Disaster Detection and Reporting System through Analysis of Twitter Tweets

GRADUATE PROJECT REPORT

Submitted to the Faculty of
The School of Engineering & Computing Sciences
Texas A&M University-Corpus Christi
Corpus Christi, TX

In Partial Fulfillment of the Requirements for the Degree of
Master of Science in Computer Science

By

Nikhil Srinivas Komatireddy
Spring 2015

Committee Members

Dr. Longzhuang Li
Committee Chairperson

Dr. David Thomas
Committee Member
ABSTRACT

Real time events that are described as natural disasters can be investigated in Twitter and the tweets are monitored to detect a target event like a hurricane. An event in Twitter is identified as a target event by classifying the tweets based upon some of the key parameters like the keywords in a tweet, the number of words and their context. A large number of tweets are scanned wherein these target events are involved and are stored into a database. Then the events that are of our interest only are extracted from all the data that is saved and the locations of these events are estimated. All the registered users of the notification system get a notification through their registered email IDs whenever a hurricane occurs. These valid customers can keep themselves and their loved ones safe by always being informed about this dangerous life-taking event, which occur prior to no information. This notification system is even more helpful when there is a higher risk of getting trapped where there is no availability of television or other sources of information.

Unlike other reporting systems, this project can distinguish between a storm, a hurricane and a tsunami based on the key characteristics like the wind speed, rainfall, rise in sea level and time period. It just simply does not notify the registered users without checking for trustworthiness. It only notifies if the message in the tweet is true. This project can be improved by sending message alerts through the telecommunication network companies.
# TABLE OF CONTENTS

Abstract .......................................................................................................................... ii

Table of Contents ........................................................................................................ iii

List of Figures ........................................................................................................ vi

List of Tables ........................................................................................................ vii

1. Background and Rationale ..................................................................................1

   1.1 History and Evolution of Twitter ................................................................. 1

   1.2 Evolution of Data Mining ............................................................................ 3

   1.3 Introduction to Proposed System ............................................................... 4

   1.4 Existing Systems ........................................................................................ 5

   1.5 NetBeans IDE 8.0.2 .................................................................................. 7

   1.6 XAMPP ....................................................................................................... 8

2. Narrative ................................................................................................................. 9

   2.1 Problem Statement ...................................................................................... 9

   2.2 Motivation .................................................................................................. 9

   2.3 Project Objective ........................................................................................ 10

   2.4 Project Scope ............................................................................................ 10

3. System Design ....................................................................................................... 12

   3.1 System Design and Architecture ............................................................... 12

   3.2 Desktop Twitter Application ...................................................................... 15

      3.2.1 Data Flow – Desktop Twitter Application ......................................... 16

   3.3 Web Twitter Application ............................................................................ 17

      3.3.1 Data Flow – Web Twitter Application ............................................... 17
3.4 System Requirements

3.4.1 Functional Requirements

3.4.2 Software Requirements

3.4.3 Hardware Requirements

4. System Implementation

4.1 Environment

4.1.1 Twitter 4j

4.1.2 NetBeans IDE

4.1.3 XAMPP

4.2 User Interface

4.2.1 Desktop Twitter Application

4.2.2 Checking the Connection

4.2.3 Retrieving the Tweets

4.2.4 Retrieving from News Accounts

4.2.5 Storing the Tweets

4.2.6 Filtering the Tweets

4.2.7 Comparing with News Channels

4.2.8 Mail Alert

4.2.9 Web Twitter Application

5. Testing and Evaluation

5.1 Software Testing

5.2 Unit Testing

5.3 System Testing
5.4 Test Cases for Tweet Tracking.................................................................32

5.5 Test Results ...........................................................................................33
  5.5.1 Storing of Tweets..............................................................................34
  5.5.2 Testing with Trustworthy Accounts..................................................35
  5.5.3 Testing of Mail Alert.........................................................................37

6. Conclusion..................................................................................................39

Bibliography and References.........................................................................40
# LIST OF FIGURES

| Figure 1.1 | Content of Tweets According to Pear Analytics | 2 |
| Figure 3.1 | System Architecture | 14 |
| Figure 3.2 | Data Flow of Desktop Twitter Application | 16 |
| Figure 3.3 | Data Flow of Web Twitter Application | 17 |
| Figure 4.1 | Desktop Twitter Application Screen | 21 |
| Figure 4.2 | Checking the Connection | 22 |
| Figure 4.3 | Retrieving the Tweets | 23 |
| Figure 4.4 | Success Message for Retrieving the Tweets | 23 |
| Figure 4.5 | Retrieving from News Accounts | 24 |
| Figure 4.6 | Storing the Tweets | 25 |
| Figure 4.7 | Filtering the Tweets | 26 |
| Figure 4.8 | Comparing with News Channels | 27 |
| Figure 4.9 | Login Screen of Web Twitter Application | 28 |
| Figure 4.10 | List of Tweets | 29 |
| Figure 4.11 | User Registration Screen | 30 |
| Figure 5.1 | Tweets of Admin Account | 36 |
| Figure 5.2 | Tweets of CNN Account | 36 |
| Figure 5.3 | All Positive Tweets | 37 |
| Figure 5.4 | Corrected Tweets | 37 |
| Figure 5.5 | Registered User Notification | 38 |
LIST OF TABLES

Table 5.1  Test Cases for Tweet Tracking...........................................................32
Table 5.2  First Retrieved Tweets.................................................................35
Table 5.3  Second Retrieved Tweets..............................................................35
1. BACKGROUND AND RATIONALE

1.1 History and evolution of Twitter

According to Oxford dictionary, the word twitter means a series of calls or sounds that are not too long and the meaning of tweet is to make a sound which is generally produced by a bird. This is what the social networking site “Twitter” reflects even by keeping its logo as a small bird. Twitter is just like Facebook, a social networking site or application that can be used on a desktop, laptop, phone, tablet, or even on a phablet. All the Twitter users who are registered can post and read short messages through online services like internet and Wi-Fi. These messages are called “Tweets” and a word short is added in front of it because they are not more than 140 characters. Using the twitter account, tweets are posted as a user status. Only registered user of twitter can post their tweets in their status, others can only read the tweets.

It was founded in the year 2006 by Evan Williams, Noah Glass, Jack Dorsey and Biz Stone and head quartered in San Francisco, California. This service became popular by 2012 with about hundred million registered users with 340 million tweets per day and 1.6 billion search queries per day. By the end of 2014, Twitter had more than five hundred million users of which more than fifty percentages of them are very active [12].

Considering all the above mentioned numbers, Twitter is described as the Short Message Sending (SMS) of the internet. Besides this, twitter service is also called as micro blogging service as it describes what a user is doing in the form of a small text message or an image or an audio message [1]. At the same time tweeter users can post
many tweets in a day. The number of tweets that a twitter user can tweet is not limited. So it provides a real-time nature by making users to be connected with their family and friends residing throughout the world. Twitter is so user friendly that it does not restrict a user on the content [2]. Users can update their status like weather details and their feelings through tweets. These details are very helpful for others. They can share images about the weather details in their tweets.

“Trending topic” is a word, phrase or a topic that is used at a greater rate than other tags. Trending topic becomes more popular because of an event that makes people talk about a more specific topic. In this project most of the target events will be trending topics because it is these topics that help Twitter users understand what is going around in the world. Trending topic word is used just to make a clear idea for the user about the target event. Target event in this project or paper is a substitute for trending topic. As many users are used to trending topic which is a twitter word, it is used as a reference to target events to make a clear idea for the users.

Figure 1.1: Content of tweets according to Pear Analytics [12]
1.2 Evolution of Data Mining

Data mining is the analysis of raw data to turn it into useful information and knowledge. The extracted information will be used in making predictions and decision making for the benefit of organizations or applications for that matter. The information industry is more attracted towards data mining because of the availability of wide range of data and the need of turning this available data into useful information and knowledge. This information gathered can be used in applications irrespective of their field in which they are found. Data mining is now being widely used in business fields such as insurance, banking and retail, in science research areas such as astronomy and medicine and in government security such as detection of criminals and terrorists.

The evolution of databases and information technology started from early 1960’s and they systematically evolved from an under developing file processing systems to complicated databases. The research and development of database systems from the next decade has progressed from hierarchical database systems to relational database management systems where data is stored in a relational table structure [15]. Data can be accessed through query languages, optimized query processing and transaction management which are convenient and flexible to users.

The evolution of information technology resulted in data mining. Database industry has witnessed an evolutionary path in the development of data collection, data creation, data management and data analysis and understanding. Query and transaction processing is being offered by numerous database systems as a common practice, so data mining and data warehousing has become the future targets. Data mining uses software
techniques to find patterns and regularities in sets of data and analyzes this data. A set of data is analyzed and an optimal representation of the structure of the data is developed using a methodology [13].

Data mining nowadays has become increasingly common in both the private and public sectors. Industries such as banking, medicine, insurance and retailing commonly use data mining to enhance research in cost reduction and sales increment. In the public sector, initially data mining applications were used as a means to detect fraud and waste, but now have grown to also be used for the purposes such as measuring and improving program performance [19].

In this project, all the twitter data is mined and is filtered for some specific keywords.

1.3 Introduction to Present System

This project is a reporting system which notifies its registered users through an email about the details of the hurricane. There are many online social networking services that provide information that can be shared. This information will not be enough to know about particular events in particular place. They only give the generic information about the event. We cannot identify all the information about the event from this data. Here the event is a hurricane. To get all the information, the data needs to be fetched from twitter into this application and filter with the keywords from the integrated data of twitter.
There are many other applications and various platforms which are used to communicate about various disasters happening in and around the world. Twitter is one of the most important medium which has a large number of active users in it. So getting the information from twitter is easier than other social networking sites which are difficult to extract data [5].

A user can post or tweet about the disaster details in the social media. There are a large number of the Internet users available in this world and every user of the internet will know very little information about the disaster because they cannot access or find the full information about the disaster in the social network. So in order to notify the users for disaster alert there are no reporting system online. The information about the disasters will be notified to the users via an email in a short span. By building such an application users can take safety measures for all their needs, if they are notified within minutes after any disaster.

1.4 Existing Systems

There are few proposals such as “Using Clustering and Sentiment Analysis on Twitter” [18] and “Tweet Analysis for Real-Time Event Detection and Earthquake Reporting System Development” [6] which are similar to this kind of system but they are targeting to only one event at a time. They are not capable of differentiating between two similar kinds of events by which they can conclude at a wrong decision [14]. They do not have any mechanism to find out whether what they are depending on is correct or
incorrect. In other words they believe all the users on whom they are depending to be one hundred percent sure without performing any kind of check [5].

In “Using Clustering and Sentiment Analysis on Twitter”, popular electronic devices are searched on twitter and their information is stored in a document but in this proposed system, the data collected from twitter is stored in a database. The already existing system is executed using C# and Java and the environments that are setup are Microsoft Visual C# and Net beans IDE. But the suggested system uses only Core Java for the execution and Net beans IDE 6.9.1is the programming environment. In the clustering and semantic analysis paper, at the time of clustering, the k-means algorithm is used to deal with the tweets, and k is set to 3. In this reporting system, Pattern Matching Technique algorithm is used to differentiate and match the data [18].

Unlike many other reporting systems [4],[5],[6],[15], Disaster Detection and Reporting System through Analysis of Twitter Tweets can distinguish between a storm, a hurricane and a tsunami depending on some key characteristics like the wind speed, rainfall, rise in sea level and time period. It does not notify the registered users without checking for the trustworthiness about the tweets. It compares with other trustworthy tweets or news channels and decides whether the information provided is true or not. Then only it notifies its users about the disaster. By this a user can have complete confidence of not being misguided or providing false information on the reporting system that is being developed [6].
1.5 NetBeans IDE 8.0.2

NetBeans Integrated Development Environment (IDE) version 8.0.2 is used in the development of this project. NetBeans IDE 8.0.2 provides out-of-the-box code analyzers and editors for working with the latest Java 8 technologies--Java SE 8, Java SE Embedded 8, and Java ME Embedded 8. The IDE also supports Maven and Java EE with Prime Faces, new tools for HTML 5 in particular for AngularJS, C and C++ support and improvements to PHP. NetBeans IDE 8.0.2 is available in many other languages besides English [20]. In addition to the capabilities expected, NetBeans IDE 8.0.2 also offers:

i. Integration with JavaFX Scene Builder is improved.

ii. Flexibility for JPA hints is increased.

iii. Projects window shows additional web resources.

iv. Navigation of AngularJS from view to controllers via hyperlinking.

v. Android 4.4 WebKit debugging supported for Cordova 3.3+.

vi. Supports PHP 5.6

vii. New GDB Console to enable usage of command line GDB mode.

viii. Flexibility in Windows system is enhanced like split of editor via drag and drop and to customize editor tabs, new API.

ix. For bug tracking, task schedule and other new features in Tasks window [20].
1.6 XAMPP

XAMPP is very popular for providing PHP environment for development. It provides a web server and a database along with it. It is an integrated server package of Apache, MySQL, PHP and Perl. The full form of the abbreviation for XAMPP is X – cross platform, A – Apache, M – MySQL, PHP and P - Perl (the AMPP in XAMPP). The database and the server are accessed from XAMPP Control Panel in this project. The web server that is used here is Apache server and the application server is Tomcat server. MySQL database is the database that is used to store the data in this project [21].

XAMPP is used here to make the installation process of Apache server along with MySQL easy.
2. NARRATIVE

2.1 Problem Statement

Most of the existing systems can target only one event at a time [6]. They can detect either a hurricane or a storm or a tsunami only but not all of them at the same time. They lack in differentiating between their target event and any other similar events. These systems completely believe on the tweets of all the users. There is no functionality to check whether the tweet posted is correct or incorrect. If a user is misled, he has to bear his time and this in turn may lead him to bear his money. A user has to take bold and intelligent decisions in case of these types of emergencies if needed. So a false notification can cost him in huge loss. An application should not be a source for all this. So a registered user on this reporting system should have complete confidence on it and should not be a prey to false notifications as there is a possible chance of his life being at risk [5].

2.2 Motivation

Nothing in this world is more precious than life. All the essentials like food, clothing and shelter including money are for a human being to keep alive. This reporting system keeps its users informed about dangerous life taking events by which they can escape from being trapped and even help their families or friends by informing them. The information about the disasters will be notified to the users through an email in a short span of time after an event like a hurricane occurs. By building such an application users can take safety measures for all their needs, if they are notified within minutes after any
disaster. Integrating data from twitter into a new web based application is a motivating task for the developer as well.

2.3 Project Objective

From the previous observations, there should be reporting systems that should serve its users to take safety measures for all their needs if they are notified within minutes after any disaster occurs. This reporting system identifies the target event that is set from all the twitter data and filters for the positive and negative tweets from it. It will be developed in such a way so that it can distinguish between a hurricane, a storm and a tsunami. Even if some of the users tweet it wrong it is corrected and notified to its users properly. It checks for the trustworthiness of the tweets by comparing it with some of its trustworthy messages and there after it notifies its registered users about it. By this a user can completely believe on this system without having any confusion whether it is correct or not and without waiting for information from other sources. Waiting means elapse of time. In these types of situations every second and every decision taken is important for which correct information need to be provided which can be attained by this proposed system [6].

2.4 Project Scope

The primary task of this project is to keep its users informed of the hurricanes or storms or tsunamis as early as possible after the event occurred so that users can take safety measures for all their needs. They are kept updated by sending an email to their registered email addresses. Twitter is one of the most important medium which has many
numbers of active users in it along with its user friendly interface and backend. This is the reason why twitter data is considered over other social networking sites. User can only avail the services of this system when he is connected to internet.
3. SYSTEM DESIGN

3.1. System Design and Architecture

In Figure 3.1, the system architecture describes how the data is fetched from the twitter website and the user is notified about the disasters. The system first mines all the twitter data and stores it in the form of a table in a database. All the trustworthy news channel data is also retrieved into the database. Then the system filters this data with the required keywords and stores it in another table in MySQL database. Then the system identifies for positive and negative tweets by matching for all the possible combinations. The system even tries to locate the place of the event where it occurred. Then it combines the event and the location and sends a notification email to its registered users [6].

In this project, first the internet connection and the database connection are checked. Once users log in to their twitter page, they go to their twitter main page and refresh all content in the main page. The system detects all tweets of the account and can fetch the data logs depending on the refresh time that is set in the twitter 4j technique we use. This 4j technique is an Application Program Interface that is written in Java programming language. The main functionality of the twitter 4j technique is to establish a connection with the twitter account using four tokens provided by twitter. All the data that is retrieved from twitter main page is stored in a MySQL database. All the data that is collected from the twitter site and stored in the database is now filtered using specific keywords like ‘Hurricane’ or ‘Storm’. This filtered data is now stored in two different tables. In one table, all the positive filtered tweets are stored, and in another table, all the negative filtered tweets are stored. The Pattern Matching Technique algorithm is used to
identify the positive and negative tweets in the database to retrieve the disaster details. Even the negative tweets are differentiated based on the negative keywords that are there in the negative words table. The differentiation of positive and negative tweets are done in the same way. An additional functionality of not storing a negative tweet in the positive tweet table is also implemented. A negative tweet is something like a ‘hurricane conference’ for the keyword ‘hurricane’. Negative tweets are ignored here as the main concern of this system is to notify a user about a disaster.

Once the positive tweets are identified, events and locations are identified from them. Events and locations can be checked for their correctness by comparing them with the trustworthy accounts. Trustworthy accounts are assumed to be always correct in this system. The trustworthy accounts that are followed are BBC World News, CNN, and New York Times. All the positive tweets are compared with the positive tweets from these accounts and are concluded either correct or wrong. If a tweet is declared as wrong by the mismatch of location or event, then a priority is given to the trustworthy accounts and these false tweets are replaced with the tweets of the trustworthy accounts, which are assumed to be correct always. A location table is also stored in the database to make sure that a user in any case is not notified with wrong location about the disaster in the mail. A location table has all the locations in it.

High confidence among the users can be achieved by this. First the positive tweet is checked for the correctness by comparing it with the trustworthy account. These accounts never give false information unless they are hacked. Hacking these accounts is also very difficult as they are news accounts and these types of accounts are much secured. If by any chance they are hacked, then also this reporting system can correct the
location by checking for the place in the database whether this place exists in reality or not.

Users are notified about the disaster immediately after its occurrence through an email. In this project, email services provided by Google are used to notify its users.

Figure 3.1 System Architecture [5]
This Reporting System can be divided into two different applications based on their functionality.

1. Desktop Twitter Application
2. Web Twitter Application

3.2 Desktop Twitter Application

A desktop twitter application is mainly used to differentiate between tweets and then inform a registered user about this positive tweet by sending an email to his google mail account. Through this application, an administrator checks for the internet connection and connection to the database. In this application, tweets which are retrieved from Twitter account are differentiated into positive and negative tweets. Once the admin logs in into his twitter account, it goes to the twitter main page, refresh all the content in the main page and all the tweets of the following accounts can be seen. The fetching of all the tweets from the twitter main page into MySQL database which is in XAMPP is done in this application. Along with these tweets, the tweets of the trustworthy accounts which are news channels in this case are also retrieved. All the tweet messages are stored in the database.

The positive tweets and negative tweets are differentiated now with the existing data collected from the twitter account to identify the disaster events. After identifying the positive tweets and storing them in the table, they are checked with the trustworthy accounts for their correctness. After knowing the correct positive tweets, they are all retweeted in the main page of web twitter application. All the registered users are sent an
email with the free email service provided by Google. Sending the alert message to the users is done through this application which is developed through NetBeans environment.

3.2.1 Data Flow – Desktop Twitter Application

Figure 3.2 shows us the data flow of a twitter desktop application.

Figure 3.2: Data Flow of Desktop Twitter Application
3.3 Web Twitter Application:

Web Twitter Application is mainly for the functionalities of the user. A user who needs to register with the “Disaster Alert System” does so by this application. All the positive tweets are re tweeted to this application so that all users who are registered can see these tweets.

3.3.1 Data Flow – Web Twitter Application

Figure 3.3 shows us the flow of a twitter web application.

![Data Flow Diagram](image)

Figure 3.3: Data Flow of Web Twitter Application
3.4 System Requirements

3.4.1 Functional Requirements

- User needs a registered account in twitter online social network.
- User need to connect my application with twitter account.
- Application integrates all the data from the twitter account.
- Admin specifies the positive and negative keywords for application.
- Application detects all the positive and negative keywords from the integrated data of twitter.
- Compares with the trustworthy accounts data for the correctness of the tweets.
- Users are alerted through an email if a disaster occurs.

3.4.2 Software Requirements

- Language : Core Java
- Version : JDK 1.5
- IDE : Net-beans IDE 8.0.2
- Back-end : MySQL
- Server : XAMPP Control Panel

3.4.3 Hardware Requirements

- Processor : PENTIUM IV
- Clock Speed : 2.5 GHZ
- Ram Capacity : 1 GB
- Hard Disk Drive : 250 GB
4. System Implementation

From the architectural diagram, twitter 4j search API is implemented by me to get all the data logs of twitter into the database. Tokens are mapped for authentication in this API. Pattern Matching Technique Algorithm is also implemented to differentiate positive tweets from negative tweets and to locate for target events and location of these events. Checking for the correctness with the trustworthy accounts is also done with this algorithm. Users are notified about the disaster through an email. The sending of emails to all the registered user is done here. All the backend database SQL transactions are also used in collecting and maintaining the data.

4.1 Environment

The reporting system is implemented using Java in NetBeans IDE environment as they are more suitable for programming. JFrame form is used for the user interface of the desktop application. The web services are provided by Apache using XAMPP control panel along with MySQL database to store all the retrieved twitter data and other information got from data mining. Twitter 4j was the main program used in extracting the data from the twitter account. Alert notifications are sent to the users through google account.

4.1.1 Twitter4j

Twitter4j is an unofficial Java library for the Twitter API. Twitter services can be integrated with the Java application with Twitter4j [23].
4.1.2 NetBeans IDE

NetBeans is an integrated development environment (IDE) that is mainly used for Java and used with other languages also like PHP, C or C++ and HTML 5. NetBeans is an application platform framework not only for Java desktop applications, but for others also. NetBeans IDE is developed in Java and can run on Linux, Windows, Solaris, OS X and on other platforms supporting a compatible JVM [24].

4.1.3 XAMPP

XAMPP is a control panel which is developed by Apache. It provides a web server called apache and an application server called tomcat. Along with these services, it also provides a MySQL database to store all the data in it. XAMPP is mainly used to develop projects in PHP but as it is providing these extra services along with it, it is used in this project.

4.2 User Interface

The main activities performed in this reporting system are checking the connection, retrieving the tweets, retrieving the data from news channel, storing the tweets, filtering the tweets, comparing with news channels and mail alert.

4.2.1 Desktop Twitter Application

In this application, all the admin functionalities are executed. When this application is run, the user interface of it looks like the Figure 4.1
In the Figure 4.1, there are seven different buttons which on their click action performs their respective functionalities. All these functions are explained below.

![Figure 4.1: Desktop Twitter Application Screen](image)

**4.2.2 Checking the Connection**

When checking the connection button is clicked, the actions performed are, it checks for the internet connection and database connections. Internet connection should be active to retrieve all the tweets from the twitter account and the database connection which is java database connection (JDBC) checks for the connection of the application to the database to perform all the functions related with the database. If both the connections are successful, then a message is shown to the admin. This message is shown in Figure 4.2.
4.2.3 Retrieving the Tweets

This function is used to extract the data from the twitter account which is given in the Fetch.java class of the code. The twitter account that is used for this reporting system is Nikhil_TAMUCC. What all tweets that are got to this account are retrieved and stored in the database.

Figure 4.3 shows all the tweets that are tweeted by that particular account. In this case, the account is the Nikhil_TAMUCC and the tweets that are tweeted from this account are shown in the Figure 4.3.
Figure 4.3: Retrieving the Tweets

Figure 4.4 message is shows when the internet connection is established properly and tweets are retrieved from the twitter account provided, then a successful message showing “Tweets are fetched from the account” is displayed. What tweets are fetched from the account are also displayed if the fetching is done successfully.

Figure 4.4: Success Message for Retrieving the Tweets
4.2.4 Retrieving from News Account

When this button is clicked in the desktop twitter application, the same above process is followed and tweets are fetched from the news accounts that are followed by the admin twitter account. If this is successful, then a message confirming that the tweets are fetched successfully is displayed.

Figure 4.5 shows all the tweets that are fetched from the trustworthy account BBC World News and display what all tweets are retrieved from this account. All the tweets are displayed for the accounts that are being followed.

Figure 4.5: Retrieving from news accounts
4.2.5 Storing the Tweets

To store all the tweets that are fetched from the twitter account into the database, this button is used. These tweets are stored into their respective tables in the database. After they are stored in the database, a message showing that the tweets are successfully stored in the database is shown. This message is shown in Figure 4.6.

![Figure 4.6: Storing the Tweets](image)
4.2.6 Filtering the Tweets

This functionality does the filtering part in this reporting system. All the tweets those are fetched from the twitter account and stored in the database are filtered for the positive and negative tweets. This is done by comparing the tweets with the positive and negative words that are stored in the m_negative_word and m_positive_word tables in the database. The positive and negative words are given by the administrator in the tables. Depending on these words, all the tweets are differentiated into positive tweets and negative tweets and these tweets are again stored in two different tables, t_positive_tweet and t_negative_tweet respectively.

![Figure 4.7: Filtering the Tweets](image)

Figure 4.7 shows a list showing what the negative tweets are and what the positive tweets are after clicking on this button. They are differentiated based on the keywords of positive word list and negative word list that are stored in the database.
4.2.7 Comparing with News Channels

This functionality is to check for the correctness of the tweets that are fetched from the twitter account so that no user is notified with wrong information. This is implemented by comparing the positive tweets of the admin account with the tweets of the trustworthy accounts; here the trustworthy tweets are the tweets of the news channels. If both these tweets are matching, then the users are notified about the disaster but if both these are not matching, then a user is notified based on the tweet from the trustworthy account if there are any.

Figure 4.8 shows all the tweets that are resulted after comparison is done when this functionality is executed.
4.2.8 Mail Alert

All the registered users are notified about the hurricane through an email. Here the admin uses a google account to send emails to users. It is the free Gmail service provided by google.

4.2.9 Web Twitter Application

A web twitter application concentrates mainly about users. It is in this application where the user needs to register to get alert notifications. A user can view all the tweets that are re tweeted in this application.

![Figure 4.9: Login Screen of Web Twitter Application](image)
Figure 4.9 shows the login screen of the users of web twitter application. A first time user needs to register from this login page. A user who already got registered to this application just gives his email id and password to login to this application.

From the Figure 4.10, a user can see all the tweets that are re tweeted in this application along with the location from where the tweet is coming from. By displaying the location column in this screen, it helps a user to identify from where this tweet is coming and this in turn helps the user to know where that particular event occurred in case when the location is not specified in the tweet. By this if a user misses to see the email notification sent, can come into this application and view all the tweets and can come to know what happened about a disaster.

Figure 4.10: List of Tweets
A user can register with this application by going into the register link. It is necessary for a user to register with this application to get the alert notifications. Figure 4.11 displays the registration screen and what all criteria a user needs to get registered with this application to get the notifications.

For a user to have an account with this application, he needs to register first and then can receive email notifications. To get registered with this application, a user needs to give his name so that at the time of sending an email it refers to his name in the content of the email. A user needs to specify an email ID for the obvious reason to send him an alert notification. There is a password column to set a password for his account for the purpose of authentication. The location where the user is from also need to be provided. There is a register button at the end to save all the users account information.

Figure 4.11: User Registration Screen
5. TESTING AND EVALUATION

5.1 Software Testing

The process of evaluating a software product to identify the differences between given input and expected output is called as software testing. The quality of a product can be judged by testing.

5.2 Unit Testing

Unit testing is a type of testing where a single unit or groups of related units are tested. It is tested for the expected outputs for given inputs of the unit that is implemented [16]. In this reporting system, keywords selected tweets step can be tested for its accuracy. The input value for this is all the tweets that we get from the twitter 4j search API. Pattern Matching Technique algorithm can be tested for correct positive and negative tweets. Other related units like event detector and location detector can also be tested under unit testing.

5.3 System Testing

System testing usually verifies that the system that is being delivered meets the scope of the product along with the specifications and its purpose [17]. In this project one of the system testing tasks could be, how many users can be registered with this product?

The testing part for this project can be evaluated by trying with different keywords for the target event based on the expected output. The matching algorithm that will be used to compare the tweets of the users with that of the tweets of trustworthy
accounts can also be tested for them both tweeting different messages for the occurrence of same event. The matching of the combinations for positive and negative tweets can be tested by the results of the positive and negative tweets after they are differentiated.

For the mined data from the twitter, it should filter with the keywords for the target event like hurricanes and differentiate the tweets into positive and negative tweets depending on the possible matching combinations set and then check for the trustworthiness by comparing the positive tweets with the trusted accounts like CNN or BBC for example. The location of the positive tweets is identified and users are notified through their registered email ids if the system executes as it is expected to work.

### 5.4 Test Cases for Tweet Tracking

<table>
<thead>
<tr>
<th>S.No</th>
<th>Name</th>
<th>Description</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Checking the internet connection</td>
<td>The internet connection will be checked to get connect with the twitter account</td>
<td>Success</td>
</tr>
<tr>
<td>2</td>
<td>Fetching the twitter application detail</td>
<td>The secret token is generated while creating the twitter application. The secret tokens will be stored in the database.</td>
<td>Success</td>
</tr>
<tr>
<td>3</td>
<td>Collecting the tweets from the user account</td>
<td>By using the secret tokens, our project will be connected with the twitter and tweets are fetched from the desired account</td>
<td>Success</td>
</tr>
<tr>
<td>4</td>
<td>Fetching from news channel</td>
<td>As same as collecting the tweets from the twitter account, the tweets of the news channel are collected for further comparison.</td>
<td>Success</td>
</tr>
</tbody>
</table>
Table 5.1 gives a description of all the possible cases that can be tested. The status of these cases that are tested is also given whether they are functioning properly or not. For all the cases that are tested, they were functioning properly as expected and the results that are obtained after the testing were also correct. So a status message showing success is in the last column of the table.

**5.5 Test Results**

Some of the cases are tested in this section and their results are shown so that they can be concluded whether they are behaving as expected or not.

<table>
<thead>
<tr>
<th>S.No</th>
<th>Name</th>
<th>Description</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>Storing the tweets</td>
<td>The tweets retrieved from user account will be stored in the database</td>
<td>Success</td>
</tr>
<tr>
<td>6</td>
<td>Filtering the tweets</td>
<td>The stored tweets will be filtered with the positive and negative keywords.</td>
<td>Success</td>
</tr>
<tr>
<td>7</td>
<td>Comparing with the news channel</td>
<td>The filtered positive tweets will be compared with the news channel and resultant tweets will be displayed on the screen.</td>
<td>Success</td>
</tr>
<tr>
<td>8</td>
<td>Mail Alert</td>
<td>Admin will send mail alert to all the registered user from the web application.</td>
<td>Success</td>
</tr>
<tr>
<td>9</td>
<td>List Tweets in web application</td>
<td>Stored tweets in the database will be displayed in the web application.</td>
<td>Success</td>
</tr>
<tr>
<td>10</td>
<td>Register new user</td>
<td>Users will be registered in the web application.</td>
<td>Success</td>
</tr>
</tbody>
</table>
5.5.1 Storing of Tweets

Tweets fetched from the accounts are stored into the database when “store the tweets” button on the admin panel is clicked. Here it is test for the non-duplicate tweets being stored or not in the table. As twitter is a large social networking site with more than 300 million active users, it is going to occupy a lot of space in the database if all these tweets are stored in it. Of all these tweets, most of them will not be that useful because they all do not come under positive tweets. So storing of this unused data is a waste of space.

To test for the uniqueness of the content in the database, an example is considered. Tweets are fetched and stored in the database for the first time on 4/18/2015 at 11:46 AM for that day. There were 103 tweets that were stored from 12:01 AM on 4/18/2015 till 11:46 AM on 4/18/2015. Again on the same day after 6 minutes at 11:52 AM, all the tweets were fetched and stored into the database. But this time only 22 tweets were added to the table t_tweets because these were the only tweets that were unique from the previous storage.
In Table 5.2, all the tweets that are fetched at 11:46 AM are shown. The tweets started from row number 623 in Table 5.2 and ended at row number 725 in Table 5.3. So a total of 103 tweets are stored when the tweets are first fetched. Table 5.3 shows that the next tweet fetched and stored in the table is at row number 726, and this tweet is stored at 11:52 when the tweets are again fetched for the second time. There were about 22 tweets stored when they were fetched second time. This shows that there were no duplicate data stored in the table.

### 5.5.2 Testing with Trustworthy Accounts

The positive tweets of the Twitter data are tested with the trustworthy accounts to check for their correctness so that a user is not notified with wrong information. A tweet

<table>
<thead>
<tr>
<th>Table 5.2: First Retrieved Tweets</th>
<th>Table 5.3: Second Retrieved Tweets</th>
</tr>
</thead>
<tbody>
<tr>
<td>81) Look deep into his dog's eyes. Are you in love y... Not Found</td>
<td>72) RT @nytimes: Video: The timeless charms of P... Not Found</td>
</tr>
<tr>
<td>91) RT @sydneymourn: Hillary: Steps short of impe... Not Found</td>
<td>73) Decrees killed in suicide bombing at Bank in E... Not Found</td>
</tr>
<tr>
<td>601) Why Americans don't want to soak the rich http://... Not Found</td>
<td>74) Was this man in Connecticut, not the Wright brother... Not Found</td>
</tr>
<tr>
<td>61) RT @Quddhash: Next week. @JoshNYT turns one yea... Not Found</td>
<td>75) RT @Brencher: Many of us can trace our kids cre... Not Found</td>
</tr>
<tr>
<td>22) RT @JoshNYT: The itch are getting richer, but th... Not Found</td>
<td>76) Australian PM Tony Abbott moves bear in 7 seconds... Not Found</td>
</tr>
<tr>
<td>62) Williams of purple jellyfish are washing up in Ca... Oregon</td>
<td>77) Joseph Goodenoot: artist using public's Random Has... Not Found</td>
</tr>
<tr>
<td>63) M injured in California gas line explosion: http... California</td>
<td>78) Ceremony held at Everest base camp, a year after... Not Found</td>
</tr>
<tr>
<td>64) Oklahoma approves nitrogen gas as a backup somew... Oklahoma</td>
<td>79) CPS uninves of Lord Jenner claims <a href="http://cbs.cm/">http://cbs.cm/</a>... Not Found</td>
</tr>
<tr>
<td>65) As the climbing season resumes on Mount Everest... Not Found</td>
<td>80) RT @BBC: Xiaolu: &quot;I am his hands, he is my eyes&quot;... Not Found</td>
</tr>
<tr>
<td>66) RT @BBCSport: Hamilton takes a 4th straight pole... Not Found</td>
<td>81) RT @BBC: Xiaolu: Police call Melbourne terror plot... Not Found</td>
</tr>
<tr>
<td>67) RT @bbcnews1: More than 1800 clothing workers... Not Found</td>
<td>82) Time: Republican candidates plot their coup http://... Not Found</td>
</tr>
<tr>
<td>68) RT @bbcnews1: The @MercedesAMGP1 boys are happy... @Low... Not Found</td>
<td>83) Police call: Melbourne terror plot <a href="http://cbs.cm/">http://cbs.cm/</a>... Not Found</td>
</tr>
<tr>
<td>69) RT @BBC: CNN: Political satire in #Mexico has a... Not Found</td>
<td>84) I am in love and he is my arms <a href="http://cbs.com.br/">http://cbs.com.br/</a>... Not Found</td>
</tr>
<tr>
<td>70) RT @BBCNewsMagazine: Einstein's death 68 years ago... Not Found</td>
<td>85) The strange afterlife of Einstein's brain http://... Not Found</td>
</tr>
<tr>
<td>71) RT @BBC: John Olav: &quot;I saw death with my eyes so many ti... Not Found</td>
<td>86) 10 things we didn't know last week <a href="http://cbs.com.br/">http://cbs.com.br/</a>... Not Found</td>
</tr>
<tr>
<td>72) RT @BBCNewsMagazine: Einstein's death 68 years ago... Not Found</td>
<td>87) RT @CNN: Yankees Stadium opens on this day... Not Found</td>
</tr>
<tr>
<td>73) Senior saguaro lichen in eastern #Arizona varie... Not Found</td>
<td>88) A road beef sandwich the way the doll makes it... Not Found</td>
</tr>
<tr>
<td>74) RT @BBC: Xiaolu: &quot;I saw death with my eyes so many ti... Not Found</td>
<td>89) RT @NYtimes: Yankees Stadium opens on this day... Not Found</td>
</tr>
<tr>
<td>75) Ceremony held at Everest base camp, a year after... Not Found</td>
<td>90) A road beef sandwich the way the doll makes it... Not Found</td>
</tr>
<tr>
<td>76) Polkman proud to surpass Bellhorn <a href="http://cbs.com.br/">http://cbs.com.br/</a>... Not Found</td>
<td>91) A road beef sandwich the way the doll makes it... Not Found</td>
</tr>
<tr>
<td>77) Decrees killed in suicide bombing at Bank in E... Not Found</td>
<td>92) A road beef sandwich the way the doll makes it... Not Found</td>
</tr>
<tr>
<td>78) Ceremony held at Everest base camp, a year after... Not Found</td>
<td>93) A road beef sandwich the way the doll makes it... Not Found</td>
</tr>
<tr>
<td>79) CPS uninves of Lord Jenner claims <a href="http://cbs.cm/">http://cbs.cm/</a>... Not Found</td>
<td>94) Why Americans don't want to soak the rich http://... Not Found</td>
</tr>
<tr>
<td>80) RT @BBC: Xiaolu: Police call Melbourne terror plot... Not Found</td>
<td>95) Australian PM Tony Abbott moves bear in 7 seconds... Not Found</td>
</tr>
<tr>
<td>81) RT @nytimes: Video: The timeless charms of P... Not Found</td>
<td>96) RT @nytimes: Video: The timeless charms of P... Not Found</td>
</tr>
<tr>
<td>82) Decrees killed in suicide bombing at Bank in E... Not Found</td>
<td>97) Was this man in Connecticut, not the Wright brother... Not Found</td>
</tr>
<tr>
<td>83) Australian PM Tony Abbott moves bear in 7 seconds... Not Found</td>
<td>98) RT @Brencher: Many of us can trace our kids cre... Not Found</td>
</tr>
<tr>
<td>84) Australian PM Tony Abbott moves bear in 7 seconds... Not Found</td>
<td>99) Police call: Melbourne terror plot <a href="http://cbs.cm/">http://cbs.cm/</a>... Not Found</td>
</tr>
<tr>
<td>85) The strange afterlife of Einstein's brain http://... Not Found</td>
<td>100) I am in love and he is my arms <a href="http://cbs.com.br/">http://cbs.com.br/</a>... Not Found</td>
</tr>
<tr>
<td>86) 10 things we didn't know last week <a href="http://cbs.com.br/">http://cbs.com.br/</a>... Not Found</td>
<td>101) The strange afterlife of Einstein's brain http://... Not Found</td>
</tr>
<tr>
<td>87) RT @CNN: Yankees Stadium opens on this day... Not Found</td>
<td>102) A road beef sandwich the way the doll makes it... Not Found</td>
</tr>
<tr>
<td>88) A road beef sandwich the way the doll makes it... Not Found</td>
<td>103) A road beef sandwich the way the doll makes it... Not Found</td>
</tr>
<tr>
<td>89) A road beef sandwich the way the doll makes it... Not Found</td>
<td>104) A road beef sandwich the way the doll makes it... Not Found</td>
</tr>
<tr>
<td>90) A road beef sandwich the way the doll makes it... Not Found</td>
<td>105) A road beef sandwich the way the doll makes it... Not Found</td>
</tr>
<tr>
<td>91) A road beef sandwich the way the doll makes it... Not Found</td>
<td>106) A road beef sandwich the way the doll makes it... Not Found</td>
</tr>
<tr>
<td>92) A road beef sandwich the way the doll makes it... Not Found</td>
<td>107) A road beef sandwich the way the doll makes it... Not Found</td>
</tr>
<tr>
<td>93) A road beef sandwich the way the doll makes it... Not Found</td>
<td>108) A road beef sandwich the way the doll makes it... Not Found</td>
</tr>
<tr>
<td>94) Why Americans don't want to soak the rich http://... Not Found</td>
<td>109) A road beef sandwich the way the doll makes it... Not Found</td>
</tr>
<tr>
<td>95) Australian PM Tony Abbott moves bear in 7 seconds... Not Found</td>
<td>110) A road beef sandwich the way the doll makes it... Not Found</td>
</tr>
<tr>
<td>96) RT @nytimes: Video: The timeless charms of P... Not Found</td>
<td>111) A road beef sandwich the way the doll makes it... Not Found</td>
</tr>
<tr>
<td>97) Was this man in Connecticut, not the Wright brother... Not Found</td>
<td>112) A road beef sandwich the way the doll makes it... Not Found</td>
</tr>
<tr>
<td>98) RT @Brencher: Many of us can trace our kids cre... Not Found</td>
<td>113) A road beef sandwich the way the doll makes it... Not Found</td>
</tr>
<tr>
<td>99) Police call: Melbourne terror plot <a href="http://cbs.cm/">http://cbs.cm/</a>... Not Found</td>
<td>114) A road beef sandwich the way the doll makes it... Not Found</td>
</tr>
<tr>
<td>100) I am in love and he is my arms <a href="http://cbs.com.br/">http://cbs.com.br/</a>... Not Found</td>
<td>115) A road beef sandwich the way the doll makes it... Not Found</td>
</tr>
<tr>
<td>101) The strange afterlife of Einstein's brain http://... Not Found</td>
<td>116) A road beef sandwich the way the doll makes it... Not Found</td>
</tr>
<tr>
<td>102) A road beef sandwich the way the doll makes it... Not Found</td>
<td>117) A road beef sandwich the way the doll makes it... Not Found</td>
</tr>
<tr>
<td>103) A road beef sandwich the way the doll makes it... Not Found</td>
<td>118) A road beef sandwich the way the doll makes it... Not Found</td>
</tr>
<tr>
<td>104) A road beef sandwich the way the doll makes it... Not Found</td>
<td>119) A road beef sandwich the way the doll makes it... Not Found</td>
</tr>
<tr>
<td>105) A road beef sandwich the way the doll makes it... Not Found</td>
<td>120) A road beef sandwich the way the doll makes it... Not Found</td>
</tr>
</tbody>
</table>
is tested for its correctness by comparing it with trustworthy accounts. Here the trustworthy accounts that are followed in twitter are BBC World News, CNN and New York Times. It is tested by taking an example of an event occurring at Corona, California but a user tweeting the same event being occurred at Dallas.

In Figure 5.1, the tweets that are fetched from the user account are shown. Here the tweet says “Wildfire in the forests near Dallas” where the event is wildfire and the location is Dallas. But in contradiction to this, Figure 5.2 shows all the tweets fetched from CNN account where the wildfire tweet says “Wildfire near Corona, California”. Here the event is wildfire and the location is California. This is in contradiction to the users tweet. So a registered user needs to be notified with the correct information from the above two tweets of which only one being correct.

![Figure 5.1: Tweets of Admin Account](image1)
![Figure 5.2: Tweets of CNN Account](image2)
Figure 5.3 shows all the positive tweets that are filtered from all the tweets that are fetched. This figure shows that there are two different locations for the same disaster event. The next Figure 5.4 shows the corrected information of the tweet. This tweet is the correct tweet and is got from the trustworthy channel. The user tweet is wrong and the trustworthy tweet is correct.

5.5.3 Testing of Mail Alert

Here it is checked for an email notification of a user when a disaster occurred. In Figure 5.5, an email is seen sent to a registered user after a disaster occurred.
Figure 5.5: Registered User Notification

A user is notified using the email service provided by Google. The subject of this email is Disaster Alert. In this email, the location of where the event happened is given along with the tweet from which it is concluded that a disaster occurred.
6. CONCLUSION

This project proposes the idea of creating an application of a real time nature of twitter using the event detection as a target. Using the pattern matching technique algorithm we can find the positive and negative tweets from the user twitter main page. A disaster reporting system is created with the user tweets being analyzed for positive and negative tweets and these are stored in the database. The positive tweets are compared with the trustworthy tweets for their correctness. Using the mailing option here users are notified of this event just after its occurrence or sometimes even before the event occurred at that particular place. Micro-blogging is a real time supporting service. Using the real-time supporting services like twitter, social problems like natural disasters can be reduced.

Apart from the work done in this reporting system, the enhancement for this could be sending a message through the communication network. In this project an email is sent as an alert notification.
BIBLIOGRAPHY AND REFERENCES


[20] https://netbeans.org/community/releases/80/


