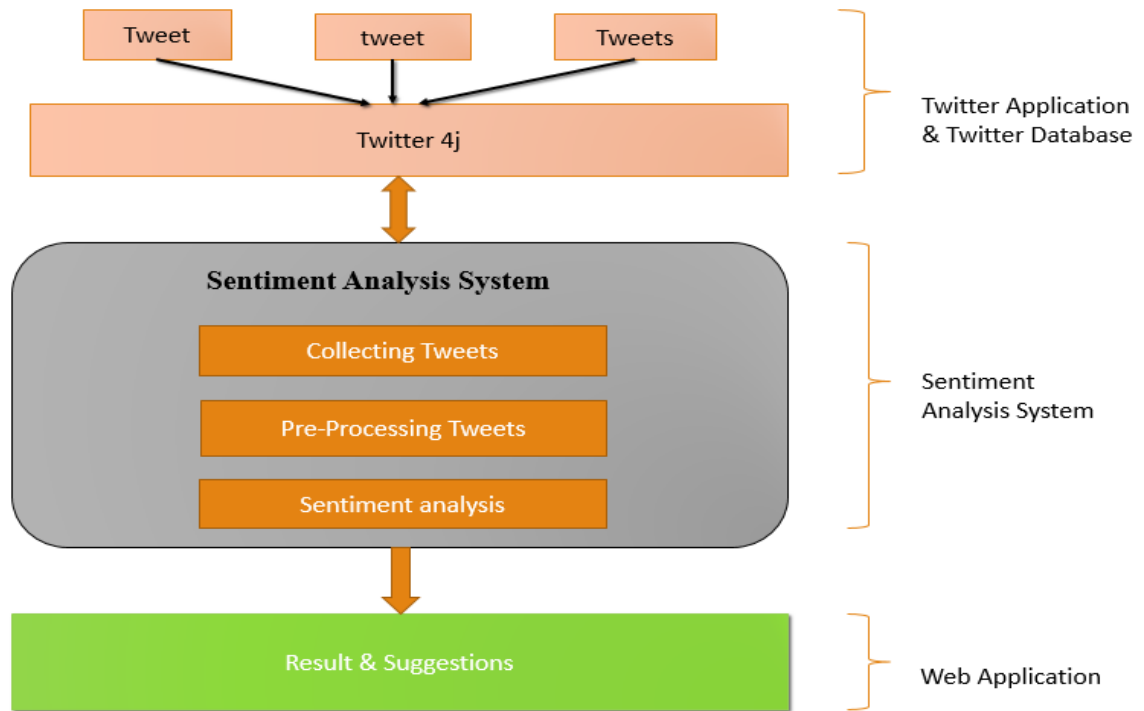


Figure 3.1 Project Architecture

Figure 3.2 shows the dataflow diagram of the implemented system. At first, a twitter application is created and tweets are collected from the twitter database. Collected tweets are stored as data set and is pre-processed and parsed by removing common unwanted words, symbols, characters, numbers and converts the upper case letters to lower case letters. After pre-processing, the sentiments will be analyzed by using Natural language processing tool. Each sentence is provided with sentiment value, based on this sentiment value the data is cataloged as positive or negative. Both positive and negative data are analyzed and similar data are identified. Then by using a web application, the result are displayed to the users. In addition, users are provided with few suggestions.

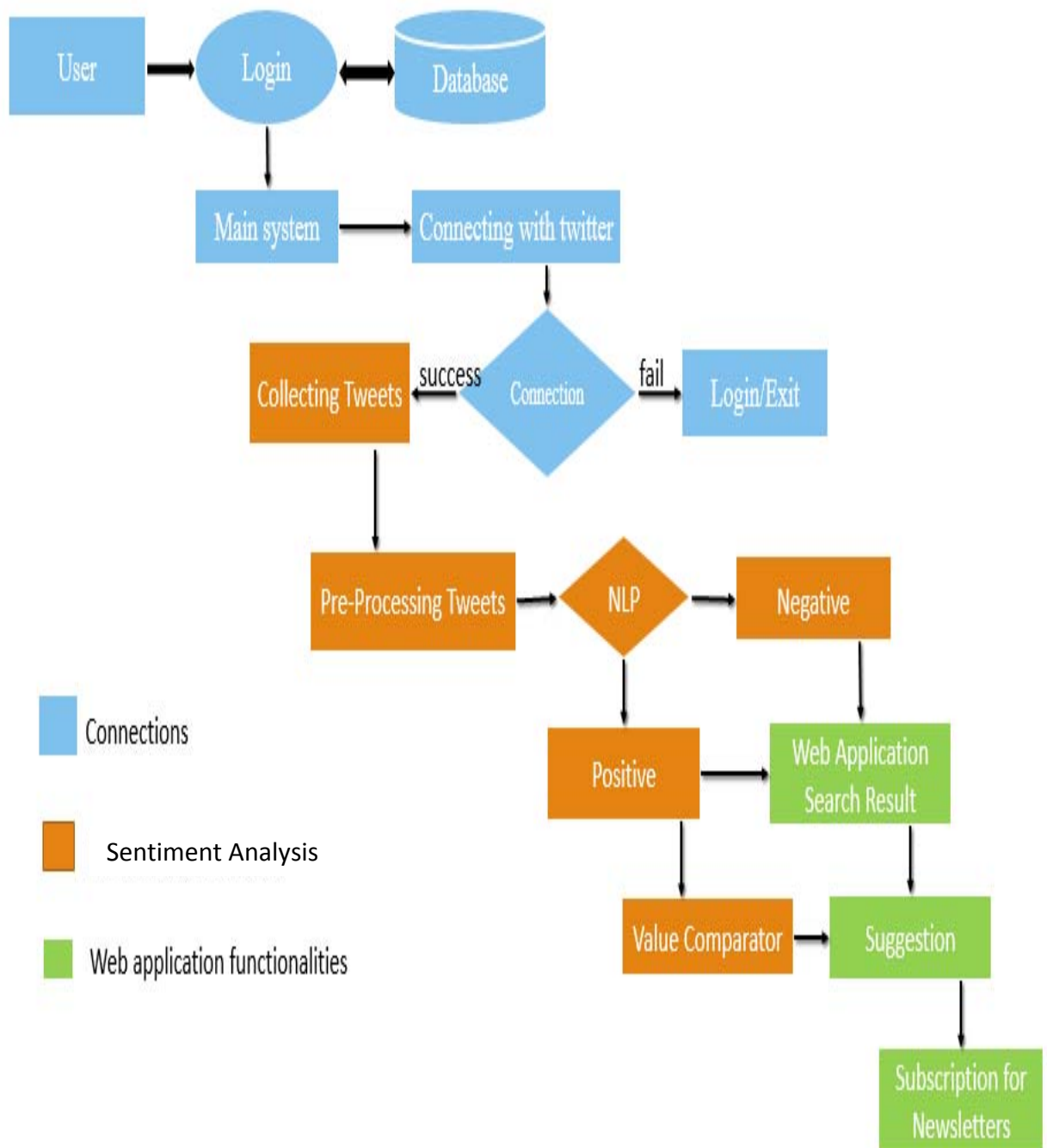


Figure 3.2 Data Flow Diagram

## 3.2 Creating Twitter Application

On an average, 6,000 tweets are tweeted on twitter for every second, which implies 360,000 tweets are tweeted per minute and 500 million tweets per day [16]. To retrieve tweets from twitter, we need to create a twitter application. Twitter application can be created by login into user's twitter account and by visiting the website <https://apps.twitter.com>. Once the twitter application is created, we need to obtain a request token. As shown in Figure 3.2, the request token can be obtained by sending a signed message to POST oauth/request\_token.

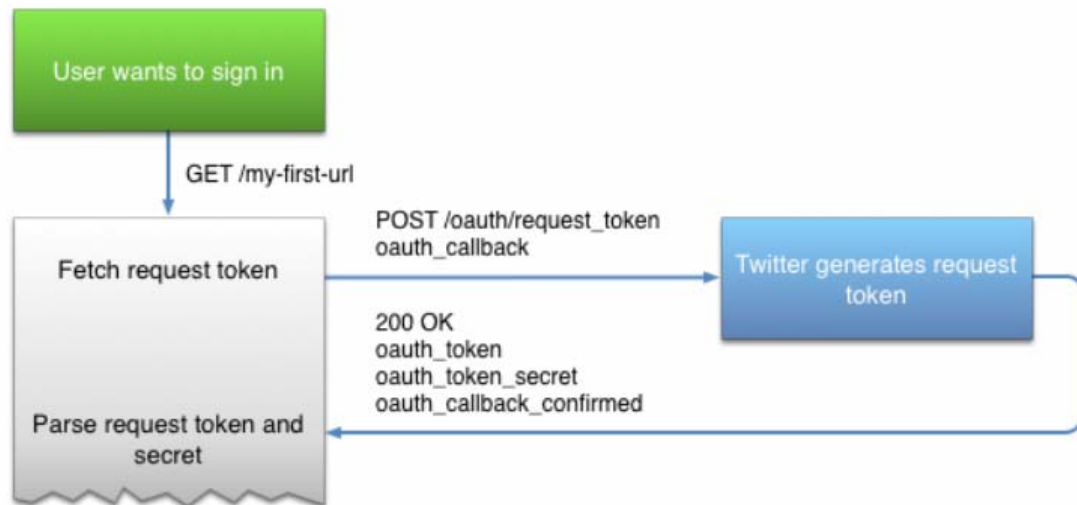


Figure 3.2 Obtaining a Request Token

The next step is to direct the user to GET oauth/authenticate, where the obtained request token is passed as oauth\_token parameter. User receives a oauth\_verifier value from the second step. This oauth\_verifier value is passed in the next step to get a access token.



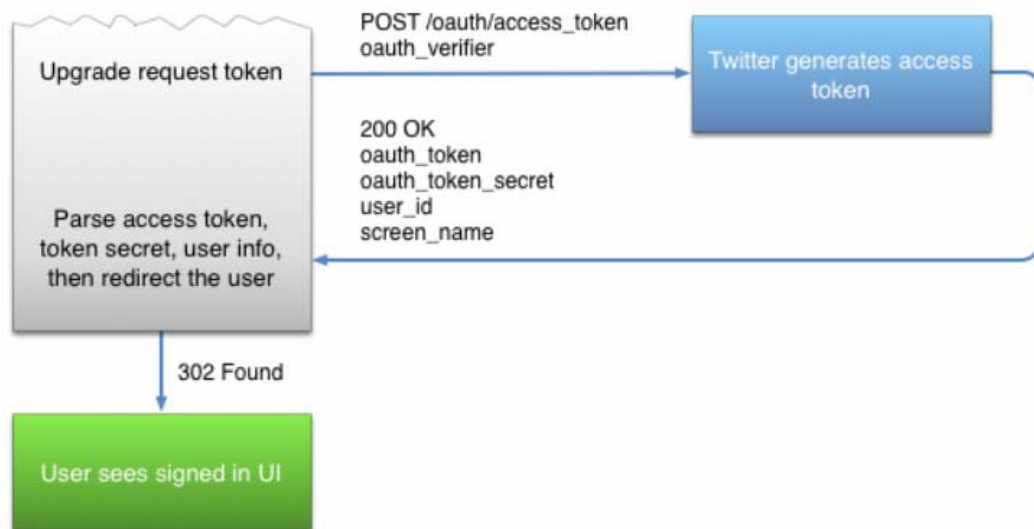


Figure 3.3 Obtaining a Secret Token

As shown in Figure 3.3, a successful response generate secret tokens which helps in having access to the twitter database. Figure 3.4 shows the consumer key and consumer secret generated by twitter for the application which acts as credentials for the created application.

Application Management

---

## HealthCare by Param

[Details](#)
[Settings](#)
[Keys and Access Tokens](#)
[Permissions](#)

Test OAuth

### Application Settings

*Keep the "Consumer Secret" a secret. This key should never be human-readable in your application.*

Consumer Key (API Key)	KXVdBHxBIUH07M6mowMzclSgV
Consumer Secret (API Secret)	cBQIZEkrhRNsnyxW76tmE1ch4oHOErVg4xxCUOkVYQjdnTJFK
Access Level	Read and write ( <a href="#">modify app permissions</a> )
Owner	paramjoth
Owner ID	3244379821

Figure 3.4 Consumer Key and Consumer Secret Generated by Twitter

### **3.3 Establishing Connections**

There are two connections which are established for implementing the proposed system. The first connection is between the healthcare application and the healthcare database created for backend. The application uses “com.mysql.jdbc.Driver” driver to establish connection with the database. The developer provides localhost port number, database name, user ID and password to get connected with the database created for the application. The second connection is between the healthcare application and the twitter database. This connection is established by providing OAuth consumer key, OAuth consumer secret, OAuth access token and OAuth access token secrets, which are generated while creating the twitter application.

### **3.4 Collecting Tweets**

Once the connection is established then the developer can add keyword and category to the healthcare care application. Whenever a keyword or category is added it gets update in the healthcare database. For a particular category, a user can add any number of keywords. When the user wants to collect tweets, it needs to select a file in which it wants to store the collected tweets. Once the file is selected, the user can start collecting the tweets.

### **3.5 Parsing Tweets**

Parsing is nothing but syntactic analysis. It is a process of analyzing a string of symbols in natural language according to rules of grammar [17]. Once the tweets are

collected, the developer arranges the tweets in a particular manner. Collected tweets may start at first line and then may end up at any line number, in order to differentiate one tweets from another, tweeter adds a comment “EOT” at the end of each tweet. While parsing the developer removes all those ending words. The developer then parse the tweets which have many blank spaces, or empty newline and make it as a single line tweet. The parser even replaces the abusive words with “\*\*\*\*\*” indicating the word is abusive and then removes the “\*\*\*\*\*” from the sentences.

### **3.6 Pre-processing:**

In the pre-processing step, the parsed tweets are collected and it removes unwanted words, numbers, symbols, special characters. In pre-processing, the complete data is changed to lower case letters. If there are any uppercase, bold letters or words in the collected data, they are converted into lower case letters. The output of pre-processed tweets becomes more meaningful and readable when compared to the collected tweets.

### **3.7 Natural Language Processing**

Sentiment analysis is a process which determines the intended emotion of the data. In sentiment analysis, the polarity of each sentence in the given data set is identified as positive or negative [18]. In this project, sentiment analysis is performed by using an Natural Language Processing (NLP) algorithm. Natural Language Processing is an interaction between human languages and computers. The NLP algorithm is based on statistical machine learning. In the NLP algorithm the machine actually understands the context, sentence arrangement and focuses mainly on the succession of a string of words.

The NLP algorithm makes a probabilistic decision based on sentiment value of each input [19] [20].

## **4 REQUIREMENTS**

### **4.1 Environment**

The proposed system is implemented in Java as front end and MySQL as back end. Twitter4j, XAMPP and Java swing are mainly used. NetBeans IDE is used as the software development platform.

### **4.2 System Requirements**

System requirements include both software requirements and hardware requirements used in developing the project.

#### **4.2.1 Software Requirements**

Software requirement is the description of the software needed by the developer before implementing the proposed system.

- Operating System : Windows 7 or higher version
- Language : Java
- Version : 7
- IDE : Netbeans 8.0
- Back-end : MySQL
- Package : XAMPP
- Server : Glassfish

### **4.2.2 Hardware Requirements**

Hardware requirements describe the basic version of hardware needed to run the application. In case of high speed of execution, developer can use updated versions.

- Processor : Intel core i3
- RAM : 2GB
- Hard Disk Drive : 256 GB

### **4.3 XAMPP**

XAMPP is a free open source cross-platform developed by Apache Friends. The word XAMPP is an acronym and its initials are represented as :

X – Cross-platform

A – Apache HTTP Server

M – MySQL

P – PHP

P – Perl

A developer needs XAMPP to create a local web server and access MySQL database system [21].

### **4.4 Netbeans IDE**

Netbeans Integrated Development Environment (IDE) is a open source software development platform [22]. Basically, it is intended for development of java applications. Netbeans IDE can run on any operating system platform which can support a compatible

Java Virtual Machine (JVM). Netbeans IDE 8.0.2 is used in the implementation of the project. Neatbeans platform provides various features such as project management, user interface management, storage management and window management [23].

## **4.5 Java Swing**

Java swing is a Graphical User Interface (GUI) toolkit, released by Oracle for java programs. Swing provides various features such as extensible, customizable and configurable which makes which makes this platform more user friendly. This platform includes components such as labels, buttons, checkboxes etc, which allows the features of drag and drop, making it further easy to use [24]. Class names of these components begin with 'J' such as JLabel, JButton, and JFrame .

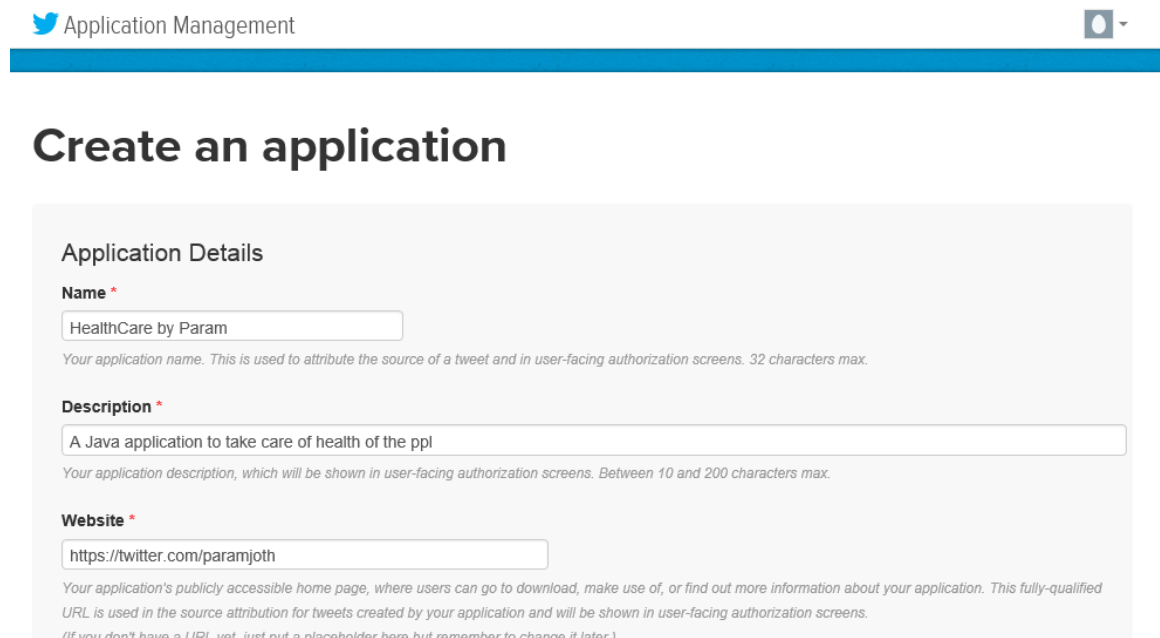
## **4.6 Glassfish server**

Glassfish is an application server for Java EE platform. It is an open-source server sponsored by Oracle Corporation. Glassfish supports implementation of JavaBeans, JMS,JPA, JavaServer Pages, servlets, JavaServer Faces, etc. It uses derivative of Apache Tomcat as Servlet container for serving web contents [25]. In this project, Glassfish server is used to run the web application.

## 5 IMPLEMENTATION

### 5.1 Inductive Content Analysis

First, the developer needs to login to the twitter account, then to get access to the java project, to create an application. As shown in Figure 4.1, twitter4j jar collection provides a class known as Configuration Builder class, which acts as an interface between the twitter account and the application. This configuration Builder class is responsible for collecting the data from twitter.



The screenshot displays the 'Application Management' interface on Twitter. At the top, there is a header with the Twitter logo and the text 'Application Management'. Below this, a blue bar contains the title 'Create an application'. The main content area is a light gray box titled 'Application Details'. It contains three sections: 'Name \*' with a text input field containing 'HealthCare by Param' and a note 'Your application name. This is used to attribute the source of a tweet and in user-facing authorization screens. 32 characters max.'; 'Description \*' with a text input field containing 'A Java application to take care of health of the ppl' and a note 'Your application description, which will be shown in user-facing authorization screens. Between 10 and 200 characters max.'; and 'Website \*' with a text input field containing 'https://twitter.com/paramjoth' and a note 'Your application's publicly accessible home page, where users can go to download, make use of, or find out more information about your application. This fully-qualified URL is used in the source attribution for tweets created by your application and will be shown in user-facing authorization screens. (If you don't have a URL, use just a placeholder here but remember to change it later.)'.

Figure 5.1 Creating Twitter Application

The application is created successfully and it returns Consumer Key, Consumer Secret, Access Token and Access Token Secret to the developer [26]. These Keys and Tokens are used as Parameters of Authentication to configure twitter4j.

The developer needs a database to collect the data from the twitter database and maintain it. XAMPP is a free open source cross-platform which provides MySQL database. The developer needs to download and run XAMPP exe file. XAMPP Control panel consists of Apache HTTP Server, MySQL database, PHP and Perl programming language. Figure 5.2 shows the XAMPP control panel and its components which need to be started before running the application [27].

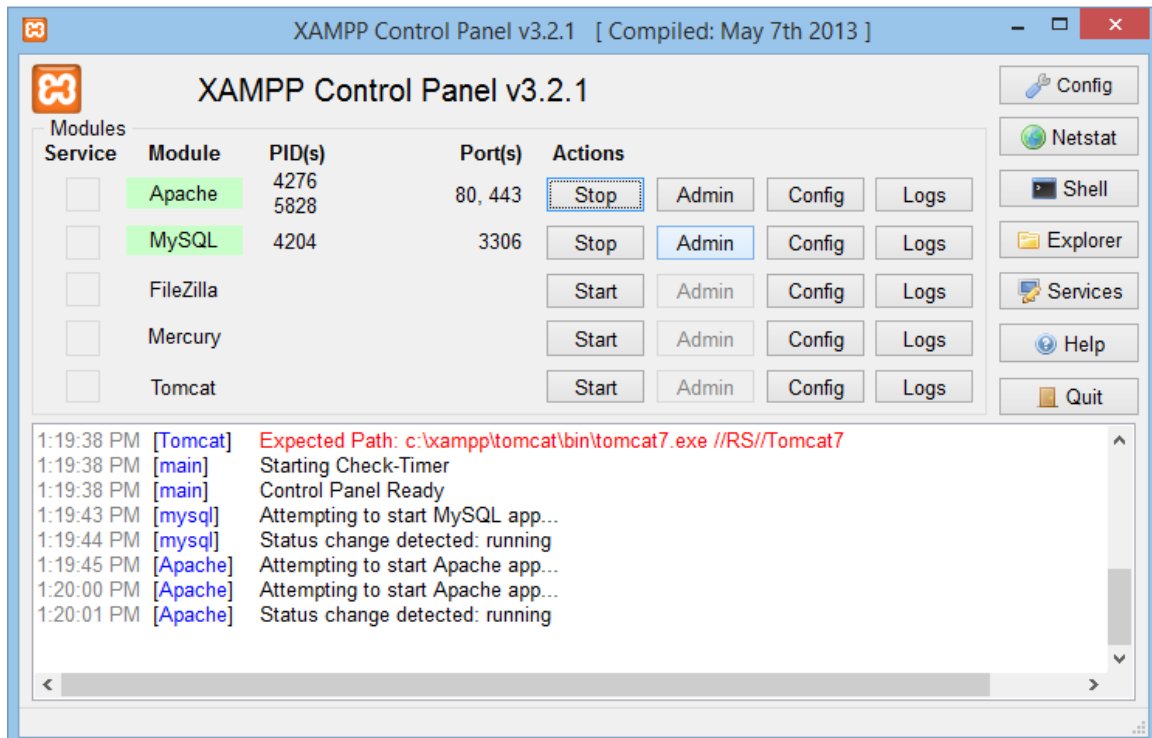


Figure 5.2 XAMPP Control Panel



Various databases are available in MySQL admin. Developer can create the database by clicking on the database tab available on the homepage of the localhost. Figure 5.3 shows the database and the tables created by the developer for this project.

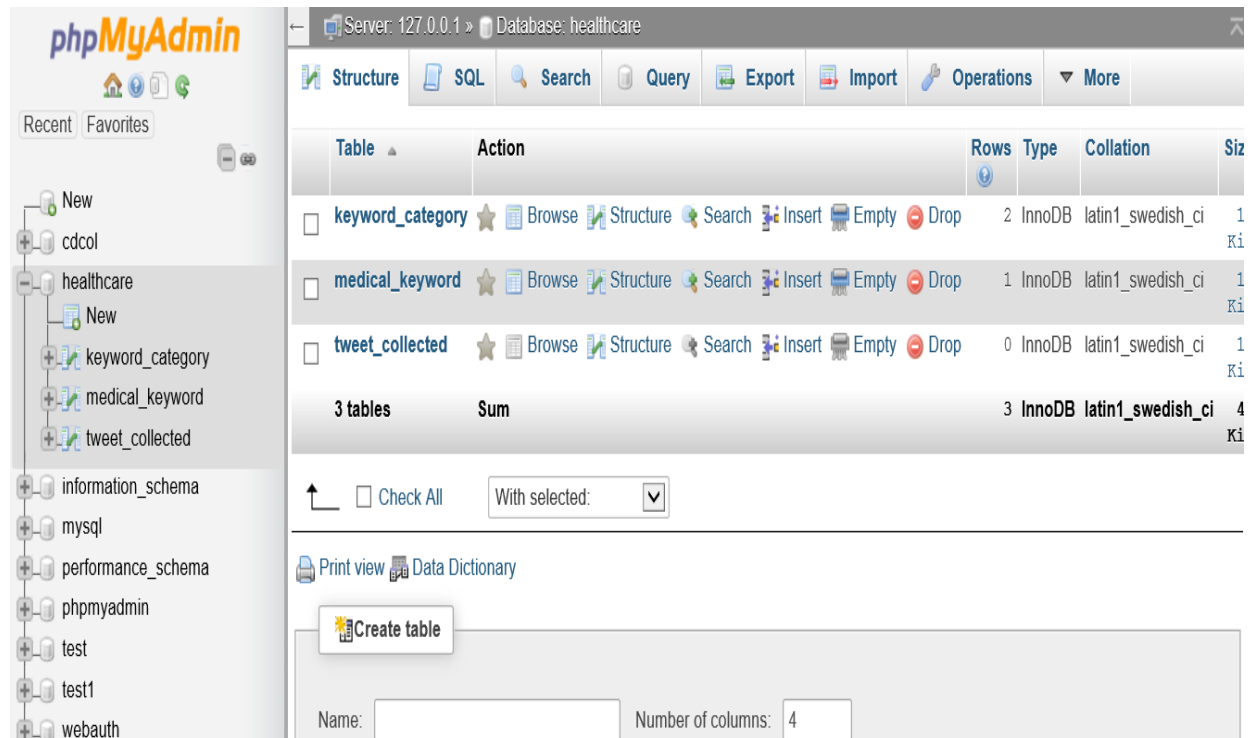


Figure 5.3 Healthcare Database

The functionalities provided by the desktop application are shown in Figure 5.4.

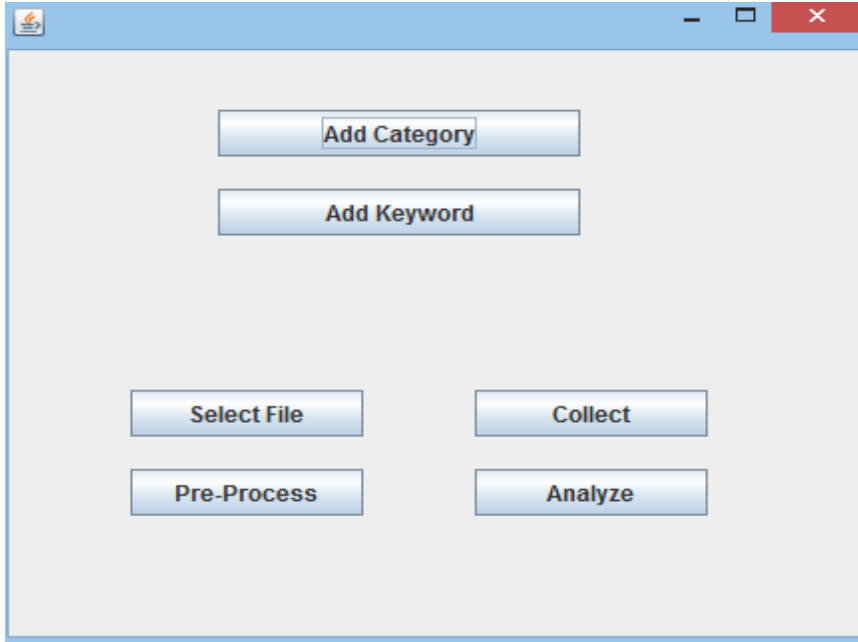


Figure 5.4 Functionalities of the Application

## 6 TESTING AND EVALUATION

Data mining from twitter will provide a lot of information based on our need. In the project, twitter data is collected, made as a data set, pre-processed, categorized and analyzed to provide information on medicines available for diabetes. The project can even suggest few medicines based on the number of positive user tweets available for the medicines for diabetes. Various test cases are generated based on the flow of the design and implementation, like collecting the data, pre-processing the collected data etc.

### 6.1 Add Category

Once the user login to the system, the flow of the application should be tested as shown in the data flow diagram. As shown in Figure 6.1, first testing should be done to check whether category is added in proper manner or not. Once the category name is entered, the user must click on the submit button to store the new category in the

database, which has a list of all the categories added so far. If the submitted category name matches the structure of the keyword\_category table available in the database, a window will appear, which displays the message “Category added successfully” else the window will display the message “ category not added” As the project is to provide a better Healthcare for diabetic patients, the category added should be diabetes.

Figure 6.1 Add Category

## 6.2 Add Keyword

As shown in Figure 6.2, the second test is to check whether a keyword is added in combination with the category which is added in the first step. Once the submit button is clicked a window will appear which displays the message either “Keyword added successfully” or “Keyword is not added” based on the structure of the medical\_keyword table available in the database. The keywords added should be the medicine names available for diabetes.

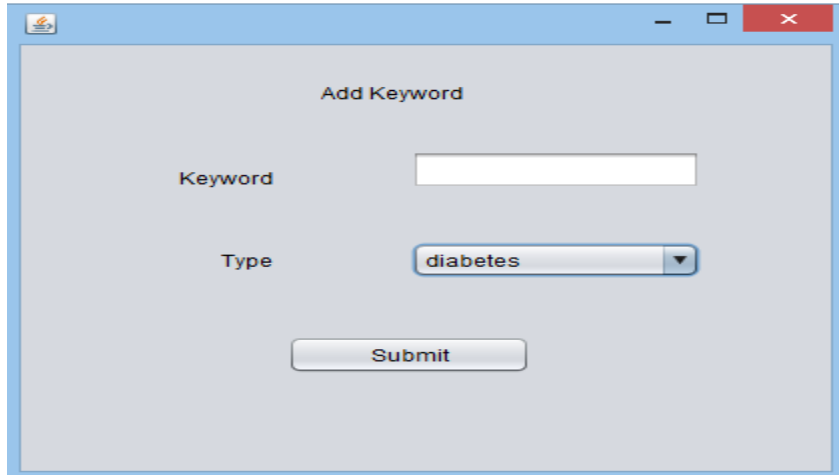


Figure 6.2 Add Keyword

### 6.3 Collect Tweets

The next test is to collect the tweets from the twitter based on the keywords provided for the category. Before collecting the tweets, the user must first specify a file where he/she will store the collected tweets. As shown in Figure 6.3, the collected tweets must be based on diabetes.

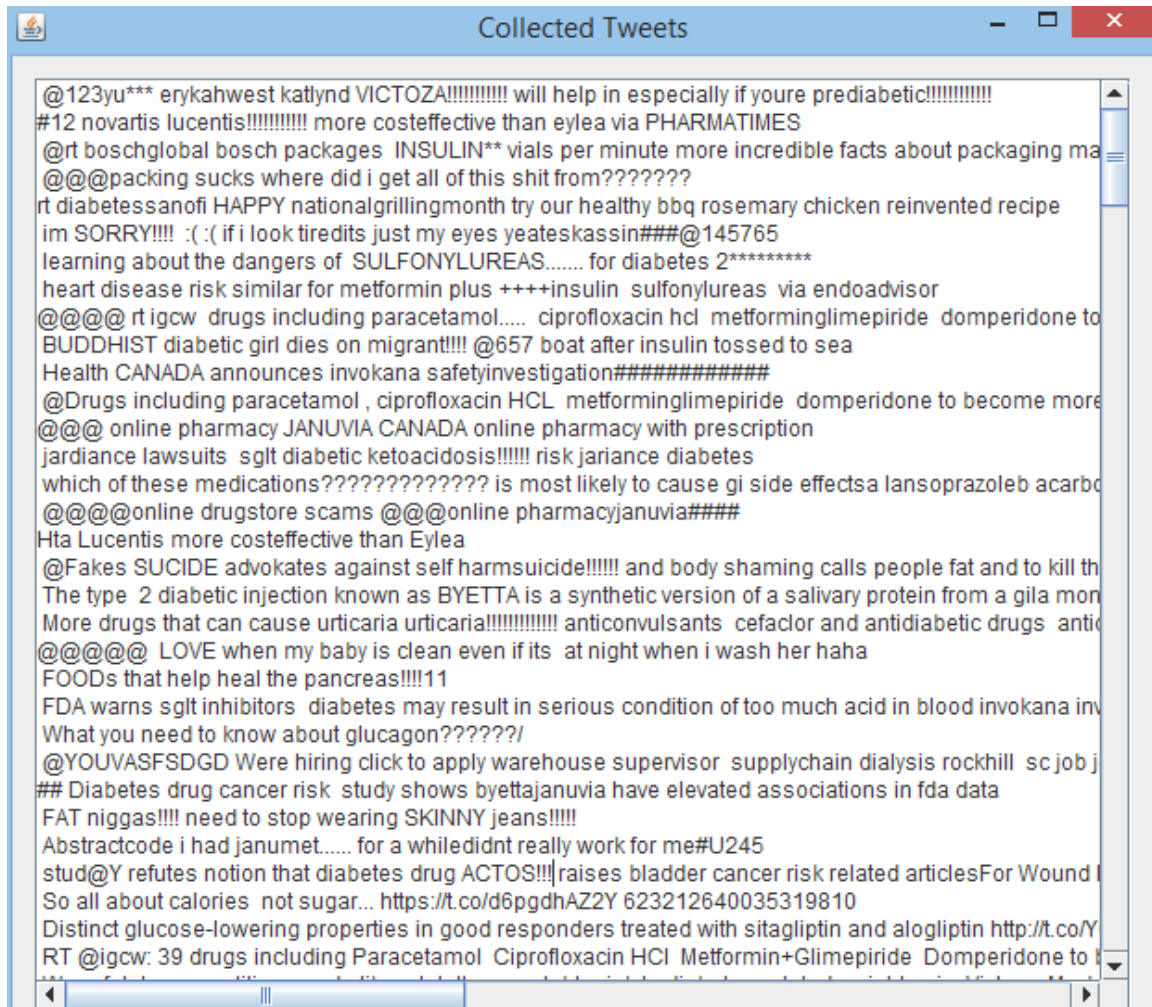


Figure 6.3 Collecting Tweets from Twitter

## 6.4 Pre-process Tweets

The next test is to check whether pre-processing the tweets is performed correctly or not. In this test, the user checks whether unnecessary words, numbers, symbols and special characters are removed from the collected tweets. As shown in Figure 6.4, the pre-processed tweets should be readable easily.

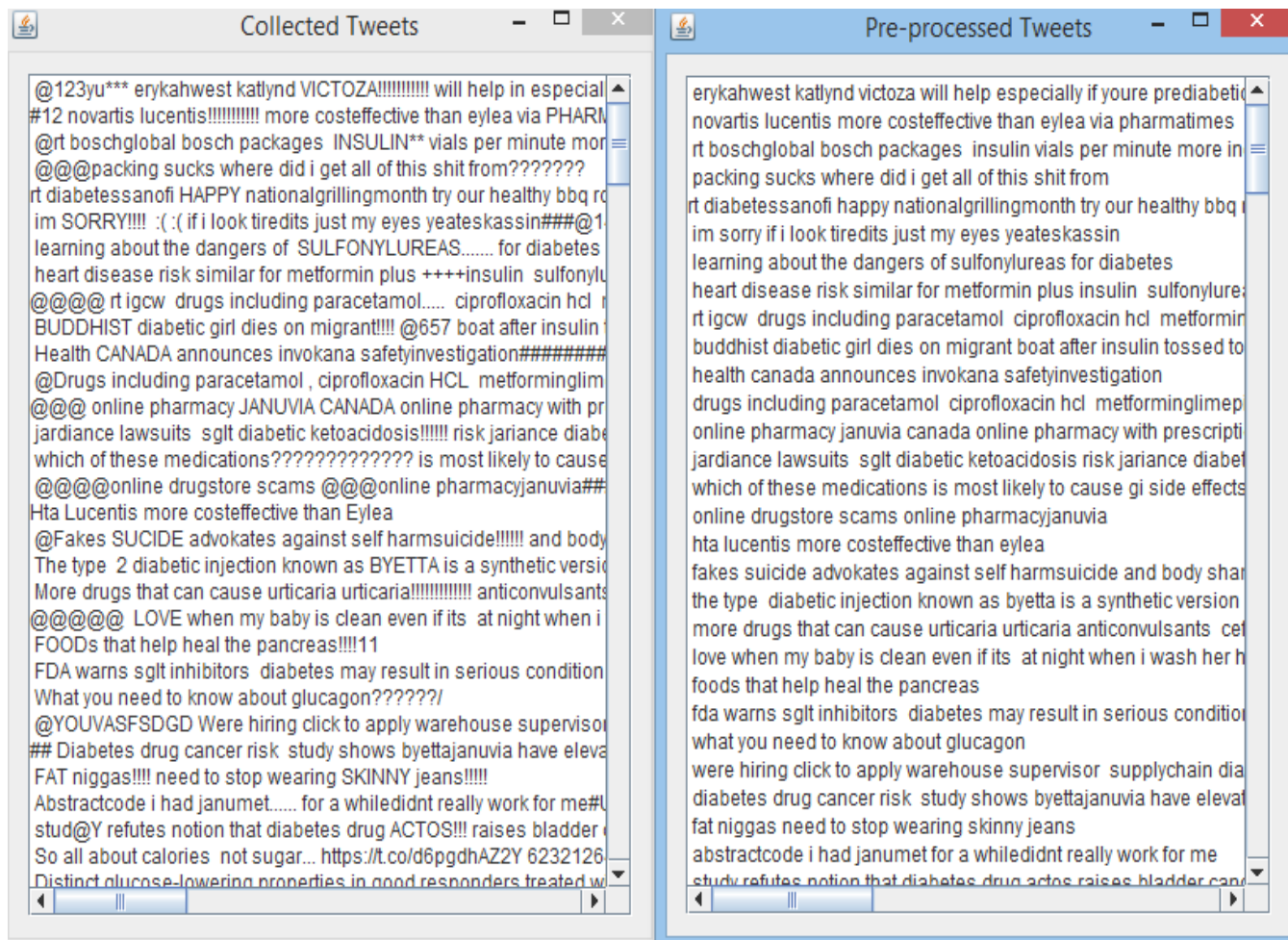


Figure 6.4 Pre-Processed Tweets

## 6.5 Positive Tweets

The next test is to analyze the pre-processed tweets either as positive tweets or negative tweets based on the sentiment associated with them. Running the NLP algorithm on the filtered tweets will assign a sentiment value to each tweet. If the Tweet sentiment value generated by NLP algorithm is greater than 1, then the sentiment associated with the tweet is categorized as positive. As shown in Figure 6.5, the pre-processed tweets, which has value greater than 1 are collected and displayed in Positive Tweets window.

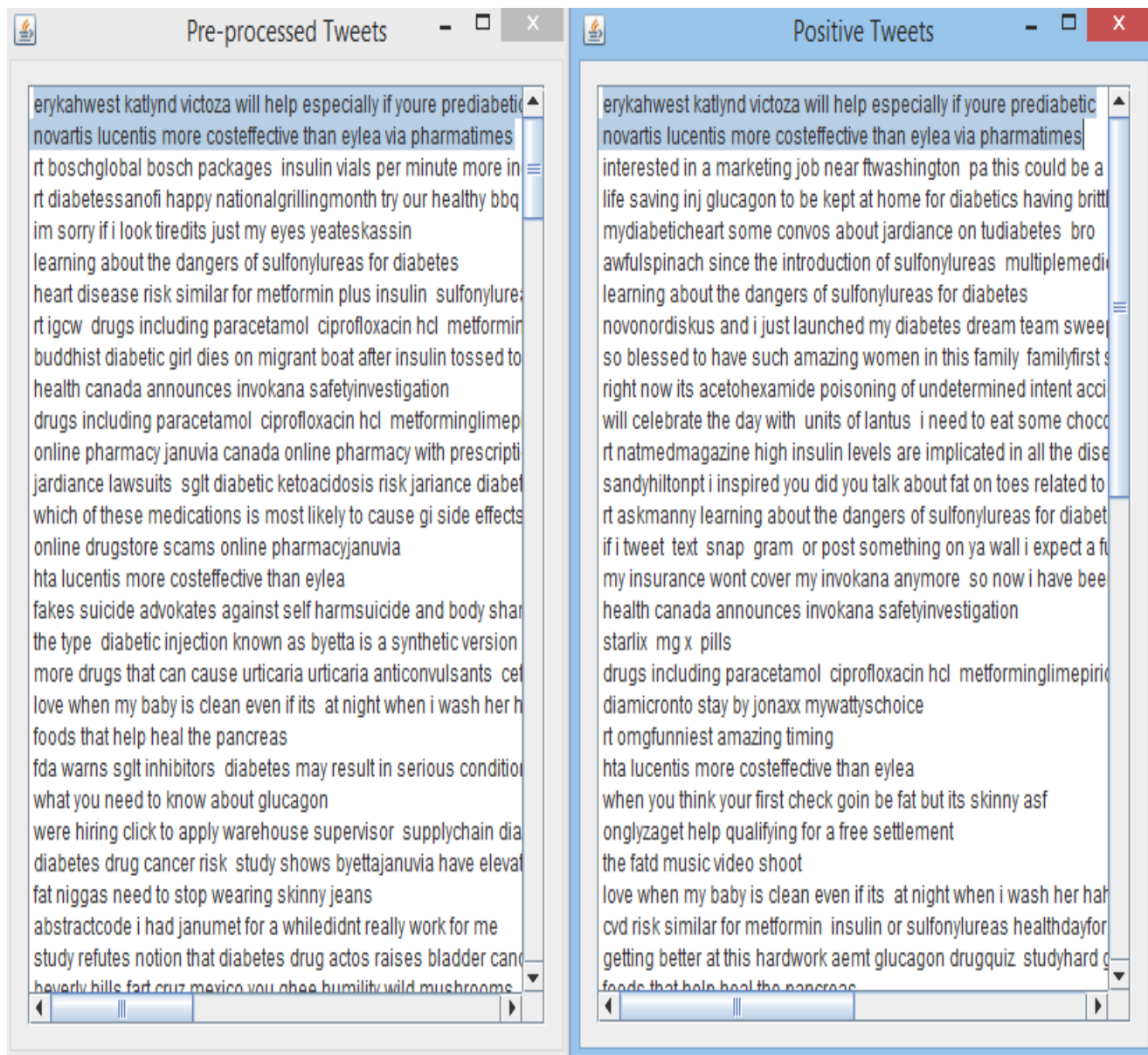


Figure 6.5 List of Positive Tweets

## 6.6 Negative Tweets

If the Tweet sentiment value generated by the NLP algorithm is either 0 or 1, then the sentiment associated with the tweet is categorized as negative. As shown in Figure 6.6, the pre-processed tweets, which has value either 0 or 1 are collected and displayed in Negative Tweets window.



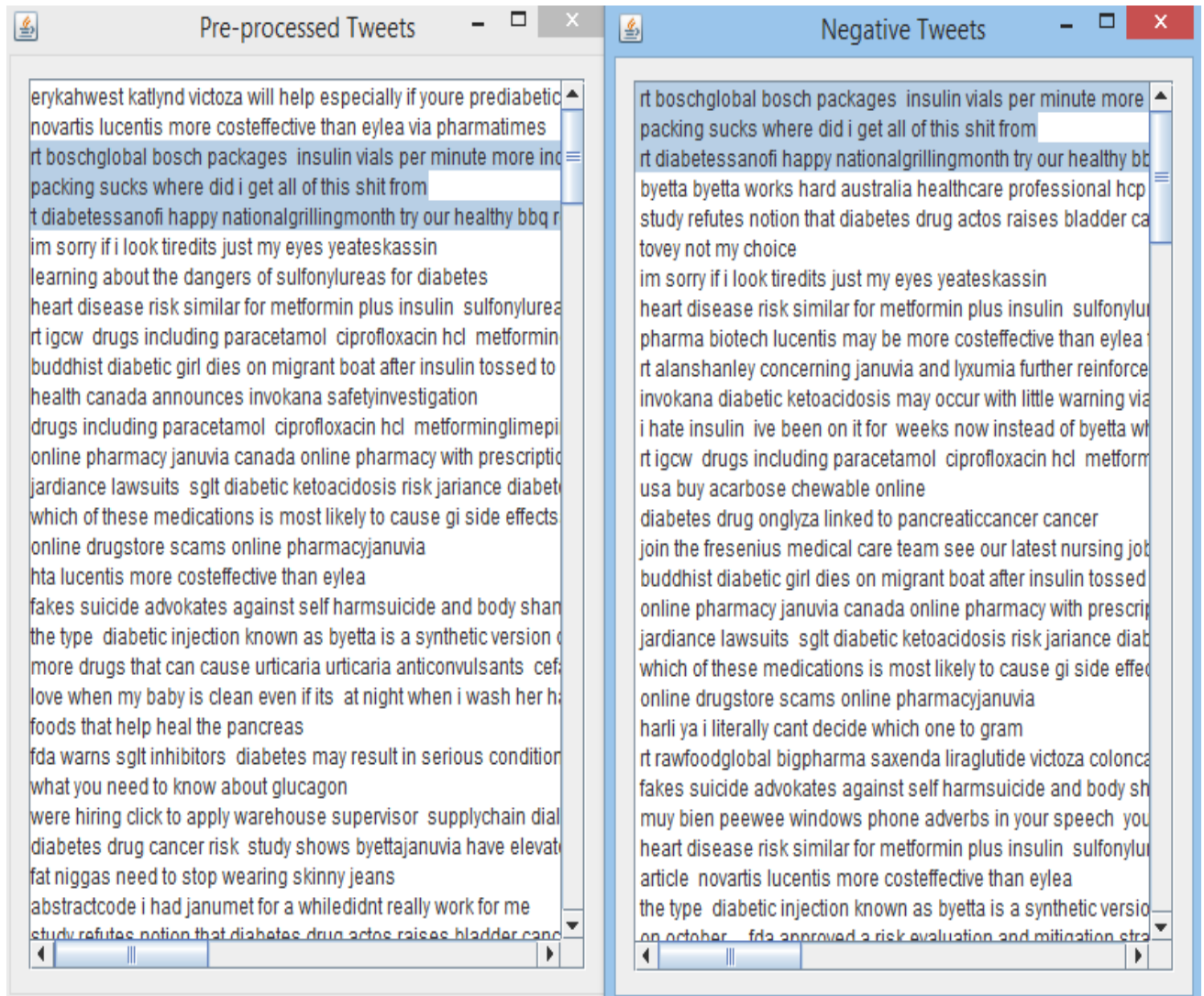


Figure 6.6 List of Negative Tweets

## 6.7 Designing the Forum and Suggesting the Tweets

A web application is created which will help the user to get information on the medicines available for diabetes. This web application will even suggest the user with better medicines available in the market. The user can even subscribe to this forum. There are four functionalities in this web application, they are: Home, About, Tweets and Subscribe



## 6.8 Home and About

As shown in Figure 6.8, the Home option welcomes the user to the home page of the web forum, when the About option is clicked, It provides the user with the information about the web forum, its importance and the advantages associated with it.

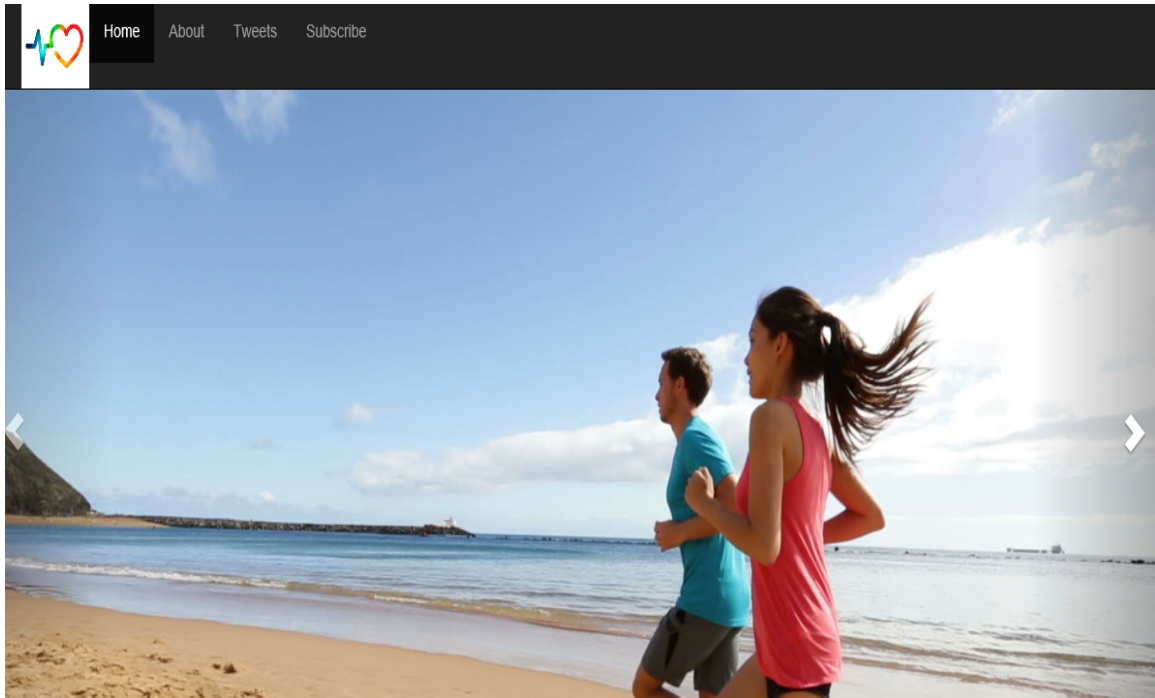


Figure 6.8 Home Page of Web Application

## 6.9 Tweets and Suggestions

In Tweets option, the user needs to enter the name of medicine, for which he/she wants to know the opinions from the consumers. When the user mentions the medicine name and click the retrieve suggestions button, it gives two results to the user. As shown in Figure 6.9, the first result displays percentage of positive tweets and negative tweets obtained for the mentioned medicine. The second result displays the suggestions to the user.

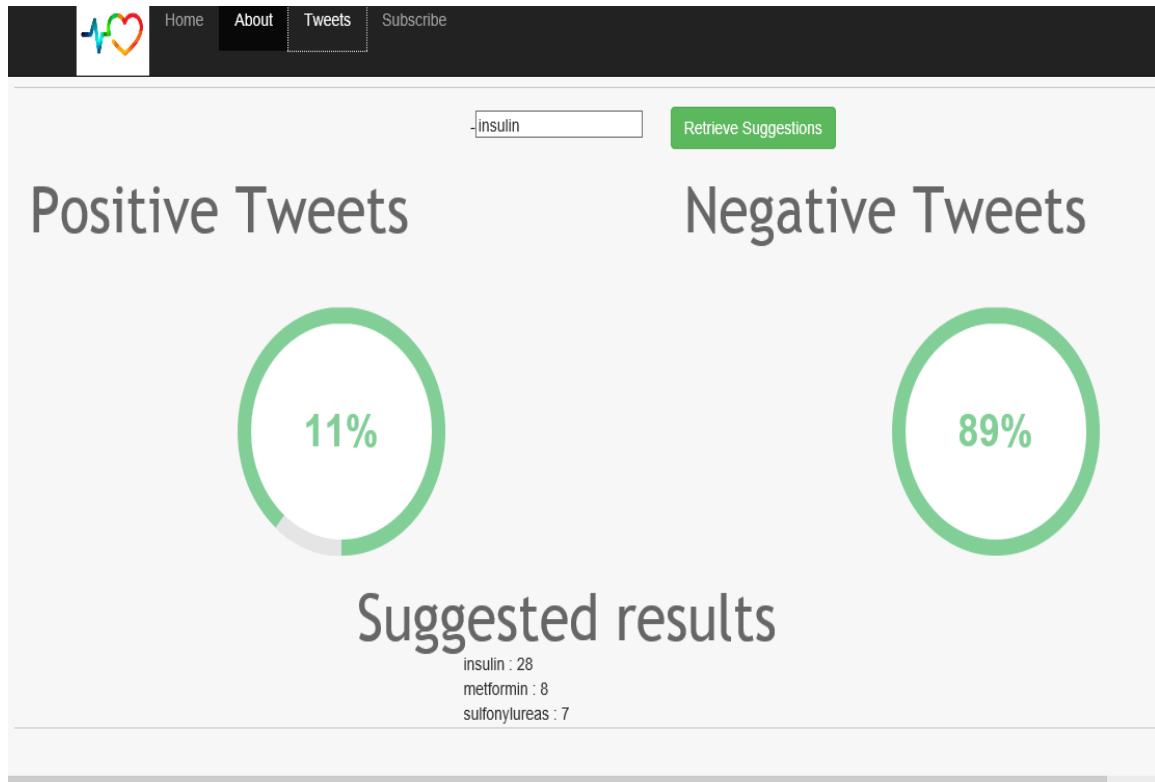


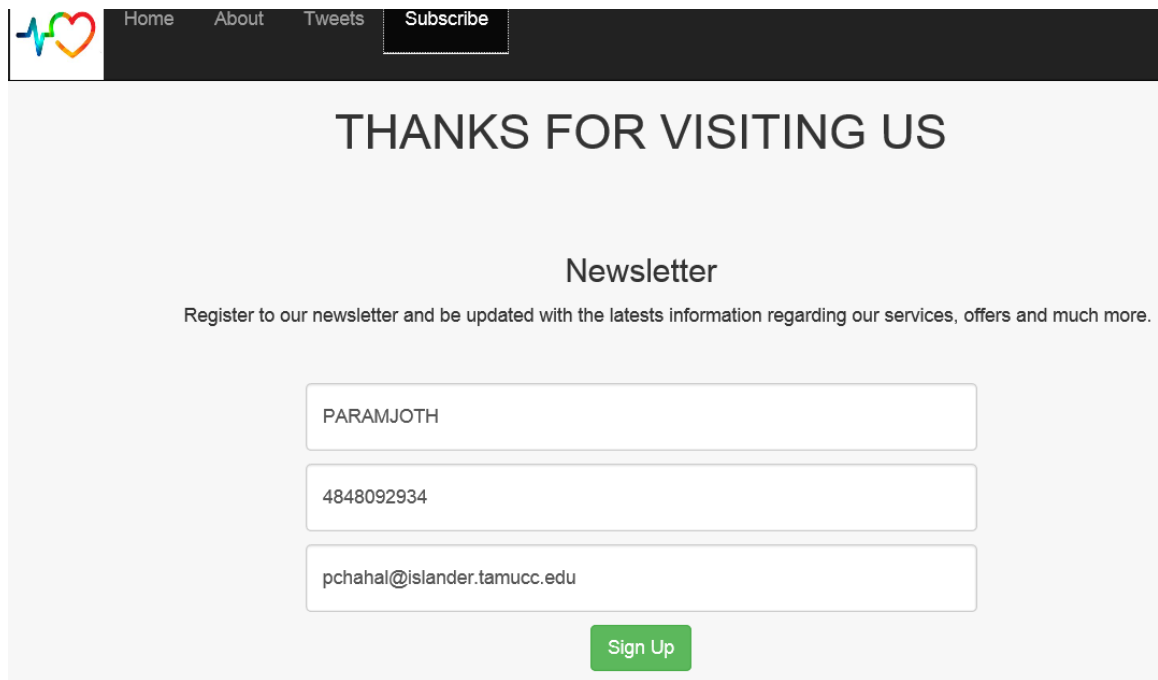
Figure 6.9 Retrieving Results and Suggestions

These suggested results are obtained from the Positive Tweets Analysis. The suggested results consist of a list of medicine names, which is in the decreasing order of the number of positive tweets for that medicine. The medicines which have the maximum number of positive tweets, will be suggested on the top of the list. In this way, top three medicines which have maximum number of positive tweets will be suggested.

## 6.10 Subscription for Newsletter

Users can subscribe for Newsletters, which will help them in keeping themselves updated with the latest medicines and their effectiveness. As shown in Figure 6.10 user needs to enter his/her name, phone number and email address, and hit the submit button

to complete the subscription. Once the user clicks the submit button, a window appears, which will display the message “Registered Successful!!”.



The screenshot shows a website's navigation bar with a logo (a heart with a pulse line) and links for Home, About, Tweets, and a highlighted Subscribe button. Below the navigation bar, the text "THANKS FOR VISITING US" is displayed in large, bold, black capital letters. Underneath this, the word "Newsletter" is centered. A line of text reads: "Register to our newsletter and be updated with the latests information regarding our services, offers and much more." Below this text are three stacked input fields. The first field contains the text "PARAMJOTH". The second field contains the text "4848092934". The third field contains the text "pchahal@islander.tamucc.edu". At the bottom right of the form is a green button with the text "Sign Up" in white.

Figure 6.10 Subscribing for Newsletter

## 6.11 Result without Suggestions

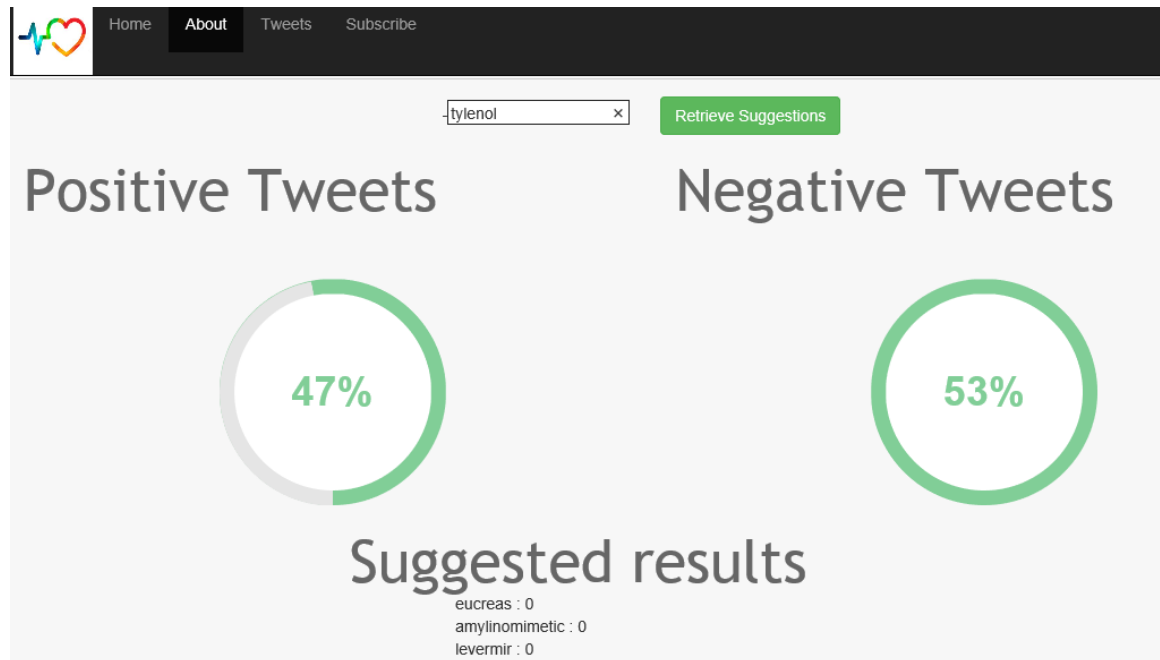


Figure 6.11 Partial Output

As shown in Figure 6.11 there are few medicines like Tylenol, for which tweets are available on twitter, but in combination with this medicine, no other medicine have a positive comment. Therefore, the user is provided with partial output, where a user is provided with the results of sentiment analysis but, the user is not provided with any suggestions.

## **7 LIMITATIONS**

Twitter is a good resource of data. Large amount of data can be obtained from twitter. But, there are cases where even twitter is unable to provide any data. Such cases occur when the data to be retrieved is not enough popular in the society. While performing test cases on this project, it is observed that, there are few medicines which are not very famous and does not contain any data on twitter. When such medicines are given to collect the data from twitter, the search query does not provide any output. While performing the test case for “Byedroun” and “Glipzide”, it is observed that, the search query does not provide any result. Collecting data for such medicines acts as limitations of this project.

## **8 CONCLUSION**

Sentiment analysis of data collected from social media is beneficial for mankind in providing them better healthcare. The workflow of analyzing healthcare content in the social media helps to overcome the limitations of large scale data analysis and manual analysis of user generated textual content in social media.

This work can help the users to be updated with the effectiveness of the medicines and it can even suggest them with few better medications available. This project can provide feedback to the healthcare system organization and pharmaceutical companies for the available treatments and medicines. With the help of this project, pharmaceutical companies and healthcare providers can work on the feedback and try to come up with the improvised medicines and treatments for diabetes. Users are provided with the resources of social media for the corresponding field of healthcare.

## **9 FUTURE WORK**

In future, one can collect large healthcare related data from multiple social networking sites which may provide better results by overcoming the limitations of the project. In future, one can even collect data which includes videos and images for analyzing the effectiveness of the medicines and new treatments.

## References

- [1] [Online]. Available: [www.medicalnewstoday.com/articles/44967.php](http://www.medicalnewstoday.com/articles/44967.php). [Accessed 13 /6/ 2015].
- [2] [Online]. Available: <http://www.idf.org/latest-diabetes-figures-paint-grim-global-picture>. [Accessed 10 /6/ 2015].
- [3] Mohammad Ali Abbasi, Huan Liu and Reza Zafarani, "Social Media Mining - An introduction, Arizona state university", April 2014.
- [4] [Online]. Available: <http://www.socialmediaexaminer.com/5-new-studies-show-facebook-a-marketing-powerhouse/>. [Accessed 10 /6/ 2015].
- [5] [Online]. Available: [https://en.wikipedia.org/wiki/Social\\_media\\_mining](https://en.wikipedia.org/wiki/Social_media_mining). [Accessed 11 /6/ 2015].
- [6] [Online]. Available: [http://train.ed.psu.edu/WFED-543/SocNet\\_TheoryApp.pdf](http://train.ed.psu.edu/WFED-543/SocNet_TheoryApp.pdf). [Accessed 11 /6/ 2015].
- [7] Altug Akay, Bjorn Erilc Erlandsson, Andrei Dragonier "Network-Based Modeling and Intelligent Data Mining of Social Media for Improving Care," *IEEE Journal Of Biomedical and Health Informatics*, vol. 19, no. 1, january 2015.
- [8] X. Dong, X. Feng, M. Feng, A.Cai, J. Inciardi, T.Woldemariam, N.S> Bhutada and W. Chaing, "Assessing pancreatic cancer risk associated with dipeptidyl peptidase 4 inhibitors: data mining of FDA adverse event reporting system (FAERS)," *J.Pharmacovigilance*, vol. 1, 2013.
- [9] A. Mikler, K. Singh, D. Cook and C.Corley, "Text and structural data mining of influenza mentions in web and social media," *Int. J. Environ*, vol. 7, pp. 596-615, 2010.
- [10] Christopher D. Manning, Prabhakar Raghavan & Hinrich Schutze, "Scoring, term Weighting & the vector space model," in *Introduction to Information Retrieval*, 2008.
- [11] Michael, R. Dittenbach, "Scoring and Raking Techniques - tf-idf term weighting and cosine similarity," March 31, 2010.
- [12] Ruilin, Xu, "POS Weighted TF-IDF algorithm and its application for an MOOC



- search engine," in *2014 International Conference*, shanghai, 2014.
- [13] [Online]. Available: <http://www.aclweb.org/anthology/C10-2028>. [Accessed 11 /6/ 2015].
- [14] [Online]. Available: <http://www.sci.tamucc.edu/~cams/GraduateProjects/view.php?view=431>. [Accessed 11 /6/ 2015].
- [15] Robert A, Hanneman and Mark Riddle, "Using matrices to represent social relations," in *Introduction to social network methods*, Riverside, CA, published in digital form at <http://faculty.ucr.edu/~hanneman/>, 2005.
- [16] [Online]. Available: <http://www.internetlivestats.com/twitter-statistics/>. [Accessed 12 /6/ 2015].
- [17] [Online]. Available: <https://en.wikipedia.org/wiki/Parsing>. [Accessed 10 /7/ 2015].
- [18] [Online]. Available: <http://nlp.stanford.edu/software/>. [Accessed 8 /7/ 2015].
- [19] Tyschuk Y and Wallace, "The Nature of Communications and Emerging Communities on Twitter Following the 2013 Syria Sarin Gas Attacks," in *Intelligence and Security Informatics Conference (JISIC)*, 2014.
- [20] Tian, Yuan and Lo D., "A comparative study on the effectiveness of part-of-speech tagging techniques on bug reports," in *Software Analysis, Evolution and Reengineering (SANER), 2015 IEEE 22nd International Conference*, 2015.
- [21] [Online]. Available: <https://en.wikipedia.org/wiki/XAMPP>. [Accessed 2 /6/ 2015].
- [22] [Online]. Available: <https://en.wikipedia.org/wiki/NetBeans>. [Accessed 5 /6/ 2015].
- [23] [Online]. Available: <https://netbeans.org/features/>. [Accessed 5 /6/ 2015].
- [24] Memon A, Banerjee I, Nguyen B.N and Et Al, "The first decade of GUI ripping: extension,application, and broader impacts," in *Reverse Engineering (WCRE), 2013 20th Working Conference*, 2013.
- [25] [Online]. Available: <https://en.wikipedia.org/wiki/GlassFish>. [Accessed 25 /6/ 2015].
- [26] [Online]. Available: <http://blog.twitter4j.org/2013/06/13/59>. [Accessed 5 /6/ 2015].

[27] [Online]. Available:

[http://www.textpattern.net/wiki/?title=Using\\_XAMPP\\_\(Apache-MySQL-PHP-Perl\)\\_for\\_Windows](http://www.textpattern.net/wiki/?title=Using_XAMPP_(Apache-MySQL-PHP-Perl)_for_Windows). [Accessed 2 /6/ 2015].